Digenic descent measures for finite populations*

BY C. CLARK COCKERHAM AND B. S. WEIR

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

(*Received* 1 *March* 1977)

SUMMARY

The development of a set of two-locus descent measures is reviewed. The three digenic measures, inbreeding coefficient and parental and recombinant descent coefficients, are considered in detail. The derivations of these three in pedigrees, fixed mating systems, and random mating in monoecious or dioecious populations are given. General expressions for digenic frequencies and disequilibria functions at any time are found by applying the three digenic descent measures to two types of initial populations. The final or equilibrium status of the population is also given. As the inbreeding coefficient is the same as the recombinant descent coefficient in the case of complete linkage, avoidance or promotion of early inbreeding has similar effects on the two coefficients. Estimable components of linkage disequilibrium and other measures of association within and among populations are elaborated.

1. INTRODUCTION

There is currently a great deal of interest in, and need for, two-locus theory in population genetics since modern techniques for collecting data allow frequencies of alleles at several loci to be found simultaneously. This data, when analysed on the basis of an adequate theory, will provide a picture of the interactions between loci and should indicate to what extent such interactions are the result of linkage or the constraints of finite population size and/or the system of mating.

We have been concerned with developing a set of parameters on which a theory allowing the analysis of two-locus data could be based. It is the purpose of the present paper to clarify our approach (Cockerham & Weir, 1968, 1973; Weir & Cockerham, 1969*a,b*, 1973, 1974). To aid this review, we concentrate on digenic measures in finite populations and show, for example, how they affect linkage disequilibrium.

2. ONE-LOCUS PARAMETERS

The study of one-locus systems is well served by the inbreeding coefficient, F. This one-locus measure gives the probability of identity by descent of genes within an individual, while an analogous quantity, the coancestry coefficient θ, is for genes chosen at random from two individuals.

In the first place, the inbreeding coefficient summarizes information about the mating system. Different systems can be compared on the basis of the appropriate inbreeding coefficients. When F is known, a knowledge of population gene frequencies, p_i for allele a_i, allows the expression of population genotypic frequencies as

$$P_{a_i}^{a_i} = p_i^2 + p_i(1-p_i)F, \quad P_{a_j}^{a_i} = (1-F)p_ip_j \quad (i \neq j). \tag{1}$$

(Note that $P_{a_j}^{a_i} = P_{a_i}^{a_j}$ so that $2P_{a_j}^{a_i}$ is the total heterozygote frequency.) It is well known that there is difficulty with such expressions for populations that are initially heterozygous, and we return to this later.

Because F, and θ, find use in frequency expressions, they also find use in analyses of variation (Cockerham, 1969, 1973). Expectations of various quadratic forms may be phrased in terms of these measures or their analogues, and so clarify estimable parameters and testable hypotheses. In these situations, the parameters do need to change from being measures of identity by descent to being correlations.

Among the analogues of F and θ mentioned above are those which deal with population structure. Wright (1951) introduced F_{IT}, F_{IS} and F_{ST} and there are others. Cockerham (1967) also utilized similar measures that apply to groups of individuals.

The inbreeding coefficient or its analogues are often estimated from data sets, generally by appropriate use of the results in (1). Knowledge of the sampling distribution of F under any particular set of assumptions may then lead to tests of hypotheses about such assumptions.

Not the least of the advantages of the ubiquitous inbreeding coefficient is the ease with which it may be calculated. For pedigrees, Wright (1922) showed that the inbreeding coefficient for an individual I whose parents had a common ancestor A, with n_1 and n_2 generations between parents and common ancestor, is given by

$$F_I = \left(\frac{1}{2}\right)^{n_1+n_2} \left(\frac{1+F_A}{2}\right). \tag{2}$$

This expression is summed over all pathways from the parents to the common ancestor, and then summed over all common ancestors, pathways always stopping when they meet at a common ancestor.

Evaluation of F in regular systems of inbreeding was aided by the introduction of the concept of identity by descent and associated probability treatments (Malécot, 1948). Recurrence formulas can be established and are well known.

We have found it helpful to look at F in a slightly different way. If allelic genes a, a' are on gametes uniting to form individual I, we say that F_I is the probability that a, a' descended from one gene in an initial population. Alternatively, a and a' are both descended from a gene on one initial gamete. This viewpoint does not change the value of F but is helpful in relating present frequencies of gene combinations to those in an initial population. Furthermore, we write the one-locus inbreeding and coancestry coefficients as F_1 and θ_1, respectively, and use the equivalence sign $\equiv$ to refer to genes descending from one initial gamete. Such genes are said to be equivalent by descent. If individual I receives genes a, a', then

$$F_{1I} = \text{prob}\,(a \equiv a'), \tag{3}$$

while if genes a, a' are random genes from individuals B, C, respectively, then

$$\theta_{1_{BC}} = \text{prob}\,(a \equiv a'), \tag{4}$$

and $F_{1_I} = \theta_{1_{BC}}$ if I is an offspring of B and C.

3. TWO-LOCUS PARAMETERS

Our first attempt at a two-locus parameterization was to consider identity by descent at two loci (Cockerham & Weir, 1968; Weir & Cockerham, 1969 a, b). If gametes uniting to form individual I carry genes a, b, and a', b', respectively at two loci, then the two-locus inbreeding coefficient is

$$F_{11_I} = \text{prob}\,(a \equiv a' \quad \text{and} \quad b \equiv b').$$

The pair of genes a, a' may or may not have descended from a gene on the same initial gamete as the gene from which b, b' are both descended. This two-locus coefficient has been defined several times in the literature (e.g. Haldane, 1949) and we were able to establish an algorithm which enabled F_{11} to be calculated for any non-assortative system of mating. The algorithm required the introduction of trigametic and quadrigametic measures in addition to the digametic F_{11}. We discussed an identity disequilibrium function

$$\eta_{11} = F_{11} - F_1^2$$

but found really very little application for F_{11}. The difficulty was that this coefficient did not allow the expression of joint genotypic frequencies at two loci with linkage disequilibrium.

We then went quite a bit further (Cockerham & Weir, 1973; Weir & Cockerham, 1973, 1974) and defined a class of measures which gave the probabilities of equivalence by descent for all (6) possible pairs, all (4) possible triples, all (3) possible two distinct pairs, and the (1) quadruple of the two uniting gene pairs ab and $a'b'$. For each of these 14 combinations of genes, the corresponding descent measure gives the probability that all these genes descend from genes on one initial gamete. Since we seldom need to distinguish between the descent status of two loci, symmetry allows us to reduce the number of measures to eight: three digenic, one trigenic, three for two distinct pairs and one quadrigenic. Once again we were able to establish an algorithm for the calculation of these measures in all non-assortative systems of mating and showed how, with a knowledge of the constitution of the initial population, the measures led to two-locus genotypic frequencies. A consequence of this was the ability to determine various disequilibrium functions, and we derived expressions for the means, variances and covariances of sample linkage disequilibria.

Our notation has apparently led to some difficulty in understanding the procedures. We were anxious that the need for 14 digametic measures, and similar numbers of trigametic and quadrigametic ones, would not lead to a plethora of symbols and so retained F for the digametic case within individuals and θ for the digametic case between individuals. Six suffixes are used, one for each possible pair

9

of four genes, and combinations of these six accommodated triples and quadruples of genes. In line with existing notation (e.g. Shikata, 1965), we used F_1 and F_{11} for one- and two-locus inbreeding coefficients, i.e. for the non-gametic but allelic pairs a, a' and b, b' within individuals. The probability of equivalence by descent for either or both of the gametic but non-allelic pairs a, b and a', b' is written as F^1 or F^{11}, respectively. For either or both of the non-gametic and non-allelic pairs, a, b' or a', b, we use $_1F$ or $_{11}F$, respectively.

4. DIGENIC DESCENT MEASURES

There are three types of gene pairs for the uniting gametes ab and $a'b'$ and the three digenic descent measures are

inbreeding coefficient: $F_1 = \frac{1}{2}[\text{prob}\,(a \equiv a') + \text{prob}\,(b \equiv b')],$

recombinant coefficient: $_1F = \frac{1}{2}[\text{prob}\,(a \equiv b') + \text{prob}\,(a' \equiv b)],$

parental coefficient: $F^1 = \frac{1}{2}[\text{prob}\,(a \equiv b) + \text{prob}\,(a' \equiv b')].$

The quantities averaged in each of the three cases are often equal but it is the average with which we are concerned. Both F_1 and $_1F$ are digametic measures since they refer to two genes, one on each of two uniting gametes, and this similarity is emphasized by the similarity of notation. If the two loci are completely linked, so that the pairs ab and $a'b'$ are always transmitted between generations as single units, there can be no difference between the inbreeding and recombinant coefficients and $F_1 = {}_1F$. We regard F_1 as a special case of $_1F$ then. The third measure, F^1, is composed of monogametic ingredients but it is the digametic average that is useful. In the special case of completely linked loci, we must have $F^1 = 1$. If the two gametes are from distinct individuals, B and C, we replace any descent measure F with θ, the corresponding coancestry between B and C. As mentioned previously, if I is an offspring of B and C, then $F_I = \theta_{BC}$.

Between them the digenic measures contain a great deal of information about the two-locus structure of a population, but before elaborating on this use of the measures we establish a procedure for calculating them. By way of comparison to classical theory, we start with a consideration of pedigrees.

5. PEDIGREES

For any individual I we trace the gametes received by I from its parents back to a common ancestor A of these parents. In the calculation of $_1F$, just as for F_1, we determine the contribution of A to $_1F_I$ and see that a factor of $\frac{1}{2}$ is required for each generation between the parents of I and A. Having traced genes a, b' (or a', b) back to two gametes from A, we recognize that with equal probabilities they are descended from genes received on either one or two gametes by A and so are equivalent by descent with probabilities F^1_A or $_1F_A$, respectively. We have argued, then, that if n_1 and n_2 generations separate the parents of I from A,

$$_1F_I = \left(\frac{1}{2}\right)^{n_1+n_2} \left(\frac{F^1_A + {}_1F_A}{2}\right), \tag{5}$$

and this expression is to be summed over all paths from the parents of I to A, and over all common ancestors A. The similarity to the classical result (2) for the inbreeding coefficient is evident, and indeed (5) reduces to (2) for completely linked loci, as it should.

As the evaluation of $_1F$ involves knowledge of F^1 we now consider that measure. Regardless of the mating scheme, any gamete received by individual I carries two genes that are descended from two genes carried on either one or both of the
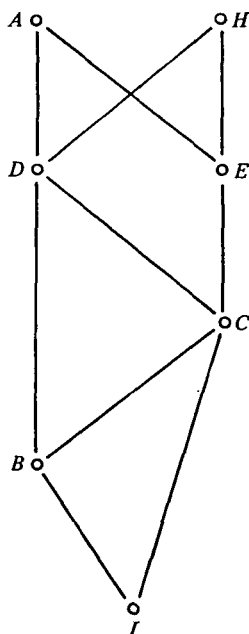

Fig. 1. A pedigree.

gametes received by a parent B of I. If the two loci are linked to an extent λ, where λ is one minus twice the recombination fraction, then for the gamete received from parent B

$$F^1_{IB} = \frac{1+\lambda}{2} F^1_B + \frac{1-\lambda}{2} {}_1F_B. \tag{6}$$

Since F^1 refers to the average of gene pairs ab, $a'b'$ received by I,

$$F^1_I = \frac{1+\lambda}{4}(F^1_B + F^1_C) + \frac{1-\lambda}{4}({}_1F_B + {}_1F_C),$$

where B, C are the two parents of I.

In pedigrees then, equations (5) and (6) are used in conjunction and expansions made back to common ancestors for (5) and the initial individuals for (6). As an illustration we consider the pedigree in Fig. 1, where the initial pair of individuals A, H are non-inbred and unrelated. In Table 1 we show the three digenic measures for all seven individuals. Descent status is given relative to the initial individuals so that the initial gametes for which equivalence by descent is defined are those gametes which united to form the initial individuals A, H.

9·2

Table 1. *Digenic descent measures for pedigree of Fig. 1.*

Individual	F_1	$_1F$	F^1
A	0	0	1
H	0	0	1
D	0	0	$\dfrac{1+\lambda}{2}$
E	0	0	$\dfrac{1+\lambda}{2}$
C	$\frac{1}{4}$	$\frac{1}{4}$	$\left(\dfrac{1+\lambda}{2}\right)^2$
B	$\frac{3}{8}$	$\dfrac{2+\lambda}{8}$	$\dfrac{4+6\lambda+5\lambda^2+\lambda^3}{16}$
I	$\frac{1}{2}$	$\dfrac{4+3\lambda+\lambda^2}{16}$	$\dfrac{16+16\lambda+21\lambda^2+10\lambda^3+\lambda^4}{64}$

6. REGULAR SYSTEMS WITH SPECIFIED MATINGS

Regular inbreeding systems, when all matings are specified so that there is no choice of mates, are also dealt with quite simply. We will consider only systems where every individual leaves exactly two gametes to the next generation. Inbreeding coefficients for populations of size N with two gametes per parent and undergoing various 'early' or 'late' inbreeding mating patterns have been considered in the past (Kimura & Crow, 1963; Cockerham, 1970).

In the various cousin, or maximum avoidance, systems least related individuals are mated in any generation. A general recurrence formula in the case of $N = 2^v$ was given for the inbreeding coefficient by Cockerham (1970) as

$$F_{1_{(t+v+1)}} = \frac{1}{2N}\left(\sum_{u=0}^{v} 2^u F_{1_{(t+u)}} + 1\right), \tag{7}$$

where $F_{1_{(t)}}$ is the coefficient for each individual in the tth generation, and initially there is no inbreeding: $F_{1_{(0)}} = F_{1_{(1)}} = \ldots = F_{1_{(v)}} = 0$.

The last term, 1, in equation (7) was introduced by the first common ancestor, $v + 1$ generations back. From our previous discussion then we see that

$$_1F_{(t+v+1)} = \frac{1}{2N}\left(\sum_{u=0}^{v} 2^u {}_1F_{(t+u)} + F^1_{(t)}\right) \tag{8}$$

while equation (6) may be written as

$$F^1_{(t+1)} = \frac{1+\lambda}{2}F^1_{(t)} + \frac{1-\lambda}{2} {}_1F_{(t)} \tag{9}$$

and symmetry obviates the need for separate accounting of the two parents of an individual. Equation (9) suggests that F^1 can be eliminated from (8) by taking $_1F_{(t+v+2)} - \frac{1}{2}(1+\lambda)\,_1F_{(t+v+1)}$. This provides the homogeneous equation

$$_1F_{(t+v+2)} = \frac{2+\lambda}{2}\,_1F_{(t+v+1)} - \frac{\lambda}{4N}\sum_{u=0}^{v} 2^u\,_1F_{(t+u)} - \frac{\lambda}{4N}\,_1F_{(t)},$$

which can be rearranged to show the following equivalence for all t,

$$_1F_{(t+1+v+1)} - \frac{1}{2N} \sum_{u=0}^{v} \lambda 2^u \, {}_1F_{(t+1+u)} = {}_1F_{(t+v+1)} - \frac{1}{2N} \sum_{u=0}^{v} \lambda 2^u \, {}_1F_{(t+u)}.$$

Using the initial condition that the first $v+1$ values of $_1F$ are zero, the common value of these two expressions is $1/2N$ and the lowest order recurrence formula for F is seen to be

$$_1F_{(t+v+1)} = \frac{1}{2N} \left(\sum_{u=0}^{v} \lambda 2^u \, {}_1F_{(t+u)} + 1 \right). \tag{10}$$

We can show that $F^1_{(t)}$ will also satisfy a recurrence formula of exactly the same form as equation (10) with initial values of $F^1_{(t)} = [\frac{1}{2}(1+\lambda)]^t$; $t = 0, 1, \ldots, v$. In fact, F^1 and $_1F$ both satisfy one equation in any inbreeding system. If λ is set equal to 1, this is always the equation for F_1 as well. The different initial values for the three digenic measures ensure their different values in later generations, and the main consequence of a common recurrence formula is a common rate of approach to equilibrium values.

As an example of an early inbreeding system, we consider circular mating. Individuals may be regarded as being arranged on the circumference of a circle and each individual mated to both of its neighbours, each mating producing one offspring. We restrict attention to the case of $N = 2n$.

We need a series of recombinant coancestry coefficients $_1\theta(k)_{(t)}$ referring to the equivalence of non-allelic genes on two gametes taken from individuals k apart ($k = 1, 2, \ldots, 2n-1$) in generation t. Since adjacent individuals mate

$$_1F_{(t+1)} = {}_1\theta(1)_{(t)} \tag{11}$$

and the usual tracing of gametes back one generation gives

$$_1\theta(1)_{(t+1)} = \frac{1}{4}[{}_1\theta(0)_{(t)} + 2\,{}_1\theta(1)_{(t)} + {}_1\theta(2)_{(t)}]$$

while, since $_1\theta(0)$ refers to a common ancestor

$$_1\theta(1)_{(t+1)} = \frac{1}{8}(F^1_{(t)} + {}_1 F_{(t)}) + \frac{1}{2}\,{}_1\theta(1)_{(t)} + \frac{1}{4}\,{}_1\theta(2)_{(t)}. \tag{12}$$

We also see that

$$_1\theta(k)_{(t+1)} = \frac{1}{4}[{}_1\theta(k-1)_{(t)} + 2\,{}_1\theta(k)_{(t)} + {}_1\theta(k+1)_{(t)}] \quad (2 \leqslant k \leqslant n-1), \tag{13}$$

$$_1\theta(n)_{(t+1)} = \frac{1}{2}[{}_1\theta(n-1)_{(t)} + {}_1\theta(n)_{(t)}]. \tag{14}$$

Equation (9) holds, and we have

$$_1\theta(k)_{(t)} = {}_1\theta(2n-k)_{(t)} \quad (k = 1, 2, \ldots, n). \tag{15}$$

Now equations (11) to (15) may be amalgamated and some matrix manipulations similar to those of Kimura & Crow (1963) used to provide a recurrence formula for $_1F_{(t)}$, and hence for $F^1_{(t)}$ and $F_{1_{(t)}}$. The details are given in Appendix A, and here we just present the first two cases, for $N = 2$ and $N = 4$:

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

$$N = 4 \colon {}_1F_{(t+3)} = \frac{1+\lambda}{2}\,{}_1F_{(t+2)} - \frac{\lambda}{16}\,{}_1F_{(t)} + \frac{1}{16},$$

$$_1F_{(0)} = {}_1F_{(1)}0, \quad {}_1F_{(2)} = \tfrac{1}{8}.$$

These results reduce to those given by Kimura & Crow (1963) for the inbreeding coefficient when $\lambda = 1$. Another early inbreeding system known as circular pair mating (Kimura & Crow, 1963) may be treated similarly and details are given in Appendix A.

7. FINITE POPULATIONS WITH RANDOM MATING

The next level of complexity in our discussion is for cases of finite populations of constant size when matings are not specified. Various gametic sampling plans need to be considered and there is the complication that all N members of a generation do not now have the same pedigree. It is necessary to work with average measures, denoted by bars; for example, $\overline{F}_{1_{(t)}}$ is the average inbreeding coefficient for members of generation t. Such averaging is additional to the expectations implied in all uses of inbreeding or descent measures. These measures are probabilities of identity or equivalence by descent and so they are the proportions of all possible replicates of the population under study in which the corresponding genes have that descent status. In a particular population, for example, two allelic genes are either identical by descent or not, but for all possible replicates of that population we can talk about the probability of them being identical by descent.

The first mating system we discuss in this section is for monoecious populations with random mating including self-fertilization. Gametic output of parents is such that there is a probability of $1/N_e$ that two gametes taken at random from a generation are from the same individual. In other words, $1/N_e$ is the probability of selfing and the gametic variance effective number varies from $N_e = N$ for equal chance of each parent contributing each gamete to $N_e = 2N - 1$ for exactly two gametes per parent.

The two non-allelic genes for which $_1\overline{F}_{(t+1)}$ is defined are descended from one member of generation t with probability $1/N_e$, and so are descended equally often from genes on one or two gametes received by that individual. With probability $1 - 1/N_e$ the two genes are descended from genes on gametes taken from any two distinct members of generation t. Such gametes have the same descent status as gametes uniting to form a member of generation t. We have argued then that

$$_1\overline{F}_{(t+1)} = \frac{1}{N_e} \frac{\overline{F}^1_{(t)} + {}_1\overline{F}_{(t)}}{2} + \left(1 - \frac{1}{N_e}\right) {}_1\overline{F}_{(t)}$$

$$= \frac{1}{2N_e} \overline{F}^1_{(t)} + \left(1 - \frac{1}{2N_e}\right) {}_1\overline{F}_{(t)}. \tag{16}$$

We use (16) to provide the well-known result

$$\overline{F}_{1_{(t+1)}} = \frac{1}{2N_e} + \left(1 - \frac{1}{2N_e}\right) \overline{F}_{1_{(t)}}, \tag{17}$$

and again equation (9) provides the transition for $\overline{F}^1$ (with all measures now written as average measures). Explicit solutions to the three simultaneous transition equations are

$$\overline{F}_{1_{(t)}} = 1 - \left(1 - \frac{1}{2N_e}\right)^t, \tag{18}$$

$$_1\overline{F}_{(t)} = \left[1 - \left(\frac{1+\lambda}{2} - \frac{1}{2N_e}\right)^t\right]\bigg/[1 + N_e(1-\lambda)], \tag{19}$$

$$\overline{F}^1_{(t)} = \left[1 + N_e(1-\lambda)\left(\frac{1+\lambda}{2} - \frac{1}{2N_e}\right)^t\right]\bigg/[1 + N_e(1-\lambda)], \tag{20}$$

assuming the initial population to consist of non-inbred and unrelated individuals ($\overline{F}_{1_{(0)}} = {}_1\overline{F}_{(0)} = 0$, $\overline{F}^1_{(0)} = 1$). These solutions may be seen more easily by obtaining separate equations for each measure. Eliminating $\overline{F}^1$ from equations (9) and (16) gives

$$_1\overline{F}_{(t+2)} = \left(1 - \frac{1}{2N_e} + \frac{1+\lambda}{2}\right){}_1\overline{F}_{(t+1)} + \left(\frac{1}{2N_e} - \frac{1+\lambda}{2}\right){}_1\overline{F}_{(t)} \tag{21}$$

or

$$_1\overline{F}_{(t+1)} = \left(\frac{1+\lambda}{2} - \frac{1}{2N_e}\right){}_1\overline{F}_{(t)} + \frac{1}{2N_e},$$

which again reduces to (17) when $\lambda = 1$. The parental coefficient also must satisfy these last two equations.

The inbreeding coefficient is seen to increase monotonically from zero to one, while the recombinant and parental coefficients approach a limiting value of

$$_1\overline{F}_{(\infty)} = \overline{F}^1_{(\infty)} = \frac{1}{1 + N_e(1-\lambda)}$$

from below and above, respectively.

Since uniting gametes are chosen at random for this monoecious mating system, $\overline{\theta}_{(t)} = \overline{F}_{(t)}$ for each of the three digenic measures.

Our final mating system is for dioecious mating with N_m males and N_f females every generation. Male and female gametes are united at random. Gametic production is such that $1/N_{em}$, $1/N_{ef}$ are the probabilities that two gametes taken at random from the group of males or the group of females, respectively, in any generation are from the same member of that group. Then,

$$\frac{1}{N_e} = \frac{1}{4}\left(\frac{1}{N_{em}} + \frac{1}{N_{ef}}\right)$$

is the probability that any member of generation $t+2$ receives two non-gametic genes from one member of generation t. The effective number N_e varies from $4N_m N_f/(N_m + N_f)$ with equal chance of any member of a group contributing any gamete from that group to $2(N-1)$ for each parent providing exactly two gametes to offspring when $N_m = N_f = N/2$.

The two non-allelic genes for which $_1\overline{F}_{(t+2)}$ is defined are now descended from one member of generation t with probability $1/N_e$, and so are descended equally often from genes on one or two gametes received by that individual. With probability $1 - 1/N_e$ the two genes are descended from genes on gametes taken from any two distinct members of generation t. Such gametes have the same descent status as gametes uniting to form a member of generation $t+1$,

$$_1\overline{F}_{(t+2)} = \frac{1}{N_e}\frac{\overline{F}^1_{(t)} + {}_1\overline{F}_{(t)}}{2} + \left(1 - \frac{1}{N_e}\right){}_1\overline{F}_{(t+1)} \tag{22}$$

with the special, and usual, case of

$$\overline{F}_{1_{(t+2)}} = \frac{1}{N_e}\frac{1+\overline{F}_{1_{(t)}}}{2} + \left(1 - \frac{1}{N_e}\right)\overline{F}_{1_{(t+1)}}. \tag{23}$$

Equation (9), for average measures, remains appropriate. Notice that the three average digenic measures apply to a random member of the whole of a generation, whether that member is male or female.

Combining equations (9) and (22) provides

$$_1\overline{F}_{(t+3)} = \left(\frac{3+\lambda}{2} - \frac{1}{N_e}\right)_1\overline{F}_{(t+2)} - \left(\frac{1+\lambda}{2} - \frac{2+\lambda}{2}\frac{1}{N_e}\right)_1\overline{F}_{(t+1)} - \frac{\lambda}{2}\frac{1}{N_e}\,_1\overline{F}_{(t)}$$

or

$$_1\overline{F}_{(t+2)} = \left(\frac{1+\lambda}{2} - \frac{1}{N_e}\right)_1\overline{F}_{(t+1)} + \frac{\lambda}{2}\frac{1}{N_e}\,_1\overline{F}_{(t)} + \frac{1}{2N_e}, \tag{24}$$

which again reduces to (23) for complete linkage. The parental coefficient must satisfy these same equations, and (24) provides the limiting values

$$_1\overline{F}_{(\infty)} = \overline{F}^1_{(\infty)} = 1/[1 + (1-\lambda)(1+N_e)].$$

For this dioecious scheme, the digenic coancestry coefficients are the same as the coefficients for uniting gametes in the following generation, $\overline{\theta}_{(t)} = \overline{F}_{(t+1)}$.

The results obtained for monoecious and dioecious populations reduce to some of those found previously (Cockerham & Weir, 1973). For $N = N_e = 1$, the results in equations (18) to (20) are just those given previously for self mating while if $N_m = N_f = 1$ and $N_e = 2$, equation (24) is the previous result for sib mating. In this treatment of monoecious and dioecious systems, we have given a very simple derivation of results and avoided the introduction of gametic set measures (Weir & Cockerham, 1969b, 1974). While the three average digenic measures $\overline{F}_1$, $_1\overline{F}$, $\overline{F}^1$ were defined for gametes identified by the offspring generation, gametic set measures also take into account the parent generation. This extra complexity is not necessary in the present case, but it does allow the discussion of more restrictive gametic sampling plans and is necessary for the trigenic and quadrigenic descent measures.

8. TWO GAMETES PER PARENT

We have made explicit mention a couple of times of inbreeding systems with two gametes per parent. This situation is interesting because it offers a basis for comparing various early and late inbreeding schemes with various random mating schemes. All such schemes have the same final value for parental and recombinant coefficients. Certainly the same value

$$_1\overline{F}_{(\infty)} = \overline{F}^1_{(\infty)} = 1/[1 + (2N-1)(1-\lambda)] \tag{25}$$

is seen to hold for the systems considered above: maximum avoidance, circular and circular pair mating and the types of monoecious and dioecious mating considered. The result can be extended to all inbreeding systems with two gametes per parent provided the population does not subdivide into reproductively isolated units. The form of (25) shows that it is also necessary for the population eventually to become completely inbred.

To prove the result in general for the non-trivial case $\lambda \neq 1$, we let $X_{(t)}$ be a function defined for non-allelic genes carried on gametes uniting to form generation t. It is the average probability of equivalence by descent for all $4N^2$ pairs of such genes. Now $2N$ pairs are gametic and equivalent by descent with probability $\overline{F}^1_{(t)}$. From now on we will use bars on the digenic measures so that expressions will apply whether or not all members of a generation have the same pedigree. A further $2N$ pairs are carried on pairs of gametes from single individuals and equivalent with probability $\overline{Y}_{(t)}$, say. This number $2N$ made use of the assumption of two gametes per parent. We write the probability of equivalence for the remaining $4N(N-1)$ pairs on pairs of gametes from distinct individuals as $\overline{Z}_{(t)}$, so that

$$X_{(t)} = \frac{1}{2N}\overline{F}^1_{(t)} + \frac{1}{2N}\overline{Y}_{(t)} + \frac{N-1}{N}\overline{Z}_{(t)}.$$

Tracing the genes back one generation gives

$$X_{(t)} = \frac{1}{2N}\left(\frac{1+\lambda}{2}\overline{F}^1_{(t-1)} + \frac{1-\lambda}{2}\,_1\overline{F}_{(t-1)}\right) + \frac{1}{2N}(\tfrac{1}{2}\overline{F}^1_{(t-1)} + \tfrac{1}{2}\,_1\overline{F}_{(t-1)}) + \frac{N-1}{N}\overline{W}_{(t-1)}$$

$$= \frac{1}{2N}\overline{F}^1_{(t-1)} + \frac{1}{2N}\,_1\overline{F}_{(t-1)} + \frac{N-1}{N}\overline{W}_{(t-1)} + \frac{\lambda}{4N}(\overline{F}^1_{(t-1)} - \,_1\overline{F}_{(t-1)})$$

where $\overline{W}_{(t)}$ refers to non-allelic genes on two gametes going to two individuals. The first three terms in this last equation are now seen to be an alternative definition for $X_{(t-1)}$, if genes are identified by the recipients rather than the donors of the gametes on which they are located. In other words

$$X_{(t)} = X_{(t-1)} + \frac{\lambda}{4N}(\overline{F}^1_{(t-1)} - \,_1\overline{F}_{(t-1)}),$$

so that

$$X_{(\infty)} = X_{(0)} + \frac{\lambda}{4N}\sum_{t=1}^{\infty}(\overline{F}^1_{(t-1)} - \,_1\overline{F}_{(t-1)}),$$

which, as (9) provides

$$\overline{F}^1_{(t)} - \overline{F}^1_{(t-1)} = -\left(\frac{1-\lambda}{2}\right)(\overline{F}^1_{(t-1)} - \,_1\overline{F}_{(t-1)}),$$

leads to

$$X_{(\infty)} = X_{(0)} + \frac{\lambda}{2N(1-\lambda)}(\overline{F}^1_{(0)} - \overline{F}^1_{(\infty)}).$$

Since $X_{(0)} = \dfrac{1}{2N}$ and $X_{(\infty)} = \overline{F}^1_{(\infty)} = \,_1\overline{F}_{(\infty)}$, then (25) follows.

If the transition equation for any particular digenic measure by itself is regarded as a difference equation, then one minus the largest root of the corresponding characteristic equation gives the final rate of approach to equilibrium for that measure (Cockerham & Weir, 1968). These rates are given in Table 2 for five mating schemes and three population sizes, and with $\lambda = 0\cdot0(0\cdot1)1\cdot0$. The values for $\lambda = 1$ have been given previously by Kimura & Crow (1963). Any system in which the first non-zero value of $_1\overline{F}$ is $\dfrac{1}{2N}$ needs care for $\lambda = 0$. In these cases, this first non-zero value is also the final value of $_1\overline{F}$ so that final rates of approach to equilibrium apply only to $\overline{F}^1$ and such cases are marked with an asterisk in Table 2.

Table 2. *Final rates of approach to equilibrium of digenic descent measures for two gametes per parent.*

						λ					
	0·0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9	1·0
						$N = 4$					
Dioecious mating	0·6666	0·5960	0·5311	0·4695	0·4102	0·3522	0·2956	0·2399	0·1849	0·1304	0·0764
Monoecious mating	0·5714	0·5314	0·4714	0·4314	0·3714	0·3314	0·2714	0·2314	0·1714	0·1314	0·0714
Circular pair mating (= maximum avoidance)	(1·0)*	0·7122	0·6088	0·5254	0·4515	0·3831	0·3185	0·2564	0·1963	0·1377	0·0804
Circular mating	(0·5)*	0·4725	0·4398	0·4025	0·3613	0·3170	0·2705	0·2223	0·1731	0·1232	0·0727
						$N = 8$					
Maximum avoidance	(1·0)*	0·6435	0·5418	0·4614	0·3907	0·3255	0·2638	0·2047	0·1473	0·0912	0·0362
Dioecious mating	0·5714	0·5141	0·4582	0·4035	0·3495	0·2961	0·2431	0·1905	0·1383	0·0863	0·0344
Monoecious mating	0·5333	0·4833	0·4333	0·3833	0·3333	0·2833	0·2333	0·1833	0·1333	0·0833	0·0333
Circular pair mating	(0·5)*	0·4711	0·4346	0·3916	0·3444	0·2945	0·2433	0·1914	0·1392	0·0870	0·0348
Circular mating	0·1465	0·1441	0·1411	0·1372	0·1321	0·1253	0·1157	0·1022	0·0832	0·0575	0·0249
						$N = 16$					
Maximum avoidance	(1·0)*	0·5970	0·5000	0·4238	0·3568	0·2948	0·2360	0·1794	0·1243	0·0702	0·0176
Dioecious mating	0·5333	0·4801	0·4275	0·3753	0·3235	0·2719	0·2205	0·1493	0·1182	0·0673	0·0164
Monoecious mating	0·5161	0·4661	0·4161	0·3661	0·3161	0·2661	0·2161	0·1661	0·1161	0·0661	0·0161
Circular pair mating	0·1465	0·1450	0·1432	0·1407	0·1371	0·1319	0·1236	0·1098	0·0874	0·0549	0·0143
Circular mating	0·0387	0·0378	0·0374	0·0370	0·0364	0·0356	0·0345	0·0326	0·0294	0·0226	0·0076

* Equilibrium is reached at first non-zero value.

9. FREQUENCIES OF DIGENIC COMBINATIONS

With the details of evaluating the digenic descent measures established, we can turn to the applications of these quantities. Data analysis is concerned with frequencies rather than probability relations and, in general, we write the population frequency of an individual formed by the union of gametes $a_i b_j$ and $a_k b_l$ as P^{ij}_{kl}. Note that $P^{ij}_{kl} = P^{kl}_{ij}$. The subscripts on a and b now refer to particular allelic types. If dots denote summation over all possible values of that suffix, then the following types of digenic frequency are obtained

$$P^{ij}_{..} = P^{ij}_{ij} + \sum_{l \neq j} P^{ij}_{il} + \sum_{k \neq i} P^{ij}_{kj} + \sum_{k \neq i, l \neq j} P^{ij}_{kl}$$

$$= \text{gametic frequency for } a_i b_j,$$

$$P^{i}_{.j} = P^{ij}_{ij} + \sum_{l \neq j} P^{il}_{ij} + \sum_{k \neq i} P^{ij}_{kj} + \sum_{k \neq i, l \neq j} P^{il}_{kj}$$

$$= \text{recombinant frequency for } a_i b_j,$$

$$P^{i}_{k.} = \sum_{j,l} P^{ij}_{kl}$$

$$= \text{zygotic frequency for } a_i a_k,$$

$$P^{.j}_{.l} = \sum_{i,k} P^{ij}_{kl}$$

$$= \text{zygotic frequency for } b_j b_l.$$

As our descent measures relate the present generation to the initial one, it is natural to relate present frequencies to initial frequencies (always denoted by script letters). Two types of initial populations are considered. The first type is considered to be formed by the random union of a specific set of $2N$ gametes. This specific set has n_{ij} gametes carrying alleles $a_i b_j$ and is characterized by

$$\mathscr{P}^{ij}_{..} = \frac{n_{ij}}{2N} = \hat{p}_i \hat{q}_j + \hat{\mathscr{D}}_{ij}, \quad \hat{p}_i = \frac{1}{2N} \sum_j n_{ij}, \quad \hat{q}_j = \frac{1}{2N} \sum_i n_{ij}.$$

These relations define $\hat{\mathscr{D}}_{ij}$ and show that specific initial frequencies carry hats. The initial ancestors are considered to be a randomization of the $2N$ gametes into pairs so that initial genotypic frequencies are not specified, and frequencies in later generations are averages over all possible pairings.

The second type of initial population is when the initial set of gametes is considered to be a random sample from an infinite randomly mating population characterized by

$$\mathscr{P}^{ij}_{..} = p_i p_j + \mathscr{D}_{ij}$$

and frequencies in later generations are averages over all possible samples of gametes as well as all possible pairings of gametes into genotypes.

The two sets of initial conditions are related in that the 'random' gametic frequencies are expected values for the 'specific' gametes. Expectation, which is also written as $\mathscr{E}$, is over all samples of gametes here and we have that

$$\mathscr{E}(\hat{p}_i) = p_i, \quad \mathscr{E}(\hat{q}_j) = q_j, \quad \mathscr{E}(\hat{\mathscr{P}}^{ij}_{..}) = \mathscr{P}^{ij}_{..},$$

$$\mathscr{E}(\hat{\mathscr{D}}_{ij}) = \frac{2N-1}{2N} \mathscr{D}_{ij}, \quad \mathscr{E}(\hat{p}_i \hat{q}_j) = p_i q_j + \mathscr{D}_{ij}/2N.$$

Other differences can be seen between the two initial conditions

$$\hat{\mathscr{P}}^i_{\cdot j} = \hat{p}_i \hat{q}_j - \hat{\mathscr{D}}_{ij}/(2N-1), \qquad \mathscr{P}^i_{\cdot j} = \mathscr{E}(\hat{\mathscr{P}}^i_{\cdot j}) = p_i q_j,$$

$$\hat{\mathscr{P}}^i_{i \cdot} = \hat{p}^2_i - \hat{p}_i(1-\hat{p}_i)/(2N-1), \quad \mathscr{P}^i_{i \cdot} = \mathscr{E}(\hat{\mathscr{P}}^i_{i \cdot}) = p^2_i,$$

$$\hat{\mathscr{P}}^i_{k \cdot} = \hat{p}_i \hat{p}_k 2N/(2N-1), \qquad \mathscr{P}^i_{k \cdot} = \mathscr{E}(\hat{\mathscr{P}}^i_{k \cdot}) = p_i p_k.$$

Now consider the population frequency of $a_i b_j$ gametes in a later generation. Not only are such frequencies averaged over initial frequencies as described above, but also they are averages over all possible replicate populations with the particular mating system. This latter averaging process allows us to make use of the descent measures, and we say with probability $\bar{F}^1$ a gamete is a copy of an initial gamete (at the two loci of interest) and with probability $1 - \bar{F}^1$ it has descended from distinct initial gametes.

For specific initial gametes, later gametic frequencies carry hats and

$$\hat{P}^{ij}_{\cdot \cdot} = \bar{F}^1 \hat{\mathscr{P}}^{ij}_{\cdot \cdot} + (1-\bar{F}^1)\hat{\mathscr{P}}^i_{\cdot j}$$

$$= \hat{p}_i \hat{q}_j + \left(\bar{F}^1 \frac{2N}{2N-1} - \frac{1}{2N-1}\right)\hat{\mathscr{D}}_{ij}. \tag{26}$$

A similar argument for the recombinant frequency leads to

$$\hat{P}^i_{\cdot j} = {}_1\bar{F}\,\hat{\mathscr{P}}^{ij}_{\cdot \cdot} + (1 - {}_1\bar{F})\hat{\mathscr{P}}^i_{\cdot j}$$

$$= \hat{p}_i \hat{q}_j + \left({}_1\bar{F} \frac{2N}{2N-1} - \frac{1}{2N-1}\right)\hat{\mathscr{D}}_{ij}. \tag{27}$$

Zygotic frequencies make use of the inbreeding coefficient in an analogous way and we recall that $\hat{\mathscr{P}}^i_{\cdot \cdot} = \hat{p}_i$ so that

$$\hat{P}^i_{i \cdot} = \bar{F}_1 \hat{p}_i + (1-\bar{F}_1)\hat{\mathscr{P}}^i_{i \cdot}$$

$$= \hat{p}^2_i + \left(\bar{F}_1 \frac{2N}{2N-1} - \frac{1}{2N-1}\right)\hat{p}_i(1-\hat{p}_i). \tag{28}$$

$$2\hat{P}^i_{k \cdot} = 2(1-\bar{F}_1)\hat{\mathscr{P}}^i_{k \cdot} = 2(1-\bar{F}_1)\,\hat{p}_i \hat{p}_k \frac{2N}{2N-1} \quad (i \neq k). \tag{29}$$

For random initial gametes we can take expectations in equations (26) to (29) or go directly to the frequencies of the infinite population from which the initial gametes are drawn to obtain

$$P^{ij}_{\cdot \cdot} = p_i q_j + \bar{F}^1 \mathscr{D}_{ij}, \quad P^i_{\cdot j} = p_i q_j + {}_1\bar{F}\mathscr{D}_{ij}.$$

$$P^i_{i \cdot} = p^2_i + \bar{F}_1 p_i(1-p_i), \quad 2P^i_{k \cdot} = 2(1-\bar{F}_1)p_i p_k \quad (i \neq k). \tag{30}$$

The last two equations have brought us back to equation (1), but now the assumptions about initial gametes needed for the application of (1) have been revealed.

Two-locus genotypic frequencies in the final population, where the descent status is $\bar{F}_{1(\infty)} = 1$, $\bar{F}^1_{(\infty)} = {}_1\bar{F}_{(\infty)}$, are given by the gametic frequencies since $P^{ij}_{ij(\infty)} = P^{ij}_{\cdot \cdot (\infty)}$, $\hat{P}^{ij}_{ij(\infty)} = \hat{P}^{ij}_{\cdot \cdot (\infty)}$. These quantities are often referred to as the probabilities of fixation of the gametes.

10. LINKAGE DISEQUILIBRIUM

We have already defined $\mathscr{D}_{ij}, \hat{\mathscr{D}}_{ij}$ for the initial gametes and now consider linkage disequilibria in later generations. An experimenter often has a single population, or line, for which he measures the linkage disequilibrium. Even so, there is more than one form of linkage or non-allelic disequilibrium. If tildes denote within-line or within-population values, there is linkage disequilibrium pertaining to $\tilde{P}^{ij}_{..}$ or genes on the same gamete, to $\tilde{P}^{i}_{.j}$ or genes on different gametes in the same individual and to $\tilde{P}^{i+}_{+j}$ or genes in different individuals. ($\tilde{P}^{i+}_{+j}$ is the frequency of $a_i b_j$ in all possible pairs of non-allelic genes between individuals within populations.) Only two components of linkage disequilibrium are estimable within populations

$$\tilde{D}^{ij}_w = \tilde{P}^{ij}_{..} - \tilde{P}^{i}_{.j} = (\tilde{P}^{ij}_{ij} - \tilde{P}^{i\bar{j}}_{ij}), \quad \bar{i} = \text{not } i, \quad j \stackrel{=}{=} \text{not } j,$$

for genes within individuals, and

$$\tilde{D}^{ij}_b = \tilde{P}^{i}_{.j} - \tilde{P}^{i+}_{+j}$$

for genes between individuals, with a total within-population value of $\tilde{D}^{ij}_w + \tilde{D}^{ij}_b$. With regular systems of mating such as maximum avoidance and circular mating, or other forms of structured populations, frequencies $\tilde{P}^{i+}_{+j}$ can be computed for each class of relatives, i.e. $\tilde{P}^{i+(1)}_{+j}, \tilde{P}^{i+(2)}_{+j}, \dots, \tilde{P}^{i+(n)}_{+j}$ in descending order of relationship and a component of disequilibrium estimated for each step of increase in relationship

$$\tilde{D}^{ij(k)} = \tilde{P}^{i+(k)}_{+j} - \tilde{P}^{i+(k+1)}_{+j} \quad (k = 1, 2, \dots, n-1)$$

and the total between individuals is $\tilde{P}^{i}_{.j} - \tilde{P}^{i+(n)}_{+j}$ or

$$\tilde{D}^{ij}_b = \tilde{P}^{i}_{.j} - \tilde{P}^{i+(1)}_{+j} + \sum_{k=1}^{n-1} \tilde{D}^{ij(k)}.$$

Most populations or groups are not so structured, however.

Since

$$\tilde{p}_i \tilde{q}_j = \frac{1}{2N} \tilde{P}^{ij}_{..} + \frac{1}{2N} \tilde{P}^{i}_{.j} + \frac{N-1}{N} \tilde{P}^{i+}_{+j},$$

the usual measure of linakge disequilibrium within populations is

$$\tilde{P}^{ij}_{..} - \tilde{p}_i \tilde{q}_j = \frac{2N-1}{2N} (\tilde{P}^{ij}_{..} - \tilde{P}^{i}_{.j}) + \frac{N-1}{N} (\tilde{P}^{i}_{.j} - \tilde{P}^{i+}_{+j})$$

$$= \frac{2N-1}{2N} \tilde{D}^{ij}_w + \frac{N-1}{N} \tilde{D}^{ij}_b,$$

which is almost the sum of the two, or the total within populations.

Theory cannot predict particular within-line values, so we take expectations over lines to obtain

$$\mathscr{E}\tilde{D}^{ij}_w = \hat{D}^{ij}_w = \hat{P}^{ij}_{..} - \hat{P}^{i}_{.j} = \frac{2N}{2N-1} (\bar{F}^1 - {}_1\bar{F})\hat{\mathscr{D}}_{ij},$$

$$\mathscr{E}\tilde{D}^{ij}_b = \hat{D}^{ij}_b = \hat{P}^{i}_{.j} - \hat{P}^{i+}_{+j} = \frac{2N}{2N-1} ({}_1\bar{F} - {}_1\bar{\theta})\hat{\mathscr{D}}_{ij} \tag{31}$$

for lines from a specific set of gametes, and

$$\mathscr{E}\tilde{D}_w^{ij} = D_w^{ij} = P_{\cdot\cdot}^{ij} - P_{\cdot j}^i = (\overline{F}^1 - {}_1\overline{F})\,\mathscr{D}_{ij},$$
$$\mathscr{E}\tilde{D}_b^{ij} = D_b^{ij} = P_{\cdot j}^i - P_{+j}^{i+} = ({}_1\overline{F} - {}_1\overline{\theta})\,\mathscr{D}_{ij} \qquad (32)$$

for lines from random samples of gametes. (Recall that ${}_1\overline{\theta}$ is the average of the recombinant coefficients for all pairs of individuals.)

With random union of gametes as in a monoecious population, ${}_1\overline{F} = {}_1\overline{\theta}$ or $P_{\cdot j}^i = P_{+j}^{i+}$ and there is only one component of linkage disequilibrium within populations. In the dioecious population considered, ${}_1\overline{\theta}_{(t)} = {}_1\overline{F}_{(t+1)}$ and there is a small component that disappears with fixation. Of course all components within lines disappear with fixation since $\overline{F}_{(\infty)}^1 = {}_1\overline{F}_{(\infty)} = {}_1\overline{\theta}_{(\infty)}$. The rates of approach to zero, being the rates at which these coefficients approach their equilibrium values, differ among types of populations however, and are much dependent on λ.

The full complement of linkage disequilibrium is estimable only with independent replicate lines or populations. With a sample of such independent populations, one can calculate the frequency $\tilde{P}_{-j}^{i-}$ of all possible pairs of non-alleles between populations leading to an additional component of disequilibrium,

$$\tilde{D}_a^{ij} = \tilde{P}_{+j}^{i+} - \tilde{P}_{-j}^{i-},$$

which has expectations,

$$\mathscr{E}\tilde{D}_a^{ij} = \hat{D}_a^{ij} = \left({}_1\overline{\theta}\,\frac{2N}{2N-1} - \frac{1}{2N-1}\right)\hat{\mathscr{D}}_{ij}$$

for lines from a specific set of gametes and

$$\mathscr{E}\tilde{D}_a^{ij} = D_a^{ij} = {}_1\overline{\theta}\,\mathscr{D}_{ij}$$

for lines from random samples of gametes. The total linkage disequilibrium, either $\hat{D}_T^{ij}$ or D_T^{ij},

$$D_T^{ij} = D_w^{ij} + D_b^{ij} + D_a^{ij}$$

includes that between and within replicate populations.

The final disequilibrium is just that between populations, since the components within populations are zero. For a specific set of initial gametes

$$\hat{D}_{T(\infty)}^{ij} = \hat{D}_{a(\infty)}^{ij} = \left({}_1\overline{\theta}_{(\infty)}\,\frac{2N}{2N-1} - \frac{1}{2N-1}\right)\hat{\mathscr{D}}_{ij} \qquad \text{general}$$

$$= \frac{N(1+\lambda)-1}{(2N-1)[1+N(1-\lambda)]}\,\hat{\mathscr{D}}_{ij} \qquad \text{monoecious equal chance}$$

$$= \frac{(N+1)(1+\lambda)-3}{(2N-1)[1+(N+1)(1-\lambda)]}\,\hat{\mathscr{D}}_{ij} \qquad \begin{array}{l}\text{dioecious equal chance}\\ N_m = N_f = N/2\end{array}$$

$$= \frac{\lambda}{1+(2N-1)(1-\lambda)}\,\hat{\mathscr{D}}_{ij} \qquad \text{two gametes per parent.} \quad (33)$$

Note in the latter case that there is no final linkage disequilibrium for genes which recombine freely. Also, this case and monoecious equal chance with $N = 1$ generates the classical result of Wright (1933) for self-fertilization. With $N = 2$, the final

disequilibrium for dioecious equal chance and two gametes per parent are the same, as they should be, for full sib mating.

For random samples of initial gametes, or the average over particular initial sets

$$D^{ij}_{T(\infty)} = D^{ij}_{a(\infty)} = {}_1\bar{\theta}_{(\infty)}\mathscr{D}_{ij} \qquad\qquad \text{general}$$

$$= \frac{1}{1+N(1-\lambda)}\,\mathscr{D}_{ij} \qquad\qquad \text{monoecious equal chance}$$

$$= \frac{1}{1+(N+1)(1-\lambda)}\,\mathscr{D}_{ij} \qquad \begin{array}{l}\text{dioecious equal chance}\\ N_m = N_f = N/2\end{array}$$

$$= \frac{1}{1+(2N-1)(1-\lambda)}\,\mathscr{D}_{ij} \quad \text{two gametes per parent.}$$

These last values for the averages over all particular initial sets of gametes may be partitioned into two components, one for lines from the same initial set

$$\mathscr{E}\hat{D}^{ij}_{a(\infty)} = \left({}_1\bar{\theta}_{(\infty)}\frac{2N}{2N-1} - \frac{1}{2N-1}\right)\mathscr{E}\hat{\mathscr{D}}_{ij}$$

$$= \left({}_1\bar{\theta}_{(\infty)} - \frac{1}{2N}\right)\mathscr{D}_{ij}$$

and one for different initial sets of gametes,

$$D^{ij}_{a(\infty)} - \mathscr{E}\hat{D}^{ij}_{a(\infty)} = \frac{1}{2N}\,\mathscr{D}_{ij}.$$

This last term is just the initial sampling covariance and is independent of the subsequent mating system.

11. DISCUSSION

We discuss implications of the results found with digenic measures first before relating this paper to the more general theory of descent measures. We do note here, however, that parental and recombinant descent coefficients permit the same simple elaboration of gametic frequencies and linkage disequilibrium as the inbreeding coefficient does for one-locus genotypic frequencies. The determination of these two coefficients closely parallels that of the inbreeding coefficient. Results have been obtained for a variety of inbreeding systems which would have required a formidable effort by usual approaches and these results have the advantage of applying to any number of alleles at a locus.

(i) *Final populations*

The final descent status of populations for two gametes per parent every generation was shown to be independent of the particular inbreeding system. With free recombination, $\lambda = 0$, and two gametes per parent moreover, the final lines from a specific set of gametes were shown to be in linkage equilibrium (equation (33)). This is a generalization of the well-known result for continued self-fertilization from a single doubly heterozygous individual.

While the final descent status does not vary with mating system for two gametes per parent, the rates of approach to final status do vary. For any descent measure F the rate of approach to equilibrium is defined as

$$r(t) = \frac{F_{(t)} - F_{(t-1)}}{F_{(\infty)} - F_{(t-1)}}$$

and the final value of this rate is $r(\infty) = 1 - R$, where R is the largest characteristic root of the transition equation for F. For monoecious mating these rates are constant over time, $r_1 = 1/2N_e$ for $\bar{F}_1$ and $r^1 = {}_1r = (1-\lambda)/2 + 1/2N_e$ for $\bar{F}^1$ and ${}_1\bar{F}$ [equations (18) to (20)], and this mating system serves as a standard for comparing the effects of mating closer or more distant relatives. As expected the effects on the recombinant coefficients are similar to the well-known (Kimura & Crow, 1963; Cockerham, 1970) effects on the inbreeding coefficient. Greater degrees of avoidance of mating relatives result in lower early values of the recombinant coefficient and higher final rates of approach to equilibrium (Table 2).

There are large differences between the rates r_1 and $r^1 = {}_1r$ unless the population size is very small or the linkage parameter is very large. For monoecious mating, for example, $r^1 = (1-\lambda)/2 + r_1$ is dominated by λ and is much greater than r_1 unless λ is close to 1.

(ii) *Initial gametes*

We considered two separate starting conditions for gametes and illustrated the differences for final linkage disequilibria. Probably the most illustrative difference is for one-locus genotypic frequencies. If we write

$$\hat{\bar{F}}_1 = \frac{2N}{2N-1}\bar{F}_1 - \frac{1}{2N-1}, \tag{34}$$

then the expressions (28), (29) for specific initial gametes become

$$\hat{P}^i_{i:(t)} = \hat{p}_i^2 + \hat{\bar{F}}_{1_{(t)}}\hat{p}_i(1-\hat{p}_i), \tag{35}$$

$$2\hat{P}^i_{k:(t)} = 2(1 - \hat{\bar{F}}_{1_{(t)}})\,\hat{p}_i\,\hat{p}_k \quad (i \neq k), \tag{36}$$

which have the same form as (30) for random initial gametes, but involve the exact allelic frequencies in the founder population.

The difference between $\bar{F}_1$ and $\hat{\bar{F}}_1$ has concerned plant geneticists for some time, particularly with reference to self-fertilization. There is doubt as to whether genotypic frequencies should be written with $\bar{F}_{1_{(0)}} = 0$, $\bar{F}_{1_{(1)}} = \frac{1}{2}, \ldots$ or with $\hat{\bar{F}}_{1_{(1)}} = 0$, $\hat{\bar{F}}_{1_{(2)}} = \frac{1}{2}, \ldots$ for selfing from a single individual ($N = 1$). The latter set requires $\hat{\bar{F}}_{1_{(0)}} = -1$ to express the initial frequencies correctly and although this appears strange, it is a valid consequence of (34). We have started out arbitrarily with $\bar{F}_{1_{(0)}} = 0$ and then allowed $\bar{F}_{1_{(t)}}$ to be dictated by the mating system. When there is a specific set of initial gametes with specified frequencies, the expressions (28), (29) or (35), (36) must be used. Hardy–Weinberg frequencies are obtained in such systems when $\hat{\bar{F}}_1 = 0$, which is the first (not initial) generation for selfing. Specific gametic

sets also constrain allelic frequencies to be integral multiples of $1/2N$ so that for one specific initial individual $\hat{p}_i = 0, \frac{1}{2}, 1$ with $\frac{1}{2}$ being the only one of interest usually.

As further illustration of the difference between the two types of initial gametes, we list $\bar{F}_1$ and $\hat{\bar{F}}_1$ for full sib mating (S), monoecious mating with equal chance gamete production (M_{cc}), and monoecious mating with two gametes per parent (M_{tg}), all with $N = 2$, in Table 3. For specific initial gametes we expect full sib mating to have Hardy–Weinberg frequencies in the second generation, M_{ec} in the first generation, and M_{tg} never actually attains the frequencies but passes through $\hat{\bar{F}}_1 = 0$ between the first and second generation. Because of the nature of the mating system, M_{ec} is always expected to have Hardy–Weinberg frequencies in the first generation for all values of N. By contrast, for random initial gametes we average over all random samples of gametes and the initial population has the same frequencies as the infinite population from which the gametes are drawn. In other words, the initial generation always has Hardy–Weinberg frequencies and we just use $\bar{F}_1$ in equation (30). This discussion amplifies the comment following equation (1).

Table 3. *Values of $\bar{F}_1$ and $\hat{\bar{F}}_1$ for three mating schemes with $N = 2$*

Mating scheme	Generation				
	0	1	2	3	4
Full sib					
$\bar{F}_1$	0	0	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{2}$
$\hat{\bar{F}}_1$	$-\frac{1}{3}$	$-\frac{1}{3}$	0	$\frac{1}{6}$	$\frac{1}{3}$
Monoecious (equal chance)					
$\bar{F}_1$	0	$\frac{1}{4}$	$\frac{7}{16}$	$\frac{37}{64}$	$\frac{175}{256}$
$\hat{\bar{F}}_1$	$-\frac{1}{3}$	0	$\frac{1}{4}$	$\frac{7}{16}$	$\frac{37}{64}$
Monoecious (two gametes per parent)					
$\bar{F}_1$	0	$\frac{1}{6}$	$\frac{11}{36}$	$\frac{91}{216}$	$\frac{671}{1296}$
$\hat{\bar{F}}_1$	$-\frac{1}{3}$	$-\frac{1}{9}$	$\frac{2}{27}$	$\frac{37}{162}$	$\frac{347}{972}$

(iii) *Initial individuals*

We have already remarked that our frequency expressions do not apply to a specific set of initial genotypic frequencies, although the random initial gametes case essentially assumes Hardy–Weinberg frequencies initially. This is in contrast to the approach to Kimura (1963). Kimura's probability method for inbreeding systems has many similarities to the digenic descent approach. In essence he worked with frequencies $P^i_{k\cdot}$, $P^{ij}_{\cdot\cdot}$ and $P^i_{\cdot j}$ with the first and last being fractionated into analogues for various categories of gametes as required. The transitional arguments for the digenic P's are, of course, identical to those for $\bar{F}_1$ and its analogues.

The difference between our approach and that of Kimura is that we have divorced the descent measure arguments from any frequency considerations and then

10

GRH 30

applied them to (two sets of) initial gametes. Kimura argues back to the exact constitution of genes in the founder individuals.

We have avoided the complication of arguing back to exactly specified founders to preserve simplicity but we can extend our measures to handle these cases. We demonstrated this earlier for sib mating (Cockerham & Weir, 1973) and now give the extension for monoecious populations.

For a monoecious population, in addition to $\bar{F}^1$ we let $\bar{F}_x^1$ be the probability that non-allelic genes on one gamete descend from genes on two gametes within one founder individual and $\bar{F}_y^1$ be the probability that they descend from two gametes in different founder individuals. Evidently $\bar{F}^1$, $\bar{F}_x^1$ and $\bar{F}_y^1$ sum to one since they are the probabilities of the only three possible descent histories of the two genes. For non-allelic and non-gametic genes we define $_1\bar{F}_x$, $_1\bar{F}_y$ in the same way with $_1\bar{F}$, $_1\bar{F}_x$ and $_1\bar{F}_y$ summing to one.

Among the founder individuals we let $\mathscr{P}^i_{\cdot j}$, $\mathscr{P}^{i+}_{+j}$ be the recombinant frequency of genes within and among individuals so that gametic population frequencies for a specific set of initial individuals are now written as

$$\hat{\bar{P}}^{ij}_{\cdot\cdot} = \bar{F}^1 \mathscr{P}^{ij}_{\cdot\cdot} + \bar{F}_x^1 \mathscr{P}^i_{\cdot j} + \bar{F}_y^1 \mathscr{P}^{i+}_{+j}. \tag{37}$$

Since $\bar{F}_x^1$, $_1\bar{F}_x$ and $\bar{F}_y^1$, $_1\bar{F}_y$ have the same joint one-step transition equations as do $\bar{F}^1$, $_1\bar{F}$, each of the six measures satisfies the same second degree homogeneous recurrence equation (21). The measures have different initial values, however, as shown in Table 4, so that (21) is satisfied for $\bar{F}^1$, $_1\bar{F}$ for $t \geqslant 0$ while $t \geqslant 1$ is required for the other four measures. The different initial conditions mean that measures satisfy different non-homogeneous recurrence equations of first order. These equations differ in their constant terms and have the following solutions

$$\bar{F}^1_{x(t)} = \frac{1-\lambda}{1+N_e(1-\lambda)}\left[1 + \frac{N_e[1-N_e(1-\lambda)]}{1-N_e(1+\lambda)} R^t\right],$$

$$_1\bar{F}_{x(t)} = \frac{1}{1+N_e(1-\lambda)}\left[(1-\lambda) - \frac{1-N_e(1-\lambda)}{1-N_e(1+\lambda)} R^t\right],$$

$$\bar{F}^1_{y(t)} = \frac{(N_e-1)(1-\lambda)}{1+N_e(1-\lambda)}\left[1 + \frac{2N_e}{1-N_e(1+\lambda)} R^t\right],$$

$$_1\bar{F}_{y(t)} = \frac{(N_e-1)}{1+N_e(1-\lambda)}\left[(1-\lambda) - \frac{2}{1-N_e(1+\lambda)} R^t\right],$$

where $R = \dfrac{1+\lambda}{2} - \dfrac{1}{2N_e}$.

Final values of the measures follow from these equations by setting R^t to zero. With a founder population of all doubly heterozygous individuals AB/ab

$$\mathscr{P}^{AB}_{\cdot\cdot} = \tfrac{1}{2}, \quad \mathscr{P}^A_{\cdot B} = 0, \quad \mathscr{P}^{A+}_{+B} = \tfrac{1}{4}$$

and equation (37) leads to

$$\hat{\bar{P}}^{AB}_{\cdot\cdot(\infty)} = \frac{2+(N_e-1)(1-\lambda)}{4[1+N_e(1-\lambda)]}$$

as obtained by Kimura (1963) and Wright (1933) with $N_e = N$. For other founder populations, we need simply modify the $\mathscr{P}$'s.

It is interesting that Kimura (1963) had the correct transitional factor, $(1+\lambda)/2 - 1/2N$, for monoecious mating, although he did not apply it to the decay of linkage disequilibrium. Much later, Galley & Curnow (1972) pointed out its correctness in contrast to $(1+\lambda)(2N-1)/4N$ found by Hill & Robertson (1966) and Karlin & McGregor (1968).

Table 4. *Initial values of descent measures for monoecious mating*

Measure	t = 0	t = 1	t = 2
$\bar{F}^1$	1	$\dfrac{1+\lambda}{2}$	$\dfrac{(1+\lambda)^2}{4}+\dfrac{1-\lambda}{4N_e}$
$_1\bar{F}$	0	$\dfrac{1}{2N_e}$	$\dfrac{1+\lambda}{4N_e}+\dfrac{2N_e-1}{4N_e^2}$
$\bar{F}^1_x$	0	$\dfrac{1-\lambda}{2}$	$\dfrac{1-\lambda^2}{4}+\dfrac{1-\lambda}{4N_e}$
$_1\bar{F}_x$	1	$\dfrac{1}{2N_e}$	$\dfrac{1-\lambda}{4N_e}+\dfrac{2N_e-1}{4N_e^2}$
$\bar{F}^1_y$	0	0	$\dfrac{1-\lambda}{2}\dfrac{N_e-1}{N_e}$
$_1\bar{F}_y$	0	$\dfrac{N_e-1}{N_e}$	$\dfrac{2N_e-1}{2N_e}\dfrac{N_e-1}{N_e}$

(iv) *Estimation of linkage disequilibrium*

The analogy between the partitioning of linkage disequilibrium into components and the partitioning of the variance of gene frequencies into components (Cockerham, 1969, 1973) may have been noted. If we let s index the gamete, n the individual and m the population, then the random variables

$$x_{mns} = 1 \text{ if the gamete carries } a_i,$$
$$= 0 \text{ otherwise};$$

$$y_{mns} = 1 \text{ if the gamete carries } b_j,$$
$$= 0 \text{ otherwise}$$

can each be analysed into three components of variance; that within individuals, between individuals within populations, and between populations. There is an exact analogy for products, xy, as for squares and products of x or of y. The details of analysis will not be repeated here, but any of the estimation procedures (Cockerham, 1973) may be utilized. The components of covariance to be estimated are the D's which are set out on page 142, along with the components of variance for random samples of initial gametes.

	Components of variance		Components of covariance
	xx	yy	xy
Between populations	$\sigma_{ai}^2 = \bar{\theta}_1 p_i(1-p_i)$	$\sigma_{aj}^2 = \bar{\theta}_1 q_j(1-q_j)$	$D_a^{ij} = {}_1\bar{\theta}\mathscr{D}_{ij}$
Between individuals	$\sigma_{bi}^2 = (\bar{F}_1 - \bar{\theta}_1)\, p_i(1-p_i)$	$\sigma_{bj}^2 = (\bar{F}_1 - \bar{\theta}_1)\, q_j(1-q_j)$	$D_b^{ij} = ({}_1\bar{F} - {}_1\bar{\theta})\,\mathscr{D}_{ij}$
Within individuals	$\sigma_{wi}^2 = (1 - \bar{F}_1)\, p_i(1-p_i)$	$\sigma_{wj}^2 = (1 - \bar{F}_1)\, q_j(1-q_j)$	$D_w^{ij} = (\bar{F}^1 - {}_1\bar{F})\,\mathscr{D}_{ij}$
Total	$\sigma_{Ti}^2 = p_i(1-p_i)$	$\sigma_{Tj}^2 = q_j(1-q_j)$	$D_T^{ij} = \bar{F}^1\mathscr{D}_{ij}$

Given the appropriate data, unbiased estimates of all the components are available. Various intraclass correlations from the components of variance are of interest

$$\bar{\theta}_1 = \frac{\sigma_a^2}{\sigma_T^2}, \quad \bar{F}_1 = \frac{\sigma_a^2 + \sigma_b^2}{\sigma_T^2}, \quad f_1 = \frac{\sigma_b^2}{\sigma_w^2 + \sigma_b^2} = \frac{\bar{F}_1 - \bar{\theta}_1}{1 - \bar{\theta}_1},$$

the far right term being the correlation for alleles within individuals within populations. In practice these correlations bear definitions of a general nature rather than just accounting for drift. The components of covariance may be viewed also as fractions of the total

$$\frac{{}_1\bar{\theta}}{\bar{F}^1} = \frac{D_a^{ij}}{D_T^{ij}}, \quad \frac{{}_1\bar{F}}{\bar{F}^1} = \frac{D_a^{ij} + D_b^{ij}}{D_T^{ij}}, \quad {}_1f = \frac{D_b^{ij}}{D_w^{ij} + D_b^{ij}} = \frac{{}_1\bar{F} - {}_1\bar{\theta}}{\bar{F}^1 - {}_1\bar{\theta}}.$$

While free of the unknown parameter $\mathscr{D}_{ij}$, they involve ratios of the parental and recombinant descent coefficients. The ratio ${}_1\bar{\theta}/\bar{F}^1$ approaches one as the populations approach fixation. The portion ${}_1f$ is reflective of the mating system just as is f_1 and both are zero with random union of gametes. Selection or other perturbing forces may affect ${}_1f$ and f_1 differently, however.

Correlations of effects by category are available

$$\rho_{aij} = \frac{D_a^{ij}}{\sigma_{ai}\sigma_{aj}} = \frac{{}_1\bar{\theta}}{\bar{\theta}_1}\, K_{ij}, \quad \rho_{bij} = \frac{D_b^{ij}}{\sigma_{bi}\sigma_{bj}} = \frac{{}_1\bar{F} - {}_1\bar{\theta}}{\bar{F}_1 - \bar{\theta}_1}\, K_{ij},$$

$$\rho_{wij} = \frac{D_w^{ij}}{\sigma_{wi}\sigma_{wj}} = \frac{\bar{F}_1 - {}_1\bar{F}}{1 - \bar{F}_1}\, K_{ij}, \quad \rho_{Tij} = \frac{D_T^{ij}}{\sigma_{Ti}\sigma_{Tj}} = \bar{F}^1 K_{ij},$$

where $K_{ij} = \mathscr{D}_{ij}/\sigma_{Ti}\sigma_{Tj}$. These involve the unknown parameter $\mathscr{D}_{ij}$, and only with historical information can one extract information other than that provided in the context of the general formulations.

Consider now the estimation of linkage disequilibrium in a single population. There is no problem in estimating D_w^{ij} and D_b^{ij} or their sum if double heterozygotes can be distinguished. Often, however, genotypic classification is such that the two types of double heterozygotes are not distinguished and in this case $\tilde{D}_w^{ij}$ and $\tilde{D}_b^{ij}$ are not separable. Dr Peter Burrows (unpublished) considered the estimator

$$\tilde{\Delta}_{ij} = \frac{N}{N-1}(\tilde{P}_{..}^{ij} + \tilde{P}_{.j}^i - 2\tilde{p}_i\tilde{q}_j)$$

which is actually the mean product (corresponding to the mean square) for between individuals in the analysis of covariance. The equivalent expressions

$$\tilde{\Delta}_{ij} = \tilde{D}_w^{ij} + 2\tilde{D}_b^{ij}$$

have expectation $$\mathscr{E}\tilde{\Delta}_{ij} = \Delta_{ij} = D_w^{ij} + 2D_b^{ij}.$$

Based only on the assumption that the observed genotypic proportions were multinomially distributed, Burrows showed that $\tilde{\Delta}$ was the 'best' unbiased estimator for Δ and produced its variance which involved three and four gene disequilibrium functions in addition to two gene disequilibria. However, if one assumes random union of gametes, then $D_b^{ij} = 0$, $\Delta_{ij} = D_w^{ij}$ and $\tilde{\Delta}_{ij}$ is an unbiased estimator of D_w^{ij} with variance

$$\frac{p_i(1-p_i)\,q_j(1-q_j)}{N-1} + \frac{(2p_i-1)\,(2q_j-1)}{2N}\,D_w^{ij} + \frac{1}{N(N-1)}\,(D_w^{ij})^2.$$

Under the assumption of random mating, there is no unique 'best' unbiased estimator of D_w^{ij}. The maximum likelihood estimator given by Hill (1974), a solution to a cubic equation, has an asymptotic variance which tends to $p_i(1-p_i)\,q_j(1-q_j)/N$ as D_w^{ij} tends to zero, which is not substantially different from that for $\tilde{\Delta}_{ij}$.

(v) *Testing hypotheses about linkage disequilibrium*

Before turning to tests of significance it is useful to identify the factors which can affect linkage disequilibrium. Note that we are considering inferences based on zygotic frequencies of individuals which survived until the frequencies were recorded.

We use the term 'random union of gametes' to imply that non-gametic genes are united at random at the time of zygote formation. In addition to non-random mating, many factors can lead to non-random union of gametes. Differential fecundity and fertility in monogamous matings and maternal–paternal incompatibilities in general can lead to non-random union of gametes even when there is random mating. Gametes are not united at random when subdivisions are amalgamated or when there is gametic or zygotic migration in single subdivisions. It was pointed out previously in this paper that separate sexes with random mating but without other perturbing forces also cause a slight but trivial deviation from random union of gametes. Different gene frequencies in the two sexes can have a large effect. All of these factors can affect D_b, and all such effects as well as zygotic selection can lead to $D_b \neq 0$. Cancellations can occur among the various effects however, to result in no net effect on D_b.

With random union of gametes and no zygotic selection, D_w is a measure of the linkage disequilibrium in gametes that form the zygotes. It is equal to one-half the difference between the frequencies of coupling and repulsion double heterozygotes. If the two types of double heterozygotes have the same survival values, then zygotic selection will not introduce a non-zero value of D_w, and the nature of the test of the hypothesis that $D_w = 0$ is not affected. The power of the test may be affected though. Otherwise, any of the factors leading to non-random union of

gametes or linkage disequilibrium in the gametic constitution of the parents can have an effect on D_w. These effects of course are not necessarily the same for D_w as for D_b.

Without zygotic selection, the manner in which gametes unite does not affect the total disequilibrium $D_w + D_b = P^{ij}_{..} - P^{i+}_{+j}$. This is because, in this case, neither $P^{ij}_{..}$ nor P^{i+}_{+j} reflects the system of gametic union. However, the effects of the factors which led to non-random union of gametes and which contributed to D_w and to D_b are now summed together into the one measure, $D_b + D_w$.

Several different hypotheses about linkage disequilibrium within a single line or population may be tested. The first three hypotheses require the complete identification of double heterozygotes so that gametic frequencies can be determined.

Random association of non-allelic genes on uniting gametes is tested by considering the hypothesis

$$H_1: D^{ij}_b = 0 \Leftrightarrow H_1: P^i_{.j} = p_i q_j.$$

Under this hypothesis, the quantity

$$\overline{p_i q_{j_1}} = \frac{2N}{2N-1} \left(\tilde{p}_i \tilde{q}_j - \frac{1}{2N} \tilde{P}^{ij}_{..} \right)$$

is unbiased for $p_i q_j$ and so is compared to $\tilde{P}^i_{.j}$ values. The unbiasedness follows from taking expectations in the identity

$$\tilde{p}_i \tilde{q}_j - \frac{1}{2N} \tilde{P}^{ij}_{..} = \frac{1}{2N} \tilde{P}^i_{.j} + \frac{N-1}{N} \tilde{P}^{i+}_{+j}.$$

Under H_1, $\tilde{P}^i_{.j}$ has expected value $p_i q_j$, while $\tilde{P}^{i+}_{+j}$ always has expected value $P^{i+}_{+j} = p_i q_j$. A chi-square goodness of fit test is performed on the two by two table with rows i, $\bar{\imath}$ and columns j, $\bar{\jmath}$. The test statistic

$$X^2_1 = 2N(\tilde{P}^i_{.j} - \overline{p_i q_{j_1}})^2 \left[\frac{1}{\overline{p_i q_{j_1}}} + \frac{1}{\tilde{p}_i - \overline{p_i q_{j_1}}} + \frac{1}{\tilde{q}_j - \overline{p_i q_{j_1}}} + \frac{1}{1 - \tilde{p}_i - \tilde{q}_j + \overline{p_i q_{j_1}}} \right]$$

is distributed as χ^2 with one degree of freedom when H_1 is true.

The hypothesis

$$H_2: D^{ij}_b + D^{ij}_w = 0 \Leftrightarrow H_2: P^{ij}_{..} = p_i q_j$$

states that non-allelic genes are associated at random within gametes. The sample gametic frequencies $\tilde{P}^{ij}_{..}$ are compared to

$$\overline{p_i q_{j_2}} = \frac{2N}{2N-1} \left(\tilde{p}_i \tilde{q}_j - \frac{1}{2N} \tilde{P}^i_{.j} \right)$$

which are unbiased estimators for $p_i q_j$ when H_2 is true. The unbiasedness follows from the same type of argument used for $\overline{p_i q_{j_1}}$. The test statistic X^2_2 is obtained from X^2_1 by replacing $\tilde{P}^i_{.j}$ with $\tilde{P}^{ij}_{..}$ and $\overline{p_i q_{j_1}}$ with $\overline{p_i q_{j_2}}$.

A test of the hypothesis that D^{ij}_w is zero is equivalent to testing the equality of double heterozygote frequencies.

$$H_3: D^{ij}_w = 0 \Leftrightarrow H_3: P^{ij}_{ij} = P^{i\bar{\jmath}}_{ij}.$$

Coupling and repulsion double heterozygotes are compared to their mean frequency to give the one degree of freedom chi-square test statistic

$$X_3^2 = \frac{N(\tilde{P}_{ij}^{ij} - \tilde{P}_{ij}^{i\bar{j}})^2}{\tilde{P}_{ij}^{ij} + \tilde{P}_{ij}^{i\bar{j}}}.$$

We can also regard this as a test of non-allelic genes being associated more strongly within gametes than between uniting gametes. When H_1 is true and there is no association between non-allelic genes on uniting gametes, H_3 becomes equivalent to H_2. It is possible for H_2 to be true when H_1 and H_3 are both false. The usual procedure would be to test H_1 and H_3 and proceed to a test of H_2 only if these two were both rejected.

Without complete classification of double heterozygotes, any test for associations between non-allelic genes must involve some composite hypothesis such as

$$H_4: D_w^{ij} + 2D_b^{ij} = 0 \Leftrightarrow H_4: P_{..}^{ij} + P_{.j}^i = 2p_i q_j.$$

As in all composite hypotheses this includes $P_{..}^{ij} = p_i q_j$ and $P_{.j}^i = p_i q_j$, but also other alternatives involving non-zero values of $(P_{..}^{ij} - p_i q_j) = -(P_{.j}^i - p_i q_j)$. Disregarding these latter alternatives, $\tilde{p}_i \tilde{q}_j$ is an unbiased estimator of $p_i q_j$ and

$$X_4^2 = \frac{N(\tilde{P}_{..}^{ij} + \tilde{P}_{.j}^i - 2\tilde{p}_i \tilde{q}_j)^2}{\tilde{p}_i(1 - \tilde{p}_i)\tilde{q}_j(1 - \tilde{q}_j)}$$

is the one degree of freedom chi-square statistic.

We are making in this treatment the usual assumptions that allow X_i^2, $i = 1, 2, 3, 4$ to be treated as chi-square variables when the null hypotheses are true. These assumptions, such as the asymptotic normality of the various sample values $\tilde{P}, \tilde{p}, \tilde{q}$, generally require large sample sizes N. A more detailed discussion of testing for disequilibria, including reference to small samples, will be presented elsewhere.

While no explicit expressions can be given, the maximum likelihood estimator of Hill (1974) involves both D_w and D_b and tests of significance involving this statistic are of composite hypotheses.

(vi) *General descent theory*

The present treatment of digenic descent measures had a threefold purpose. In the first place we wanted to elucidate the descent status of pairs of genes in finite populations and discuss the behaviour of digenic frequencies and disequilibria. If that was all it might be questioned if it was necessary to erect a structure different from that of Kimura (1963). We offer our system as part of a more general two-locus theory. A second purpose for this paper then is that it serves as a review and a clarification of our general theory. Also, the descent coefficients lead naturally into the elaboration and clarification of the components of linkage disequilibrium, whatever the cause, and serve as an interpretive basis in the case of drift alone.

Descent measures are based on the very simple notion of from which gametes, amongst an initial set of gametes, various pairs, triples or quadruples of genes, are descended. The probabilities of the various descent relations may be found by methods reminiscent of those used for the inbreeding coefficient. The measures

are then combined with frequencies for the founder gametes to provide frequencies and disequilibrium functions for various combinations of genes.

Although the basic ideas are simple, the details of evaluation do sometimes lead to obscurity, and so in the present paper we considered only two genes at a time and did not employ the gametic set measures otherwise needed for group mating schemes. Such restrictions may provide clarity but prevent the full potential of the measures being used. For example, digenic measures allow a discussion of expected values of within-line linkage disequilibria, but the full set of measures, up to quadrigenic, are necessary to evaluate variances. The full set of measures also lead to a parameterization of two-locus genotypic frequencies, and the means and variances of two-locus quantitative characters.

REFERENCES

COCKERHAM, C. CLARK (1967). Group inbreeding and coancestry. *Genetics* **58**, 89–104.

COCKERHAM, C. CLARK (1969). Variance of gene frequencies. *Evolution* **23**, 72–84.

COCKERHAM, C. CLARK (1970). Avoidance and rate of inbreeding. In *Mathematical Topics in Population Genetics* (ed. K. Kojima), pp. 104–127. New York: Springer-Verlag.

COCKERHAM, C. CLARK (1973). Analyses of gene frequencies. *Genetics* **74**, 679–700.

COCKERHAM, C. CLARK & WEIR, B. S. (1968). Sib mating with two linked loci. *Genetics* **60**, 629–640.

COCKERHAM, C. CLARK & WEIR, B. S. (1973). Descent measures for two loci with some applications. *Theoretical Population Biology* **4**, 300–330.

GALLEY, S. J. & CURNOW, R. N. (1972). The effects of finite population size and selection on the correlation between gene frequency changes at two different loci and on the amount of linkage disequilibrium. *Theoretical and Applied Genetics* **42**, 335–345.

HALDANE, J. B. S. (1949). The association of characters as a result of inbreeding and linkage. *Annals of Eugenics* **15**, 15–23.

HILL, W. G. (1974). Estimation of linkage disequilibrium in randomly mating populations. *Heredity* **33**, 229–239.

HILL, W. G. & ROBERTSON, A. (1966). The effect of linkage on limits to artificial selection. *Genetical Research* **8**, 269–294.

KARLIN, S. & McGREGOR, J. (1968). Rates and probabilities of fixation for two locus random mating finite populations without selection. *Genetics* **58**, 141–159.

KIMURA, M. (1963). A probability method for treating inbreeding systems, especially with linked genes. *Biometrics* **19**, 1–17.

KIMURA, M. & CROW, J. F. (1963). On the maximum avoidance of inbreeding. *Genetical Research* **4**, 399–415.

MALÉCOT, M. (1948). *Les mathématiques de l'hérédité*. Masson, Paris.

SHIKATA, M. (1965). A generalization of the inbreeding coefficient. *Biometrics* **21**, 665–681.

WEIR, B. S. & COCKERHAM, C. CLARK (1969a). Pedigree mating with two linked loci. *Genetics* **61**, 923–940.

WEIR, B. S. & COCKERHAM, C. CLARK (1969b). Group inbreeding with two linked loci. *Genetics* **63**, 711–742.

WEIR, B. S. & COCKERHAM, C. CLARK (1973). Mixed self and random mating at two loci. *Genetical Research* **21**, 247–262.

WEIR, B. S. & COCKERHAM, C. CLARK (1974). Behavior of pairs of loci in finite monoecious populations. *Theoretical Population Biology* **6**, 323–354.

WRIGHT, S. (1922). Coefficients of inbreeding and relationship. *American Naturalist* **56**, 330–338.

WRIGHT, S. (1933). Inbreeding and recombination. *Proceedings of the National Academy of Sciences, U.S.A.* **19**, 410–433.

WRIGHT, S. (1951). The genetical structure of populations. *Annals of Eugenics* **15**, 323–354.

APPENDIX A

General treatment of circular and circular pair mating

For circular mating, if we add equation (9) to (11)–(14), we have a set of $(n+2)$ simultaneous transition equations in $_1F$, F^1 and $_1\theta(k)$ $(k=1,2,\ldots,n)$. If these $(n+2)$ quantities are written as a vector $w_{(t)}$ then the transition equations become

$$w_{(t+1)} = \Xi w_{(t)}.$$

The characteristic equation $|\Xi - xI| = 0$ of the matrix Ξ is found to be

$$x(1+\lambda-2x)(1-2x)\,\Phi_{n-1}(2-4x) + [4x^2 - (2+3\lambda)\,x + \lambda]\,\Phi_{n-2}(2-4x)$$
$$+ (x-\lambda)\,\Phi_{n-3}(2-4x) = 0$$

where $\Phi_n(z)$ is the $n \times n$ Wolstenholme determinant

$$\Phi_n(z) = \begin{vmatrix} z & 1 & 0 & \ldots & 0 & 0 & 0 \\ 1 & z & 1 & \ldots & 0 & 0 & 0 \\ 0 & 1 & z & \ldots & 0 & 0 & 0 \\ \multicolumn{7}{c}{\dotfill} \\ 0 & 0 & 0 & \ldots & 0 & 1 & z \end{vmatrix}.$$

Using $\Phi_n(z) = z\Phi_{n-1}(z) - \Phi_{n-2}(z)$ and removing the factor $(x-1)$ the above equation reduces to the $(n+1)$th degree equation

$$(4x^2 - 2x\lambda - \lambda)\,\Phi_{n-1}(2-4x) + \lambda\Phi_{n-2}(2-4x) = 0.$$

From previous theory (Cockerham & Weir, 1968) we know that this last equation furnishes the recurrence relation for $_1F$. The first two instances are:

$$N = 2\colon x^2 - \frac{\lambda}{2}x - \frac{\lambda}{4} = 0$$

$$N = 4\colon x^3 - \frac{1+\lambda}{2}x^2 + \frac{\lambda}{16} = 0,$$

and, together with the appropriate initial values, these lead to the recurrence formulas for $_1F$ given in the text.

The mating system known as circular pair mating can be treated similarly. A population of size $N = 4n$ is imagined as being arranged in $2n$ cages on the circumference of a circle. In each generation a male is mated with a female in the cage to his right. One male and one female offspring from this mating constitute a cage in the next generation. Progress of the digenic descent measures is governed by the characteristic equation

$$(8x^3 - 4\lambda x^2 - 2\lambda x - \lambda)\,\Phi_{n-1}(2-4x) + \lambda\Phi_{n-2}(2-4x) = 0.$$