Definition and Properties of Disequilibrium Statistics for Associations Between Nuclear and Cytoplasmic Genotypes

Marjorie A. Asmussen, Jonathan Arnold[1] and John C. Avise

Department of Genetics, University of Georgia, Athens, Georgia 30602

Manuscript received August 18, 1986
Revised copy accepted January 21, 1987

ABSTRACT

We define and establish the interrelationships of four components of statistical association between a diploid nuclear gene and a uniparentally transmitted, haploid cytoplasmic gene: an allelic (gametic) disequilibrium (D), which measures associations between alleles at the two loci; and three genotypic disequilibria (D_1, D_2, D_3), which measure associations between two cytotypes and the three respective nuclear backgrounds. We also consider an alternative set of measures, including D and the residual disequilibrium (d). The dynamics of these disequilibria are then examined under three conventional models of the mating system: (1) random mating; (2a) assortative mating without dominance (the "mixed-mating model"); and (2b) assortative mating with dominance ("O'DONALD's model"). The trajectories of gametic disequilibria are similar to those for pairs of unlinked nuclear loci. The dynamics of genotypic disequilibria exhibit a variety of behaviors depending on the model and the initial conditions. Procedures for statistical estimation of cytonuclear disequilibria are developed and applied to several real and hypothetical data sets. Special attention is paid to the biological interpretations of various categories of allelic and genotypic disequilibria in hybrid zones. Genetic systems for which these statistics might be appropriate include nuclear genotype frequencies in conjunction with those for mitochondrial DNA, chloroplast DNA, or cytoplasmically inherited microorganisms.

BECAUSE mitochondrial DNA (mtDNA) is cytoplasmically housed, and maternally inherited in most animals and many plants, it can be discussed as an asexual, haploid genome within otherwise sexually reproducing, diploid species (AVISE 1986; BIRKY, MARUYAMA and FUERST 1983; NEIGEL and AVISE 1986; TAKAHATA and SLATKIN 1983). Other cytoplasmic genomes include chloroplast DNA in diploid plants (BIRKY, MARUYAMA and FUERST 1983; CURTIS and CLEGG 1984; DEWEY, LEVING and TIMOTHY 1986), and certain intracellular microorganisms in metazoa (WADE and STEVENS 1985; HOFFMAN, TURELLI and SIMMONS 1986). Frequencies of nuclear and cytoplasmic genotypes (*e.g.*, from restriction site maps) are currently being gathered for many organisms. It is important to ask under what biological conditions departures from random association between nuclear and cytoplasmic genotypes might exist, and to develop statistics to describe such disequilibria. Although we will specifically couch discussion in terms of mtDNA (because more is known about this cytoplasmic genome), most statements or results should also apply to other cytoplasmic-nuclear associations.

Apart from historical sampling of gametes in finite populations, two other classes of phenomena could in principle generate genetic disequilibria between nuclear and cytoplasmic genotypes:

1. Epistatic effects on fitness. Functionally, the nuclear and mitochondrial genomes are interdependent in complex ways (BROWN 1983; GRIVELL 1983; MULLER *et al.* 1984). Most of the structural and functional proteins of mitochondria are encoded by nuclear genes, including more than 90 proteins required to form mitochondrial ribosomes, the RNA subunits of which are encoded by mtDNA (O'BRIEN *et al.* 1980). Probably all mitochondrially encoded proteins form components of metabolic pathways or enzyme complexes whose remaining constituents are nuclear-encoded (BROWN 1983; CHOMYN *et al.* 1985). As noted by GRIVELL (1983), "the mitochondrial genetic system is maintained only through a considerable investment on the part of the nucleus and the cell's own protein-synthesizing machinery." In return, mitochondria are the sites of oxidative phosphorylation, the main source of cellular energy. The varied interactions between products of nuclear and mitochondrial genotypes could provide many opportunities for epistatic interactions on fitness, and hence for cytonuclear disequilibria.

2. Nonrandom mating. For mtDNA and nuclear DNA, which often exhibit uniparental and biparental transmission, respectively, to what extent can nonrandom mating generate cytonuclear disequilibria? For example, some secondary hybrid zones represent rather extreme situations of nonrandom mating in

[1] To whom correspondence and reprints should be addressed. This paper is number IV in the series "Statistics of Natural Populations."

which various degrees of association between nuclear and mitochondrial genotypes have been observed (AVISE *et al.* 1984; FERRIS *et al.* 1983; LAMB and AVISE 1986; SPOLSKY and UZZELL 1984). Furthermore, it is of interest to consider the effects on disequilibria of directionalities in hybridization—that is, situations in which hybrid mating propensities are different for males and females of a given species (LAMB and AVISE 1986).

Much attention has previously been given to the description of disequilibria between nuclear genes (*e.g.*, CHARLESWORTH and CHARLESWORTH 1973; CLEGG, KIDWELL and HORCH 1980; HILL 1974; LANGLEY, TOBARI and KOJIMA 1974; SMOUSE 1974; WEIR 1979), and some progress has also been made in the mathematical analysis of cytonuclear interactions (WATSON and CASPARI 1960; CLARK 1984, 1985; GREGORIUS and ROSS 1984; ROSS and GREGORIUS 1985). Here we define and examine the properties of several additional disequilibrium measures for haploid-diploid genome associations, and present their application to some real and hypothetical data sets. This treatment will lay the foundation for later analyses of the dynamical behavior of cytonuclear disequilibria in special situations.

MEASURES OF CYTONUCLEAR DISEQUILIBRIA

There are a number of ways by which the statistical association between a nuclear and cytoplasmic locus could be measured (WEIR and WILSON 1986). We introduce several measures which arise naturally in a wide class of biological models. Consider a diploid population, which has two alleles A and a at a nuclear locus and two other alleles M and m at, for example, a mitochondrial locus. There are six possible genotypes with frequencies as given in Table 1.

The mitochondrial genotypes M and m have frequencies x and y; the nuclear locus has genotype frequencies u, v, and w. At the level of genotypes, nuclear-cytoplasmic disequilibria can be measured by the departures of genotypic frequencies from expectations under random association. Define the *genotypic disequilibrium* for AA/M to be D_1 = freq. (AA/M) − freq.(AA)freq.(M), or equivalently, $D_1 = u_1 - ux$. Altogether there are three such measures

$$D_1 = u_1 - ux; \tag{1a}$$

$$D_2 = v_1 - vx; \tag{1b}$$

$$D_3 = w_1 - wx. \tag{1c}$$

The frequencies of genotypes can then be written in terms of these three genotypic disequilibria as in Table 2. Since the genotypic disequilibria must allow the genotypic frequencies to sum to the marginal values u, v, w, x, and y, then

$$D_3 = -(D_1 + D_2), \tag{2}$$

TABLE 1

Genotypic frequencies

	Nuclear genotype			
Cytoplasm	*AA*	*Aa*	*aa*	Total
M	u_1	v_1	w_1	x
m	u_2	v_2	w_2	y
Total	u	v	w	1.0

TABLE 2

Genotypic disequilibria

	AA	*Aa*	*aa*	Total
M	$ux + D_1$	$vx + D_2$	$wx + D_3$	x
m	$uy - D_1$	$vy - D_2$	$wy - D_3$	y
Total	u	v	w	1.0

and there is no need to explicitly define disequilibria for the m genotypes. For valid genotypic frequencies (nonnegative and no more than one), there are additional bounds on the genotypic disequilibria:

$$-ux \le D_1 \le uy;$$

$$-vx \le D_2 \le vy;$$

$$-wx \le D_3 \le wy.$$

It is also possible to measure cytonuclear associations at the level of alleles. In the population there are four possible allelic combinations A/M, A/m, a/M, and a/m. Denote their frequencies by e_1, e_2, e_3, and e_4, respectively. These can be defined in terms of the genotypic frequencies as in Table 3. Here, p and q are the frequencies of alleles A and a. In the absence of selection, e_1, e_2, e_3, and e_4 represent gametic frequencies (in the sex transmitting the cytoplasmic gene) and for simplicity will be referred to as such in what follows.

The allelic association between cytoplasmic and nuclear markers can be described by the *gametic disequilibrium* parameter D, which measures the departure of gametic frequencies from expectations under random association (HEDRICK 1983). Define D as freq.(A/M) − freq.(A)freq.(M), or equivalently, $D = u_1 + \frac{1}{2}v_1 - px$. The gametic disequilibrium can also be defined as:

$$D = e_1 - px \tag{3}$$

or $D = e_1e_4 - e_2e_3$. This is a traditional measure of gametic disequilibrium, and the one introduced by CLARK (1984) in the context of nuclear-cytoplasmic relationships. Gametic frequencies can then be expressed in terms of the allele frequencies and the gametic disequilibrium as in Table 4. From these relationships the following constraints on the gametic

TABLE 3

Frequencies of allelic combinations

Cytoplasm	Nuclear allele: A	Nuclear allele: a	Total
M	$e_1 = u_1 + \frac{1}{2}v_1$	$e_3 = w_1 + \frac{1}{2}v_1$	x
m	$e_2 = u_2 + \frac{1}{2}v_2$	$e_4 = w_2 + \frac{1}{2}v_2$	y
Total	p	q	1.0

TABLE 4

Gametic (allelic) disequilibrium

	A	a	Total
M	$px + D$	$qx - D$	x
m	$py - D$	$qy + D$	y
Total	p	q	1.0

disequilibrium can be derived:

$$-px,\ -qy \leq D \leq py,\ qx. \tag{4}$$

The gametic and genotypic disequilibria are closely related since:

$$D = D_1 + \tfrac{1}{2}D_2 = -D_3 - \tfrac{1}{2}D_2 = \tfrac{1}{2}D_1 - \tfrac{1}{2}D_3. \tag{5}$$

The allelic association thus can be partitioned into (any) two genotypic components. Furthermore, equations (2) and (5) show that the values of any two disequilibria determine all four, and that if any two of the disequilibria equal zero, then all are zero. Thus, all possible interrelationships between gametic and genotypic disequilibria can be grouped into the following six categories:

$$D = 0,\ D_1 = 0,\ D_2 = 0,\ D_3 = 0; \tag{6a}$$

$$D_1 = 0, \quad D = \tfrac{1}{2}D_2 = -\tfrac{1}{2}D_3 \neq 0; \tag{6b}$$

$$D_2 = 0, \quad D = D_1 = -D_3 \neq 0; \tag{6c}$$

$$D_3 = 0, \quad D = \tfrac{1}{2}D_1 = -\tfrac{1}{2}D_2 \neq 0; \tag{6d}$$

$$D = 0, \quad D_1 = D_3 = -\tfrac{1}{2}D_2 \neq 0; \tag{6e}$$

$$D \neq 0,\ D_1 \neq 0,\ D_2 \neq 0,\ D_3 \neq 0. \tag{6f}$$

This set of measures has practical utility for several reasons. For instance, one central application in studies of mitochondrial-nuclear associations has been to exploit the maternal inheritance of a cytoplasmic gene to infer the directionality of matings in hybrid zones. The genotypic disequilibrium D_2 can be a direct measure of the directionality of matings. Furthermore, as shown in the DISCUSSION, the various categories of disequilibria listed in (6) allow simple biological interpretations in terms of the mating system within a hybrid zone. The allelic-genotypic disequilibrium measures also arise as the natural coordinates in a variety of models of the mating system and selection.

TABLE 5

Alternative set of disequilibrium measures[a]

AA	Aa	aa	Total
$M\ ux + 2pD + d$	$vx + 2(q\text{-}p)D - 2d$	$wx - 2qD + d$	x
$m\ uy - 2pD - d$	$vy - 2(q\text{-}p)D + 2d$	$wy + 2qD - d$	y
u	v	w	1.0

[a] Compare with Table 2.

They serve to decouple and/or linearize a model's dynamical behavior. The genotypic disequilibria also serve to partition the gametic disequilibrium as in (5), and thereby explain the allelic disequilibrium.

Nonetheless, as discussed in WEIR and WILSON (1986), a given parameterization of the genotypic frequencies in terms of disequilibrium measures must be entered into delicately. There are several alternative parameterizations. One such set of disequilibrium measures motivated by B. S. WEIR and C. C. COCKERHAM (unpublished results) involves the gametic disequilibrium D, as defined in (3), and the *residual disequilibrium*, d, defined by $v_1 = vx + 2(q - p)D - 2d$, or

$$\begin{aligned} d &= (q - p)D - \tfrac{1}{2}(v_1 - vx) \\ &= (q - p)D - \tfrac{1}{2}D_2. \end{aligned} \tag{7}$$

The analog of Table 2 is given in Table 5.

While the biological interpretation of this particular parameterization is unclear, its statistical interpretation is quite simple. The gene frequencies can be viewed as linear effects on the genotypic frequencies, and the disequilibria D and d can be viewed as the interaction effects. The allelic disequilibrium corresponds to a linear × linear interaction, and the residual disequilibrium to a linear × quadratic interaction. This pair of measures generates an alternative classification of the pattern of disequilibria: (i) $D = d = 0$; (ii) $d = 0$, $D \neq 0$; (iii) $D = 0$, $d \neq 0$; or (iv) $D \neq 0$, $d \neq 0$.

DYNAMICS OF CYTONUCLEAR DISEQUILIBRIA

In this section, we describe and compare the dynamics of these measures of cytonuclear disequilibria under three conventional deterministic models of the mating system: (1) random mating; and (2) positive assortative mating: (a) without dominance (the "mixed-mating" model); and (b) with dominance [O'DONALD's (1960) model]. In each case, the cytoplasmic locus is assumed to be uniparentally inherited, and the genotypic frequencies in the two sexes are assumed equal. All models considered are "neutral" in that there are no fitness differences among genotypes, nor do allele frequencies change at either the nuclear or cytoplasmic locus.

Random mating: The dynamics of gametic dis-

equilibrium have been derived by CLARK (1984). To develop some notation useful for later models, we present a slightly different derivation of the dynamics of D, and extend results to a consideration of the genotypic and residual disequilibria. The latter results are new.

Consider an A/M gamete produced by an individual in the next generation. This gamete is equilikely to arise in two ways: (i) both its alleles are inherited from the mother of the individual, who carries A and M with probability e_1; or (ii) the A allele is inherited from the father, and the M cytotype from the mother of the individual, independently with probabilities p and x, respectively. Putting these possibilities together and repeating the arguments for the other three gametes yields, after one generation:

$$\begin{aligned} e_1' &= \tfrac{1}{2}e_1 + \tfrac{1}{2}px; \\ e_2' &= \tfrac{1}{2}e_2 + \tfrac{1}{2}py; \\ e_3' &= \tfrac{1}{2}e_3 + \tfrac{1}{2}qx; \\ e_4' &= \tfrac{1}{2}e_4 + \tfrac{1}{2}qy. \end{aligned} \tag{8}$$

Adding together $p' = e_1' + e_2' = p$ and $x' = e_1' + e_3' = x$ establishes the constancy of allele frequencies. We can thus treat p, q, x, and y as parameters.

Any expression in (8) can be used to derive the recursion for gametic disequilibrium D' in the next generation in terms of its value in the current generation, D. For example, subtract px from both sides of the recursion for e_1':

$$e_1' - px = \tfrac{1}{2}(e_1 - px).$$

From (3), this result shows that:

$$D^{(t)} = \tfrac{1}{2}D^{(t-1)} = (\tfrac{1}{2})^t D^{(0)}, \tag{9}$$

where t is time in generations. The departure D from gametic equilibrium is halved in each generation of random mating. Gametic disequilibrium thus decays geometrically, at the same rate as for two unlinked, nuclear genes.

Using (9) and Table 4 we can then write explicit solutions for the trajectories of gametic frequencies:

$$\begin{aligned} e_1^{(t)} &= px + (\tfrac{1}{2})^t D^{(0)}; \\ e_2^{(t)} &= py - (\tfrac{1}{2})^t D^{(0)}; \\ e_3^{(t)} &= qx - (\tfrac{1}{2})^t D^{(0)}; \\ e_4^{(t)} &= qy + (\tfrac{1}{2})^t D^{(0)}. \end{aligned} \tag{10}$$

Gametic frequencies cease to change once gametic equilibrium (*i.e.*, $D = 0$) is obtained.

The behavior of the genotypic disequilibria is also readily obtained. By reasoning similar to that leading to the gametic recursions in (8), it is possible to derive the genotypic frequencies in the next generation:

$$\begin{aligned} u_1' &= e_1 p; & u_2' &= e_2 p; \\ v_1' &= e_3 p + e_1 q; & v_2' &= e_4 p + e_2 q; \\ w_1' &= e_3 q; & w_2' &= e_4 q. \end{aligned} \tag{11}$$

Using Table 3, one can verify the consistency of these recursions with those of the gametic frequencies in (8). More importantly, we can obtain recursions for the genotypic disequilibria. By definition, from (1a), we have

$$D_1' = u_1' - u'x' = u_1' - u'x. \tag{12}$$

After one generation, Hardy-Weinberg equilibrium is achieved at any nuclear locus, so $u' = p^2$ for all time. The expression for D_1' reduces to:

$$D_1' = u_1' - p^2 x = e_1 p - p^2 x, \tag{13}$$

where we have substituted $u_1' = e_1 p$ from (11). By rearrangement of terms,

$$D_1' = p(e_1 - px) = pD. \tag{14}$$

The dynamics of the genotypic disequilibria are then:

$$D_1^{(t)} = pD^{(t-1)} = p(\tfrac{1}{2})^{t-1} D^{(0)}; \tag{15a}$$

$$D_2^{(t)} = (q - p)D^{(t-1)} = (q - p)(\tfrac{1}{2})^{t-1} D^{(0)}; \tag{15b}$$

$$D_3^{(t)} = -qD^{(t-1)} = -q(\tfrac{1}{2})^{t-1} D^{(0)}. \tag{15c}$$

[Notice that the recursion for D can also be derived from (15) using (5).]

The genotypic disequilibria are all coupled to the gametic disequilibrium in the previous generation. After Hardy-Weinberg equilibrium is achieved in the first generation, the signs of the genotypic disequilibria are fixed by the initial D. When $D^{(0)} \neq 0$, the signs of D_1 and D_3 are always opposite. When $p < 0.5$, D_1 and D_2 have the same sign; when $p > 0.5$, D_3 and D_2 are of the same sign; and when $p = 0.5$, D_2 is zero thereafter.

Since $|D|$ decays monotontically to zero by ½ per generation, it follows from (15) that $|D_1|$, $|D_2|$, and $|D_3|$ do so also after one generation. Another feature of the random mating model is that *all* disequilibria are either decaying, or all are fixed at 0. Thus, (i) $D = 0$ if and only if any $D_i = 0$; and (ii) gametic equilibrium implies complete genotypic equilibrium, and *vice versa*. These will not be features of all other models. There is an exception to the results in this paragraph when $p = 0.5$ in which case $D_2 \equiv 0$.

From the recursions for the allelic and genotypic disequilibria and definition (7) we may also deduce the dynamics of the residual disequilibrium:

$$\begin{aligned} d' &= (q - p)D' - \tfrac{1}{2}D_2' \\ &= \tfrac{1}{2}(q - p)D - \tfrac{1}{2}(q - p)D = 0. \end{aligned}$$

Thus, after the first generation the residual disequilibrium is zero for all time under random mating.

Positive assortative mating without dominance (the "mixed-mating" model): Many hermaphroditic animals and plants can self-fertilize as well as outcross. The well-known mixed-mating model (CLEGG 1980) considers such situations by distinguishing the mating events due to self-fertilization (with probability α), from those due to random outcrossing (with probability $\bar{\alpha} = 1 - \alpha$). As noted by ENDLER (1977, p. 143), the mixed-mating model can also be applied to hybrid zones (or other situations) in which mating preferences are influenced by all genotypes (*AA*, *Aa*, *aa*) at a diallelic nuclear locus under consideration. In this context, α is the probability that an individual in the population prefers to mate with like nuclear genotype, while the probability of mating at random is $\bar{\alpha}$ ($= 1 - \alpha$). Additional assumptions of this "narcissistic" model of assortative mating are: (1) the fraction α is constant across individuals, (2) no fertility or viability differences exist among matings, (3) the hybrid population is closed to further outside recruitment from parental species' gene pools, and (4) all individuals mate so that a 1:1 sex ratio is maintained. Here we will examine the nuclear-cytoplasmic disequilibria under this assortative mating model.

As before, consider an *A/M* gamete in a population. This is carried by progeny from random matings (probability $\bar{\alpha}$), by (8), with probability $\frac{1}{2}e_1 + \frac{1}{2}px$. *A/M* is carried by progeny from assortative matings (probability α) with probability e_1. Combining these possibilities, and repeating the argument for the other gametes, yields:

$$\begin{aligned} e_1' &= \alpha e_1 + \bar{\alpha}[\tfrac{1}{2}e_1 + \tfrac{1}{2}px]; \\ e_2' &= \alpha e_2 + \bar{\alpha}[\tfrac{1}{2}e_2 + \tfrac{1}{2}py]; \\ e_3' &= \alpha e_3 + \bar{\alpha}[\tfrac{1}{2}e_3 + \tfrac{1}{2}qx]; \\ e_4' &= \alpha e_4 + \bar{\alpha}[\tfrac{1}{2}e_4 + \tfrac{1}{2}qy]. \end{aligned} \tag{16}$$

As under random mating, allele frequencies remain constant over time.

By the same argument used for the random mating model, we can derive the trajectory of D:

$$D^{(t)} = [\tfrac{1}{2}(1 + \alpha)]D^{(t-1)} = [\tfrac{1}{2}(1 + \alpha)]^t D^{(0)}. \tag{17}$$

The qualitative behavior of D is identical to that under random mating, that is, a geometric rate of decay to zero. However, the decay rate is decelerated relative to the random mating model, and is the same as that for gametic disequilibrium between two unlinked, nuclear loci (WEIR, ALLARD and KAHLER 1972). The reduction in frequency of heterozygotes under positive assortative mating results in a lessened opportunity for recombination which generates the decay in disequilibrium.

To obtain the dynamics of the genotypic disequilibria, we need recursions for the genotypic frequencies:

$$\begin{aligned} u_1' &= \alpha(e_1 - \tfrac{1}{4}v_1) + \bar{\alpha}e_1p; \\ u_2' &= \alpha(e_2 - \tfrac{1}{4}v_2) + \bar{\alpha}e_2p; \\ v_1' &= \tfrac{1}{2}\alpha v_1 + \bar{\alpha}(e_3p + e_1q); \\ v_2' &= \tfrac{1}{2}\alpha v_2 + \bar{\alpha}(e_4p + e_2q); \\ w_1' &= \alpha(e_3 - \tfrac{1}{4}v_1) + \bar{\alpha}e_3q; \\ w_2' &= \alpha(e_4 - \tfrac{1}{4}v_2) + \bar{\alpha}e_4q. \end{aligned} \tag{18}$$

These can be derived by reasoning analogous to that leading to (16). By adding together $u_1' + u_2'$ etc., we recover the usual one-locus mixed-mating model (HEDRICK 1983, p. 90):

$$\begin{aligned} u' &= \alpha(p - \tfrac{1}{4}v) + \bar{\alpha}p^2; \\ v' &= \tfrac{1}{2}\alpha v + \bar{\alpha}2pq; \\ w' &= \alpha(q - \tfrac{1}{4}v) + \bar{\alpha}q^2. \end{aligned} \tag{19}$$

As under the random mating model, the genotypic equations (18) can be used with Table 3 to derive the gametic recursions (16). The recursions for the genotypic disequilibria are:

$$D_1' = (\alpha + \bar{\alpha}p)D - \tfrac{1}{4}\alpha D_2; \tag{20a}$$

$$D_2' = \bar{\alpha}(q - p)D + \tfrac{1}{2}\alpha D_2; \tag{20b}$$

$$D_3' = -(\alpha + \bar{\alpha}q)D - \tfrac{1}{4}\alpha D_2. \tag{20c}$$

If we set $\alpha = 0$, the dynamics reduce to those of the random mating model (equations 15a, 15b, and 15c), and we can recover the recursion for D in (17) using (5).

Explicit solutions for the behavior of genotypic disequilibria through time can be obtained from equations (17) and (20):

$$\begin{aligned} D_1^{(t)} = {} & D^{(0)}[\tfrac{1}{2}(1 + \alpha)]^{t-1}[\alpha + \bar{\alpha}p - \tfrac{1}{2}\alpha\bar{\alpha}(q - p)] \\ & + (\tfrac{1}{2}\alpha)^t[\bar{\alpha}(q - p)D^{(0)} - \tfrac{1}{2}D_2^{(0)}]; \end{aligned} \tag{21a}$$

$$\begin{aligned} D_2^{(t)} = {} & 2D^{(0)}\{[\tfrac{1}{2}(1 + \alpha)]^t - (\tfrac{1}{2}\alpha)^t\}\bar{\alpha}(q - p) \\ & + (\tfrac{1}{2}\alpha)^t D_2^{(0)}; \end{aligned} \tag{21b}$$

$$\begin{aligned} D_3^{(t)} = {} & D^{(0)}[\tfrac{1}{2}(1 + \alpha)]^{t-1}[-\alpha - \bar{\alpha}q - \tfrac{1}{2}\alpha\bar{\alpha}(q - p)] \\ & + (\tfrac{1}{2}\alpha)^t[\bar{\alpha}(q - p)D^{(0)} - \tfrac{1}{2}D_2^{(0)}]. \end{aligned} \tag{21c}$$

Two qualitative points can be made in comparing these results to those of the random-mating model. First, unlike under random mating, $|D_1|$, $|D_2|$, and $|D_3|$ do not approach zero monotonically from all starting conditions under the mixed-mating model ($0 < \alpha < 1$). Second, under the mixed-mating model it is possible to be in gametic phase equilibrium ($D^{(t)} \equiv 0$ for all t) while all genotypic disequilibria remain nonzero until equilibrium is achieved.

Using (7), (17), and (20) the recursion for the residual disequilibrium is

$$d' = \alpha(q - p)D - \tfrac{1}{4}\alpha D_2 = \tfrac{1}{2}\alpha(q - p)D + \tfrac{1}{2}\alpha d.$$

The trajectory for $d^{(t)}$ can be obtained from (17) using (21b).

Positive assortative mating with dominance: This is the classic assortative mating model introduced by O'DONALD (1960) and treated subsequently in several texts [*e.g.*, HEDRICK (1983), pp. 113–116]. All assumptions of the mixed-mating model are in place, except that nuclear allele A is dominant to a, and the mating rules are changed to be as follows: (1) with probability α, the aa genotype prefers to mate with aa, and $A-$ genotypes prefer to mate with $A-$; (2) with probability $\bar{\alpha} = 1 - \alpha$, matings take place at random.

As in the previous models, consider an A/M gamete. This is carried by progeny from random matings (probability $\bar{\alpha}$), by (8), with probability $\tfrac{1}{2}e_1 + \tfrac{1}{2}px$. Alternatively, A/M is carried by progeny from assortative matings (probability α) with (i) probability 1 in AA/M progeny and (ii) probability $\tfrac{1}{2}$ in Aa/M progeny. Now, AA/M offspring have two possible assortative mating sources. They are produced by: (1a) AA/M ♀ $\times A-$ ♂ matings with probability $p/(u + v)$, the mating having (conditional) probability u_1; or (1b) Aa/M ♀ $\times A-$ ♂ matings with probability $\tfrac{1}{2}p/(u + v)$, the mating having (conditional) probability v_1. Similarly, Aa/M progeny are produced by (2a): AA/M ♀ $\times Aa$ ♂ matings with probability $\tfrac{1}{2}$, the mating having (conditional) probability $u_1v/(u + v)$; or (2b) Aa/M ♀ $\times A-$ ♂ matings with probability $\tfrac{1}{2}$, the mating having (conditional) probability v_1. Putting these cases together and repeating the arguments for the other gametes yields:

$$\begin{aligned}
e_1' &= \bar{\alpha}[\tfrac{1}{2}e_1 + \tfrac{1}{2}px] + \alpha\left[\frac{e_1p}{p + \tfrac{1}{2}v} + \frac{\tfrac{1}{4}u_1v}{p + \tfrac{1}{2}v} + \tfrac{1}{4}v_1\right];\\
e_2' &= \bar{\alpha}[\tfrac{1}{2}e_2 + \tfrac{1}{2}py] + \alpha\left[\frac{e_2p}{p + \tfrac{1}{2}v} + \frac{\tfrac{1}{4}u_2v}{p + \tfrac{1}{2}v} + \tfrac{1}{4}v_2\right];\\
e_3' &= \bar{\alpha}[\tfrac{1}{2}e_3 + \tfrac{1}{2}qx] + \alpha\left[\frac{\tfrac{1}{4}v_1v}{p + \tfrac{1}{2}v} + \frac{\tfrac{1}{4}u_1v}{p + \tfrac{1}{2}v} + \tfrac{1}{4}v_1 + w_1\right];\\
e_4' &= \bar{\alpha}[\tfrac{1}{2}e_4 + \tfrac{1}{2}qy] + \alpha\left[\frac{\tfrac{1}{4}v_2v}{p + \tfrac{1}{2}v} + \frac{\tfrac{1}{4}u_2v}{p + \tfrac{1}{2}v} + \tfrac{1}{4}v_2 + w_2\right].
\end{aligned} \tag{22}$$

In the above equations, and below, we have substituted $p + \tfrac{1}{2}v$ for $u + v$.

As under the earlier models, allele frequencies remain constant over time. Using expression (3), we can derive the recursion for D:

$$D' = D\left[\tfrac{1}{2}(1 + \alpha) - \frac{\tfrac{1}{4}\alpha v}{p + \tfrac{1}{2}v}\right] + D_2\left[\frac{\tfrac{1}{4}\alpha p}{p + \tfrac{1}{2}v}\right]. \tag{23}$$

Unlike the random- or mixed-mating models, the dynamics of the gametic disequilibrium directly involve a genotypic disequilibrium.

To obtain the dynamics of the genotypic disequilibria, we can either use the reasoning used in the previous models or a table of all two-locus mating combinations and offspring produced, similar to that in HEDRICK (1983, Table 3.14). Both approaches lead to the following recursions for the genotype frequencies:

$$\begin{aligned}
u_1' &= e_1p\left[\bar{\alpha} + \frac{\alpha}{p + \tfrac{1}{2}v}\right];\\
u_2' &= e_2p\left[\bar{\alpha} + \frac{\alpha}{p + \tfrac{1}{2}v}\right];\\
v_1' &= \bar{\alpha}[e_1q + e_3p] + \tfrac{1}{2}\alpha\left[\frac{u_1v}{p + \tfrac{1}{2}v} + v_1\right];\\
v_2' &= \bar{\alpha}[e_2q + e_4p] + \tfrac{1}{2}\alpha\left[\frac{u_2v}{p + \tfrac{1}{2}v} + v_2\right];\\
w_1' &= \bar{\alpha}e_3q + \alpha\left[\frac{\tfrac{1}{4}v_1v}{p + \tfrac{1}{2}v} + w_1\right];\\
w_2' &= \bar{\alpha}e_4q + \alpha\left[\frac{\tfrac{1}{4}v_2v}{p + \tfrac{1}{2}v} + w_2\right].
\end{aligned} \tag{24}$$

If the nuclear genotypic frequencies are summed (*e.g.*, $u' = u_1' + u_2'$), we recover O'DONALD's model. These expressions can also be used to calculate directly the gametic frequencies just given in (22).

We obtain the recursions for the genotypic disequilibria in the usual way by working with the definitions such as $D_1' = u_1' - u'x$. We find:

$$D_1' = \left[\bar{\alpha}p + \frac{\alpha p}{p + \tfrac{1}{2}v}\right]D; \tag{25a}$$

$$D_2' = \left[\bar{\alpha}(q - p) + \frac{\tfrac{1}{2}\alpha v}{p + \tfrac{1}{2}v}\right]D + \left[\frac{\tfrac{1}{2}\alpha p}{p + \tfrac{1}{2}v}\right]D_2; \tag{25b}$$

$$D_3' = -(\bar{\alpha}q + \alpha)D - \left[\frac{\tfrac{1}{2}\alpha p}{p + \tfrac{1}{2}v}\right]D_2. \tag{25c}$$

From these expressions we can also obtain another derivation of (23) using $D' = D_1' + \tfrac{1}{2}D_2'$.

The derivation of the recursion for the residual disequilibrium is totally analogous to that in the previous two models:

TABLE 6

Qualitative results on the dynamics of nuclear-cytoplasmic disequilibria under various mating models

Behavior	Random mating	Assortative mating: Without dominance	Assortative mating: With dominance
D alone sets sign of:	D_1' D_2'[a] D_3'	No D_i'	D_1'
D can change sign	No	No	Yes
D_i can change sign	No	Yes	Yes
$D = 0$ iff D_1, D_2, and $D_3 = 0$	Yes[b]	No	No
D must decay monotonically to zero	Yes	Yes	No[c]
D_i's must decay monotonically to zero	Yes[b]	No[c]	No[c]
Dynamics of D and D_i independent of u, v, and w	Yes	Yes	No

[a] In conjunction with p.
[b] This statement holds only from the first generation on.
[c] Sometimes decays monotonically.

$$d' = \frac{\alpha p}{p + \frac{1}{2}v}\{[q(q - p) - \frac{1}{2}v]D + pd\}.$$

The O'DONALD model shares some features with both the random- and mixed-mating models. As with the other two models, all the disequilibria ultimately decay to zero ($0 < \alpha < 1$). The sign of D sets the sign of D_1 in the next generation, as for random mating, whereas the genotypic and residual disequilibria are not necessarily monotonic, as for mixed-mating. A distinctive feature of the O'DONALD model is that D itself can behave nonmonotonically.

Some of the major qualitative results concerning the dynamics of disequilibria are summarized in Table 6 for the random mating and positive assortative mating models. Examples of the dynamics of cytonuclear disequilibria are plotted in Figure 1. Points of particular interest include the following: (i) decelerated decay of disequilibria under assortative mating, and (ii) change of sign of D_2 under O'DONALD's model.

With the exception of complete assortative mating, a common feature of all three models above is the ultimate decay to zero of all disequilibria (see *e.g.*, Figure 1). It is natural to ask under what conditions disequilibria can be permanently nonzero. This would of course necessitate a joint nuclear and cytoplasmic polymorphism. We have suggested in the introduction that selection may be one mechanism to maintain disequilibria. This is precluded in the cytonuclear viability selection models of CLARK (1984), for they have no stable interior equilibria (unless there is a limit cycle). But, in the models of fertility selection in a partially selfing, gynodioecious population of ROSS and GREGORIUS (1985), stable interior equilibria are found associated with permanent disequilibria. In particular, all the disequilibria along the trajectories shown in Figure 1 of their paper are permanently nonzero.

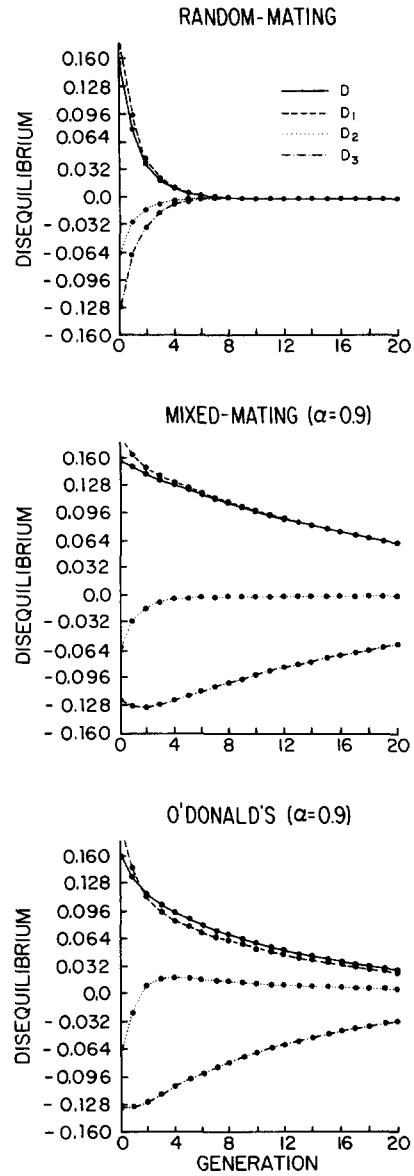

FIGURE 1.—Examples of the dynamics of cytonuclear disequilibria under the random mating, mixed-mating, and O'DONALD models of mating systems. Initial genetic conditions were those actually observed (see Table 9) in a hybrid population of treefrogs: $\hat{D}_1 = 0.190$; $\hat{D}_2 = -0.063$; $\hat{u} = 0.479$; $\hat{v} = 0.213$; $\hat{x} = 0.465$. For the mixed-mating and O'DONALD models, an assortative mating rate of $\alpha = 0.9$ was utilized. Each model displays a qualitatively different kind of behavior for the allelic-genotypic disequilibria. The dynamics of d (not shown) are qualitatively indistinguishable for the two assortative mating models; $d^{(t)}$ first decreases below zero, then increases to zero.

As a cautionary note, if stable disequilibria are observed, it would be tempting to invoke selection as the explanation. This inference is weakened by the fact that in extensions to the mating system models of this section, large transient disequilibria are observed lasting several hundred generations. To distinguish permanent disequilibria from the slowly changing,

transient disequilibria under some mating systems will prove difficult.

STATISTICAL INFERENCE ABOUT CYTONUCLEAR DISEQUILIBRIA

Our major concern is to select the pattern of disequilibria in (6) most descriptive of a data set, and secondarily to ascertain the magnitudes of the disequilibria. Each category in (6) determines a statistical hypothesis, which can be fitted to the data by the method of maximum likelihood. Because tests of goodness of fit depend on the maximum likelihood estimates (MLEs) obtained, we first describe the estimation of disequilibria and then the selection of a category which best fits the data.

Suppose an individual drawn at random from a population has one of the six genotypes in Table 1 with probabilities listed in the vector $\eta = (u_1, v_1, w_1, u_2, v_2, w_2)^T$, where the T means transpose. These genotypic probabilities can be expressed (using Table 2) in terms of the independent parameters required for any of the six categories (6a–6f) listed earlier. The number of independent parameters varies from three (u, v, x for (6a)) to five (u, v, x, D_1, D_2 for (6f)), depending on which disequilibria are zero. The independent parameters are written in the column vector β for each category.

In a population survey, sampling is repeated N times to generate a list of the six genotypic counts, $\mathbf{N} = (N_{11}, N_{12}, N_{13}, N_{21}, N_{22}, N_{23})^T$. The probability of a particular list of genotypic counts is multinomial and proportional to:

$$u_1^{N_{11}} v_1^{N_{12}} \cdots w_2^{N_{23}}. \tag{26}$$

Denote this probability by $\Pr(\mathbf{N} \mid \beta)$. We fix the expression $\Pr(\mathbf{N} \mid \beta)$ at the observed counts, $\mathbf{N}$ and consider $\Pr(\mathbf{N} \mid \beta)$ as a function of the parameters β, thus yielding the likelihood function $L(\beta \mid \mathbf{N})$. The MLEs are the values of the parameters ($\hat{\beta}$) which maximize the likelihood function.

One method of finding MLEs involves solving the likelihood equations $\partial \ln L/\partial \beta_i = 0$ for each parameter β_i, or simply $\partial \ln L/\partial \beta = \mathbf{0}$. Using the chain rule, these become:

$$\frac{\partial \ln L}{\partial \beta} = \frac{\partial \ln L}{\partial \eta} \frac{\partial \eta}{\partial \beta} = \mathbf{0}. \tag{27}$$

These equations can be solved in a variety of ways, including the traditional numerical method of maximum likelihood scoring. We use this method as a springboard to a novel approach, which involves rewriting the likelihood equations as normal equations to a weighted, least squares regression problem (GREEN 1984; BURN 1982). The new approach allows us to (i) relax easily the multinomial sampling assumption; (ii) make use of standard statistical packages (*e.g.*, BMDP or GLIM) to do the computing; (iii) use resistant and robust methods of fitting; (iv) shed ourselves of the computational burden of repeated matrix inversion; (v) enlarge the domain of convergence; and (vi) utilize well known techniques for accelerating convergence in least squares problems.

Maximum likelihood scoring usually begins with the calculation of the derivatives of the genotypic frequencies η with respect to the parameters β. For each of the six categories, a *derivative matrix* $X = \partial \eta/\partial \beta$ is presented in Table 7. We also define a score vector $\mathbf{S}$, which is the derivative of the loglikelihood with respect to the genotypic frequencies, where for each category in (6):

$$\ln L = N_{11} \ln u_1 + N_{12} \ln v_1 + \cdots + N_{23} \ln w_2. \tag{28}$$

The score vector $\mathbf{S} = \partial \ln L/\partial \eta$ can be written as $\mathbf{S} = (N_{11}/u_1, \cdots, N_{23}/w_2)^T$. With the derivative matrix $\mathbf{X}$ and the score vector $\mathbf{S}$, we can more simply write the likelihood equations in (27) as:

$$\mathbf{X}^T\mathbf{S} = \mathbf{0}. \tag{29}$$

The derivative matrix $\mathbf{X}$ summarizes the particular structure of each category in (6).

Maximum likelihood scoring and ultimately the computation of the variance-covariance matrix for the MLEs $\hat{\beta}$ require two *information matrices*:

$$\mathbf{A} = \begin{bmatrix} u_1^{-1} & & & & & \\ & v_1^{-1} & & 0 & & \\ & & w_1^{-1} & & & \\ & & & u_2^{-1} & & \\ & 0 & & & v_2^{-1} & \\ & & & & & w_2^{-1} \end{bmatrix}, \tag{30}$$

which summarizes the information in the sample about the genotypic frequencies; and

$$\mathbf{I} = \mathbf{X}^T\mathbf{A}\,\mathbf{X}, \tag{31}$$

which summarizes information about the parameters β. Given a provisional estimate $\hat{\beta}$, the likelihood equations (29) can be solved iteratively for an updated estimate $\hat{\beta}^*$ according to:

$$N\mathbf{I}\,\hat{\beta}^* = N(\mathbf{X}^T\,\mathbf{A}\,\mathbf{X})\hat{\beta}^* = \mathbf{X}^T(\mathbf{S} + N\,\mathbf{A}\,\mathbf{X}\,\hat{\beta}), \tag{32}$$

where N is the sample size.

We depart from maximum likelihood scoring in realizing (32) is a set of *normal equations* solving a weighted regression problem. The regression problem has $\mathbf{Y} = \mathbf{S} + N\mathbf{A}\,\mathbf{X}\,\hat{\beta}$ as the provisional dependent variable; $\mathbf{X}$ as the *design matrix*; and $\mathbf{A}$ as the *weight matrix*. The computation of MLEs is obtained by: (i) using the current estimates $\hat{\beta}$ to evaluate $\mathbf{S}$, $\mathbf{X}$, $\mathbf{A}$, and $\mathbf{Y}$; (ii) solving the normal equations (32) of this

TABLE 7

Derivative matrices for various categories (models) of disequilibria[a]

Category[b]	Parameter	$\mathbf{X}^T$					
		u_1	v_1	w_1	u_2	v_2	w_2
(6a)	u	x	0	$-x$	y	0	$-y$
	v	0	x	$-x$	0	y	$-y$
	x	u	v	w	$-u$	$-v$	$-w$
(6b)	u	x		$-x$	y	0	$-y$
	v	0	0	$-x$	0	y	$-y$
	x	u	x	w	$-u$	$-v$	$-w$
	D_2	0	v	-1	0	-1	1
			1				
(6c)	u	x	0	$-x$	y	0	$-y$
	v	0	x	$-x$	0	y	$-y$
	x	u	v	w	$-u$	$-v$	$-w$
	D_1	1	0	-1	-1	0	1
(6d)	u	x	0	$-x$	y	0	$-y$
	v	0	x	$-x$	0	y	$-y$
	x	u	v	w	$-u$	$-v$	$-w$
	D_1	1	-1	0	-1	1	0
(6e)	u	x	0	$-x$	y	0	$-y$
	v	0	x	$-x$	0	y	$-y$
	x	u	v	w	$-u$	$-v$	$-w$
	D_1	1	-2	1	-1	2	-1
(6f)	u	x	0	$-x$	y	0	$-y$
	v	0	x	$-x$	0	y	$-y$
	x	u	v	w	$-u$	$-v$	$-w$
	D_1	1	0	-1	-1	0	1
	D_2	0	1	-1	0	-1	1

[a] If one of the genotype frequencies vanishes in the information matrix **A**, then the corresponding row and column must be deleted and the derivatives in the **X**-matrices above, recomputed subject to the constraint of the observed frequency being zero. Because of the lost degree of freedom, attention must be restricted to the categories (6a)–(6e).

[b] Refers to equations (6a)–(6f) in the text.

weighted regression problem for the new estimate $\hat{\beta}^*$ and (iii) cycling back to (i) until convergence of $\hat{\beta}^* - \hat{\beta}$ to zero is achieved. As a caveat, it is computationally useful to note that $\mathbf{S} = \mathbf{A}\,\mathbf{N}$ and that at the end of this iterative procedure the variance-covariance matrix of $\hat{\beta}$ is approximately $\mathbf{I}^{-1}/N$.

The procedure above must be performed separately for each category's derivative matrix in Table 7, where at each iteration **S**, **X**, **A**, and **Y** are evaluated using the current estimates $\hat{\beta}$ together with Table 2 and the appropriate entry in (6). Note that the MLEs of genotypic frequencies under (6a) and (6f) are obtained simply from observed genotypic frequencies, $\hat{u}_1 = N_{11}/N$, etc., where $\hat{u} = \hat{u}_1 + \hat{u}_2$, $\hat{v} = \hat{v}_1 + \hat{v}_2$, and $\hat{x} = \hat{u}_1 + \hat{v}_1 + \hat{w}_1$. Under (6f) the MLEs for the disequilibria are computed by the formulas in Table 9, also using the observed frequencies.

It remains to determine which of the six categories best fits the data. While the associated standard errors can provide a rough guide to the significance of disequilibria, the G-statistic (FIENBERG 1977) is likely to provide a better test of goodness of fit. This is done for each category by: (i) using its MLEs $\hat{\beta}$ with Table 2 and (6) to compute the expected number of each genotype: $\mathbf{E} = (E_{11}, E_{12}, \cdots, E_{23}) = N\hat{\eta}$, and (ii) comparing these expected counts **E** with the observed counts **N** through:

$$G = 2[N_{11}\ln(N_{11}/E_{11}) + N_{12}\ln(N_{12}/E_{12}) + \cdots + N_{23}\ln(N_{23}/E_{23})]. \quad (33)$$

If the category reflects the true state of the population, G will have an approximate χ^2 distribution with d.f. = 5 − (no. of independent parameters).

In order to compare the categories by their fit, it is helpful to arrange them in order of their complexity as in Figure 2. Starting at the top of the hierarchy (6f), we move sequentially down until the loss of a parameter results in a significant decrease in fit. The decrease in fit by moving from category A to B ($df_A < df_B$) is measured by $G_B - G_A$. This difference has a χ^2 distribution on $df_B - df_A$ degrees of freedom. When $G_B - G_A$ is significant, we accept A as providing a parsimonious fit to the data. Categories on the same tier are not comparable by this method.

If the measures D and d were used to describe

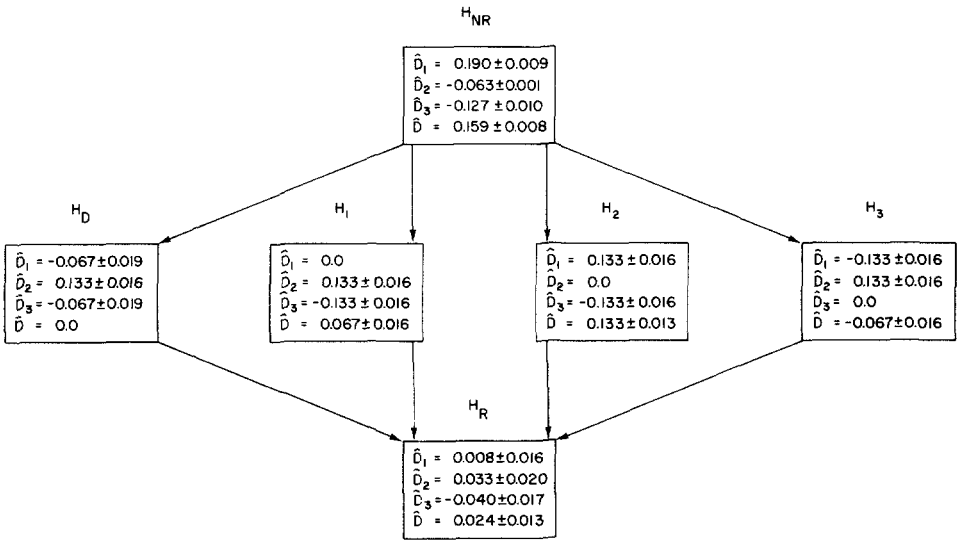

FIGURE 2. Categories of cytonuclear disequilibria arranged as a series of hypotheses about mating patterns in a hybrid zone. The disequilibria indicated for the nonrandom mating hypothesis (H_{NR}) are those empirically observed in the Hyla treefrog hybrid population (see Table 9); disequilibria under the random-mating hypothesis (H_R) are those observed in the hybrid population involving subspecies of bluegill sunfish *L. macrochirus* (see Table 10); disequilibria for the middle tier of hypotheses were calculated from the hypothetical data sets in Table 11. All disequilibria and their standard errors were calculated as in Table 9. The data in this figure is a composite summary of six such figures. All the disequilibria presented for each distinct data set were computed as if (6f) were true for the sake of comparability.

cytonuclear associations, the hierarchy of six categories in (6) would simplify to four categories: (i) $D = d = 0$; (ii) $d = 0$, $D \neq 0$; (iii) $D = 0$, $d \neq 0$; (iv) $D \neq 0$, $d \neq 0$. As stressed by B. S. WEIR and C. C. COCKERHAM (personal communication), it is important to consider the higher order disequilibrium d as well as the allelic disequilibrium D in describing departures from the no association hypothesis (i). This can be done via the procedure above, with the derivatives of the genotypic probabilities with respect to D_1 and D_2 in the derivative matrix **X** replaced by the derivatives of the genotypic probabilities with respect to D and d. (In the case of category (i) or (iv), the maximum likelihood estmates $\hat{\beta}$ are simple functions of the observed frequencies of genotypes). Using G-tests, one would attempt to move down the new hierarchy from (iv), until loss of fit is significant.

DISCUSSION

We have introduced and analyzed the dynamical behavior of four measures of disequilibria between a nuclear and a cytoplasmic gene. One set of measures decomposes departures from a no-association model into an allelic (gametic) component, D, and three genotypic components, D_1, D_2, and D_3. The other set decomposes associations into a linear × linear component, D, and a linear × quadratic component, d. Departures from random associations, as indicated by the signs and magnitudes of the disequilibrium measures, could arise from any of several evolutionary forces, including founder effect and genetic drift, epistatic selection, and nonrandom mating. The same measures could be extended to a broader context of haplo-diploid systems or to associations in the heterogametic sex between a sex-linked gene and an autosomal locus.

To illustrate the calculation and conceptual application of these disequilibrium statistics to real and hypothetical data sets involving nuclear and cytoplasmic genes, we will now consider D and D_i values in a class of commonly encountered evolutionary settings—secondary hybrid zones. As noted by WEIR, ALLARD and KAHLER (1972), the mating system itself can often provide parsimonious hypotheses about genetic disequilibria. In this spirit, we ask what kinds of nonrandom mating are sufficient to explain various patterns of cytonuclear disequilibria in a hybrid population.

Cytonuclear disequilibria in hybrid zones: Table 8 summarizes possible explanations involving the mating system for various categories of relationships between allelic and genotypic disequilibria. For example, all disequilibria could be significantly different from zero, and this might arise if there were directional and strong assortative mating in a hybrid population of fairly recent origin. At the other end of the continuum, all disequilibria could be zero, if the hybrid population was random mating and fairly old.

LAMB and AVISE (1986) provide an empirical example of the former situation. The treefrogs *Hyla cinerea* and *Hyla gratiosa* hybridize in a series of artificial ponds near Auburn, Alabama. Normally, these species are behaviorally isolated, in part because of mating call site preferences. *H. cinerea* males typically call from elevated perches in shoreline vegetation, while *H. gratiosa* males call from the water surface. At the Auburn site, frequent mowing of the pond perimeters has eliminated the preferred perches for *H. cinerea*, and as a consequence many males call at

TABLE 8

Hypotheses about nuclear-cytoplasmic disequilibria in a hybrid population

Hypothesis	Disequilibria	Possible explanation[a]
H_R	$D = D_1 = D_2 = D_3 = 0$	Random mating; fairly old
H_1	$D_1 = 0$; $D = \frac{1}{2}D_2 = -\frac{1}{2}D_3 \neq 0$	Strong directionality to interspecific matings; hybrids preferentially backcross to less discriminating species
H_2	$D_2 = 0$; $D = D_1 = -D_3 \neq 0$	Species mate assortatively; no directionality to interspecific matings
H_3	$D_3 = 0$; $D = \frac{1}{2}D_1 = -\frac{1}{2}D_2 \neq 0$	Strong directionality to interspecific matings; hybrids preferentially backcross to less discriminating species
H_D	$D = 0$; $D_1 = D_3 = -\frac{1}{2}D_2 \neq 0$	Nuclear allele frequencies identical in the two cytotypes; mixed-mating
H_{NR}	$D \neq 0$; $D_1 \neq 0$; $D_2 \neq 0$; $D_3 \neq 0$	Nonrandom mating; directionality to interspecific matings; fairly young

[a] The listed explanations are by no means exhaustive or definitive.

ground level. Gravid females of both species approach the ponds from surrounding woods to mate. The expectation is that males of *H. cinerea* would intercept *H. gratiosa* females, while crosses in the opposite direction (*H. gratiosa* ♂ × *H. cinerea* ♀) would seldom take place. LAMB and AVISE (1986) tested this hypothesis by simultaneously surveying protein products of five diagnostic nuclear loci, and the mitochondrial genotypes *M* of *H. cinerea* and *m* of *H. gratiosa*, in 305 individuals. As a test of our reasoning, we can apply our measures of cytonuclear association to their data. Methods of calculating gametic and genotypic disequilibria and their standard errors under category (6f) are exemplified in Table 9, and results for the five nuclear loci are summarized in Table 10. None of the other categories fits the data by the *G*-test. All nuclear-mitochondrial disequilibria are thus highly significant. Results are consistent with the hypothesis of nonrandom mating (limited hybridization) with strong directionality such that those interspecific matings which do occur involve primarily *H. cinerea* males with *H. gratiosa* females.

An empirical example more closely approximating a random mating situation involves two geographic subspecies of bluegill sunfish (*Lepomis macrochirus macrochirus* and *L. m. purpurescens*) which hybridize in parts of Georgia. In one small north-Georgia lake, a sample of 151 bluegill was assayed for allozyme genotype at two unlinked and diagnostic nuclear loci, and for the distinctive *macrochirus* and *purpurescens*

TABLE 9

Estimation of disequilibria between the nuclear albumin locus and mitochondrial genotypes in a hybrid population of treefrogs

Mitochondria	Albumin *AA*	Albumin *Aa*	Albumin *aa*	Total
M	$\hat{u}_1 = 0.413$	$\hat{v}_1 = 0.036$	$\hat{w}_1 = 0.016$	$\hat{x} = 0.465$
m	$\hat{u}_2 = 0.066$	$\hat{v}_2 = 0.177$	$\hat{w}_2 = 0.292$	$\hat{y} = 0.535$
	$\hat{u} = 0.479$	$\hat{v} = 0.213$	$\hat{w} = 0.308$	1.000

$\hat{p} = \hat{u} + \frac{1}{2}\hat{v} = 0.585$; $\hat{q} = 1 - \hat{p} = 0.415$

$\hat{D}_1 = \hat{u}_1 - \hat{u}\hat{x} = 0.413 - 0.223 = 0.190 \pm 0.009$

$\hat{D}_2 = \hat{v}_1 - \hat{v}\hat{x} = 0.036 - 0.099 = -0.063 \pm 0.011$

$\hat{D}_3 = \hat{w}_1 - \hat{w}\hat{x} = 0.016 - 0.143 = -0.127 \pm 0.010$

$\hat{D} = \hat{u}_1 + \frac{1}{2}\hat{v}_1 - \hat{p}\hat{x} = 0.413 - 0.273 = 0.159 \pm 0.008$

(check: $\hat{D} = \hat{D}_1 + \frac{1}{2}\hat{D}_2 = 0.159$)

Standard Errors (SE)

$I^{-1}/N = (X^TAX)^{-1}/N$:

$\hat{u}$	$\hat{v}$	$\hat{x}$	$\hat{D}_1$	$\hat{D}_2$
0.00082	−0.00033	0.00062	0.00003	−0.00003 $\hat{u}$
	0.00055	−0.00021	−0.00003	−0.00012 $\hat{v}$
		0.00082	0.00004	−0.00001 $\hat{x}$
			0.00009	−0.00005 $\hat{D}_1$
				0.00012 $\hat{D}_2$

$\text{estSE}(\hat{D}_1) = (\text{Var}(\hat{D}_1))^{1/2} = (0.00009)^{1/2} = 0.009$

$\text{estSE}(\hat{D}_2) = (\text{Var}(\hat{D}_2))^{1/2} = (0.00012)^{1/2} = 0.011$

$\text{estSE}(\hat{D}_3) = (\text{Var}(\hat{D}_1) + \text{Var}(\hat{D}_2) + 2\,\text{Cov}(\hat{D}_1, \hat{D}_2))^{1/2}$
$= (0.00009 + 0.00012 - 2(0.00005))^{1/2} = 0.010$

$\text{estSE}(\hat{D}) = (\text{Var}(\hat{D}_1) + \frac{1}{4}\,\text{Var}(\hat{D}_2) + 2/2\,\text{Cov}(\hat{D}_1, \hat{D}_2))^{1/2}$
$= (0.00009 + \frac{1}{4}(0.00012) - 0.00005)^{1/2} = 0.008$

The body of the table consists of genotype frequencies in 305 individuals [from Lamb and Avise (1986)]. *AA/M* is characteristic of "pure" *H. cinerea*; *aa/m* of *H. gratiosa*.

mitochondrial genotypes (AVISE *et al.* 1984). Nuclear-mitochondrial disequilibria calculated from these data are summarized in Table 10. The *Es-3* and *Got-2* loci showed small but marginally significant values of D_2 and D_3, respectively. Overall, in *G*-tests of goodness-of-fit to the random-mating expectations (6a), probability levels were 0.05 and 0.06 for the two loci. Thus neither locus provides strong evidence against the random-mating hypothesis.

The remaining categories of disequilibria listed in Table 8 require other forms of nonrandom mating. We know of no real nuclear-cytoplasmic data sets available to exemplify these outcomes, but hypothetical cases can be imagined. For example, in a hybrid zone in which *AA/M* is characteristic of one species and *aa/m* of the other, we might realistically observe partial assortative matings of parentals and no directionality to those interspecific crosses which do occur. Then D_2 alone would be zero, under the appropriate initial conditions. It is also possible that females of one of the species have developed strong premating isolating barriers, while females of the other species mate nearly at random, and furthermore, that hybrids preferentially backcross to the less discriminating species.

TABLE 10

Empirical nuclear-cytoplasmic disequilibria in hybrid populations of *Hyla* treefrogs[a] and *Lepomis macrochirus* sunfish[b]

Taxa	Nuclear locus	Nuclear-cytoplasmic disequilibria			
		D	D_1	D_2	D_3
H. cinerea/H. gratiosa hybrid population	*Alb*	0.159 ± 0.008	0.190 ± 0.009	−0.063 ± 0.011	−0.127 ± 0.010
	Pgi	0.187 ± 0.007	0.221 ± 0.007	−0.067 ± 0.011	−0.154 ± 0.010
	Ldh	0.176 ± 0.008	0.202 ± 0.009	−0.053 ± 0.011	−0.150 ± 0.010
	Pep	0.178 ± 0.008	0.210 ± 0.008	−0.063 ± 0.011	−0.147 ± 0.010
	Mdh	0.172 ± 0.007	0.206 ± 0.008	−0.067 ± 0.011	−0.139 ± 0.005
L. macrochirus macrochirus/L. m. purpurescens hybrid population	*Es-3*	−0.001 ± 0.014	−0.026 ± 0.016	0.050 ± 0.020	−0.024 ± 0.018
	Got-2	0.024 ± 0.013	0.008 ± 0.016	0.033 ± 0.020	−0.040 ± 0.017

[a] Data from LAMB and AVISE (1986).
[b] Data from AVISE *et al.* (1984).

TABLE 11

Hypothetical examples of data structures producing other categories of disequilibrium outcomes

Disequilibria	Mitochondrial genotype	Nuclear genotype		
		AA	*Aa*	*aa*
(H_1) $D_1 = 0$; $D = \frac{1}{2}D_2 = -\frac{1}{2}D_3 \neq 0$	*M*	25	45	5
	m	25	5	45
(H_2) $D_2 = 0$; $D = D_1 = -D_3 \neq 0$	*M*	45	25	5
	m	5	25	45
(H_3) $D_3 = 0$; $D = \frac{1}{2}D_1 = -\frac{1}{2}D_2 \neq 0$	*M*	5	45	25
	m	45	5	25
(H_D) $D = 0$; $D_1 = D_3 = -\frac{1}{2}D_2 \neq 0$	*M*	15	45	15
	m	35	5	35

The body of each table consists of genotypic counts in samples of 150 individuals, in which $x = y = 0.5$, and $u = v = w = 0.33$. Actual disequilibria and their standard errors are presented in Figure 2.

Then we conjecture that only D_1 (or D_3) and D_2 might be large in magnitude. Finally, a situation could arise, in principle, in which genotypic disequilibria exist but gametic disequilibrium is zero. This would arise under the mixed-mating model if nuclear allele frequencies were identical in the two cytotypes. Numerical examples of these various possibilities are presented in Table 11.

The categories of disequilibria listed in Table 8 can thus be viewed as a series of hypotheses about the mating system in hybrid zones. As pictured in Figure 2, there is a natural hierarchy to these hypotheses, beginning with the simplest of random mating, and ending with nonrandom mating in which females of only one species tend to hybridize. For each hypothesis, a G-test of goodness-of-fit can be computed (FIENBERG 1977). For example, under H_R there are five independent counts and three independent parameters estimated, leaving two degrees of freedom. For H_{NR}, there are no degrees of freedom because five parameters are estimated from the genotypic counts. For the middle tier of hypotheses, there is one degree of freedom (four parameters estimated from genotypic counts). Thus, taking the differences of G-statistics between tiers of hypotheses (and differences in their corresponding degrees of freedom) allows sequential testing of the hypotheses by their complexity. Such a conceptual design parallels that developed for selection component analysis by CHRISTIANSEN and FRYDENBERG (1973).

Other considerations about cytonuclear disequilibria: Although static descriptions of cytonuclear disequilibria may lead to inferences about the evolutionary forces, including mating system responsible, it must also be remembered that the magnitudes (and in some cases signs) of the allelic and genotypic disequilibria can change in time-dependent fashion under a given set of evolutionary forces (Figure 1). Thus for example, D and D_i could all be near zero in a very young, random-mating hybrid swarm, or in a much older hybrid population with very strong but imperfect positive assortative mating. Furthermore, the dynamical behavior of the cytonuclear disequilibria are to a considerable extent influenced by the particular models assumed for the genetic basis of the mating system. In the case of hybrid zones, additional relevant concerns (which we will pursue elsewhere) include the pattern of disequilibria at the outset of hybridization, whether the hybrid population was closed to new recruitment from the parental species, and whether differential viability and/or fertility selection were also at work.

In general, associations between nuclear and cytoplasmic genotypes will be generated continually as gene pools differentiate, either among spatially subdivided conspecific populations, or among species. Epistatic selection involving interactions between particular nuclear genes and the cytoplasm may further contribute to disequilibria. The effects on disequilibria of gene pool differentiation due to drift or historical considerations, or to mating patterns, might in principle be distinguishable from effects due to epistatic selection *per se*—the former would be ex-

pected to generate concordance in the patterns of cytonuclear disequilibria for many unlinked (and functionally unrelated) nuclear genes (as in our Hyla example; Table 10), while the latter might generate consistent disequilibrium involving only the target nuclear gene (and loci linked to it) with particular cytotypes. Nonetheless, epistasis and gene pool differentiation may seldom provide mutually exclusive explanations for observed disequilibria.

On the other hand, as is true for pairs of unlinked, nuclear genes, cytonuclear disequilibria will also tend to decay, at rates that are importantly influenced by the pattern of extinction and recolonization of populations in subdivided species, and by the mating system in populations or species exchanging genes. Consequently, at any point in time, observed cytonuclear associations will depend on the particular blend of forces acting to generate and decay disequilibria.

Work was supported by NSF grants BSR-8420803, BSR-8315821, and BSR-8603775. We wish to thank CHRIS WILLIAMS for writing the IRLS programs for estimating the disequilibria, and ANDY CLARK and BRUCE WEIR for helpful comments on an earlier draft. We also wish to acknowledge an anonymous reviewer for suggesting the *d* measure.

LITERATURE CITED

AVISE, J. C., 1986 Mitochondrial DNA and the evolutionary genetics of higher animals. Philos. Trans. R. Soc. **312:** 325–342.

AVISE, J. C., E. BERMINGHAM, L. G. KESSLER and N. C. SAUNDERS, 1984 Characterization of mitochondrial DNA variability in a hybrid swarm between subspecies of bluegill sunfish (*Lepomis macrochirus*). Evolution **38:** 931–941.

BIRKY, C. W., JR., T. MARUYAMA and P. A. FUERST, 1983 An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. Genetics **103:** 513–527.

BROWN, W. M., 1983 Evolution of animal mitochondrial DNA. pp. 62–88. In: *Evolution of Genes and Proteins*, Edited by M. NEI and R. K. KOEHN. Sinauer, Sunderland, Massachusetts.

BURN, R., 1982 Loglinear models with composite link functions in genetics. pp. 144–154. In: *Proceedings of the International Conference on Generalized Linear Models*, Edited by R. GILCHRIST. Springer-Verlag, New York.

CHARLESWORTH, B. and D. CHARLESWORTH, 1973 A study of linkage disequilibrium in populations of *Drosophila melanogaster*. Genetics **73:** 351–359.

CHOMYN, A., P. MARIOTTINI, M. W. J. CLEETER, C. I. RAGAN, A. MATSUNO-YAGI, Y. HATEFI, R. F. DOOLITTLE and G. ATTARDI, 1985 Six unidentified reading frames of human mitochondrial DNA encode components of the respiratory-chain NADH dehydrogenase. Nature **314:** 592–597.

CHRISTIANSEN, F. B. and O. FRYDENBERG, 1973 Selection component analysis of natural polymorphisms using population samples including mother-offspring combinations. Theor. Popul. Biol. **4:** 425–445.

CLARK, A. G., 1984 Natural selection with nuclear and cytoplasmic transmission. I. A deterministic model. Genetics **107:** 679–701.

CLARK, A. G., 1985 Natural selection with nuclear and cytoplasmic transmission. II. Tests with Drosophila from diverse populations. Genetics **111:** 97–112.

CLEGG, M. T., 1980 Measuring plant mating systems. BioScience **30:** 814–818.

CLEGG, M. T., J. F. KIDWELL and C. R. HORCH, 1980 Dynamics of correlated genetic systems. V. Rates of decay of linkage disequilibria in experimental populations of *Drosophila melanogaster*. Genetics **94:** 217–234.

CURTIS, S. E. and M. T. CLEGG, 1984 Molecular evolution of chloroplast DNA sequences. Mol. Biol. Evol. **1:** 291–301.

DEWEY, R. E., C. S. LEVING III and D. H. TIMOTHY, 1986 Novel recombinations in the maize mitochondrial genome produce a unique transcriptional unit in the Texas male-sterile cytoplasm. Cell **44:** 439–449.

ENDLER, J. A., 1977 *Geographic Variation, Speciation, and Clines.* Princeton University Press, Princeton, New Jersey.

FERRIS, S. D., R. D. SAGE, C.-M. HUANG, J. T. NIELSEN, U. RITTE and A. C. WILSON, 1983 Flow of mitochondrial DNA across a species boundary. Proc. Natl. Acad. Sci. USA **80:** 2290–2294.

FIENBERG, S. E., 1977 *The Analysis of Cross-Classified Categorical Data.* MIT Press, Cambridge, Massachusetts.

GREEN, P. J., 1984 Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives (with Discussion). J. R. Statist. Soc. **46B:** 149–192.

GREGORIUS, H.-R. and M. D. ROSS, 1984 Selection with gene-cytoplasm interactions. I. Maintenance of cytoplasm polymorphisms. Genetics **107:** 165–178.

GRIVELL, L. A., 1983 Mitochondrial DNA. Sci. Am. **248:** 78–89.

HEDRICK, P. W., 1983 *Genetics of Populations.* Van Nostrand Reinhold, New York.

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

HOFFMAN, A. A., M. TURELLI and G. M. SIMMONS, 1986 Unidirectional incompatibility between populations of *Drosophila simulans*. Evolution **40:** 692–701.

LAMB, T. and J. C. AVISE, 1986 Directional introgression of mitochondrial DNA in a hybrid population of treefrogs: the influence of mating behavior. Proc. Natl. Acad. Sci. USA **83:** 2526–2530.

LANGLEY, C. H., Y. N. TOBARI and K.-I. KOJIMA, 1974 Linkage disequilibrium in natural populations of *Drosophila melanogaster*. Genetics **78:** 921–936.

MULLER, P. P., M. K. REIF, S. ZONGHOU, C. SENGSTAG, T. L. MASON and T. D. FOX, 1984 A nuclear mutation that post-transcriptionally blocks accumulation of a yeast mitochondrial gene product can be suppressed by a mitochondrial gene rearrangement. J. Mol. Biol. **175:** 431–452.

NEIGEL, J. E. and J. C. AVISE, 1986 Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. pp. 515–534. In: *Evolutionary Processes and Theory*, Edited by E. NEVO and S. KARLIN. Academic Press, New York.

O'BRIEN, T. W., N. D. DENSLOW, T. O. HARVILLE, R. A. HESSLER and D. E. MATTHEWS, 1980 Functional and structural roles of proteins in mammalian mitochondrial ribosomes. pp. 301–305. In: *The Organization and Expression of the Mitochondrial Genome*, Edited by A. M. KROON and C. SACCONE. Elsevier/North Holland, New York.

O'DONALD, P., 1960 Assortative mating in a population in which two alleles are segregating. Heredity **15:** 389–396.

ROSS, M. D. and H.-R. GREGORIUS, 1985 Selection with gene-cytoplasm interactions. II. Maintenance of gynodioecy. Genetics **109:** 427–439.

SMOUSE, P. E., 1974 Likelihood analysis of recombinational disequilibrium in multiple-locus gametic frequencies. Genetics **76:** 557–565.

SPOLSKY, C. and T. UZZELL, 1984 Natural interspecies transfer

of mitochondrial DNA in amphibians. Proc. Natl. Acad. Sci. USA **81:** 5802–5805.

TAKAHATA, N. and M. SLATKIN, 1983 Evolutionary dynamics of extranuclear genes. Genet. Res. **42:** 257–265.

WADE, M. J. and L. STEVENS, 1985 Microorganism mediated reproductive isolation in flour beetles (genus *Tribolium*). Science **227:** 527–528.

WATSON, G. S. and E. CASPARI, 1960 The behavior of cytoplasmic pollen sterility in populations. Evolution **14:** 56–63.

WEIR, B. S., 1979 Inferences about linkage disequilibrium. Biometrics **35:** 235–254.

WEIR, B. S. and S. R. WILSON, 1986 Log-linear models for linked loci. Biometrics **42:** 665–669.

WEIR, B. S., R. W. ALLARD, and A. L. KAHLER, 1972 Analysis of complex allozyme polymorphisms in a barley population. Genetics **72:** 505–523.

Communicating editor: B. S. WEIR