A Statistical Test of a Neutral Model Using the Dynamics of Cytonuclear Disequilibria

Susmita Datta,* Mike Kiparsky,[†] David M. Rand[†] and Jonathan Arnold[‡]

*Department of Biostatistics, Rollins School of Public Health, Emory University, Atlanta, Georgia 30322, [†]Department of Ecology and Evolutionary Biology, Brown University, Providence, Rhode Island 02912 and [‡]Department of Genetics, University of Georgia, Athens, Georgia 30602

Manuscript received February 26, 1996

Accepted for publication September 11, 1996

ABSTRACT

In this paper we use cytonuclear disequilibria to test the neutrality of mtDNA markers. The data considered here involve sample frequencies of cytonuclear genotypes subject to both statistical sampling variation as well as genetic sampling variation. First, we obtain the dynamics of the sample cytonuclear disequilibria assuming random drift alone as the source of genetic sampling variation. Next, we develop a test statistic using cytonuclear disequilibria via the theory of generalized least squares to test the random drift model. The null distribution of the test statistic is shown to be approximately chi-squared using an asymptotic argument as well as computer simulation. Power of the test statistic is investigated under an alternative model with drift and selection. The method is illustrated using data from cage experiments utilizing different cytonuclear genotypes of *Drosophila melanogaster*. A program for implementing the neutrality test is available upon request.

CYTONUCLEAR disequilibrium measures provide new inferential tools in analyzing hybrid zone data. In recent years, there has been increasing attention on studying the association or interaction between a nuclear gene or genotype and maternally inherited cytoplasmic components such as mitochondria (see LAMB and AVISE 1986; ASMUSSEN *et al.* 1987; ARNOLD 1993). Questions have arisen concerning whether or not these cytonuclear associations can be explained without invoking natural selection.

Since a significant amount of work in the last decade is based on neutrality of mitochondrial DNA (mtDNA) markers, a number of researchers have recently designed experiments to test the neutrality of mtDNA markers (CLARK and LYCKEGAARD 1988; MACRAE and ANDERSON 1988; FOS *et al.* 1990; POLLAK 1991; SCRIBNER and AVISE 1994a,b; S. T. KILPATRICK and D. M. RAND, unpublished results; HUTTER and RAND 1995). We have taken a different approach to test the neutrality of a mtDNA marker with measures of association of mtDNA markers with nuclear genes or genotypes termed *cytonuclear disequilibria*.

Stochastic behavior of allelic disequilibria was addressed by FU and ARNOLD (1992). The dynamics of cytonuclear genotypic disequilibria over different generations have been described by DATTA *et al.* (1996). Stochastic trajectories of cytonuclear disequilibria calculated in FU and ARNOLD (1992) and DATTA *et al.* (1996) were used to test the neutrality of mitochondrial

DNA markers in a vertebrate cage experiment involving two species of mosquito fish (SCRIBNER and AVISE 1994). In this experiment an artificial hybrid zone composed of two competing species of mosquito fish that interbreed was established. Frequencies of cytonuclear genotypes and associated cytonuclear disequilibria were monitored over time and compared with their expectations under drift.

In such an experiment there are two potential sources of variation in cytonuclear frequencies, *statistical sampling variation* and *genetic sampling variation* (WEIR, 1990). Statistical sampling variation arises from sampling individuals from a population to estimate frequencies of cytonuclear genotypes (ASMUSSEN *et al.*, 1987). Genetic sampling variation arises from genetic drift, the sampling of gametes from a finite breeding pool of individuals in nature to constitute the next generation (FU and ARNOLD 1992; DATTA *et al.* 1996). In the experiment of SCRIBNER and AVISE (1994) in every generation, all individuals from an entire population were sampled. As a consequence statistical sampling variation was eliminated, and only genetic sampling variation (*i.e.*, drift) remained. This unusual design to a cage experiment permitted DATTA and ARNOLD (1996) to develop a simple neutrality test of a mtDNA marker using the observed trajectories of cytonuclear disequilibria. It is the purpose of this report to extend the domain of applicability of our neutrality test of an mtDNA marker to include cage experiments with both statistical and genetic sampling variation in cytonuclear disequilibria.

In Figure 1, we present a schematic diagram for the design of a more general sampling scheme in a cage

Corresponding author: Jonathan Arnold, Department of Genetics, University of Georgia, Athens, GA 30602.
E-mail: arnold@bscr.uga.edu

FIGURE 1.—Format of experiment.

experiment than in DATTA and ARNOLD (1996). We note that a statistical test of neutrality of a single genetic locus was carried out by SCHAFFER *et al.* (1977) under a similar sampling scheme. We follow a similar approach in testing whether or not the dynamics (over time) of the observed cytonuclear disequilibria are consistent with those expected under a random drift model. Tests of hypotheses about cytonuclear disequilibria in the absence of genetic sampling variation have been considered by ASMUSSEN *et al.* (1987), FU and ARNOLD (1992), and ASMUSSEN and BASTEN (1994). The focus of this article is on a test of fit to a hypothesized drift process using sample estimates of cytonuclear disequilibria over time.

THE METHOD

Cytonuclear disequilibria: Let us suppose we observe a hybrid population at a nuclear locus with two alleles A and a , and at a cytoplasmic locus with alleles M and m simultaneously as in SCRIBNER and AVISE (1994a). The frequencies of the respective cytonuclear genotypic classes are denoted by $p_1, \dots, p_6$, respectively. For example, p_1 is the frequency of the class AA/M (see Table 1). Note that the same population can be represented in terms of the frequencies of their allelic combinations (A/M , A/m , a/M , a/m) in Table 2.

TABLE 1
Frequencies of cytonuclear genotypes

Cytoplasm	Nuclear genotype			Total
	AA	Aa	aa	
M	p_1	p_2	p_3	q
m	p_4	p_5	p_6	$1 - q$
Total	u	v	w	1

TABLE 2
Frequencies of allelic combinations

Cytoplasm	Nuclear allele		Total
	A	a	
M	$e_1 = p_1 + p_2/2$	$e_3 = p_3 + p_2/2$	q
m	$e_2 = p_4 + p_5/2$	$e_4 = p_6 + p_5/2$	$1 - q$
Total	p	$1 - p$	1

Cytonuclear disequilibrium is defined to be the association of nuclear genotypes or alleles with cytoplasmic alleles (ASMUSSEN *et al.* 1987). The following are the cytonuclear disequilibria corresponding to the homozygote AA/M and the heterozygote Aa/M , respectively:

$$D_1 = p_1 - uq, \quad (1)$$

$$D_2 = p_2 - vq, \quad (2)$$

where $u = p_1 + p_4$ and $v = p_2 + p_5$. Similarly, the allelic disequilibrium is given by

$$D = e_1 - pq, \quad (3)$$

where $p = e_1 + e_2$, and $q = e_1 + e_3$, and the e_i 's are the frequencies of the allelic combinations (Table 3). For detailed definitions and their moment dynamics we refer to DATTA *et al.* (1996).

All the above quantities can be defined at the sample level, as well. We will supplement all of these variables with carets to denote their sample counterpart. For example, $\hat{D}_1 = \hat{p}_1 - \hat{u}\hat{q}$ denotes the sample disequilibrium corresponding to AA/M .

A statistical test based on disequilibrium dynamics: From now on, we will consider a cage experiment with the sampling scheme in Figure 1. The experiment starts with an initial base population with nonzero allelic disequilibrium $D = d_0$. Generations are discrete and non-overlapping, and only a fixed number of randomly selected offspring of the previous generation are introduced into the cage, leading to genetic drift (*i.e.*, genetic sampling variation). The offspring are allowed to mature, mate, and have offspring, and are removed from the cage to make room for the next generation. From the adults of generation t removed from the cage, a total of $n(t)$ adults are randomly sampled to estimate frequencies of cytonuclear genotypes in Table 1 (see Figure 1), leading to statistical sampling variation. Con-

TABLE 3
Gametic (allelic) disequilibrium

Cytoplasm	Nuclear allele		Total
	A	a	
M	$e_1 = pq + D$	$e_3 = (1 - p)q - D$	q
m	$e_2 = p(1 - q) - D$	$e_4 = (1 - p)(1 - q) + D$	$1 - q$
Total	p	$1 - p$	1

sequently, the sample genotypic counts are multinomial with parameters equal to the cytonuclear genotypic frequencies $p_1(t), \dots, p_6(t)$ of the t th generation of the population. In each generation these genotypic frequencies are the realization of a drift process. The model used here to describe the genetic sampling variation is termed the random union of zygotes (RUZ) model (WATTERSON 1970, 1972).

Genetic sampling variation: Let $X_{fm}(t)$ be the number of individuals receiving gamete f from the father and gamete m from the mother at time t , and let $\mathbf{X}(t) = (X_{11}(t), \dots, X_{44}(t))$ be the vector of such counts. The probability distribution of the counts $\mathbf{X}(t)$ given the gametic combination counts at time $t-1$, $\mathbf{X}(t-1)$ is multinomial and is given by

$$\Pr(\mathbf{X}(t) = \mathbf{x}(t) | \mathbf{X}(t-1)) = \frac{N(t)}{x_{11}(t)! \cdots x_{44}(t)!} \prod_{f,m} (e_f(t-1)e_m(t-1))^{x_{fm}(t-1)}, \quad (4)$$

where $N(t) = \sum_{f,m} x_{fm}(t)$, $f, m = 1, 2, 3, 4$ (WATTERSON, 1970) and e_f and e_m are frequencies of allelic combinations in fathers and mothers. It is not hard to see that the cytonuclear genotypic frequencies are linear combinations of the gametic combination counts X_{fm} (see *e.g.*, DATTA *et al.* 1996). Therefore, one can find the conditional moment generating function of $p_1, \dots, p_6$ at time t given the frequencies at time $t-1$, which in turn extends to the calculation of the moments of the cytonuclear disequilibria, D_1 and D_2 . See DATTA *et al.* (1996) for details.

Statistical sampling variation: The estimated disequilibria based on the statistical sample from the population at time t are given by

$$\hat{D}_1(t) = \hat{p}_1(t) - \hat{u}(t)\hat{q}(t), \quad (5)$$

$$\hat{D}_2(t) = \hat{p}_2(t) - \hat{v}(t)\hat{q}(t), \quad (6)$$

where

$$\hat{u}(t) = \hat{p}_1(t) + \hat{p}_4(t), \quad \hat{q}(t) = \hat{p}_1(t) + \hat{p}_2(t) + \hat{p}_3(t), \quad \hat{p}_j(t) = n_j(t)/n(t), \quad j = 1, \dots, 6, \quad (7)$$

and $n_1(t) - n_6(t)$ are the genotypic counts in the sample from generation $t = 0, 1, \dots$.

Here the estimates are subject to multinomial sampling variation, conditional on the genotypic frequencies in Table 1, at time t generated by the drift process.

Total sampling variation: The total (statistical + genetic) variance and covariances of the cytonuclear disequilibria can now be calculated from the above model. These expectations are necessary to construct a neutrality test comparing observed and expected disequilibria.

The (unconditional) variance of $\hat{D}_i(t)$, $i = 1, 2$, can be calculated by the formula

$$\text{Var}(\hat{D}_i(t)) = \text{Var}_s(E_s(\hat{D}_i(t))) + E_s(\text{Var}_s(\hat{D}_i(t))), \quad i = 1, 2, \quad (8)$$

where $\mathcal{G}$ and $\mathcal{S}$ stand for the genetic and the statistical sampling, respectively. Now ignoring smaller terms of the order $O(n^{-1}(t)N^{-1}(t))$ and $O(n^{-2}(t))$, (*i.e.*, terms like $1/n(t)N(t)$ and $1/n^2(t)$), we can obtain an approximation for the first term in (8):

$$\text{Var}_s(E_s(\hat{D}_i(t))) \cong \text{Var}_s(D_i(t)), \quad (9)$$

which can be calculated from the formulae given in DATTA *et al.* (1996). Note that $E_s(\hat{D}_i(t))$ is calculated by differentiating the moment generating function of a multinomial distribution with parameters $n(t)$ and $p_1(t), \dots, p_6(t)$. Direct calculation using the moment generating function and ignoring the higher order terms shows that

$$\begin{aligned} E_s(\text{Var}_s(\hat{D}_1(t))) &= \frac{1}{n(t)} \{ -(p_0^4 q_0^2 + 2d_{t-1}p_0^3 q_0 - p_0^2 q_0 \\ &+ d_{t-1}^2 p_0^2 - d_{t-1}p_0) + p_0(-d_{t-1}^3 p_0^2 + 2d_{t-1}^2 p_0^3 - d_{t-1}p_0^4 \\ &+ 2d_{t-1}p_0^2 q_0 + p_0^3 q_0 - 3d_{t-1}^2 p_0^3 q_0 + 4d_{t-1}p_0^4 q_0 - p_0^5 q_0 \\ &+ p_0 q_0^2 - 2p_0^3 q_0^2 - 3d_{t-1}p_0^4 q_0^2 + 2p_0^5 q_0^2 - p_0^5 q_0^3) - 2d_{t-1}^2 p_0^2 \\ &+ d_{t-1}p_0^3 q_0 + d_{t-1}p_0 q_0 + p_0^2 q_0^2 - d_{t-1}p_0^3 q_0 - p_0^4 q_0^2) \}, \quad (10) \end{aligned}$$

where $p_0 = p(0)$, $q_0 = q(0)$ are the initial gene frequencies and $d_{t-1} = E_s(D(t-1))$. In deriving the above formula, $\text{Var}_s(\hat{D}_1(t))$ is calculated by arguments similar to those for $E_s(\hat{D}_1(t))$ from the multinomial statistical sampling. For calculating the second term, $E_s(\text{Var}_s(\hat{D}_1(t)))$, in (8), we used the δ -method and the fact that expectations of p and q remain constant over generations. Using a result of FU and ARNOLD (1992), for the random drift model we find

$$d_{t-1} = E(D_s(t-1)) = \prod_{i=1}^{t-1} \left(\frac{N_i - 1}{2N_i} \right) d_0, \quad (11)$$

where $N_i = N(i)$, $i \geq 1$.

In the same way, one obtains the second term in (8):

$$\begin{aligned} E_s(\text{Var}_s(\hat{D}_2(t))) &= \{(d_{t-1} - 2p_0 d_{t-1} + 2p_0 q_0 - 2p_0^2 q_0) \\ &\times (1 - d_{t-1} + 2p_0 d_{t-1} - 2p_0 q_0 + 2p_0^2 q_0) \\ &- 2(d_{t-1} - 2d_{t-1}p_0 + 2p_0 q_0 - 2p_0^2 q_0)(2p_0 - 2p_0^2 \\ &+ q_0 - 4p_0 q_0 + 4p_0^2 q_0) + 2(2d_{t-1}^4 p_0^2 - 8d_{t-1}^3 p_0^3 \\ &- 8d_{t-1}^4 p_0^3 + 8d_{t-1}^2 p_0^4 + 24d_{t-1}^3 p_0^4 + 8d_{t-1}^4 p_0^4 \\ &- 16d_{t-1}^2 p_0^5 - 16d_{t-1}^3 p_0^5 + 8d_{t-1}^2 p_0^6 + 2d_{t-1}p_0 q_0 \\ &+ 2p_0^2 q_0 - 6d_{t-1}p_0^2 q_0 - 4p_0^3 q_0 + 4d_{t-1}p_0^3 q_0 \\ &+ 12d_{t-1}^3 p_0^3 q_0 + 2p_0^4 q_0 - 32d_{t-1}^2 p_0^4 q_0 - 40d_{t-1}^3 p_0^4 q_0 \\ &+ 16d_{t-1}p_0^5 q_0 + 80d_{t-1}^2 p_0^5 q_0 + 32d_{t-1}^3 p_0^5 q_0 - 32d_{t-1}p_0^6 q_0 \\ &- 48d_{t-1}^2 p_0^6 q_0 + 16d_{t-1}p_0^7 q_0 + p_0 q_0^2 - 5p_0^2 q_0^2 + 8p_0^3 q_0^2 \\ &- 4p_0^4 q_0^2 + 26d_{t-1}^2 p_0^4 q_0^2 - 40d_{t-1}p_0^5 q_0^2 - 72d_{t-1}^2 p_0^5 q_0^2 \\ &+ 8p_0^6 q_0^2 + 88d_{t-1}p_0^6 q_0^2 + 48d_{t-1}^2 p_0^6 q_0^2 - 16p_0^7 q_0^2) \} \end{aligned}$$

$$\begin{aligned}
& -48d_{t-1}p_0^7q_0^2 + 8p_0^8q_0^2 + 24d_{t-1}p_0^5q_0^3 - 16p_0^6q_0^3 \\
& -56d_{t-1}p_0^6q_0^3 + 32p_0^7q_0^3 + 32d_{t-1}p_0^7q_0^3 \\
& -16p_0^8q_0^3 + 8p_0^6q_0^4 - 16p_0^7q_0^4 + 8p_0^8q_0^4\}. \quad (12)
\end{aligned}$$

For a complete understanding of the system we also need to calculate the total covariance between the sample disequilibria as well. Covariances can occur between disequilibria in the same generation or disequilibria in different generations. Both are now calculated. The total covariance between $\hat{D}_1(t)$ and $\hat{D}_2(t)$ ($\text{Cov}(\hat{D}_1(t), \hat{D}_2(t))$) in the same generation is given by

$$\begin{aligned}
& \text{Cov}_v(E_s(\hat{D}_1(t)), E_s(\hat{D}_2(t))) \\
& + E_v(\text{Cov}_s(\hat{D}_1(t), \hat{D}_2(t))). \quad (13)
\end{aligned}$$

Using similar calculations as before and ignoring higher order terms, we find

$$\begin{aligned}
& \text{Cov}(\hat{D}_1(t), \hat{D}_2(t)) = \text{Cov}_v(D_1(t), D_2(t)) \\
& + \frac{1}{n(t)} \{p_0(8d_{t-1}^4p_0 - 16d_{t-1}^3p_0^2 - 24d_{t-1}^4p_0^2 + 32d_{t-1}^3p_0^3 \\
& + 16d_{t-1}^4p_0^3 - 16d_{t-1}^3p_0^4 + 3d_{t-1}p_0q_0 - 8d_{t-1}^3p_0q_0 \\
& + 2p_0^3q_0 - 4d_{t-1}^2p_0^2q_0 + 16d_{t-1}^3p_0^2q_0 + 64d_{t-1}^3p_0^2q_0 \\
& - 2p_0^3q_0 - 80d_{t-1}^3p_0^3q_0 - 120d_{t-1}^3p_0^3q_0 + 112d_{t-1}^2p_0^4q_0 \\
& + 64d_{t-1}^3p_0^4q_0 - 48d_{t-1}^2p_0^5q_0 - 4p_0^2q_0^2 - 32p_0^2q_0^2 \\
& - 32d_{t-1}^2p_0^2q_0^2 + 4p_0^3q_0^2 + 32d_{t-1}p_0^3q_0^2 + 152d_{t-1}^2p_0^3q_0^2 \\
& - 112d_{t-1}p_0^4q_0^2 - 216d_{t-1}^2p_0^4q_0^2 + 128d_{t-1}p_0^5q_0^2 \\
& + 96d_{t-1}^2p_0^5q_0^2 - 48d_{t-1}p_0^6q_0^2 - 40d_{t-1}p_0^3q_0^3 \\
& + 16p_0^4q_0^3 + 144d_{t-1}p_0^4q_0^3 - 48p_0^5q_0^3 - 168d_{t-1}p_0^5q_0^3 \\
& + 48p_0^6q_0^3 + 64d_{t-1}p_0^6q_0^3 - 16p_0^7q_0^3 - 16p_0^4q_0^4 \\
& + 48p_0^5q_0^4 - 48p_0^6q_0^4 + 16p_0^7q_0^4\}, \quad (14)
\end{aligned}$$

where the first term $\text{Cov}_v(D_1(t), D_2(t))$ can be calculated following DATTA *et al.* (1996).

The covariance between the estimates based on samples from different generations are calculated next. For $1 \leq i, j \leq 2$, $t \geq 0$, $s \geq 1$, given $\mathcal{F}_t$ the composition of the population at time t , $\hat{D}_i(t)$ and $\hat{D}_j(t+s)$ are independent. Also up to higher order terms,

$$E(\hat{D}_i(t) | \mathcal{F}_t) = D_i(t) \quad (15)$$

and

$$\begin{aligned}
& E(\hat{D}_j(t+s) | \mathcal{F}_t) = E(E(\hat{D}_j(t+s) | \mathcal{F}_{t+s}) | \mathcal{F}_t) \\
& = E(D_j(t+s) | \mathcal{F}_t). \quad (16)
\end{aligned}$$

It can be shown using the one-step conditional moment generating function (*cf.* DATTA, FU, and ARNOLD 1996) that the above quantity equals $((N_{t+s} - 1)/N_{t+s})E(D(t+s-1)p(t+s-1) | \mathcal{F}_t)$, if $j = 1$ and $((N_{t+s} - 1)/N_{t+s})E(D(t+s-1)(1-2p(t+s-1)) | \mathcal{F}_t)$, if $j = 2$, which in turn, by the δ -method, is approximately equal

to (follows from the dynamics of D , see ARNOLD and FU 1992),

$$\frac{(N_{t+s} - 1)}{N_{t+s}} r_{j0} \prod_{i=t+1}^{t+s-1} \left(\frac{N_i - 1}{2N_i} \right) D(t),$$

where $r_{j0} = p_0$, if $j = 1$ and $= 1 - 2p_0$ if $j = 2$. Note that the above product needs to be interpreted as 1 for $s = 1$. Consequently one has, modulo higher order terms,

$$\begin{aligned}
& \text{Cov}(\hat{D}_i(t), \hat{D}_j(t+s)) = \frac{(N_{t+s} - 1)}{N_{t+s}} \\
& \times \left[E_v \left(D_i(t) \cdot r_{j0} \prod_{i=t+1}^{t+s-1} \left(\frac{N_i - 1}{2N_i} \right) D(t) \right) \right. \\
& \left. - E_v \left(D_i(t) \right) E_v \left(r_{j0} \prod_{i=t+1}^{t+s-1} \left(\frac{N_i - 1}{2N_i} \right) D(t) \right) \right] \\
& = \frac{(N_{t+s} - 1)}{N_{t+s}} r_{j0} \prod_{i=t+1}^{t+s-1} \left(\frac{N_i - 1}{2N_i} \right) \text{Cov}_v(D_i(t), D(t)). \quad (17)
\end{aligned}$$

Once again, in the above equation $\text{Cov}(D_i(t), D(t))$ can be found using the recursions of the first two moments of $D_i(t)$ in DATTA *et al.* (1996) and $D(t)$ in FU and ARNOLD (1992). Note that in obtaining the first term of the above expression we have used the independence of the samples at times t and $t+s$ given the population at time t .

Let the vector

$$\mathbf{Y} = (\hat{D}_1(1), \dots, \hat{D}_1(k), \hat{D}_2(1), \dots, \hat{D}_2(k)) \quad (18)$$

be the trajectory of sample disequilibria, $\hat{D}_1$ and $\hat{D}_2$. Note that

$$E(\hat{D}_i(t)) = E_v(E_s(\hat{D}_i(t))) = E_v(D_i(t)), \quad (19)$$

ignoring higher order terms, which is equal to $a_i p_0 d_0$ if $i = 1$ and $a_i(1 - 2p_0)d_0$ if $i = 2$. Thus, the trajectory of the sample disequilibria $\mathbf{Y}$ is approximately normally distributed with mean vector $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ and variance-covariance matrix $\boldsymbol{\Sigma}$, where $\mathbf{X} = [\mathbf{X}_1 | \mathbf{X}_2]$, $\mathbf{X}_1 = (a_1, \dots, a_k, 0, \dots, 0)'$, $\mathbf{X}_2 = (0, \dots, 0, a_1, \dots, a_k)'$, with

$$a_i = \left(\frac{N_t - 1}{N_t} \right) \prod_{i=1}^{t-1} \left(\frac{N_i - 1}{2N_i} \right),$$

$\boldsymbol{\beta} = (\beta_1, \beta_2)$, $\beta_1 = d_0 p_0$, $\beta_2 = (1 - 2p_0)d_0$. To recapitulate, the trajectory of sample cytonuclear disequilibria is denoted by $\mathbf{Y}$. The expectation of this trajectory is $\boldsymbol{\mu}$. The estimates of the trajectory are approximately jointly normal, and the variance-covariance matrix of the sample cytonuclear disequilibria, $\hat{D}_1$ and $\hat{D}_2$, along the trajectory is $\boldsymbol{\Sigma}$. The vector $\boldsymbol{\beta}$ is determined by the initial state of the cage. The calculation of the variance-covariance matrix of the disequilibria $\boldsymbol{\Sigma}$ was explained in

the earlier paragraphs. The normal approximation is justified when both $n(t)$ and $N(t)$ are large.

Neutrality test: With the mean vector $\boldsymbol{\mu}$ and the variance-covariance matrix $\boldsymbol{\Sigma}$ of the estimated trajectory for $\hat{D}_1$ and $\hat{D}_2$, we can now construct a neutrality test by comparing the inferred trajectory of cytonuclear disequilibria ($\mathbf{Y}$) with its expectation over time ($\boldsymbol{\mu}$).

If the initial values p_0 , q_0 and D_0 in the cage are known, one can use the following statistics to test the null hypothesis of a random drift model:

$$T = (\mathbf{Y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \boldsymbol{\mu}), \quad (20)$$

which will have an approximate chi-square distribution with $2k$ degrees of freedom. One would reject a random drift model if $T > \chi^2_{\alpha}(2k)$, where $\chi^2_{\alpha}(2k)$ is the upper α th quantile of a chi-square distribution with $2k$ degrees of freedom. However, often in practice, the composition of the initial population (summarized by $\boldsymbol{\beta}$) is not known and only a simple random sample from it is available. In that case, we suggest (consistently) estimating the variance covariance matrix $\boldsymbol{\Sigma}$ by substituting in the sample estimates $\hat{p}(0)$, $\hat{q}(0)$ and $\hat{D}(0)$ in places of p_0 , q_0 and d_0 , respectively, in the expression for $\boldsymbol{\Sigma}$ from (9)–(14) and (17). Next, an overall estimate of $\boldsymbol{\beta}$ is obtained by the method of weighted least squares:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}' \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{X})^{-1} \mathbf{X}' \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{Y}.$$

Finally, we could use the test statistic $T = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})' \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$, which will have an approximately chi-square distribution with $2(k - 1)$ freedom under the random drift model. This proposition is validated with a simulation study for $k = 2$. In the simulation we considered that the initial generation had the frequencies $p_1 = 0.5$ and $p_6 = 0.5$, and all the subsequent generations had constant generation size $N = 500$. In each simulation, random samples of constant sample size $n = 50$ are generated for $k + 1 = 3$ generations. The value of the test statistic is calculated for each simulation, and the entire process is repeated 5000 times. Consequently, we have 5000 independent realizations of the test statistic. A histogram was drawn with these values, which is in good agreement with an overlaid true chi-square distribution with 2 degrees of freedom (Figure 2). The agreement is emphasized with a $Q - Q$ plot also.

Note that for this neutrality test it is not necessary to have sample data from all consecutive generations. It is easy to adjust and reinterpret the test statistic in such cases. Software for carrying out this neutrality test is available from the author (datta@fox.sph.emory.edu).

Instead of using the chi-square percentiles in constructing the approximate rejection region of the test, one may resort to Monte Carlo simulation to find the exact P value given a sample. This may be preferable if the sample sizes are small and one is afraid that the large sample distribution may not be adequate. Since

FIGURE 2.—Histogram and chi-squared $Q - Q$ plot based on 5000 Monte-Carlo replications of the test statistic T . The overlaid graph in solid line is the density of chi-squared distribution with 2 degrees of freedom.

the P -value calculation using present day computing is not expensive in terms of computing time, we anticipate that this approach may be preferred by many users.

If, in practice, the entire history of the generation sizes is not known, then one may use simple interpolation. We found that the test statistic is not too sensitive with respect to misspecification of the generation sizes.

RESULTS

An example: To illustrate our test procedure, we now consider a real cage experiment constituting artificial hybrid zone data on *Drosophila melanogaster* conducted by M. KIPARSKY. The experiment uses controlled environmental conditions to study the effects of cytonuclear interaction in a population genetic context using genetically manipulated strains of *D. melanogaster*. The hybrids were formed with the crosses of initial stocks of flies collected in Ega, Denmark and Death Valley, CA. Cages were maintained in a cycle of discrete generations. In each generation, flies were allowed to mate, and the adult flies were taken out and kept frozen for genotyping by PCR analysis. Eggs were allowed to form the next generation. A random sample of adult flies were taken from the frozen adult population and using PCR-based techniques, the cytonuclear genotypes were observed at two nuclear loci (60A and DPP) and at a cytoplasmic locus simultaneously. The experiment was continued for four generations. Data were not initially collected on the third generation except for the population size. For carrying out our test procedure we have ignored the data on their initial generation as the experiment started with a purely heterozygous population in which all disequilibrium measures are zero. Thus, in

TABLE 4
Results for 60A locus

Generation	Generation size	Sample size	$\hat{D}_1$	$\hat{D}_2$	$\hat{p}(0)$	$\hat{q}(0)$	$\hat{D}(0)$	T
0		91			0.5330	0.2527	0.0466	
1	1826	73	0.0445	-0.0667				
2	694							
3	1621	81	0.0172	-0.0348				0.64

T , test statistic.

reporting our results (Table 4 and Table 5) we relabeled the generations starting from the first generation. Generation sizes, sample sizes, cytonuclear disequilibria and the value of the test statistic are reported on these two nuclear loci in Tables 4 and 5 in combination with the mtDNA marker. The value of the test statistics at locus (DPP) is significant at the significance level $\alpha = 0.05$. This suggests random drift alone cannot explain the population dynamics adequately. Some other forces (such as selection) have to be taken into consideration to explain the dynamics. Conversely, at another locus (60A) the value of the test statistic was not significant at the same level, suggesting the neutral model (random drift) is consistent with the dynamics of the cytonuclear disequilibria.

When data were obtained for the third generation, values of the test statistic in both the loci were large enough, and P values were very small (at 60A, value of the test statistic is 25.53 and the P value is 0.00027, at the DPP locus they are 18.97 and 0.00421, respectively), which leads us to reject the null hypothesis of random drift and consequently to reject the neutrality of the mtDNA marker.

The null distribution of the test statistic is simulated with the same population and sample size as the 60A locus. The values of the genotypic frequencies used for the initial generation were the same as those obtained from the initial generation sample, namely, 10/91, 13/91, 0.0, 16/91, 32/91, 20/91 for $p_1, \dots, p_6$, respectively. For this simulation, we also assume that we have no sample information at the second generation. Under this scenario the value of the test statistic is generated 5000 times and a histogram is drawn. The histogram was overlaid with a true chi-square distribution with $2(k-1) = 2$ degrees of freedom. Next, a $Q-Q$ plot is

drawn (Figure 3). Both of the plots are in good agreement with the true chi-square distribution.

A simulation study for the power of the test: We conducted a simulation to calculate the power of the test proposed in this paper. As an alternative model we consider a special type of selection model that assumes that the genotypic frequencies in the population prior to mating (and hence prior to the statistical sampling) changes according to the following selection rule:

$$p'_i = p_i w_i / \bar{w}, \quad i = 1, \dots, 6,$$

where $\bar{w} = \sum p_i w_i$, $w_1 = w_3 = w_4 = w_6 = 1$ and $w_2 = w_5 = 1 - s$ with $0 \leq s \leq 1$. Note that $s = 0$ corresponds to no selection (see CLARK 1984).

In the initial generation the population had frequencies $p_1 = 0.5$ and $p_6 = 0.5$, and in all subsequent generations the population size $N = 1000$. Random samples of size $n = 100$ from each generation were collected, and the value of the test statistic was computed based on the entire sample. The process was repeated independently 5000 times. The power is estimated by the empirical proportion of times the value of the test statistic landed in the upper 5% tail of a chi-square distribution with $2(k-1)$ degrees of freedom, where $k+1$ is the number of generations sampled.

In Figure 3, the power curves are shown as functions of s for $k = 2$ and 3. The growth of the power curve is somewhat slow at the beginning, indicating that the test is unable to detect weak selection on heterozygotes. The power however increases rapidly after a stage. The test has better power if more generations are sampled, as to be expected. Also it would increase (not documented here) if more individuals were sampled from each generation.

TABLE 5
Results for DPP locus

Generation	Generation size	Sample size	$\hat{D}_1$	$\hat{D}_2$	$\hat{p}(0)$	$\hat{q}(0)$	$\hat{D}(0)$	T
0		85			0.5118	0.2235	0.0327	
1	1826	69	-0.0256	0.0078				
2	694							
3	1621	77	0.0587	-0.0712				18.5*

T , test statistic. *Significant at 5%.

FIGURE 3.—Histogram and chi-squared $Q - Q$ plot based on 5000 Monte-Carlo replications of the test statistic T using the same population sizes and the same sample sizes as the 60A locus. Also the initial genotypic frequencies are taken to be the same as in the initial sample at the 60A locus. The overlaid graph in solid line is the density of chi-squared distribution with 2 degrees of freedom.

DISCUSSION

One of the strengths of this test is that it may be more sensitive to departures from neutrality than a test based on the history of gene frequencies at a single genetic locus, and we have found this to be the case in practice when several neutrality tests are applied to the same data set, such as the one in the example. Comparable power studies are underway. The test also takes into account both genetic and statistical sampling variation and can be applied to trajectories in which there are missing data either about the initial state of the cage population or about the state of the population at one or more time points.

Disequilibria have been used by others (MUELLER *et al.* 1985; HEDRICK and THOMSON 1986) to test neutrality. Our test is motivated by the one-locus test of FISHER and FORD (1947), but the test combines data on two markers to enhance power. HEDRICK and THOMSON (1986) constructed a two-locus test of neutrality, assuming the population is at steady state and resorted to simulation to generate a null distribution. Our test makes no steady-state assumption and requires no simulation method to simulate its null distribution except for small samples, where the asymptotic chi-square distribution may not be adequate. MUELLER *et al.* (1985) also constructed a test of neutrality, but their null hypothesis was selection with randomly varying intensity, and again they generated a null distribution by simulation. Results presented here about our neutrality test in (21) are analytical, do not resort to the simulation for a null distribution of the test statistic and can be generalized to more complicated null hypotheses, such

FIGURE 4.—The power function of the proposed test statistic under a selection model as a function of the selection coefficient s , based on samples of size $n = 100$ from each generation of size $N = 1000$.

as genetic hitchhiking of a mitochondrial DNA marker with respect to a selected nuclear locus (CLARK, 1984). We have also shown that our neutrality test has reasonable small to moderate sample properties (*i.e.*, Figures 2 and 3), which makes it useful in practice.

There are several limitations to the test. We have only shown the validity of the test when generations are discrete and nonoverlapping. Our neutrality test has only been shown to be valid under one kind of design for a cage experiment (Figure 1). When selection is weak ($s \leq 0.4$), the power to detect selection can be quite low (Figure 4). This is in contrast to the power curves under an alternative with migration alone (DATTA and ARNOLD 1996), where power curves increase much more rapidly with the migration rate m . Lastly, the history of the generation sizes is assumed to be available; although in practice, one can use the interpolated values if some are missing.

There are a number of extensions to this work that are needed. Power calculations are needed under a broader range of alternatives to the neutral model. The validity of the test in (20) needs to be examined for other sampling designs for cage experiments. Comparative studies with other neutrality tests needed to be made.

We thank ROB DEAN, RUTH FUNDYGA, and MARIA SANCHEZ for their helpful comments. Comments by two anonymous referees greatly improved the manuscript.

LITERATURE CITED

- ARNOLD, J., 1993 Cytonuclear disequilibria in hybrid zones. *Annu. Rev. Ecol. Syst.* **24**: 521–554.
- ASMUSSEN, M. A., J. ARNOLD and J. C. AVISE, 1987 Definition and properties of disequilibrium statistics for associations between nuclear and cytoplasmic genotypes. *Genetics* **115**: 755–768.
- ASMUSSEN, M. A., and C. BASTEN, 1994 Sampling theory for cytonuclear disequilibria. *Genetics* **138**: 1351–1363.

- AVISE, J. C., 1994 *Molecular Markers, Natural History and Evolution*. Chapman and Hall, New York.
- CLARK, A. G., 1984 Natural selection with nuclear and cytoplasmic transmission. I. A deterministic model. *Genetics* **107**: 679–701.
- CLARK, A. G., and E. M. S. LYCKEGAARD, 1988 Natural selection with nuclear and cytoplasmic transmission. III. Joint analysis of segregation and mtDNA in *Drosophila melanogaster*. *Genetics* **118**: 471–481.
- DATTA, S., Y. X. FU and J. ARNOLD, 1996 Dynamics and equilibrium behavior of cytonuclear disequilibria under genetic drift, mutation, and migration. *Theor. Popul. Biol.* (in press).
- DATTA, S., and J. ARNOLD, 1996 Diagnostics and a statistical test of neutrality hypotheses using the dynamics of cytonuclear disequilibria. *Biometrics* **52**: 1042–1054.
- FISHER, R. A., and E. B. FORD, 1947 The spread of a gene in natural conditions in a colony of the moth *Panaxia dominula* L. *Heredity* **1**: 143–174.
- FOS, M., M. A. DOMINGUEZ, A. LATORRE and A. MOYA, 1990 Mitochondrial DNA evolution in experimental populations of *Drosophila pseudoobscura*. *Proc. Natl. Acad. Sci. USA* **87**: 4198–4201.
- FU, Y. X., and J. ARNOLD, 1992 Dynamics of cytonuclear disequilibria in finite populations and a comparison with a two-locus nuclear system. *Theor. Popul. Biol.* **41**: 1–25.
- FU, Y. X., and J. ARNOLD, 1992 A table of exact sample sizes with Fisher's exact test for 2×2 tables. *Biometrics* **48**: 1103–1112.
- HEDRICK, P. W., and G. THOMSON, 1986 A two locus neutrality test: applications to humans, *E. Coli*, and lodgepole pine. *Genetics* **112**: 135–156.
- HUTTER, C. M., and D. M. RAND, 1994 Competition between mitochondrial haplotypes in distinct nuclear genetic environments: *Drosophila pseudoobscura* versus *Drosophila persimilis*. *Genetics* **141**: 1113–1124.
- MACRAE, A., and W. W. ANDERSON, 1988 Evidence of non-neutrality of mitochondrial DNA haplotypes in *Drosophila pseudoobscura*. *Genetics* **127**: 411–416.
- MUELLER, L. D., B. A. WILCOX, P. R. EHRLICH, D. G. HECKEL and D. G. MURPHY, 1985 A direct assessment of the role of genetic drift in determining allele frequency variation in populations of *Euphydryas editha*. *Genetics* **110**: 495–511.
- POLLAK, P. E., 1991 Cytoplasmic effects on components of fitness in tobacco hybrids. *Evolution* **45**: 785–790.
- SCHAFER, H. E., D. YARDLEY and W. W. ANDERSON, 1977 Drift or selection: test of gene frequency variation over generations. *Genetics* **87**: 371–379.
- SCRIBNER, K. T., and J. C. AVISE, 1994a Population cage experiments with a vertebrate: genetics of hybridization in *Gambusia* fishes. *Evolution* **48**: 155–171.
- SCRIBNER, K. T., and J. C. AVISE, 1994b Cytonuclear genetics of experimental fish hybrid zone inside biosphere 2. *Proc. Natl. Acad. Sci. USA* **91**: 5066–5069.
- WATTERSON, G. A., 1970 The effect of linkage in finite random-mating population. *Theor. Popul. Biol.* **1**: 72–87.
- WATTERSON, G. A., 1972 Errata. *Theor. Popul. Biol.* **3**: 117.
- WEIR, B. S., 1990 *Genetic Data Analysis*. Sinauer Assoc., Sunderland, MA.

Communicating editor: A. G. CLARK