Heredity **76** (1996) 207–214 *Received 11 January 1995*

Constraints and normalized measures for cytonuclear disequilibria

MARJORIE A. ASMUSSEN* & CHRISTOPHER J. BASTEN†
Department of Genetics, University of Georgia, Athens, GA 30602-7223, U.S.A.

The full bounds are derived for cytonuclear disequilibria in two-locus systems with an arbitrary number of alleles at the cytoplasmic and nuclear markers. The associated marginal frequencies constrain the nonrandom associations between cytoplasmic alleles and nuclear genotypes in the same way that the allele frequencies constrain the linkage disequilibrium between two nuclear loci. Additional constraints are imposed on the nonrandom associations between cytoplasmic and nuclear alleles, however, by the marginal frequencies of nuclear genotypes carrying either two or no copies of the associated nuclear allele. These bounds are analysed and used to define normalized measures of cytonuclear disequilibria, whose practical utility is illustrated through applications to two sets of recent nuclear–mitochondrial data.

Keywords: cpDNA, cytonuclear disequilibria, disequilibrium measures, mtDNA, multiallelic markers, normalized disequilibria.

Introduction

The observed pattern of nonrandom associations between nuclear and cytoplasmic genes can contain valuable information about the evolutionary history of natural populations. This is particularly true of hybrid zones, where cytonuclear data have provided estimates of rates of assortative mating and gene flow by the parental species that are more sensitive than, and often unobtainable from, nuclear systems alone (Arnold *et al.*, 1988; Asmussen *et al.*, 1989; Avise *et al.*, 1990). Cytonuclear disequilibria can also be utilized more generally to detect and estimate migration, admixture and population subdivision (Asmussen & Arnold, 1991) and to decompose gene flow in plant populations into haploid (pollen) and diploid (seed) components (Asmussen & Schnabel, 1991; Schnabel & Asmussen, 1992). To facilitate such applications, Asmussen & Basten (1994) recently developed statistical guidelines for the experimental design and analysis of population surveys seeking to draw evolutionary inferences from observed patterns of cytonuclear disequilibria. The results provide maximum likelihood disequilibrium estimators and their standard errors, simple asymptotic tests for the null hypothesis of random associations, and analytical formulae for calculating approximate minimum sample sizes to detect the cytonuclear associations defined by Asmussen *et al.* (1987).

To understand and interpret the biological significance of observed cytonuclear disequilibria, it is critical to know the range of permissible values for such nonrandom associations. These bounds were specified only indirectly by Asmussen *et al.* (1987) because of the many confounding interrelationships among the various disequilibria. The present paper provides the derivation and analysis of the complete bounds on the nonrandom associations between the alleles at a haploid cytoplasmic locus and the alleles and genotypes at a diploid nuclear locus, taking into account all the interrelations among the different measures. The original case of diallelic markers is considered first, followed by the definition of cytonuclear disequilibria and their bounds for the general case of multiallelic markers with an arbitrary number of alleles at each locus. It is then shown how these full bounds can be used to calculate the maximal levels of cytonuclear disequilibria possible in a population and to define normalized disequilibrium measures which take these constraints into account. Applications to a recent nuclear–mitochondrial survey illustrate practical implications for data analysis.

*Correspondence.
†Present address: Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203, U.S.A.

Diallelic markers

Cytonuclear variables

Consider first the original case from Asmussen *et al.* (1987) with two alleles (A, a) at the nuclear locus and two alleles (M, m) at the cytoplasmic locus. Because of its greater clarity, the present treatment uses the notation introduced by Asmussen & Basten (1994), in which P denotes a frequency and D a disequilibrium measure, with the associated nuclear allele or genotype superscripted and the cytotype subscripted. The frequencies of the six possible cytonuclear genotypes are given in Table 1 along with the marginal genotypic frequencies at the individual loci. From these one obtains the marginal frequency of nuclear allele A, $P^A = P^{AA} + \frac{1}{2}P^{Aa}$, and the joint allelic frequency, $P_M^A = P_M^{AA} + \frac{1}{2}P_M^{Aa}$. The latter can be viewed loosely as the frequency of gametes with cytotype M and nuclear allele A, but is formally defined as the probability that a random individual from the population has cytotype M and that a single (randomly sampled) allele at its nuclear locus is A.

The most natural disequilibrium measures for cytonuclear systems are the *genotypic disequilibria*

$$\begin{aligned} D_M^{AA} &= P_M^{AA} - P^{AA}P_M \\ D_M^{Aa} &= P_M^{Aa} - P^{Aa}P_M \\ D_M^{aa} &= P_M^{aa} - P^{aa}P_M \end{aligned} \tag{1}$$

which quantify the nonrandom associations between the cytotypes and each nuclear genotype (Table 1). One can also measure nonrandom associations between nuclear and cytoplasmic alleles by the *allelic disequilibrium*,

$$D_M^A = P_M^A - P^A P_M \tag{2}$$

which is the direct analogue of nuclear linkage disequilibrium. As shown by Asmussen *et al.* (1987) for the equivalent measures, $D_1 = D_M^{AA}$, $D_2 = D_M^{Aa}$, $D_3 = D_M^{aa}$ and $D = D_M^A$, these four disequilibria reduce to just two independent measures as a result of the interrelationships

$$D_M^{AA} + D_M^{Aa} + D_M^{aa} = 0 \tag{3}$$

and

$$D_M^A = D_M^{AA} + \tfrac{1}{2}D_M^{Aa}. \tag{4}$$

It is nonetheless useful to treat all four because their joint sign pattern can itself encode much useful information about the evolutionary history of a population.

Bounds on cytonuclear disequilibria

The cytonuclear disequilibria are individually constrained by the marginal frequencies, as shown in Table 2. The derivation of these complete bounds is based on the primary constraints on the genotypic disequilibria

$$\begin{aligned} -P^{AA}P_M \le D_M^{AA} \le P^{AA}(1-P_M) \\ -P^{Aa}P_M \le D_M^{Aa} \le P^{Aa}(1-P_M) \\ -P^{aa}P_M \le D_M^{aa} \le P^{aa}(1-P_M) \end{aligned} \tag{5}$$

given by Asmussen *et al.* (1987), which follow from the basic cytonuclear parameterization in Table 1. The full constraints on the genotypic associations follow by applying eqn (5) to the interrelationship in eqn (3). The disequilibrium involving AA homozygotes, for instance, can be written as $D_M^{AA} = -(D_M^{Aa} + D_M^{aa})$, which with eqn (5) yields the two additional constraints

$$D_M^{AA} \le P^{Aa}P_M + P^{aa}P_M = (1-P^{AA})P_M \tag{6}$$

and

$$\begin{aligned} D_M^{AA} &\ge -P^{Aa}(1-P_M) - P^{aa}(1-P_M) \\ &= -(1-P^{AA})(1-P_M). \end{aligned} \tag{7}$$

The constraints on the allelic association are similarly obtained by applying the primary constraints in eqn (5) to its three possible decompositions

$$D_M^A = D_M^{AA} + \tfrac{1}{2}D_M^{Aa}$$

Table 1 Basic cytonuclear parameterization for diallelic markers

	Nuclear genotypes			
Cytotype	AA	Aa	aa	Total
M	$P_M^{AA} = P^{AA}P_M + D_M^{AA}$	$P_M^{Aa} = P^{Aa}P_M + D_M^{Aa}$	$P_M^{aa} = P^{aa}P_M + D_M^{aa}$	P_M
m	$P_m^{AA} = P^{AA}P_m - D_M^{AA}$	$P_m^{Aa} = P^{Aa}P_m - D_M^{Aa}$	$P_m^{aa} = P^{aa}P_m - D_M^{aa}$	P_m
Total	P^{AA}	P^{Aa}	P^{aa}	1

Table 2 Bounds on the diallelic disequilibria from the marginal frequencies†

D	Lower bound (min D)	Upper bound (max D)
D_M^{AA}	$-\min[P^{AA}P_M, (1-P^{AA})(1-P_M)]$	$\min[P^{AA}(1-P_M), (1-P^{AA})P_M]$
D_M^{A}	$-\min[P^{A}P_M, (1-P^{A})(1-P_M), \frac{1}{2}P^{AA}P_M+\frac{1}{2}P^{aa}(1-P_M)]$	$\min[P^{A}(1-P_M), (1-P^{A})P_M, \frac{1}{2}P^{AA}(1-P_M)+\frac{1}{2}P^{aa}P_M]$

†D_M^{Aa} (D_M^{aa}) bounds are equivalent to those for D_M^{AA} *with* AA replaced by Aa (aa).

Table 3 Usage of the marginal bounds for diallelic cytonuclear disequilibria†

D	min D	Condition	Usage (%)	$D = \min D$
D_M^{AA}	$-P^{AA}P_M$	$P_M \le 1-P^{AA}$	62.5	$P_M^{AA} = 0$
	$-(1-P^{AA})(1-P_M)$	$P_M \ge 1-P^{AA}$	37.5	$P_m^{Aa} = P_m^{aa} = 0$
D_M^{Aa}	$-P^{Aa}P_M$	$P_M \le 1-P^{Aa}$	75	$P_M^{Aa} = 0$
	$-(1-P^{Aa})(1-P_M)$	$P_M \ge 1-P^{Aa}$	25	$P_m^{AA} = P_m^{aa} = 0$
D_M^{A}	$-P^{A}P_M$	$P_M \le P^{aa}$	37.5	$P_M^{AA} = P_M^{Aa} = 0$
	$-\frac{1}{2}[P^{AA}P_M+P^{aa}(1-P_M)]$	$P^{aa} \le P_M \le 1-P^{AA}$	25	$P_M^{AA} = P_m^{aa} = 0$
	$-(1-P^{A})(1-P_M)$	$P_M \ge 1-P^{AA}$	37.5	$P_m^{Aa} = P_m^{aa} = 0$
D	max D	Condition	Usage (%)	$D = \max D$
D_M^{AA}	$(1-P^{AA})P_M$	$P_M \le P^{AA}$	37.5	$P_M^{Aa} = P_M^{aa} = 0$
	$P^{AA}(1-P_M)$	$P_M \ge P^{AA}$	62.5	$P_m^{AA} = 0$
D_M^{Aa}	$(1-P^{Aa})P_M$	$P_M \le P^{Aa}$	25	$P_M^{AA} = P_M^{aa} = 0$
	$P^{Aa}(1-P_M)$	$P_M \ge P^{Aa}$	75	$P_m^{Aa} = 0$
D_M^{A}	$(1-P^{A})P_M$	$P_M \le P^{AA}$	37.5	$P_M^{Aa} = P_M^{aa} = 0$
	$\frac{1}{2}[P^{AA}(1-P_M)+P^{aa}P_M]$	$P^{AA} \le P_M \le 1-P^{aa}$	25	$P_m^{AA} = P_M^{aa} = 0$
	$P^{A}(1-P_M)$	$P_M \ge 1-P^{aa}$	37.5	$P_m^{AA} = P_m^{Aa} = 0$

†D_M^{aa} entries are equivalent to those for D_M^{AA} with AA replaced by aa in columns 1–3 and interchanged with aa in column 5.

$$= \tfrac{1}{2}D_M^{AA} - \tfrac{1}{2}D_M^{aa} \qquad (8)$$
$$= -\tfrac{1}{2}D_M^{Aa} - D_M^{aa}$$

prescribed by the interrelations in eqns (3) and (4). It can be proven analytically that the secondary constraints such as eqns (6) and (7) place no further restrictions on D_M^A, as they should not, because they reflect the interrelationships among the genotypic associations which are accounted for by the three-fold decomposition in eqn (8).

The exact conditions under which each of the multiple constraints is the actual lower or upper bound are given in the first three columns of Table 3. The far right-hand column shows that an observed genotypic disequilibrium equals a given constraint only when one or two specific cytonuclear genotypes are absent. Schematically, this corresponds to having an empty cell in the 2×2 table that remains after collapsing Table 1 to two nuclear categories: the associated nuclear genotype vs. all others. An observed allelic disequilibrium is at its minimum or maximum possible value when in Table 1 either two adjacent cells in a row are empty (i.e. a homozygote and heterozygote with the same cytotype are absent) or when two cells at diagonally opposite corners are empty (i.e. AA/M and aa/m are absent, or AA/m and aa/M are absent). When this occurs two of the genotypic disequilibria are also maximal or minimal.

Each cytonuclear disequilibrium has a maximal possible range of $[-0.25, 0.25]$, which is achieved only when the marginal frequencies of the cytotypes and the associated nuclear components (genotypes for D_M^{AA}, D_M^{Aa}, D_M^{aa}; alleles for D_M^A) are all 0.5; there is a commensurate reduction in the negative or positive range whenever either of these marginal frequencies is nearer 0 or 1. Because of the common link to 2×2 tables, the ranges of the genotypic disequilibria are determined by the associated

marginal frequencies in the same way that the linkage disequilibrium between two nuclear loci is constrained by the allele frequencies (Lewontin, 1964). The actual admissible range for the allelic disequilibrium, however, varies from the same to much smaller than that for nuclear linkage disequilibrium (even if $P^A = P_M = 0.5$), because the marginal nuclear genotypic frequencies impose additional constraints on cytonuclear allelic associations beyond the usual ones imposed by the allele frequencies.

To determine the practical significance of the additional constraints on D_M^A, its lower and upper bounds were systematically calculated for 10^9 random sets of marginal frequencies. The latter were obtained by first randomly generating P^A and P_M from the interval [0,1], and then selecting the nuclear Hardy–Weinberg disequilibrium, $D^A = P^{AA} - (P^A)^2$, at random from its permissible range (Weir, 1990, p. 73). The additional upper constraint from the nuclear genotypic frequencies was found to decrease the positive range of D_M^A for 25 per cent of the combinations of P^A, P_M and D^A; for the cases in which this occurred the positive range was reduced by 35 per cent, on average, relative to that based on the allelic constraints alone. The additional lower constraint had an equivalent effect on the negative range. The fourth column of Table 3 summarizes the usage of the relevant bounds for all four disequilibria.

The bounds in Tables 2 and 3 are the usual standard for interpreting observed cytonuclear disequilibria. For completeness, however, it should be noted that if a disequilibrium value is known in addition to the marginal frequencies, further constraints are placed on the remaining associations through the interrelations in eqns (3) and (4). If D_M^{AA} is known, for instance, the bounds on the other three disequilibria change to those in Table 4. If two or more disequilibria are known, then all others are uniquely determined.

Multiallelic markers

Cytonuclear parameterization

Consider now the general case with r alleles ($A_1, A_2, \ldots, A_r$) at the nuclear locus and m cytotypes ($M_1, M_2, \ldots, M_m$). There are then $mr(r+1)/2$ joint cytonuclear genotypes with frequencies denoted as in Table 5. Note that for simplicity the notation has been condensed somewhat by using only the indices of the nuclear alleles and cytotypes, with superscripts still denoting the nuclear component and subscripts the cytotype. For instance, P_k^{ij} denotes the frequency of A_iA_j/M_k individuals, while P_k denotes the marginal frequency of the M_k cytotype, P^{ij} the marginal frequency of A_iA_j individuals, and

$$P^i = P^{ii} + \tfrac{1}{2}\sum_{j \neq i} P^{ij} \qquad (9)$$

the frequency of nuclear allele A_i. The summation term in eqn (9) gives the frequency of A_i heterozygotes under the convention that the nuclear allele indices are unordered such that $P^{ij} = P^{ji}$ (and $P_k^{ij} = P_k^{ji}$) when i≠j.

For each cytonuclear genotype A_iA_j/M_k there is a *genotypic disequilibrium*

Table 5 Frequencies of multiallelic cytonuclear genotypes

Cytotype	A_1A_1	⋯	A_iA_j	⋯	A_rA_r	Total
M_1	P_1^{11}	⋯	P_1^{ij}	⋯	P_1^{rr}	P_1
⋮		⋱				
M_k	P_k^{11}	⋯	P_k^{ij}	⋯	P_k^{rr}	P_k
⋮			⋱			
M_m	P_m^{11}	⋯	P_m^{ij}	⋯	P_m^{rr}	P_m
Total	P^{11}	⋯	P^{ij}	⋯	P^{rr}	1.0

Table 4 Bounds on the remaining cytonuclear disequilibria when D_M^{AA} is known†

	min D		max D	
D	Bounds	Condition	Bounds	Condition
D_M^{Aa}	$-D_M^{AA} - P^{aa}(1-P_M)$	$D_M^{AA} \leq (1-P^{AA})P_M - P^{aa}$	$P^{Aa}(1-P_M)$	$D_M^{AA} \leq (1-P^{AA})P_M - P^{Aa}$
	$-P^{Aa}P_M$	$D_M^{AA} \geq (1-P^{AA})P_M - P^{aa}$	$-D_M^{AA} + P^{aa}P_M$	$D_M^{AA} \geq (1-P^{AA})P_M - P^{Aa}$
D_M^A	$\frac{1}{2}D_M^{AA} - \frac{1}{2}P^{aa}(1-P_M)$	$D_M^{AA} \leq (1-P^{AA})P_M - P^{aa}$	$D_M^{AA} + \frac{1}{2}P^{Aa}(1-P_M)$	$D_M^{AA} \leq (1-P^{AA})P_M - P^{Aa}$
	$D_M^{AA} - \frac{1}{2}P^{Aa}P_M$	$D_M^{AA} \geq (1-P^{AA})P_M - P^{aa}$	$\frac{1}{2}D_M^{AA} + \frac{1}{2}P^{aa}P_M$	$D_M^{AA} \geq (1-P^{AA})P_M - P^{Aa}$

†D_M^{aa} formulae are equivalent to those for D_M^{Aa} *with* Aa and aa interchanged.

$$D_k^{ij} = \text{freq}(A_iA_j/M_k) - \text{freq}(A_iA_j)\text{freq}(M_k) \quad (10)$$
$$= P_k^{ij} - P^{ij}P_k$$

which quantifies the nonrandom association between nuclear genotype A_iA_j and cytotype M_k. These measures provide the basic cytonuclear parameterization

$$P_k^{ij} = P^{ij}P_k + D_k^{ij} \text{ for all } k = 1, 2, \ldots, m;\ 1 \le i \le j \le r \quad (11)$$

in terms of the genotypic associations and the marginal frequencies of the cytotypes and nuclear genotypes. A set of *allelic disequilibria* can be similarly defined by

$$D_k^i = P_k^i - P^iP_k \quad (12)$$

which measure the nonrandom associations between each nuclear allele A_i and cytotype M_k, where the joint allelic frequency

$$P_k^i = P_k^{ii} + \tfrac{1}{2}\sum_{j \ne i} P_k^{ij} \quad (13)$$

is formally defined as the probability that a random individual has cytotype M_k and a single (randomly sampled) allele from its nuclear locus is A_i.

Altogether, these disequilibria reduce to $(m-1)[\frac{1}{2}r(r+1)-1]$ independent measures which are connected through three basic interrelations. Two of these are specific to the genotypic associations: for each nuclear genotype A_iA_j, the sum of genotypic associations across cytotypes is 0, as is the sum of genotypic associations across nuclear genotypes for each cytotype, M_k. That is,

$$\sum_{k=1}^{m} D_k^{ij} = 0 \text{ for all } 1 \le i \le j \le r \quad (14)$$

and

$$\sum_{i=1}^{r}\sum_{j=i}^{r} D_k^{ij} = 0 \text{ for all } k = 1,2, \ldots, m. \quad (15)$$

The third basic interrelation is that each allelic disequilibrium can be decomposed into a linear combination of the associated genotypic disequilibria

$$D_k^i = D_k^{ii} + \tfrac{1}{2}\sum_{j \ne i} D_k^{ij}. \quad (16)$$

The last two relations in eqns (15) and (16) are the multiallelic analogues of the diallelic relations in eqns (3) and (4), whereas eqn (14) corresponds to restricting attention in the diallelic case to disequilibria involving cytotype M because those involving cytotype m are their negatives (e.g. $D_m^{AA} = -D_M^{AA}$).

Bounds on cytonuclear disequilibria

The constraints on the multiallelic disequilibria are given in Table 6. These have the same form found in the diallelic case (Table 2) except for one noteworthy difference: the general formulae for multiallelic markers reveal that the added, genotypic constraints on each allelic association, D_k^i, come not just from the homozygous nuclear genotypes, which is the form for diallelic markers, but from the marginal frequencies of homozygotes for the nuclear allele in question and all nuclear genotypes without that allele.

The full bounds on each cytonuclear association are derived by a tedious but straightforward extension of the method used in the diallelic case. Those on the genotypic disequilibria are obtained by successively combining the interrelations in eqns (14) and (15) with the primary constraint on each genotypic disequilibrium

$$D_k^{ij} \ge -P^{ij}P_k, \quad (17)$$

where the latter follows from the basic parameterization in eqn (11) under the requirement that each cytonuclear genotype has non-negative frequency. The full bounds on each allelic association D_k^i then follow by applying the full constraints on each genotypic disequilibrium (Table 6) to the basic decomposition of D_k^i in eqn (16), as well as to all its secondary decompositions obtained by successively

Table 6 Bounds on multiallelic cytonuclear disequilibria from the marginal frequencies†

D	Lower bound (min D)	Upper bound (max D)
D_k^{ij}	$-\min[P^{ij}P_k, (1-P^{ij})(1-P_k)]$	$\min[P^{ij}(1-P_k), (1-P^{ij})P_k]$
D_k^i	$-\min[P^iP_k, (1-P^i)(1-P_k),$ $\frac{1}{2}P^{ii}P_k + \frac{1}{2}P^*(1-P_k)]$	$\min[P^i(1-P_k), (1-P^i)P_k,$ $\frac{1}{2}P^{ii}(1-P_k) + \frac{1}{2}P^*P_k]$

†$P^* = 1 - \sum_{j=1}^{r} P^{ij}$ is the frequency of all nuclear genotypes without an A_i allele.

rewriting each genotypic disequilibrium measure in eqn (16) via the interrelations in eqn (14) and (15). The many possible decompositions of the allelic disequilibrium D_1^1 are provided as an example in the Appendix.

Discussion

As for nuclear linkage disequilibria, the biological interpretation of cytonuclear disequilibria is complicated by the fact that their admissible range of values is highly dependent on the associated marginal frequencies. This difficulty can now be circumvented by the full cytonuclear bounds, which allow observed cytonuclear associations to be judged relative to the largest level possible for a population with the given marginal frequencies. This can be formalized by calculating normalized cytonuclear disequilibrium measures analogous to the D' measure defined by Lewontin (1964) for nuclear linkage disequilibria, in which the observed disequilibrium D is divided by the maximum possible magnitude for a disequilibrium of that sign. In the present notation, the D' value for each cytonuclear disequilibrium D is defined by

$$D' = \begin{cases} \dfrac{D}{|\min D|} & \text{if } D<0 \\ \dfrac{D}{\max D} & \text{if } D\geq 0 \end{cases} \qquad (18)$$

where min D and max D are the cytonuclear bounds, calculated from Table 2 for diallelic markers and from Table 6 for multiallelic markers. The D' measures thus have the practical advantage of ranging from -1 to 1 for all combinations of nuclear and cytoplasmic frequencies, although as in the nuclear case (Hedrick, 1987; Lewontin, 1988) the values are not truly independent of these frequencies.

The practical utility of these results is illustrated by an application to recent data from an experimental hybrid zone of mosquitofish within Biosphere 2 (Scribner & Avise, 1994a). The experiment began in September 1991, with the introduction of approximately equal numbers of *Gambusia affinis* and *G. holbrooki* into Biosphere 2 just prior to its closure. Immediately following the reopening of the facility two years (roughly 4–6 *Gambusia* generations) later, individuals were collected and assayed for species-specific mitochondrial DNA markers and nuclear genotypes at five autosomal allozyme loci. During this two-year period, dramatic changes occurred in cytonuclear genotype frequencies in a pattern consistent with experimental populations outside Biosphere 2 (Scribner & Avise, 1994b), and suggestive of some degree of interspecific hybridization coupled with strong directional selection favouring *G. holbrooki* genotypes.

A full cytonuclear disequilibrium analysis is given here for the data involving the Peptidase-A (*Pep-A*) and Adenosine deaminase (*Ada*) allozyme loci for the 97 individuals collected from the 'freshwater marsh' population of Biosphere 2. Using the notation of Asmussen & Basten (1994), the counts of the joint cytonuclear genotypes (n_M^{AA}, n_M^{Aa}, n_M^{aa}, n_m^{AA}, n_m^{Aa}, n_m^{aa}), are (71, 7, 2, 1, 3, 13) for *Pep-A* and (75, 5, 0, 0, 4, 13) for *Ada*, where in each case *AA/M* is diagnostic for *G. holbrooki* and *aa/m* is diagnostic for *G. affinis*. Table 7 provides estimates for the marginal frequencies and cytonuclear disequilibria and the statistical significance of the disequilibria, calculated from the formulae of Asmussen & Basten (1994). Also shown are the normalized disequilibrium values calculated from eqn (18) and Table 2. For *Pep-A* the heterozygote disequilibrium D_M^{Aa} is 15.1 per cent of the maximal negative level for a population with the observed marginal frequencies and does not differ significantly from zero; the other three *Pep-A* disequilibria are both highly significantly different from zero and near (84–92 per cent) their maximum possible magnitudes. The results are strongly concordant for *Ada* except that D_M^{Aa} is marginally significantly different from zero, and because of the absence of the recombinant genotypes *aa/M* and *AA/m*, D_M^{AA} and D_M^{A} are at their maximum values and D_M^{aa} is at its minimum. Similar disequilibrium patterns were observed for the other three allozyme loci, where all cytonuclear associations except D_M^{Aa} were highly significant.

This first analysis of observed and normalized cytonuclear disequilibria for the *Gambusia* data serves to reinforce and amplify previous biological interpretations from this study. In particular, the closeness of most normalized disequilibria (except for the heterozygous genotypes) to their maximum levels suggests that processes have been at work to maintain pure *G. affinis* and *G. holbrooki* genotypes in the hybrid setting, notwithstanding the occasional appearance of interspecific recombinants. Furthermore, the underlying forces appear to extend genome-wide rather than being locus-specific because of the concordance in disequilibrium levels across all five independent allozyme loci assayed.

In general, normalized disequilibrium values add a valuable dimension to the analysis of cytonuclear data. In some cases there will be a close connection between the magnitude of the normalized disequili-

Table 7 Marginal frequency ($\tilde{P}$) and cytonuclear disequilibrium ($\tilde{D}$) estimates for two nuclear–mitochondrial data sets from an experimental hybrid zone of mosquitofish (Scribner & Avise, 1994a). In both cases $\tilde{P}_M = 0.82$

Nuclear type	*Pep–A*			*Ada*		
	$\tilde{P}$	$\tilde{D}$	$\tilde{D}'$	$\tilde{P}$	$\tilde{D}$	$\tilde{D}'$
AA	0.74	0.12**	0.92	0.77	0.14**	1.0
Aa	0.10	−0.01	−0.15	0.09	−0.03*	−0.33
aa	0.15	−0.11**	−0.84	0.13	−0.11**	−1.0
A	0.79	0.11**	0.88	0.82	0.12**	1.0

*Significant ($P = 0.0476$, exact test); **highly significant ($P \ll 0.01$).

bria and the statistical significance of the observed disequilibrium estimates, such as found here. However, this will not always hold, because the latter depends strongly on the sample size in addition to the population's cytonuclear structure. In fact, normalized disequilibria can be particularly informative in the contradictory cases, where the D value has nearly maximal magnitude for the population's marginal frequencies (D' near ± 1) but one fails to reject the null hypothesis of no disequilibrium; these disequilibria may have greater biological significance than suggested by the statistical test. Normalized disequilibrium measures thus complement, rather than replace, formal statistical analyses by enhancing the biological interpretation of observed cytonuclear associations.

Acknowledgements

We thank K. T. Scribner and J. C. Avise for generously providing their data. They, with R. D. Overath and two anonymous reviewers, also provided valuable suggestions for the manuscript. This research was supported in part by National Science Foundation grant DEB 92–10895 to M.A.A. and National Institutes of Health grant GM 45344 to North Carolina State University.

References

ARNOLD, J., ASMUSSEN, M. A. AND AVISE, J. C. 1988. An epistatic mating system model can produce permanent cytonuclear disequilibria in a hybrid zone. *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 1893–1896.

ASMUSSEN, M. A. AND ARNOLD, J. 1991. The effects of admixture and population subdivision on cytonuclear disequilibria. *Theor. Pop. Biol.*, **39**, 273–300.

ASMUSSEN, M. A. AND BASTEN, C. J. 1994. Sampling theory for cytonuclear disequilibria. *Genetics*, **138**, 1351–1363.

ASMUSSEN, M. A. AND SCHNABEL, A. 1991. Comparative effects of pollen and seed migration on the cytonuclear structure of plant populations. I. Maternal cytoplasmic inheritance. *Genetics*, **128**, 639–654.

ASMUSSEN, M. A., ARNOLD, J. AND AVISE, J. C. 1987. Definition and properties of disequilibrium statistics for associations between nuclear and cytoplasmic genotypes. *Genetics*, **115**, 755–768.

ASMUSSEN, M. A., ARNOLD, J. AND AVISE, J. C. 1989. The effects of assortative mating and migration on cytonuclear associations in hybrid zones. *Genetics*, **122**, 923–934.

AVISE, J. C., NELSON, W. S., ARNOLD, J., KOEHN, R. K., WILLIAMS, G. C. AND THORSTEINSSON, V. 1990. The evolutionary genetic status of Icelandic eels. *Evolution*, **44**, 1254–1262.

HEDRICK, P. W. 1987. Gametic disequilibrium measures: proceed with caution. *Genetics*, **117**, 331–341.

LEWONTIN, R. C. 1964. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics*, **49**, 49–67.

LEWONTIN, R. C. 1988. On measures of gametic disequilibrium. *Genetics*, **120**, 849–852.

SCHNABEL, A. AND ASMUSSEN, M. A. 1992. Comparative effects of pollen and seed migration on the cytonuclear structure of plant populations. II. Paternal cytoplasmic inheritance. *Genetics*, **132**, 253–267.

SCRIBNER, K. T. AND AVISE, J. C. 1994a. Cytonuclear genetics of experimental fish hybrid zones inside Biosphere 2. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 5066–5069.

SCRIBNER, K. T. AND AVISE, J. C. 1994b. Population cage experiments with a vertebrate: the temporal demography and cytonuclear genetics of hybridization in *Gambusia* fishes. *Evolution*, **48**, 155–171.

WEIR, B. S. 1990. *Genetic Data Analysis*. Sinauer Associates, Sunderland, MA.

Appendix: decompositions of the allelic disequilibrium D_1^1

The definition of D_k^i in eqn (12) immediately leads to the primary decomposition of D_1^1 in terms of the

genotypic disequilibria involving A_1

$$D_1^1 = D_1^{11} + \tfrac{1}{2}\sum_{j\neq 1} D_1^{1j}. \qquad \text{(A1)}$$

Successively rewriting each genotypic measure in eqn (A1) as minus the sum of the other genotypic disequilibria involving the M_1 cytotype, and then as minus the sum of the other genotypic disequilibria involving its nuclear genotype yields the remaining decompositions (where r is the number of nuclear alleles)

$$\begin{aligned} D_1^1 &= -\tfrac{1}{2}\sum_{j\neq 1} D_1^{1j} - \sum_{2\leq i\leq j\leq r} D_1^{ij} \\ &= \tfrac{1}{2} D_1^{11} - \tfrac{1}{2}\sum_{2\leqslant i\leqslant j\leqslant r} D_1^{ij} \\ &= -\sum_{k\neq 1} D_k^{11} + \tfrac{1}{2}\sum_{j\neq 1} D_1^{1j} \\ &= D_1^{11} - \tfrac{1}{2}\sum_{k\neq 1} D_k^{1j} + \tfrac{1}{2}\sum_{j'\neq 1,j} D_1^{1j'}, \text{ for all } j = 2, 3, \ldots, r \end{aligned}$$