Mixed self and random mating at two loci

By B. S. WEIR

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

and C. CLARK COCKERHAM

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

(*Received* 14 *November* 1972)

SUMMARY

An infinite population practising a constant amount of selfing and random mating is studied. The effects of the mating system on two linked loci with an arbitrary number of neutral alleles are determined. Expressions are obtained for the two-locus descent measure, and hence genotypic frequencies and disequilibria functions, in any generation. The nature of the equilibrium population is deduced. The special cases of pure selfing or pure random mating and completely linked or completely unlinked loci are considered separately.

It is intended to illustrate further the evaluation and use of our recently introduced two-locus descent measures (Cockerham & Weir, 1973, referred to throughout as CW). Although several of the results derived for two-locus self and random mating models are known, some of them, especially for non-equilibrium populations, are new.

Considered separately, selfing or random mating lend themselves readily to mathematical analysis and have featured often in the literature. Mixtures of the two systems have also been considered, although discussion has generally been confined to equilibrium populations. While we do not propose to review one-locus theory, we can mention the paper of Garber (1951) which gave expressions for genotypic frequencies in all generations with two alleles at a locus.

Bennett & Binet (1956) discussed two-locus equilibrium populations for mixed self and random mating and gave expressions for genotypic frequencies. While their results contained an error they did show the existence of identity disequilibrium in the absence of linkage disequilibrium. Such a situation can occur whenever there is variation of the inbreeding among individuals, as is always the case in natural populations, and also for intermediate stages of inbreeding.

Equilibrium genotypic frequencies were given by Kimura (1958) for an arbitrary number of alleles at each locus. For two alleles per locus, Binet *et al.* (1959) exhibited some non-linear recurrence formulae which gave genotypic frequencies quite easily, but they stated that they could not extend their method to several alleles or to other mating schemes. By ignoring linkage Ghai (1964) showed how

17-2

genotypic frequencies for two alleles per locus could be found, and (1966) found the two-locus inbreeding coefficient in all generations.

The most comprehensive study to date has been given by Narain (1966). For arbitrary linkage he found the two-locus inbreeding coefficient in all generations and discussed the rate of inbreeding. For two alleles per locus he also discussed the evaluation of genotypic frequencies. As was the case with Ghai, his expressions for tth generation frequencies explicitly required frequencies for generations $0, 1, \dots, t-1$.

Another brief discussion has been given by Karlin (1969) in his review of inbreeding systems. He evaluated linkage disequilibrium and the difference in the two double heterozygote frequencies for two alleles per locus.

In this present work we derive the two-locus descent measure (CW) for self and random mating and so exhibit the descent structure of the system. Either explicit expressions or simple linear transition equations are given for the eight summary components of the measure. Genotypic frequencies are then written down, for an arbitrary number of alleles per locus, as functions of these components and the initial gametic frequencies. All known results can be recovered from this theory.

1. MATING AND LINKAGE PARAMETERS

We are concerned with a population of individuals capable both of self-fertilization and of outcrossing. We suppose that the population is infinite and that with probability s an individual chosen at random in one generation is the offspring of a single individual in the previous generation. Consequently, with probability $1-s$ it has distinct parents. The probability s of selfing is supposed to be independent of genotypes and of time.

The two loci under study may be denoted a and b with alleles a_i and b_j, respectively. They are linked to an extent λ ($0 \leqslant \lambda \leqslant 1$) so that an individual $a_1 b_1/a_2 b_2$ produces the gametic array

$$\frac{1+\lambda}{4} a_1 b_1 + \frac{1+\lambda}{4} a_2 b_2 + \frac{1-\lambda}{4} a_1 b_2 + \frac{1-\lambda}{4} a_2 b_1.$$

Disturbing forces such as mutation, migration and selection are excluded from consideration.

2. DESCENT MEASURES

The two-locus individual descent measure (CW) $\mathbf{F}(ab, a'b')$ is defined for the two gametes ab and $a'b'$ received by the individual, and it gives the probabilities of the various arrangements of a, a', b and b' on gametes in a specified initial population. For any two genes x and y we define a delta function $\delta(xy)$ such that $\delta(xy) = 1$ when x and y are copies of genes on one initial gamete and $\delta(xy) = 0$ otherwise. The components of the vector $\mathbf{F}(ab, a'b')$ are then written as $_{mn}F_{ij}^{kl}$, where

$$i = \delta(aa') \quad k = \delta(ab) \quad m = \delta(ab'),$$
$$= \delta(bb') \quad l = \delta(a'b') \quad n = \delta(a'b)$$

and the probabilities of the fifteen possible initial arrangements of the four genes are

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

There is no reason why one of the gametes received by an individual should be labelled ab and the other $a'b'$ instead of vice versa, so we have the following equivalences:

$$_{01}F^{10}_{10} = \, _{10}F^{01}_{10} = \, _{10}F^{10}_{01} = \, _{01}F^{01}_{01}, \; _{00}F^{00}_{10} = \, _{00}F^{00}_{01}, \; _{00}F^{10}_{00} = \, _{00}F^{01}_{00}, \; _{10}F^{00}_{00} = \, _{01}F^{00}_{00},$$

and the number of distinct components of $\mathbf{F}$ is seen to be nine. These nine sum to 1. For purposes of evaluating the descent measure it was found (CW) easier to work with the eight summary components

$$F^{11}_{11} = \, _{11}F^{11}_{11}, \qquad\qquad _1F^1_1 = \, _{11}F^{11}_{11} + \, _{01}F^{10}_{10},$$
$$F_{11} = \, _{11}F^{11}_{11} + \, _{00}F^{00}_{11}, \qquad F_1 = F_{11} + 2\,_{01}F^{10}_{10} + \, _{00}F^{00}_{10},$$
$$F^{11} = \, _{11}F^{11}_{11} + \, _{00}F^{11}_{00}, \qquad F^1 = F^{11} + 2\,_{01}F^{10}_{10} + \, _{00}F^{10}_{00},$$
$$_{11}F = \, _{11}F^{11}_{11} + \, _{11}F^{00}_{00}, \qquad _1F = \, _{11}F + 2\,_{01}F^{10}_{10} + \, _{10}F^{00}_{00}.$$

The expression of the original nine components in terms of these eight summary components has been given (CW).

In evaluating $\mathbf{F}$ it is necessary to trace the four genes ab, $a'b'$ for which $\mathbf{F}$ is defined back a generation when they may no longer be located on two uniting gametes. Accordingly a general descent measure $\mathbf{X}(ab, a'b')$ is defined for any four genes a, b, a' and b' and four classes of the measure are identified

$$\mathbf{F}_A = \mathbf{X}(ab, a'b' : ab, a'b' \text{ located on two distinct gametes which unite to}$$
$$\text{form individual } A),$$

$$\boldsymbol{\theta}_{BC} = \mathbf{X}(ab, a'b' : ab, a'b' \text{ located on two distinct gametes from individuals}$$
$$B \text{ and } C \text{ respectively}),$$

$$\boldsymbol{\gamma}_{B,CD} = \mathbf{X}(ab, a'b' : ab, a', b' \text{ located on three distinct gametes from individuals}$$
$$B, C \text{ and } D \text{ respectively}),$$

$$\boldsymbol{\delta}_{BC,DE} = \mathbf{X}(ab, a'b' : a, b, a', b' \text{ located on four distinct gametes from individuals}$$
$$B, C, D \text{ and } E \text{ respectively}).$$

3. TRANSITION EQUATIONS

Evaluation of the two-locus individual descent measure now proceeds according to the general algorithm established earlier (CW). Suppose A is a random member of generation $(t+1)$ and that B, B', B'' and B''' are distinct random members of generation t. Then with probability s, A has a single parent, say B, and with probability $1-s$ it has two parents, say B and B'. In all the following discussion the initial population is supposed to be an infinite random mating one so that $_{00}F^{11}_{00} = 1$ initially and all other components are zero. This gives $F^{11} = F^1 = 1$ and all other summary components zero initially.

We start with the digenic summary components. The equation for the one-locus inbreeding coefficient F_1 is well known to be

$$F_{1(t+1)} = \frac{s}{2}\{1 + F_{1(t)}\} \tag{1}$$

so that

$$F_{1(t)} = \frac{s}{2-s}\left\{1 - \left(\frac{s}{2}\right)^t\right\}. \tag{2}$$

The parental and recombinant descent coefficients, F^1 and $_1F$ respectively, need to be considered together. Note that F^1 is the probability that genes a, b on a gamete are copies of genes on one initial gamete, while $_1F$ is the probability that genes a, b, one on each of two uniting gametes, are copies of genes on one initial gamete. We first have the usual relation (CW)

$$F^1_{(t+1)} = \frac{1+\lambda}{2}F^1_{(t)} + \frac{1-\lambda}{2}{}_1F_{(t)} \tag{3}$$

and then apply the algorithm to get

$$_1F_A = s\,_1\theta_{BB} + (1-s)\,_1\theta_{BB'}$$
$$= s\{\tfrac{1}{2}F^1_B + \tfrac{1}{2}\,_1F_B\} + (1-s)\,O$$

or

$$_1F_{(t+1)} = \frac{s}{2}F^1_{(t)} + \frac{s}{2}\,_1F_{(t)}. \tag{4}$$

Equations (3) and (4) are related to those given by Bennett & Binet (1956) and by Karlin (1969) and have as solutions

$$F^1_{(t)} = \frac{1}{2D}\left\{\left(r_1 - \frac{s}{2}\right)r_1^t + \left(\frac{s}{2} - r_2\right)r_2^t\right\} \tag{5}$$

$$_1F_{(t)} = \frac{s}{4D}\{r_1^t - r_2^t\}, \tag{6}$$

where

$$D^2 = \left(\frac{1+\lambda+s}{4}\right)^2 - \frac{s\lambda}{2}, \quad r_1 = \frac{1+\lambda+s}{4} + D \quad \text{and} \quad r_2 = \frac{1+\lambda+s}{4} - D.$$

Behaviour of F^1 and $_1F$ is governed by r_1 and r_2, which are symmetric in λ and s. The cases $\lambda = 1$ and $s = 1$ must be considered separately, and throughout the discussion we will make separate mention of these cases of one locus and pure selfing respectively. For $\lambda = 1$, $r_1 = 1$, $r_2 = \frac{1}{2}s$ and $F^1_{(t)} = 1$, $_1F_{(t)} = F_{1(t)}$ for all t. When $s = 1$, $r_1 = 1$, $r_2 = \frac{1}{2}\lambda$ and

$$F^1_{(t)} = \frac{1}{2-\lambda} + \frac{1-\lambda}{2-\lambda}\left(\frac{\lambda}{2}\right)^t, \quad _1F_{(t)} = \frac{1}{2-\lambda}\left\{1 - \left(\frac{\lambda}{2}\right)^t\right\}. \tag{7}$$

When $\lambda \neq 1$, $s \neq 1$:

$$\frac{s\lambda}{2} \leqslant r_2 \leqslant \min\left(\frac{\lambda}{2}, \frac{s}{2}\right) < \frac{1}{2} < r_1 < 1 \quad \text{and} \quad r_1 > \max(s, \lambda).$$

The trigenic component $_1F_1^1$ is the probability that any three of the four genes, a, b, a' and b' on two uniting gametes are copies of genes on one initial gamete. We have

$$_1F_{1A}^1 = s\,_1\theta_{1BB}^1 + (1-s)\,_1\theta_{1BB'}^1,$$

which (CW) gives

$$_1F_{1A}^1 = s\left\{\tfrac{1}{2}\,_1F_{1B}^1 + \frac{1+\lambda}{4}F_B^1 + \frac{1-\lambda}{4}\,_1F_B\right\} + (1-s)\,O,$$

so that

$$_1F_{1(t+1)}^1 = \frac{s}{2}\,_1F_{1(t)}^1 + s\frac{1+\lambda}{4}F_{(t)}^1 + s\frac{1-\lambda}{4}\,_1F_{(t)}. \tag{8}$$

For $\lambda = 1$, $_1F_{1(t)}^1 = F_{1(t)}$ for all t. In all other cases

$$_1F_{1(t)}^1 = \frac{s}{4D}\{r_1^{t+1} - r_2^{t+1}\} - \left(\frac{s}{2}\right)^{t+1}. \tag{9}$$

Among the quadrigenic summary components, methods for calculating the two-locus inbreeding coefficient F_{11} for mating schemes such as mixed self and random mating were given by Weir & Cockerham (1969). From there we see that

$$F_{11(t+1)} = s\frac{1+\lambda^2}{4} + s\frac{1-\lambda^2}{2}F_{1(t)} + s\frac{1+\lambda^2}{4}F_{11(t)}, \tag{10}$$

which, for all values of λ and s leads to

$$F_{11(t)} = \frac{s\{2(1+\lambda^2) + s(1-3\lambda^2)\}}{(2-s)\{4-s(1+\lambda^2)\}} - \frac{2s}{2-s}\left(\frac{s}{2}\right)^t + \frac{s(3-\lambda^2)}{4-s(1+\lambda^2)}\left(s\frac{1+\lambda^2}{4}\right)^t, \tag{11}$$

agreeing with Narain (1966).

Writing the remaining three quadrigenic summary components as a vector $\bar{F}' = [F_{11}^{11}, F^{11}, F_{11}]$:

$$\bar{F}_A = s\bar{\theta}_{BB} + (1-s)\bar{\theta}_{BB'}$$

where the infinite population size gives $\bar{\theta}_{BB'}' = [0, \theta_{BB'}^{11}, 0]$. The relation between $\bar{\theta}_{BB}$ and $\bar{F}_B$ has been given previously (CW) and leads to

$$F_{11(t+1)}^{11} = s\frac{1+\lambda^2}{4}F_{11(t)}^{11} + s\frac{1-\lambda^2}{2}\,_1F_{1(t)}^1 + s\frac{(1+\lambda)^2}{8}F_{(t)}^1 + s\frac{(1-\lambda)^2}{8}\,_1F_{(t)}, \tag{12}$$

$$F_{(t+1)}^{11} = s\frac{(1+\lambda)^2}{8}\{F_{(t)}^1 + F_{(t)}^{11}\} + s\frac{1-\lambda^2}{2}\,_1F_{1(t)}^1 + s\frac{(1-\lambda)^2}{8}\{_1F_{(t)} + {}_{11}F_{(t)}\} + (1-s)\,\theta_{(t)}^{11}, \tag{13}$$

$$_{11}F_{(t+1)} = s\frac{(1+\lambda)^2}{8}\{F_{(t)}^1 + {}_{11}F_{(t)}\} + s\frac{1-\lambda^2}{2}\,_1F_{1(t)}^1 + s\frac{(1-\lambda)^2}{8}\{_1F_{(t)} + F_{(t)}^{11}\}, \tag{14}$$

where $\theta_{(0)}^{11} = \tfrac{1}{4}(1+\lambda)^2$. From these equations we see that F_{11}^{11}, the probability that all four genes a, b, a' and b' received by the individual are copies of genes on one initial gamete, may be treated separately. For $\lambda = 1$, $F_{11(t)}^{11} = F_{1(t)}$ for all t, while in all other cases:

$$F_{11(t)}^{11} = Xr_1^t + Yr_2^t - s\left(\frac{s}{2}\right)^t - (X+Y-s)\left(s\frac{1+\lambda^2}{4}\right)^t, \tag{15}$$

where
$$X\left\{r_1 - s\frac{1+\lambda^2}{4}\right\} = \frac{s}{16D}\{(1+\lambda)\,[2s(1-\lambda)+1+\lambda]\,r_1 - 2s\lambda\},$$

$$Y\left\{r_2 - s\frac{1+\lambda^2}{4}\right\} = \frac{-s}{16D}\{(1+\lambda)\,[2s(1-\lambda)+1+\lambda]\,r_2 - 2s\lambda\}.$$

Notice that setting $s = 1$ gives

$$F^{11}_{11(t)} = \frac{1}{2-\lambda} - \left(\frac{1}{2}\right)^t - \frac{\lambda}{2(2-\lambda)}\left(\frac{\lambda}{2}\right)^t - \frac{1}{2}\left(\frac{1+\lambda^2}{4}\right)^t. \tag{16}$$

For $\lambda = 1$, $\theta^{11}_{(t)} = 1$ for all t so that $F^{11}_{(t)} = 1$, $_{11}F_{(t)} = F_{1(t)}$ for all t. Note that when ab and $a'b'$ are the two gametes received by an individual, F^{11} is the probability that a and b are copies of genes on one original gamete and so are a' and b', while $_{11}F$ is the probability that a' and b are copies of genes on one original gamete as are a and b'. In other words F^{11} and $_{11}F$ are the two-locus parental and recombinant descent coefficients respectively. For $s = 1$, θ^{11} is not needed in the evaluation of these two coefficients and

$$F^{11}_{(t)} = \frac{1}{2-\lambda} - \left(\frac{1}{2}\right)^t + \frac{1-\lambda}{2-\lambda}\left(\frac{\lambda}{2}\right)^t + \left(\frac{1+\lambda^2}{4}\right)^t,$$

$$_{11}F_{(t)} = \frac{1}{2-\lambda} - \left(\frac{1}{2}\right)^t - \frac{1}{2-\lambda}\left(\frac{\lambda}{2}\right)^t + \left(\frac{1+\lambda^2}{4}\right)^t. \tag{17}$$

For all other values of λ and s we set $\theta^{11}_{(t)} = \theta^{11}_{BB'}$, $\gamma^{11}_{(t)} = \gamma^{11}_{B,B'B'}$ and $\delta^{11}_{(t)} = \delta^{11}_{BB,B'B'}$. Noting that $\gamma^{11}_{B,B'B''}$, $\delta^{11}_{BB,B'B''}$ and $\delta^{11}_{BB',B''B'''}$ are zero in the present case, the usual algorithms (CW) give

$$\begin{bmatrix} \theta^{11}_{(t+1)} \\[2mm] \gamma^{11}_{(t+1)} \\[2mm] \delta^{11}_{(t+1)} \end{bmatrix} = \begin{bmatrix} \dfrac{(1+\lambda)^2}{4} & s\dfrac{1-\lambda^2}{2} & s^2\dfrac{(1-\lambda)^2}{4} \\[2mm] \dfrac{1+\lambda}{4} & \dfrac{s}{2} & s^2\dfrac{1-\lambda}{4} \\[2mm] \dfrac{1}{4} & \dfrac{s}{2} & \dfrac{s^2}{4} \end{bmatrix} \begin{bmatrix} \theta^{11}_{(t)} \\[2mm] \gamma^{11}_{(t)} \\[2mm] \delta^{11}_{(t)} \end{bmatrix} \tag{18}$$

with
$$\gamma^{11}_{(0)} = \frac{1+\lambda}{4} \quad \text{and} \quad \delta^{11}_{(0)} = \tfrac14.$$

In all generations then

$$\theta^{11}_{(t)} = \frac{1}{16D^2}\left\{[(1+\lambda)\,r_1 - s\lambda]^2\,(r_1^2)^t + [(1+\lambda)\,r_2 - s\lambda]^2\,(r_2^2)^t + s^2\lambda(1-\lambda)\left(\frac{s\lambda}{2}\right)^t\right\}$$

$$= \left\{\frac{1}{4D}[(1+\lambda)\,r_1 - s\lambda]\,r_1^t - \frac{1}{4D}[(1+\lambda)\,r_2 - s\lambda]\,r_2^t\right\}^2. \tag{19}$$

From equations (13), (14) and (19) then we see that $F^{11}_{(t)}$ and $_{11}F_{(t)}$ are combinations of the tth powers of

$$s\frac{1+\lambda^2}{4}, \quad \frac{s\lambda}{2} = r_1 r_2, \quad s/2, \quad r_1, \quad r_2, \quad r_1^2 \quad \text{and} \quad r_2^2.$$

As a matter of interest we note that the difference between these two coefficients may be expressed as

$$F^{11}_{(t+1)} - {}_{11}F_{(t+1)} = (s\lambda/2)\{F^{11}_{(t)} - {}_{11}F_{(t)}\} + (1-s)\theta^{11}_{(t)},$$

so that F^{11} is never less than $_{11}F$, and

$$F^{11}_{(t)} - {}_{11}F_{(t)} = \left\{1 - U - V + \frac{s(1-s)(1-\lambda)t}{8D^2}\right\}\left(\frac{s\lambda}{2}\right)^t + U(r_1^2)^t + V(r_2^2)^t, \qquad (20)$$

where $\qquad U = \dfrac{1-s}{32r_1 D^3}[(1+\lambda)r_1 - s\lambda]^2, \quad V = \dfrac{-(1-s)}{32r_2 D^3}[(1+\lambda)r_2 - s\lambda]^2.$

As for computing numerical values of the descent measure, the linear transition equations (1), (3), (4), (8), (10), (12), (13), (14) and (18) together with initial values are the most useful. For $(s, \lambda) = (0\cdot0, 0\cdot0), (0\cdot5, 0\cdot5), (0\cdot5, 0\cdot9), (0\cdot9, 0\cdot5), (0\cdot9, 0\cdot9)$ and
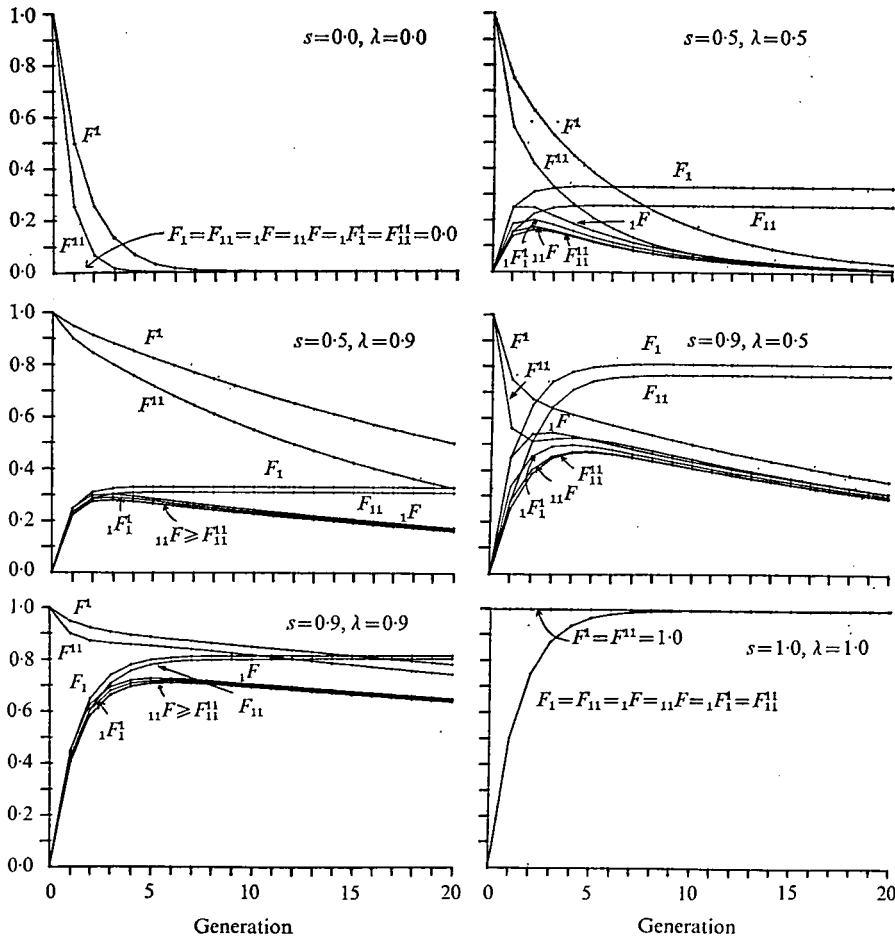
Fig. 1. Summary descent measures during the first 20 generations for $(s, \lambda) = (0\cdot0, 0\cdot0)$, $(0\cdot5, 0\cdot5), (0\cdot5, 0\cdot9), (0\cdot9, 0\cdot5), (0\cdot9, 0\cdot9), (1\cdot0, 1\cdot0)$.

(1·0, 1·0), values of the eight summary components of F for the first 20 generations are shown in Fig. 1. The explicit solutions such as those given in equations (2), (5), (6), (9), (11), (15), (19) and (20) allow such things as equilibrium values to be deduced quickly. We now apply these measures to the expression of genotypic frequencies.

4. GENOTYPIC FREQUENCIES

In our earlier discussion (CW) genotypic frequencies at two loci were expressed as functions of the descent measures and the conditions assumed for the initial ancestors. For digametic initial frequencies we used $\mathscr{P}_{kl}^{ij}$ to denote the frequency with which a random pair of gametes was $a_i b_j$ for the first gamete and $a_k b_l$ for the second,

Table 1. *Genotypic frequencies as functions of the general inbreeding measure*

	$_{11}F_{11}^{11}$	$_{00}F_{00}^{00}$	$_{00}F_{00}^{11}$	$_{11}F_{00}^{00}$	$4\,_{01}F_{10}^{10}$
P_{ij}^{ij}	p_{ij}	p_iq_j	p_{ij}^2	p_{ij}^2	$\frac{1}{2}p_{ij}(p_i+q_j)$
P_{il}^{ij}	0	0	$p_{ij}p_{il}$	$p_{ij}p_{il}$	$\frac{1}{4}(p_{ij}q_l+p_{il}q_j)$
P_{kj}^{ij}	0	0	$p_{ij}p_{kj}$	$p_{ij}p_{kj}$	$\frac{1}{4}(p_{ij}p_k+p_{kj}p_i)$
P_{kl}^{ij}	0	0	$p_{ij}p_{kl}$	$p_{il}p_{kj}$	0

	$2\,_{00}F_{10}^{00}$	$2\,_{00}F_{00}^{10}$	$2\,_{10}F_{00}^{00}$	$_{00}F_{00}^{00}$
P_{ij}^{ij}	$\frac{1}{2}p_iq_j(p_i+q_j)$	$p_{ij}p_iq_j$	$p_{ij}p_iq_j$	$p_i^2q_j^2$
P_{il}^{ij}	$\frac{1}{2}p_iq_jq_l$	$\frac{1}{2}p_i(p_{ij}q_l+p_{il}q_j)$	$\frac{1}{2}p_i(p_{ij}q_l+p_{il}q_j)$	$p_i^2q_jq_l$
P_{kj}^{ij}	$\frac{1}{2}p_ip_kq_j$	$\frac{1}{2}q_j(p_{ij}p_k+p_{kj}p_i)$	$\frac{1}{2}q_j(p_{ij}p_k+p_{kj}p_i)$	$p_ip_kq_j^2$
P_{kl}^{ij}	0	$\frac{1}{2}(p_{ij}p_kq_l+p_{kl}p_iq_j)$	$\frac{1}{2}(p_{il}p_kq_j+p_{kj}p_iq_l)$	$p_ip_kq_jq_l$

with $\mathscr{P}_{kl}^{ij} = \mathscr{P}_{ij}^{kl}$. For an individual in generation t, characterized by descent measure $\mathbf{F}_{(t)}$, the corresponding frequency is written $P_{kl(t)}^{ij}$. For a_i, a_k, b_j and b_l not alike in state, as implied by the different subscripts, four distinct initial genes are involved and these may have been arranged on two, three or four gametes. With probability $_{11}F_{00(t)}^{00}$ they were arranged $a_i b_l$ and $a_k b_j$ so that the genotypic frequency includes the term $\mathscr{P}_{kj}^{il}\,_{11}F_{00(t)}^{00}$. With this kind of argument a table (table 4 of CW) can be constructed for double homozygotes, single and double heterozygotes. Use is made also of trigametic and quadrigametic initial frequencies.

Such initial frequencies are easily found for the present case of an initial infinite random mating population (and follow from setting $N = \infty$ in table 3 of CW). If the initial frequency of gamete $a_i b_j$ is p_{ij} then, for example, $\mathscr{P}_{kl}^{ij} = p_{ij}p_{kl}$. For convenience we display each class of two-locus genotypic frequency as a function of the individual descent measure and the initial gametic frequencies in Table 1. Initial allelic frequencies for a_i and b_j are written as p_i and q_j respectively, where

$$p_i = \sum_j p_{ij}, \quad q_j = \sum_i p_{ij}.$$

These allelic frequencies are constant over time.

As we have computed the marginal components of the individual descent measure we find it more convenient to express genotypic frequencies as functions of these

marginal components. This will aid in deducing the behaviour of frequencies over time. The following expressions involve the initial linkage disequilibria Δ_{ij}, defined as the deviation of initial gametic frequencies from the products of corresponding allelic frequencies:

$$\Delta_{ij} = p_{ij} - p_i q_j.$$

Genotypic frequencies and descent measures are understood to refer to the same (tth) generation, although we often omit the t subscript.

$$
\begin{aligned}
P_{ij}^{ij} = \{p_i^2 + F_1\, p_i(1-p_i)\}\{q_j^2 + F_1 q_j(1-q_j)\} + \eta_{11} p_i(1-p_i)\, q_j(1-q_j) \\
+ \Delta_{ij}\{F_{11}^{11} + 2p_i q_j F(1) + 2(p_i + q_j) F(2)\} + \Delta_{ij}^2 F(3),
\end{aligned}
\tag{21}
$$

$$
\begin{aligned}
P_{il}^{ij} = \{p_i^2 + F_1 p_i(1-p_i)\}\{q_j q_l(1-F_1)\} - \eta_{11} p_i(1-p_i)\, q_j q_l + (q_j \Delta_{il} + q_l \Delta_{ij}) \\
\times \{p_i F(1) + F(2)\} + \Delta_{ij}\Delta_{il} F(3),
\end{aligned}
\tag{22}
$$

$$
\begin{aligned}
P_{kj}^{ij} = \{p_i p_k(1-F_1)\}\{q_j^2 + F_1 q_j(1-q_j)\} - \eta_{11} p_i p_k q_j(1-q_j) + (p_i \Delta_{kj} + p_k \Delta_{ij}) \\
\times \{q_j F(1) + F(2)\} + \Delta_{ij}\Delta_{kj} F(3),
\end{aligned}
\tag{23}
$$

$$
\begin{aligned}
P_{kl}^{ij} = p_i p_k q_j q_l(1-F_1)^2 + \eta_{11} p_i p_k q_j q_l + (p_i q_j \Delta_{kl} + p_k q_l \Delta_{ij}) F(1a) \\
+ (p_i q_l \Delta_{kj} + p_k q_j \Delta_{il}) F(1b) + \Delta_{ij}\Delta_{kl} F(3a) + \Delta_{il}\Delta_{kj} F(3b),
\end{aligned}
\tag{24}
$$

$$
\begin{aligned}
P_{kl}^{ij} + P_{kj}^{il} = 2p_i p_k q_j q_l(1-F_1)^2 + 2\eta_{11} p_i p_k q_j q_l + (p_i p_j \Delta_{kl} + p_k q_l \Delta_{ij} \\
+ p_i q_l \Delta_{kj} + p_k q_j \Delta_{il}) F(1) + (\Delta_{ij}\Delta_{kl} + \Delta_{il}\Delta_{kj}) F(3),
\end{aligned}
\tag{25}
$$

where

$$F(1a) = F^1 - 2\,_1F_1^1 + F_{11}^{11}, \quad F(1b) = \,_1F - 2\,_1F_1^1 + F_{11}^{11}, \quad F(1) = F(1a) + F(1b),$$

$$F(2) = \,_1F_1^1 - F_{11}^{11}, \quad F(3a) = F^{11} - F_{11}^{11}, \quad F(3b) = \,_{11}F - F_{11}^{11}, \quad F(3) = F(3a) + F(3b),$$

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

By taking appropriate sums of these quantities, we obtain the usual expressions for one-locus frequencies,

$$
\begin{aligned}
P_{i\cdot}^{i} &= p_i^2 + p_i(1-p_i)\, F_1, \quad & P_{k\cdot}^{i} &= p_i p_k(1-F_1), \\
P_{\cdot j}^{\cdot j} &= q_j^2 + q_j(1-q_j)\, F_1, \quad & P_{\cdot l}^{\cdot j} &= p_j q_l(1-F_1),
\end{aligned}
\tag{26}
$$

where dots denote summation over those suffices. It may be noted that the first terms in the genotypic frequencies (21)–(25) are products of these marginal locus frequencies. The usual (CW) other digenic marginal frequencies are

$$P_{\cdot\cdot}^{ij} = p_i q_j + \Delta_{ij} F^1, \tag{27}$$

$$P_{\cdot j}^{i} = p_i q_j + \Delta_{ij}\,_1F. \tag{28}$$

5. DISEQUILIBRIA FUNCTIONS

With descent measures and genotypic frequencies specified, we can now give expressions for various disequilibria functions. The deviations of digenic frequencies from products of corresponding allelic frequencies in any generation can be deduced from equations (26)–(28) and the digenic summary components in (2), (5) and (6).

In particular the linkage disequilibrium is seen to a constant multiple (its initial value) of F^1 and so behaves like this parental descent coefficient. The difference between the two two-gene non-allelic frequencies

$$P^{ij}_{..(t)} - P^i_{.j(t)} = (F^1_{(t)} - {}_1F_{(t)})\Delta_{ij} = \frac{1}{2D}\{(r_1 - s)r^t_1 + (s - r_2)r^t_2\}\Delta_{ij}, \qquad (29)$$

is always positive and disappears in time. So also does the difference in frequencies of double heterozygotes which from equation (28) we see is

$$P^{ij}_{kl(t)} - P^{il}_{kj(t)} = \{p_iq_j\Delta_{kl} + p_kq_l\Delta_{ij} - p_iq_l\Delta_{kj} - p_kq_j\Delta_{il}\}\{F^1_{(t)} - {}_1F_{(t)}\}$$
$$+ \{\Delta_{ij}\Delta_{kl} - \Delta_{il}\Delta_{kj}\}\{F^{11}_{(t)} - {}_{11}F_{(t)}\} \quad (30)$$

with values of $F^1_{(t)} - {}_1F_{(t)}$ and $F^{11}_{(t)} - {}_{11}F_{(t)}$ specified by equations (29) and (20). It will, of course, be non-zero for a time whenever there is initial linkage disequilibrium. The term $(\Delta_{ij}\Delta_{kl} - \Delta_{il}\Delta_{kj})$ is zero for only two alleles per locus.

Deviations of genotypic frequencies from products of one-locus frequencies, gametic frequencies or recombinant frequencies are also of interest. The deviations from the one-locus frequencies are given by the removal of the first term in the right-hand sides of equations (21)–(25). Deviations from gametic or recombinant frequencies are illustrated for double homozygotes:

$$P^{ij}_{ij} - P^{ij}_{..}P_{ij} = p_iq_j\{(p_i + q_j - 2p_iq_j)F_1 + (1 - p_i)(1 - q_j)F_{11}\}$$
$$+ \Delta_{ij}\{F^{11}_{11} + 2p_iq_j(F(1) - F^1) + 2(p_i + q_j)F(2)\} + \Delta^2_{ij}\{F(3) - (F^1)^2\}, \quad (31)$$

$$P^{ij}_{ij} - P^i_{.j}P^{.j}_i = p_iq_j\{(p_i + q_j - 2p_iq_j)F_1 + (1 - p_i)(1 - q_j)F_{11}\}$$
$$+ \Delta_{ij}\{F^{11}_{11} + 2p_iq_j(F(1) - {}_1F) + 2(p_i + q_j)F(2)\} + \Delta^2_{ij}\{F(3) - {}_1F^2\}. \quad (32)$$

6. SPECIAL CASES

As an aid to later discussion we now consider four special cases.

(i) *Pure selfing* ($s = 1$)

In discussing the evaluation of the descent measures we made mention of the effect of setting $s = 1$ on each component, and we have previously (CW) given genotypic frequencies and disequilibria functions for this case. For completeness we give the summary components of the individual descent measure for this case of self mating, for all t:

$$\left.\begin{aligned}
F^{11}_{11(t)} &= \frac{1}{2-\lambda} - \frac{1}{2-\lambda}\left(\frac{\lambda}{2}\right)^t - \left(\frac{1}{2}\right)^t + \frac{1}{2}\left(\frac{1+\lambda^2}{4}\right)^t, \quad &{}_1F^1_{1(t)} &= \frac{1}{2-\lambda} - \frac{1}{2-\lambda}\left(\frac{\lambda}{2}\right)^t - \left(\frac{1}{2}\right)^{t+1}, \\
F_{11(t)} &= 1 - \left(\frac{1}{2}\right)^{t-1} + \left(\frac{1+\lambda^2}{4}\right)^t, \quad &F_{1(t)} &= 1 - \left(\frac{1}{2}\right)^t, \\
F^{11}_{(t)} &= \frac{1}{2-\lambda} + \frac{1-\lambda}{2-\lambda}\left(\frac{\lambda}{2}\right)^t - \left(\frac{1}{2}\right)^t + \left(\frac{1+\lambda^2}{4}\right)^t, \quad &{}_1F_{(t)} &= \frac{1}{2-\lambda} - \frac{1}{2-\lambda}\left(\frac{\lambda}{2}\right)^t, \\
{}_{11}F_{(t)} &= \frac{1}{2-\lambda} - \frac{1}{2-\lambda}\left(\frac{\lambda}{2}\right)^t - \left(\frac{1}{2}\right)^t + \left(\frac{1+\lambda^2}{4}\right)^t, \quad &F^1_{(t)} &= \frac{1}{2-\lambda} + \frac{1-\lambda}{2-\lambda}\left(\frac{\lambda}{2}\right)^t.
\end{aligned}\right\}$$
$$(33)$$

(ii) *Random mating* $(s = 0)$

When mating is entirely at random $(s = 0)$, only F^{11} and F^1 among the marginal components of **F** are non-zero. For all t they are given by

$$F^{11}_{(t)} = \left(\frac{1+\lambda}{2}\right)^{2t}, \quad F^1_{(t)} = \left(\frac{1+\lambda}{2}\right)^t. \tag{34}$$

Gametic frequencies are given by

$$P^{ij}_{..(t)} = p_i q_j + \Delta_{ij}\left(\frac{1+\lambda}{2}\right)^t.$$

Two-locus genotypic frequencies are products of corresponding gametic frequencies, but differ from the product of one-locus frequencies. For example,

$$P^{ij}_{ij(t)} - P^{i\cdot}_{i\cdot(t)} P^{\cdot j}_{\cdot j(t)} = \Delta_{ij}\left(\frac{1+\lambda}{2}\right)^t\left\{2p_i q_j + \Delta_{ij}\left(\frac{1+\lambda}{2}\right)^t\right\}. \tag{35}$$

(iii) *Complete linkage* $(\lambda = 1)$

One-locus models follow from setting $\lambda = 1$. We pointed out the effects of this on the descent measures above: $F^{11} = F^1 = 1$ for all t, while all other summary components have the value of

$$F_{1(t)} = \frac{s}{2-s}\left\{1 - \left(\frac{s}{2}\right)^t\right\}.$$

Genotypic frequencies follow from equations (26).

(iv) *No linkage* $(\lambda = 0)$

For free recombination we set $\lambda = 0$ and present the individual descent measures for all generations after the initial one $(t \geqslant 1)$:

$$\left.\begin{aligned}
F^{11}_{11(t)} &= \frac{s(1+2s)}{2(2+s)}\left(\frac{1+s}{2}\right)^t - s\left(\frac{s}{2}\right)^t + \frac{3s}{2(2+s)}\left(\frac{s}{4}\right)^t, \\[6pt]
F_{11(t)} &= \frac{s(2+s)}{(2-s)(4-s)} - \frac{2s}{2-s}\left(\frac{s}{2}\right)^t + \frac{3s}{4-s}\left(\frac{s}{4}\right)^t, \\[6pt]
F^{11}_{(t)} &= \frac{s(1+2s)}{2(2+s)}\left(\frac{1+s}{2}\right)^t - s\left(\frac{s}{2}\right)^t + \frac{s(7+7s+4s^2)}{2(2+s)(1+s+s^2)}\left(\frac{s}{4}\right)^t \\
&\qquad\qquad\qquad + \frac{(1-s)(2+3s+2s^2)}{8(1+s+s^2)}\left(\frac{1+s}{2}\right)^{2t-2}, \\[6pt]
{}_{11}F_{(t)} &= \frac{s(1+2s)}{2(2+s)}\left(\frac{1+s}{2}\right)^t - s\left(\frac{s}{2}\right)^t + \frac{s(7+7s+4s^2)}{2(2+s)(1+s+s^2)}\left(\frac{s}{4}\right)^t + \frac{s(1-s)}{8(1+s+s^2)}\left(\frac{1+s}{2}\right)^{2t-2}, \\[6pt]
{}_1F^1_{1(t)} &= \frac{s}{2}\left(\frac{1+s}{2}\right)^t - \left(\frac{s}{2}\right)^{t+1}, \quad F_{1(t)} = \frac{s}{2-s} - \frac{s}{2-s}\left(\frac{s}{2}\right)^t, \\[6pt]
F^1_{(t)} &= \frac{1}{2}\left(\frac{1+s}{2}\right)^{t-1}, \quad {}_1F_{(t)} = \frac{s}{2}\left(\frac{1+s}{2}\right)^{t-1}.
\end{aligned}\right\} \tag{36}$$

To evaluate genotypic frequencies we use the quantities

$$(1 - F_1)^2 + \eta_{11} = \frac{4(1-s)}{4-s} + \frac{3s}{4-s}\left(\frac{s}{4}\right)^t,$$

$$F(1) = \frac{2(1-s)}{2+s}\left(\frac{1+s}{2}\right)^t + \frac{3s}{2+s}\left(\frac{s}{4}\right)^t,$$

$$F(2) = \frac{s(1-s)}{2(2+s)}\left(\frac{1+s}{2}\right)^t + \left(\frac{s}{2}\right)^{t+1} - \frac{3s}{2(2+s)}\left(\frac{s}{4}\right)^t,$$

$$F(3) = \frac{s(2+s)}{1+s+s^2}\left(\frac{s}{4}\right)^t + \frac{1-s}{1+s+s^2}\left(\frac{1+s}{2}\right)^{2t}.$$

7. EQUILIBRIUM POPULATION

It is now a straightforward matter to present the structure of a mixed self and random mating population which has been established a long time and equilibrium reached.

In the situation where neither λ nor s is 1 or 0, equations (2), (5), (6), (9), (11), (13), (14), (15) and (19) yield

$$\hat{F}_{11} = \frac{s\{2(1+\lambda^2)+s(1-3\lambda^2)\}}{(2-s)\{4-s(1+\lambda^2)\}}, \quad \hat{F}_1 = \frac{s}{2-s}, \quad \hat{\eta}_{11} = \frac{4s(1-s)\{1+\lambda^2(1-s)\}}{(2-s)^2\{4-s(1+\lambda^2)\}}, \quad (37)$$

where hats denote equilibrium values, and all other summary components are zero. The contribution of initial linkage disequilibrium to genotypic frequencies disappears and equations (21)–(24) reduce to those given by Kimura (1958). Deviations of the two-locus frequencies from products of one-locus frequencies are multiples of $\hat{\eta}_{11}$, and are given by the second terms in equations (21)–(24).

For $s = 1$ equations (33) yield:

$$\hat{F}_{11} = \hat{F}_1 = 1; \quad \hat{F}_{11}^{11} = \hat{F}^{11} = {}_{11}\hat{F} = {}_1\hat{F}_1^1 = \hat{F}^1 = {}_1\hat{F} = 1/(2-\lambda). \quad (38)$$

Only double homozygotes remain in this case of self mating, and have frequency

$$\hat{P}_{ij}^{ij} = p_i p_j + \frac{\Delta_{ij}}{2-\lambda}.$$

For $s = 0$, equations (37) hold and all genotypic frequencies reduce to the products of corresponding allelic frequencies.

For $\lambda = 1$, from above:

$$\hat{F}_{11} = \hat{F}_1 = \frac{s}{2-s}; \quad \hat{F}_{11}^{11} = \hat{F}^{11} = {}_{11}\hat{F} = {}_1\hat{F}_1^1 = \hat{F}^1 = {}_1\hat{F} = 1. \quad (39)$$

Genotypic frequencies in this one-locus case are given by

$$\hat{P}_{ij}^{ij} = p_i q_j\left\{\frac{s}{2-s} + \frac{1-s}{2-s}p_i q_j\right\}, \quad \hat{P}_{ii}^{ij} = p_i^2 q_j q_l \frac{1-s}{2-s},$$

$$\hat{P}_{kj}^{ij} = p_i p_k q_j^2 \frac{1-s}{2-s}, \qquad\qquad \hat{P}_{kl}^{ij} = p_i p_k q_j q_l \frac{1-s}{2-s}.$$

For $\lambda = 0$, equations (37) are relevant and genotypic frequencies follow from equations (21)–(24). The final identity disequilibrium now has the value of

$$\hat{\eta}_{11} = 4s(1-s)/(2-s)^2(4-s)$$

for free recombination.

8. DISCUSSION

By employing two-locus descent measures (CW) we have been able to find the descent relations in an infinite population undergoing mixed self and random mating. The two-locus structure of the population as shown by genotypic frequencies and disequilibria functions follows from the descent measures.

For each of the eight summary components of the individual descent measure a linear difference equation was found. When the equation is of order k, the component can be expressed as the sum of the tth powers of the k roots of the corresponding complementary equation in the tth generation. Treating the eight components in turn:

$$F_1 \quad \text{order 1, root } \tfrac{1}{2}s,$$
$$F^1, {}_1F \quad \text{order 2, roots } r_1 > r_2,$$
$${}_1F_1^1 \quad \text{order 3, roots } r_1 > \tfrac{1}{2}s \geqslant r_2,$$
$$F_{11} \quad \text{order 2, roots } \tfrac{1}{2}s \geqslant \tfrac{1}{4}s(1+\lambda^2),$$
$$F_{11}^{11} \quad \text{order 4, roots } r_1 > \tfrac{1}{2}s \geqslant \tfrac{1}{4}s(1+\lambda^2) \geqslant r_2,$$
$$F^{11}, {}_{11}F \quad \text{order 7, roots } r_1 \geqslant r_1^2 > \tfrac{1}{2}s \geqslant \tfrac{1}{4}s(1+\lambda^2) \geqslant r_2 \geqslant r_1 r_2 \geqslant r_2^2.$$

These relations between the roots are of particular help in explaining the long-term behaviour of the components. The two inbreeding components F_1 and F_{11} soon act like a function of $\tfrac{1}{2}s$ only. As this quantity has an upper bound of 0·5 we can deduce that inbreeding equilibrium will be reached in a few generations – certainly F_1, F_{11} will be less than 0·001 from their final values after ten generations. The remaining six summary components all have long-term behaviour determined by r_1^t. For s and λ large but not equal to 1·0, r_1 can be very close to 1·0 so that decay to equilibrium values of zero can be extremely slow. These trends are illustrated in Fig. 1. After about ten generations F_1 and F_{11} have reached their final values and the other six curves have a common slope of $-r_1$. For $\lambda = 1·0$ or $s = 1·0$, $r_1 = 1·0$ and transient behaviour is determined by $\tfrac{1}{2}s$ or $\tfrac{1}{2}\lambda$ for every component, so that equilibrium is reached quickly.

Early behaviour of the components is quite complex, as suggested by the high order of some of them and as shown by Fig. 1. The figure does show that

$$F_1 \geqslant F_{11} \geqslant {}_1F \geqslant {}_1F_1^1 \geqslant {}_{11}F \geqslant F_{11}^{11}$$

and that

$$F^{11} \geqslant {}_{11}F, \quad F^1 \geqslant {}_1F.$$

Until equilibrium then, even with free recombination, parental gametes remain more frequent than recombinant gametes $(F^1 \geqslant {}_1F)$ while for non-gametic genes, alleles are more likely to have been on one original gamete than are non-alleles

($F_1 \geqslant {}_1F$). For $\lambda \neq 1\cdot0$ the parental descent coefficients F^1, F^{11} drop down below the values of the inbreeding coefficients F_1, F_{11} at a time which increases with increasing outcrossing and linkage. Although F^1 drops monotonically from $1\cdot0$–$0\cdot0$, F^{11} has complex behaviour in the first few generations. For $s \neq 1$, $\lambda \neq 1$ original linkage blocks are finally broken up (Geiringer, 1944), while for $s = 1$ any two non-allelic genes received by an individual have probability $1/(2-\lambda)$ of having been on the same gamete originally. Original gametes remain intact of course for $\lambda = 1$.

For any particular generation then, each summary component of the individual descent measure is an increasing function of both s and λ. For any particular s and λ the behaviour over time requires more careful statements.

The equilibrium population has been discussed by many authors previously. Most of the characteristics of this population are related to the final value $\hat{\eta}_{11}$ of the identity disequilibrium, which we might also term the equilibrium constant. From definition it relates the two-locus inbreeding coefficient to the square of the one-locus coefficient, and from equations (21)–(24) it relates two-locus genotypic frequencies to products of one-locus frequencies. Discussion of the behaviour of the equilibrium constant as s and λ vary is aided by the following subdivision of $\hat{\eta}_{11}$ into two components.

In all of the present work, arguments have applied to individuals chosen at random from a generation. Although the history of each particular individual in the equilibrium population is unknown, the value $\hat{\eta}_{11}$ applies to a random member of that population, and consequently for any class of individuals. From other work involving selection (Cockerham & Rawlings, 1967) it was found useful to consider the population subdivided into classes according to the number of generations of self-fertilization. At equilibrium the frequency of individuals resulting from i generations of self-fertilization is $(1-s)s^i$ ($i = 0, 1, 2, \ldots, \infty$), for which

$$\hat{\eta}_{11i} = \left(\frac{1+\lambda^2}{4}\right)^i - \left(\frac{1}{4}\right)^i.$$

The average identity disequilibrium over classes

$$\bar{\hat{\eta}}_{11} = \frac{4s(1-s)\lambda^2}{(4-s)[4-s(1+\lambda^2)]}$$

can be alternatively formulated as

$$\bar{\hat{\eta}}_{11} = \bar{\hat{F}}_{11} - (\overline{\hat{F}_1^2})$$

which is the average of the two-locus inbreeding coefficient over classes minus the average of the square of the one-locus coefficients. By contrast

$$\hat{\eta}_{11} = \bar{\hat{F}}_{11} - (\bar{\hat{F}}_1)^2$$

so that

$$\hat{\eta}_{11} = \bar{\hat{\eta}}_{11} + \sigma_{F_1}^2.$$

The variance among the F_{1i}'s, $\sigma_{F_1}^2 = (\overline{\hat{F}_1^2}) - (\bar{\hat{F}}_1)^2$ given by

$$\sigma_{F_1}^2 = \frac{4s(1-s)}{(2-s)^2(4-s)}$$

is independent of linkage of course. Thus the equilibrium constant can be divided into two parts, one due entirely to the variation among F_1's and the other summarizing the effects of linkage. The application to genotypic frequencies is direct and points up a general phenomenon. When a population is structured into classes or subdivisions there will be genotypic disequilibrium in the population as a whole although there is none within each class.

It is now evident that the equilibrium constant is an increasing function of λ, with a maximum effect of linkage

$$\frac{\eta_{11}\,(\lambda = 1)}{\eta_{11}\,(\lambda = 0)} = \frac{4-s}{2}$$

of something less than doubling it. For any particular λ, there is an amount of selfing which maximizes $\hat{\eta}_{11}$. This value of s varies within very narrow bounds of 0·6946 for $\lambda = 0$ and 0·6667 for $\lambda = 1$. The corresponding maximum equilibrium constants are 0·1506 and 0·2500.

Final values of all disequilibrium functions considered here depend entirely on the inbreeding coefficients, generally via $\hat{\eta}_{11}$. Because of the presence of other summary components of the descent measure though, the rate of approach to these final values can be quite slow. In particular, for tightly linked loci in highly self-fertilizing species, final values of disequilibria will be attained very slowly.

This study provides a basis for the study of many natural and experimental plant populations. Departures from these results will indicate the presence of some disturbing forces.

In order that an investigator can make use of these results he must have independent knowledge of the two parameters, s and λ, used. There are standard methods for estimating these. For natural populations which are assumed to be in equilibrium then, observed genotypic frequencies can be compared to the equilibrium values given in this paper. Two-locus frequencies will differ from products of one-locus frequencies by the second terms in equations (21)–(24) if the assumptions made in this paper hold. There should be no linkage disequilibrium and, if they can be identified, both types of double heterozygote should be equally frequent.

For experimental or other non-equilibrium populations, this paper will be of most help when the investigator knows the nature of the initial population, and the number of generations since that initial population, or better yet, has information over time. In such a case he can again compare observed genotypic frequencies and/or their changes over time with those predicted in equations (21)–(24). In the absence of such knowledge, this paper offers only a qualitative guide to the behaviour of populations in the absence of disturbing forces. While inbreeding coefficients, and consequently homozygosity, reach final values quickly we have shown that descent measures, and consequently disequilibria, may reach final values very slowly. Disequilibria of various types may be observed then even in the absence of disturbing forces.

Methods of identifying disturbing forces are needed. It is to be stressed that results in this paper are for neutral genes at two loci in a population for which all other gene

18   GRH 21

variation is neutral. It does not follow that the changes in frequencies of identifiable genes or of disequilibria functions among them guarantee that these genes are the causes of the changes. For recently constituted populations selection is the most likely candidate for effecting any changes but selection operates to affect the frequencies of all genes. While population (census) size may be very large, selection would tend to make the population in time stem from a few, even one maybe, initial founders when there is a high degree of self-fertilization. Consequently, all identifiable genes would be expected to have changes in frequencies whether they played any role in selection or not. Further, as shown by Cockerham & Rawlings (1967), selection affects genotypic proportions at all loci including neutral genes. If the net effects of all genes undergoing selection is to produce an inbreeding depression then the frequencies of all heterozygotes are greater than that expected on the basis of neutral theory. Such would be the case for partially to completely recessive deleterious genes in mutation selection balance, with a concomitant effect on all genes including neutral ones.

The problem of distinguishing between *real* and *apparent selection effects* or between *real* and *apparent heterotic effects* must be solved before we can conclude anything about the role of identifiable genes from frequency data on primarily self-fertilizing species.

REFERENCES

BENNETT, J. H. & BINET, F. E. (1956). Association between Mendelian factors with mixed selfing and random mating. *Heredity* **10**, 51–55.

BINET, F. E., CLARK, A. M. & CLIFFORD, H. T. (1959). Correlation due to linkage in certain wild plants. *Genetics* **44**, 5–13.

COCKERHAM, C. CLARK & RAWLINGS, J. O. (1967). Apparent heterosis of a neutral gene with inbreeding. *Ciencia E Cultura* **19**, 89–94.

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

GARBER, M. J. (1951). Approach to genotypic equilibrium with varying percentages of self-fertilization and cross-fertilization (one autosomal gene-pair). *Journal of Heredity* **42**, 299–300.

GEIRINGER, H. (1944). On the probability theory of linkage in Mendelian heredity. *Annals of Mathematical Statistics* **15**, 25–57.

GHAI, G. L. (1964). The genotypic composition and variability in plant populations under mixed self-fertilization and random mating. *Journal of the Indian Society of Agricultural Statistics* **16**, 94–125.

GHAI, G. L. (1966). Loss of heterozygosity in populations under mixed random mating and selfing. *Journal of the Indian Society of Agricultural Statistics* **18**, 73–81.

KARLIN, S. (1969). *Equilibrium Behaviour of Population Genetic Models with Non-random Mating.* New York: Gordon and Breach.

KIMURA, M. (1958). Zygotic frequencies in a partially self-fertilizing population. *Annual Report of the National Institute of Genetics, Japan* **8**, 104–105.

NARAIN, P. (1966). Effect of linkage on homozygosity of a population under mixed selfing and random mating. *Genetics* **54**, 303–314.

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