The Effects of Assortative Mating and Migration on Cytonuclear Associations in Hybrid Zones

Marjorie A. Asmussen, Jonathan Arnold and John C. Avise

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

ABSTRACT

We examine the influence of nonrandom mating and immigration on the evolutionary dynamics of cytonuclear associations in hybrid zones. Recursion equations for allelic and genotypic cytonuclear disequilibria were generated under models of (1) migration alone, assuming hybrid zone matings are random with respect to cytonuclear genotype; and (2) migration in conjunction with refined epistatic mating, in which females of the pure parental species preferentially mate with conspecific males. Major results are as follows: (a) even the slightest migration removes the dependency of the final outcome on initial conditions, producing a unique equilibrium in which both pure parental genotypes are maintained in the hybrid zone; (b) in contrast to nuclear genes, the dynamics of cytoplasmic allele frequencies appear robust to changes in the assumed mating system, yet are particularly sensitive to gene flow; (c) continued immigration can generate permanent cytonuclear disequilibria, whether mating is random or assortative; and (d) the order of population censusing (before versus after reproduction by immigrants) can have a dramatic effect on the magnitude but not the pattern of cytonuclear disequilibria. Using the maximum likelihood method, the parameter space of migration rates and assortative mating rates was examined for best fit to observed cytonuclear disequilibria data in a hybrid population of Hyla tree frogs. An epistatic mating model with a total immigration rate of about 32% per generation produces equilibrium gene frequencies and cytonuclear disequilibria consistent with the empirical observations.

NONRANDOM associations between a diploid nuclear gene, and a haploid, uniparentally transmitted cytoplasmic gene are sometimes observed in natural populations. For example, alleles at each of five, unlinked nuclear loci exhibit strong associations with mitochondrial (mt) DNA haplotypes in a hybrid population of Hyla tree frogs (LAMB and AVISE 1986). To summarize such "cytonuclear disequilibria" statistically, we recently introduced four interrelated components of cytonuclear association: a gametic disequilibrium (D), which describes associations between nuclear and cytoplasmic alleles; and three genotypic disequilibria (D_1, D_2 and D_3), which describe associations between two cytotypes and the three genotypes at a diallelic nuclear locus (ASMUSSEN, ARNOLD and AVISE 1987). Here we continue development of a general theory for the evolutionary dynamics of these cytonuclear disequilibria, with reference to the hybrid Hyla population in particular.

Under several models in which mating propensities of individuals in a hybrid zone (closed to outside recruitment) are determined solely by genotypes at a nuclear locus with or without dominance effects, we have shown that all cytonuclear disequilibria ulti-

mately decay to zero at rates that are generally inversely related to the degree of assortative mating (ASMUSSEN, ARNOLD and AVISE 1987). Permanent nonzero cytonuclear disequilibria can be maintained when female mating preference is dictated by an epistatic interaction between a single nuclear locus and a cytoplasmic gene (ARNOLD, ASMUSSEN and AVISE 1988), but such cytonuclear disequilibria involve specifically the nuclear gene which exerts mating influence, and do not extend to other unlinked nuclear loci. When mating propensity involves an epistatic interaction between cytotype and the multilocus nuclear genotype characteristic of the "pure" parental species (refined epistatic mating), all cytonuclear disequilibria again decay to zero, as do the frequencies of both pure parentals (ARNOLD, ASMUSSEN and AVISE 1988). Thus, of the models considered to date, none predicts a continued maintenance of the pattern of disequilibria between cytotype and several nuclear genes that has been observed in the Hyla population. The results suggest moreover that other factors in addition to the mating system are probably involved.

Here we extend the refined epistatic mating system model to include the effects of migration on cytonuclear disequilibria. Three major questions will be addressed: (a) can continued migration of pure parentals into a hybrid zone generate permanent nonzero cy-

TABLE 1

Frequencies of cytonuclear genotypes

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

tonuclear disequilibria, with or without assortative mating? (b) how does the order of population censusing, before or after reproduction by immigrants, influence the magnitude and pattern of disequilibria? and (c) what level of immigration (if any) would be sufficient to generate the cytonuclear disequilibria observed in the Hyla population? A secondary issue is the role of migration in maintaining pure parentals in a hybrid zone.

THE CYTONUCLEAR SYSTEM

As in our earlier treatments (ASMUSSEN, ARNOLD and AVISE 1987; ARNOLD, ASMUSSEN and AVISE 1988), we consider a diploid population with two alleles (A, a) at an autosomal nuclear locus, and two alleles (M, m) at a haploid cytoplasmic locus. The frequencies of the six possible cytonuclear genotypes are specified in Table 1 together with the marginal genotypic frequencies at the individual loci.

The statistical association between gene frequencies at the two loci is measured by the *allelic disequilibrium*,

$$D = \text{freq. } (A/M) - \text{freq. } (A) \text{ freq. } (M) \qquad (1)$$
$$= u_1 + \tfrac{1}{2}v_1 - px,$$

where "freq." denotes "frequency of," and

$$p = u + \tfrac{1}{2}v \qquad (2)$$

is the gene frequency of allele A at the nuclear locus.

The statistical associations between the nuclear and cytoplasmic genotypic frequencies are similarly measured by the three *genotypic disequilibria*,

$$D_1 = \text{freq. } (AA/M) - \text{freq. } (AA) \text{ freq. } (M)$$
$$= u_1 - ux$$

$$D_2 = \text{freq. } (Aa/M) - \text{freq. } (Aa) \text{ freq. } (M) \qquad (3)$$
$$= v_1 - vx$$

$$D_3 = \text{freq. } (aa/M) - \text{freq. } (aa) \text{ freq. } (M)$$
$$= w_1 - wx.$$

We are interested in the dynamic and equilibrium properties of cytonuclear systems within a zone of hybridization between two genetically distinct taxa, denoted as species 1 and species 2. Species 1 (*e.g., Hyla cinerea*) is assumed to be characterized by the AA/M

cytonuclear genotype, while species 2 (*e.g., Hyla gratiosa*) is characterized by aa/m. Following our previous hybrid zone study (ARNOLD, ASMUSSEN and AVISE 1988), we distinguish between the two types of individuals having the two-locus genotype characteristic of one of the pure parental species. The group of AA/M individuals is, for instance, decomposed into pure species 1, $(AA/M)_s$, with frequency u_{1s}, and hybrids, $(AA/M)_h$, with frequency u_{1h}. The aa/m group is similarly decomposed into $(aa/m)_s$ and $(aa/m)_h$, with frequencies w_{2s} and w_{2h}, respectively. There are thus eight types in the population, with the upper left and lower right cells of Table 1 partitioned so that

$$u_1 = u_{1s} + u_{1h} \quad \text{and} \quad w_2 = w_{2s} + w_{2h}. \qquad (4)$$

The distinction between AA/M and aa/m individuals of pure species and hybrid ancestry is important because after the first generation in the hybrid zone some of these individuals represent offspring from matings involving F_1's or later generation hybrids, and thus are themselves hybrids. This partition can be made empirically by assaying individuals' joint genotypes at multiple diagnostic nuclear loci, as shown by LAMB and AVISE's (1986) study of Hyla tree frogs.

MIGRATION MODELS

It is assumed that in each generation a fixed fraction m_1 of the hybrid population consists of migrants from population 1 (species 1) and that a fixed fraction m_2 are migrants from population 2 (species 2), with the remaining fraction, $1 - m_1 - m_2$, derived from existing residents of the hybrid zone. The cytonuclear frequencies and disequilibria within the combined migrant pool are assumed to be constant over time, and are denoted by $\bar{u}_{1s}, \bar{w}_{2s}, \bar{u}_1, \bar{w}_2, \bar{p}, \bar{x}, \bar{D}$, etc.

These migrant values are readily computed given the composition of the two sources. For example, denoting the cytoplasmic gene frequency in population 1 by x_1 and that in population 2 by x_2, the overall cytoplasmic frequency in the migrant pool is simply the weighted average, $\bar{x} = (m_1 x_1 + m_2 x_2)/(m_1 + m_2)$. The other cytonuclear and marginal nuclear frequencies are computed similarly. The cytonuclear disequilibria among the migrants are then obtained from their definitions in (1)–(3) (*e.g.*, $\bar{D} = \bar{u}_1 + \tfrac{1}{2}\bar{v}_1 - \bar{p}\bar{x}$ and $\bar{D}_1 = \bar{u}_1 - \bar{u}\bar{x}$). Note that unlike the frequency variables, the joint disequilibrium measures are not simply the weighted averages of the corresponding values in the two sources.

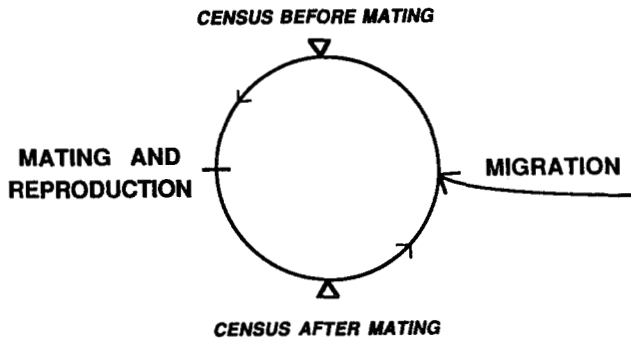For the case of interest where population 1 consists

FIGURE 1.—Alternate censusing times within the life cycle of a population receiving immigrants.

entirely of $(AA/M)_s$ individuals and population 2 is exclusively $(aa/m)_s$,

$$\bar{u}_{1s} = \bar{u}_1 = \bar{p} = \bar{x} = m_1/m$$

$$\bar{w}_{2s} = \bar{w}_2 = \bar{q} = \bar{y} = m_2/m$$

$$\bar{v}_1 = \bar{w}_1 = \bar{u}_2 = \bar{v}_2 = 0 \qquad (5)$$

$$\bar{D} = \bar{D}_1 = -\bar{D}_3 = \bar{x}\bar{y} = m_1 m_2/m^2$$

$$\bar{D}_2 = 0$$

where

$$m = m_1 + m_2 \qquad (6)$$

is the total frequency of migrants each generation, and $q = 1 - p$ is the gene frequency of nuclear allele a.

We are interested in the consequences of (i) migration alone, assuming hybrid zone matings are random with respect to cytonuclear genotype; and (ii) migration in conjunction with our refined epistatic mating model (ARNOLD, ASMUSSEN and AVISE 1988) in which pure parental females preferentially mate with conspecific males. In the latter case, $(AA/M)_s$ and $(aa/m)_s$ females preferentially mate with like males with probability α and β, respectively, and otherwise mate at random with the eight types in the hybrid population. All six hybrid females mate at random. In effect the refined epistatic mating model assumes that female mating propensities are determined by an interaction between a cytoplasmic gene and the *multilocus* nuclear genotype that distinquishes pure species from hybrids. However, the model monitors the frequencies and disequilibria involving only one of the diagnostic nuclear loci.

Common assumptions of both mating systems are that there are no viability or fertility differences among the cytonuclear genotypes, the cytoplasmic gene is maternally inherited, the hybrid population is large so that the effects of drift can be ignored, and the population has discrete, nonoverlapping generations. For each case two variations are considered, corresponding to the two possible censusing times relative to mating and migration (Figure 1). The first census scheme, developed below, involves censusing after migration and before mating, while the alternative scheme, developed in APPENDIX B, involves censusing after reproduction and before migration. Because the refined epistatic mating model contains random mating as a nested model corresponding to $\alpha = \beta = 0$, the recursions are given only for the more general, assortative mating systems.

As a final caveat, it should be emphasized that although we are primarily interested in the consequences of continued migration of two genetically distinct species into a hybrid zone, for which the migrant values are as in (5), the models below and their major conclusions apply more generally to the dynamics within any population which receives a constant fraction (m) of immigrants from a large source population having a fixed but arbitrary genetic composition.

Method 1—censusing after migration and before mating: Under this census method the frequencies in Table 1 and Equation 4 represent the values in the hybrid zone after all migrants have entered and before mating has occurred. Since the model contains no selection, censused individuals and migrants may be either juveniles or adults. A generation cycle begins with mating among all individuals in the hybrid zone, after which there is an influx of new migrants from the two parental species according to the migration scheme described above. This sequence of events completes the generation cycle (Figure 1).

The cytonuclear frequencies at the next census are then simply the weighted average of the overall frequencies among the new migrants, and the frequencies among the offspring from the (refined epistatic) matings in the hybrid zone (ARNOLD, ASMUSSEN and AVISE 1988). The recursions for the hybrid zone frequencies of the two pure species are thus

$$u_{1s}' = m\bar{u}_{1s} + (1 - m)u_{1s}[\alpha + (1 - \alpha)u_{1s}] \qquad (7a)$$

and

$$w_{2s}' = m\bar{w}_{2s} + (1 - m)w_{2s}[\beta + (1 - \beta)w_{2s}], \qquad (7b)$$

while those for the six composite cytonuclear genotypes are

$$u_1' = m\bar{u}_1 + (1 - m)(p^2 x + pD + \alpha u_{1s}q)$$

$$v_1' = m\bar{v}_1 + (1 - m)[2pqx + (q - p)D - \alpha u_{1s}q]$$

$$w_1' = m\bar{w}_1 + (1 - m)(q^2 x - qD) \qquad (8)$$

$$u_2' = m\bar{u}_2 + (1 - m)(p^2 y - pD)$$

$$v_2' = m\bar{v}_2 + (1 - m)[2pqy - (q - p)D - \beta w_{2s}p]$$

$$w_2' = m\bar{w}_2 + (1 - m)(q^2 y + qD + \beta w_{2s}p)$$

where $(')$ denotes the values at the next census. [Note that the epistatic mating components in (8) have been

rewritten to show their exclusive dependence on the frequencies of the two pure parental species (u_{1s}, w_{2s}), the gene frequencies at the two loci (p, q; x, y), the allelic disequilibrium between the loci (D), and the assortative mating rates (α, β) in the hybrid zone.]

Using (8) the recursions for the marginal single locus genotypic and gene frequencies are readily found to be

$$u' = m\bar{u} + (1 - m)(p^2 + \alpha u_{1s}q)$$

$$v' = m\bar{v} + (1 - m)(2pq - \alpha u_{1s}q - \beta w_{2s}p)$$

$$w' = m\bar{w} + (1 - m)(q^2 + \beta w_{2s}p) \qquad (9)$$

$$p' = m\bar{p} + (1 - m)[p + \tfrac{1}{2}(\alpha u_{1s}q - \beta w_{2s}p)]$$

$$x' = m\bar{x} + (1 - m)x.$$

The new cytonuclear associations can now be computed directly, using (8)–(9) plus the disequilibrium definitions in (1)–(3). The resulting disequilibrium recursions are

$$D' = m\bar{D} + (1 - m)(D + \alpha u_{1s}yq + \beta w_{2s}xp)/2$$
$$+ m(1 - m)[p + \tfrac{1}{2}(\alpha u_{1s}q - \beta w_{2s}p) - \bar{p}]$$
$$\cdot (x - \bar{x})$$

$$D_1' = m\bar{D}_1 + (1 - m)(pD + \alpha u_{1s}yq)$$
$$+ m(1 - m)(p^2 + \alpha u_{1s}q - \bar{u})(x - \bar{x})$$

$$D_2' = m\bar{D}_2 + (1 - m)[(q - p)D - \alpha u_{1s}yq \qquad (10)$$
$$+ \beta w_{2s}xp]$$
$$+ m(1 - m)(2pq - \alpha u_{1s}q - \beta w_{2s}p - \bar{v})$$
$$\cdot (x - \bar{x})$$

$$D_3' = m\bar{D}_3 - (1 - m)(qD + \beta w_{2s}xp)$$
$$+ m(1 - m)(q^2 + \beta w_{2s}p - \bar{w})(x - \bar{x}).$$

Each of these four transformations has three terms. The first two represent the weighted average of the disequilibrium in the migrants and the disequilibrium in the hybrid zone following mating. The third term is the covariance, across these two sources, between the cytoplasmic frequency and the relevant nuclear frequency.

An analysis of the combined recursions in (7)–(10) reveals several important equilibrium results that apply to both randomly mating and assortatively mating populations. First, in dramatic contrast to the assortative mating models in ARNOLD, ASMUSSEN and AVISE 1988, *even the slightest migration removes the dependency of the final outcome on initial conditions and produces one and only one equilibrium*, which is uniquely determined by the migrant composition, the total migration rate (m), and the assortative mating parameters (α and β). A second observation is that continued immigra-

tion by pure parentals maintains both species' genotypes in the hybrid zone, and, in particular, *both parental frequencies monotonically approach positive equilibrium values that are always less than those in the joint migrant pool* (provided $0 < \bar{u}_{1s}$, $\bar{w}_{2s} < 1$, $0 < m < 1$, and $0 \le \alpha$, $\beta < 1$). The derivation of these results and the explicit equilibrium frequencies, $\hat{u}_{1s}$ and $\hat{w}_{2s}$, are presented in APPENDIX A.

A third general observation is that the cytoplasmic frequency is the only variable with the same time-dependent trajectory for both mating systems. In each case, its precise value in any generation $t = 0, 1, 2, \ldots$ is given by

$$x_t = \bar{x} + (x_0 - \bar{x})(1 - m)^t \to \bar{x} \quad \text{as} \quad t \to \infty. \quad (11)$$

The cytoplasmic frequency in the hybrid zone thus monotonically approaches the overall frequency among the migrants, at a constant geometric rate of $1 - m$ per generation, whether matings are random or assortative. The hybrid zone has a constant cytoplasmic composition only if it is initially identical to the migrant pool (*i.e.*, $x_0 = \bar{x}$). It should be noted that this is the first cytonuclear model we have considered to date (ASMUSSEN, ARNOLD and AVISE 1987; ARNOLD, ASMUSSEN and AVISE 1988) in which the cytoplasmic composition can change through time.

The nuclear gene frequency has the corresponding polymorphic equilibrium value

$$\hat{p} = \frac{2m\bar{p} + (1 - m)\alpha\hat{u}_{1s}}{2m + (1 - m)(\alpha\hat{u}_{1s} + \beta\hat{w}_{2s})} \qquad (12)$$

in the hybrid zone. In the case of random mating ($\alpha = \beta = 0$) this is simply the overall nuclear gene frequency in the migrant pool ($\bar{p}$). In contrast to the cytoplasmic gene frequency the equilibrium nuclear gene frequency in (12) depends, in general, on the levels of assortative mating (*i.e.*, α and β) in the hybrid population. This result parallels our earlier finding (ARNOLD, ASMUSSEN and AVISE 1988) that the epistatic mating systems can cause nuclear (but not cytoplasmic) gene frequency changes in an isolated hybrid zone.

Turning to the four cytonuclear disequilibria, we find upon substituting the equilibrium frequencies from (A2), (A3), (11), and (12) into the recursions in (10) that at equilibrium, the allelic association is

$$\hat{D} = \frac{2m\bar{D} + (1 - m)(\alpha\hat{u}_{1s}\bar{y}\hat{q} + \beta\hat{w}_{2s}\bar{x}\hat{p})}{1 + m}, \qquad (13)$$

while the three genotypic disequilibria are

$$\hat{D}_1 = m\bar{D}_1 + (1 - m)(\hat{p}\hat{D} + \alpha\hat{u}_{1s}\bar{y}\hat{q}]$$

$$\hat{D}_2 = m\bar{D}_2 + (1 - m)[(\hat{q} - \hat{p})\hat{D} - \alpha\hat{u}_{1s}\bar{y}\hat{q} + \beta\hat{w}_{2s}\bar{x}\hat{p}] \qquad (14)$$

$$\hat{D}_3 = m\bar{D}_3 - (1 - m)(\hat{q}\hat{D} + \beta\hat{w}_{2s}\bar{x}\hat{p}).$$

The equilibrium values for the cytonuclear and mar-

ginal single locus genotypic frequencies can similarly be obtained from the remaining recursions in (8)–(9). A standard stability analysis confirms that this unique joint equilibrium is in fact locally stable whenever $0 < \bar{u}_{1s}, \bar{w}_{2s} < 1$, $0 < m < 1$, and $0 \le \alpha, \beta < 1$. We conjecture that this equilibrium may be globally stable as well, since i) the time-dependent values of u_{1s}, w_{2s}, x, p, and D determine those of all other variables; (ii) u_{1s}, w_{2s}, and x monotonically approach their equilibrium values; and (iii) once u_{1s}, w_{2s}, and x are sufficiently close to equilibrium, the recursion for p in (9) becomes (approximately) a one-dimensional monotonic transformation, as does the recursion for D in (10) once p also becomes sufficiently close to its equilibrium.

A fourth and final general conclusion is immediately evident from the formulas in (13)–(14): *continued immigration of either pure parental species can produce permanent cytonuclear disequilibrium in the hybrid zone, whether mating there is random or assortative.* Moreover, whenever the migrants consist solely of the two genetically distinct parental species, the equilibrium allelic association in the hybrid zone ($\hat{D}$) is always less than the fixed allelic disequilibrium in the migrants ($\bar{D} = \bar{x}\bar{y} > 0$). This point follows from the migrant values in (5) and the previously established equilibrium relations, $\hat{u}_{1s} < \bar{u}_{1s}$ ($= \bar{x}$) and $\hat{w}_{2s} < \bar{w}_{2s}$ ($= \bar{y}$), which show that $\alpha\hat{u}_{1s}\bar{y}\hat{q} + \beta\hat{w}_{2s}\bar{x}\hat{p} < \bar{x}\bar{y}(\hat{q} + \hat{p}) = \bar{D}$. Substitution of this inequality into the numerator of (13) immediately yields the relation, $\hat{D} < \bar{D}$. Unlike the previous results, this property fails as a general principle when the migrant pool has an arbitrary genetic composition.

Further equilibrium properties hold whenever $\bar{D}_3 < 0 < \bar{D}, \bar{D}_1$ in the migrant pool, as is true from (5) when the migrants come from the two genetically distinct parental species. For instance, these three cytonuclear disequilibria will then have the same equilibrium sign pattern under both mating systems: $\hat{D}_3 < 0 < \hat{D}, \hat{D}_1$. It is evident from (13), however, that whenever migrants have positive allelic disequilibrium, a hybrid zone with assortative mating ($\alpha > 0$ and/or $\beta > 0$) will achieve a higher allelic association at equilibrium (*i.e.*, a greater $|\hat{D}|$) than one whose residents mate at random. The sign of $\hat{D}_2$ with assortative mating, and the relative magnitudes of the genotypic disequilibria under random and assortative mating depend on the values of α, β, m, $\bar{p}$, $\bar{x}$, $\bar{u}_{1s}$, and $\bar{w}_{2s}$.

Special case of method 1—random mating populations: Additional details can be deduced when mating is random within the hybrid zone. In particular, it is possible in this case to derive and fully analyze explicit time-dependent solutions for all the cytonuclear variables, with the exception of the frequencies of the two pure parental species. (The qualitative dynamics of the latter are, nonetheless, already known

from the analysis of the general model above.)

Inspection of the fourth (p) recursion in Equation 9 shows, for instance, that the nuclear gene frequency in the hybrid zone monotonically approaches the overall nuclear frequency among the migrants at a constant geometric rate of $1 - m$ per generation, with its value in any generation $t = 0, 1, 2, \ldots$ given by

$$p_t = \bar{p} + (p_0 - \bar{p})(1 - m)^t \to \bar{p} \quad \text{as} \quad t \to \infty. \quad (15)$$

Note that this trajectory is completely analogous to that for the cytoplasmic frequency given in (11).

For the special case where the hybrid zone is initialized with just pure species 1 and 2 individuals in the frequencies m_1/m and m_2/m, and the subsequent immigrants have this same constant composition, the population dynamics are particularly simple and informative. In particular, these conditions imply from (11) and (15) that the gene frequencies are constant at both loci, with $p_t = \bar{p}$ and $x_t = \bar{x}$ for $t = 0, 1, 2, \ldots$. These facts plus (5) and the initial condition $D_0 = \bar{D}$, in turn show from (10) that the cytonuclear associations are given in any generation $t = 1, 2, \ldots$, by

$$D_t = \hat{D} + \frac{(1 - m)\bar{D}}{1 + m}\left(\frac{1 - m}{2}\right)^t \to \hat{D}$$

$$D_1{}^{(t)} = m\bar{D}_1 + (1 - m)\bar{p}D_{t-1}$$
$$\to m\bar{D} + (1 - m)\bar{p}\hat{D}$$
$$= m\bar{D}\left[1 + \frac{2(1 - m)\bar{p}}{1 + m}\right]$$

$$D_2{}^{(t)} = m\bar{D}_2 + (1 - m)(\bar{q} - \bar{p})D_{t-1} \quad (16)$$
$$\to (1 - m)(\bar{q} - \bar{p})\hat{D}$$
$$= 2(1 - m)(\bar{q} - \bar{p})\frac{m\bar{D}}{1 + m}$$

$$D_3{}^{(t)} = m\bar{D}_3 - (1 - m)\bar{q}D_{t-1}$$
$$\to -[m\bar{D} + (1 - m)\bar{q}\hat{D}]$$
$$= -m\bar{D}\left[1 + \frac{2(1 - m)\bar{q}}{1 + m}\right],$$

where

$$\hat{D} = 2m\bar{D}/(1 + m). \quad (17)$$

These solutions are relevant to other more general situations as well. For example, the trajectories in (16) are good approximations to the dynamics whenever migrants come only from the two genetically distinct parental species, whatever the initial makeup of the hybrid zone. The equilibrium genotypic associations in (16)–(17) also exactly specify those in the more general case. The formula in (17) for the equilibrium association holds in general, for all initial conditions and migrant compositions and can be used as a rule of thumb to provide a rough estimate of immigration
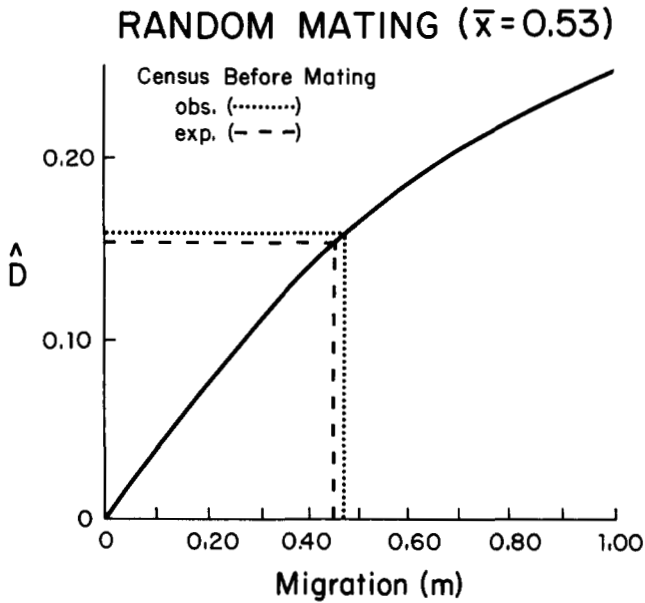
RANDOM MATING ($\bar{x}$ = 0.53)

Census Before Mating
obs. (·········)
exp. (− − −)

FIGURE 2.—Graph of the relationship in (17) showing the equilibrium allelic association as a function of total migration rate (m) into a random mating population censused prior to mating. The observed allelic association is shown for the Hyla Albumin locus (Table 3). The migrant allelic disequilibrium can be estimated from (5) and the observed cytoplasmic frequency to be $\bar{D} = \hat{x}\hat{y} = (0.53)$ $(0.47) = 0.25$. These in (17) yield a rough estimate of $m = 0.47$ for the total migration from the parental species. The actual best fit of the model to the complete cytonuclear data (left branch, middle tier of Figure 4) yields an estimate of $m = 0.45$.

rate from the observed allelic disequilibrium (Figure 2), provided $\bar{D}$ can be estimated in the immigrant pool. This approach, of course, should not be used for a final estimate since it assumes random mating and does not take all the data into account. It could, however, provide a starting point for the computation of maximum likelihood estimates.

Several important points are evident from the disequilibrium trajectories in (16). First, all four disequilibria monotonically decrease in magnitude through time, approaching their equilibrium values at a constant geometric rate of $(1 - m)/2$ per generation. Continued migration serves to accelerate the approach to equilibrium in this case, relative to that for isolated, randomly mating populations (ASMUSSEN, ARNOLD and AVISE 1987). The one exception to the behavior just described is $|D_2^{(t)}|$, which decreases for $t \geq 1$ after increasing from an initial value of 0 to $(1 - m)|\bar{q} - \bar{p}|\bar{D}$ in generation 1.

A second observation is that under the migrant and initial composition in (5), the disequilibrium values in every generation t are determined entirely by the migration rates of the two parental species. Third, the formulas in (16)–(17) show that under random mating, either all four disequilibria will reach nonzero equilibrium values or all but $\hat{D}_2$ will be nonzero. Moreover, in addition to the general sign patterns above (i.e., $\hat{D}_3 < 0 < \hat{D}, \hat{D}_1$), we see here that since $\bar{q}$

$-\bar{p} = (m_2 - m_1)/m$, $\hat{D}_2 > 0$ if $m_1 < m_2$, $\hat{D}_2 < 0$ if $m_1 > m_2$, and $\hat{D}_2 = 0$ if $m_1 = m_2$. The sign of $\hat{D}_2$ is thus a direct reflection of asymmetry in the migration rates of the two parental species into randomly mating hybrid zones (provided these represent the entire migrant pool).

A final observation is that the ratio of the equilibrium allelic association in the hybrid zone to the fixed allelic disequilibrium among the migrants, $\hat{D}/\bar{D} = 2m/(1 + m)$, is an increasing function of the total migration rate, m, increasing from 0 if there is no migration ($m = 0$) to 1 in the limiting case of complete migration ($m = 1$) (Figure 2).

Upon substitution of the gene frequency and allelic disequilibrium solutions in (11) and (15)–(17) into the righthand sides of the recursions in (8), the corresponding trajectories (with $\alpha = \beta = 0$) for the frequencies of the six cytonuclear genotypes are also readily obtained, with

$$u_1^{(t)} = m\bar{u}_1 + (1 - m)\bar{p}^2\bar{x} + (1 - m)\bar{p}D_{t-1}$$
$$v_1^{(t)} = m\bar{v}_1 + 2(1 - m)\bar{p}\bar{q}\bar{x} + (1 - m)(\bar{q} - \bar{p})D_{t-1}$$
$$w_1^{(t)} = m\bar{w}_1 + (1 - m)\bar{q}^2\bar{x} - (1 - m)\bar{q}D_{t-1}$$
$$u_2^{(t)} = m\bar{u}_2 + (1 - m)\bar{p}^2\bar{y} - (1 - m)\bar{p}D_{t-1} \quad (18)$$
$$v_2^{(t)} = m\bar{v}_2 + 2(1 - m)\bar{p}\bar{q}\bar{y} - (1 - m)(\bar{q} - \bar{p})D_{t-1}$$
$$w_2^{(t)} = m\bar{w}_2 + (1 - m)\bar{q}^2\bar{y} + (1 - m)\bar{q}D_{t-1}$$

for all $t = 1, 2, \ldots$.

These solutions reveal several final features of the special random mating situation. First, the two locus genotypes associated with the two parental species (AA/M and aa/m) will decrease in the hybrid zone to positive equilibrium levels, while the other two types of nuclear homozygotes (AA/m and aa/M) steadily increase. Second, the transient frequencies of the two heterozygote genotypes, Aa/M and Aa/m, depend on the relative magnitudes of the pure parental migration rates. If species 1 predominates (i.e., $m_1 > m_2$), heterozygotes with the cytotype of species 1 (Aa/M) steadily increase, while those with the cytotype of species 2 (Aa/m) decrease. The situation is reversed when $m_2 > m_1$.

In contrast, both heterozygotes reach equilibrium frequencies in a single generation if the two parental species have equal migration rates. A steady increase or decrease in the frequencies of the two heterozygous genotypes is thus another indication of asymmetric migration rates by the two pure species when hybrid zone matings are random with respect to cytonuclear genotype, assuming the hybrid zone begins with only the two parental species in the same frequencies as in the subsequent migrant pool.

Effect of census timing: Under the alternative census method, the hybrid zone is censused following mating and reproduction and before migration (Fig-

ure 1). Although this results in a different model, a detailed analysis (see APPENDIX B) shows that the cytonuclear variables have the same basic qualitative behavior as when censusing is before reproduction. The gene frequency at the cytoplasmic locus has the same recursion under both census schemes, as well as under both random and assortative mating. The cytoplasmic frequency in each case monotonically approaches the overall frequency in the migrants, with the exact value in any generation t given by (11). *The dynamics of cytoplasmic allele frequencies thus appear robust to changes in the assumed mating system and particularly sensitive to the pattern of gene flow.*

The census schemes differ in the precise values taken by all the other variables through time, being out of phase due to the different sequence of mating and migration within the contrasting census intervals. This phase shift is reflected by a simple relationship between the two sets of equilibrium values, which can be specified formally via the following notation. For each cytonuclear variable z, denote its overall value in the migrant pool by $\bar{z}$, and its equilibrium values under census methods 1 and 2 by $\hat{z}_1$ and $\hat{z}_2$, respectively. As shown in APPENDIX C, the two equilibrium values then satisfy

$$\hat{z}_1 = m\bar{z} + (1 - m)\hat{z}_2. \tag{19}$$

In other words, the equilibrium values for censusing after migration (census scheme 1) are each a weighted average of the migrant value and the equilibrium for censusing after mating (census scheme 2). The equilibria for the first census method thus equal the (unobserved) equilibrium values midway through the census interval under census method 2. The cytoplasmic frequency, x, of course satisfies the much stronger condition, $\hat{x}_1 = \hat{x}_2 = \bar{x}$.

A number of important points follow from the equilibrium relations in (19). First, the two equilibrium values for a given cytonuclear variable are either both higher than, both lower than, or both equal to that variable's overall value in the migrant pool. More specifically, (19) shows that for a given variable z, either $\hat{z}_2 < \hat{z}_1 < \bar{z}$, $\bar{z} < \hat{z}_1 < \hat{z}_2$, or $\hat{z}_1 = \hat{z}_2 = \bar{z}$. These comparisons of course assume both census methods are carried out under identical migrant compositions, migration rates, and mating systems.

Since we have already shown that the equilibrium hybrid zone frequencies of the two parental species are less under each census scheme than those in the migrant pool (APPENDIX A), it follows that the first method, with censusing after migration and before mating, always maintains a higher frequency of the two parental species than does the second method, where censusing is done after mating and before migration. [These results can also be verified directly from the corresponding equilibrium formulas in (A2), (A3) and (B7).] The same reasoning shows that the
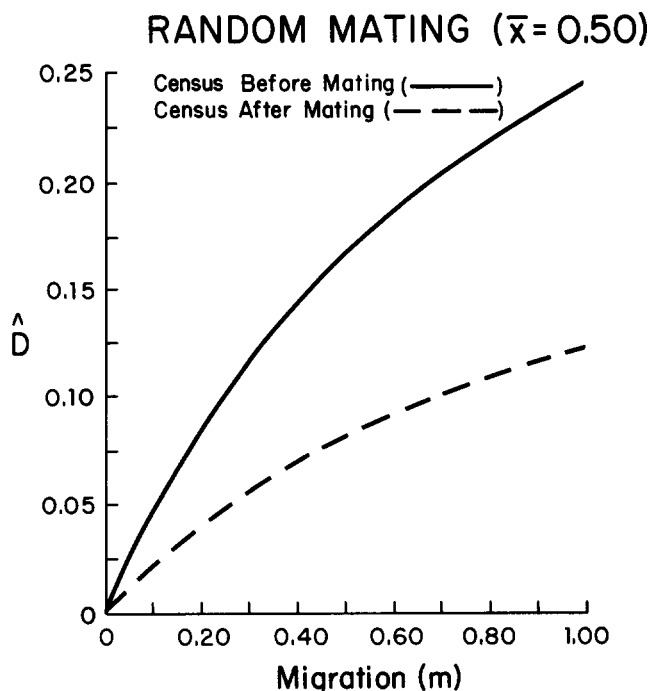

FIGURE 3.—Equilibrium allelic association $(\hat{D})$ in a randomly mating population as a function of the proportion of immigrants (m). The relation is graphed for two censusing schemes, before and after mating, assuming $\bar{x} = 0.5$ and $\bar{D} = \bar{x}\bar{y} = 0.25$.

first census method also always generates a larger allelic disequilibrium in hybrid zones whose immigrants consist only of the two genetically distinct parental species. (In general, the $\hat{D}$ order can be reversed in assortatively mating hybrid zones if the migrant allelic disequilibrium is low.)

In the case of randomly mating hybrid zones, (17) and (B13) show that censusing after migration always results in twice the equilibrium allelic association found when censusing is done after mating (Figure 3). If both census methods begin with the same values, as when the hybrid zone is initialized with just the two parental species, a gap immediately arises between the two sets of allelic associations at census $t = 1$ and steadily widens at each successive census. A comparison of (16) and (B12) shows, moreover, that whenever randomly mating hybrid zones only receive migrants from the two genetically distinct parental species, the first census method always results in greater genotypic disequilibria for nuclear homozygotes (i.e., higher $|\hat{D}_1|$ and $|\hat{D}_3|$), and smaller genotypic disequilibria for heterozygotes (i.e., lower $|\hat{D}_2|$) than does the second census scheme.

In general, the relative orders at equilibrium of the disequilibria and the other cytonuclear variables under the two census methods are not clear-cut. Whatever the order, however (i.e., whether $\bar{z} < \hat{z}_1 < \hat{z}_2$ or $\hat{z}_2 < \hat{z}_1 < \bar{z}$), the equilibrium values from census method 1 are always closer to the values in the migrant pool than are the equilibrium values under census method 2.

DISCUSSION

We have shown that continued migration of genetically distinct parentals into a hybrid zone generates permanent nonzero cytonuclear disequilibria, with or without assortative mating. We have also shown that the order of population censusing (before or after reproduction by immigrants) can have a dramatic effect on the magnitude, but not the pattern of disequilibrium. This is because genetic recombination from matings in the hybrid zone serves to undo the disequilibrium generated by admixture. In the extreme case of random mating, the equilibrium allelic association before mating is twice its value after mating. These conclusions have obvious implications for the design and interpretation of studies that monitor nuclear and cytoplasmic markers simultaneously. In particular, it is important to know when in the generation cycle censusing is done, and if possible, censusing should be after migration and before reproduction to enhance the ability to detect nonzero allelic disequilibria.

We now apply the models above to an actual hybrid population. In a series of ponds near Auburn, Alabama, the treefrog species *Hyla cinerea* and *H. gratiosa* have engaged in extensive hybridization for at least the past 30 yr (Lamb and Avise 1986; Schlefer *et al.* 1986). Normally, these species are fixed for alternate alleles at five allozyme loci and also exhibit diagnostic mitochondrial genotypes, but at the Auburn site a wide variety of recombinant genotypes is found. Furthermore, within the Auburn population there are significant nonrandom cytonuclear associations that cannot entirely be accounted for by assortative mating (Arnold, Asmussen and Avise 1988). Additional factors, such as selection against hybrids and/or migration into the population, must be at work. The issue addressed here is what (if any) amount of immigration yields equilibrium expectations for cytonuclear frequencies and associations that match the values observed.

To estimate the amount of gene flow, we fit a series of five models to the data by the method of maximum likelihood. In each case, the likelihood is the product of five factors, one for each unlinked allozyme locus. The factor for the kth allozyme locus is the standard multinomial term,

$$u_{1s}^{N_{1k}} \, u_{1h}^{N_{2k}} \ldots w_{2s}^{N_{8k}} \tag{20}$$

where $\mathbf{N}_k = (N_{1k}, \ldots, N_{8k})$ denotes the counts of the eight cytonuclear genotypes in a sample at that allozyme locus (with the partitions in (4) obtained as in Arnold, Asmussen and Avise 1988), and $\boldsymbol{\beta} = (u_{1s}, \ldots, w_{2s})$ is the corresponding vector of genotypic frequencies.

In fitting the models, it was assumed that the frequencies $\boldsymbol{\beta}$ are those achieved at equilibrium and that
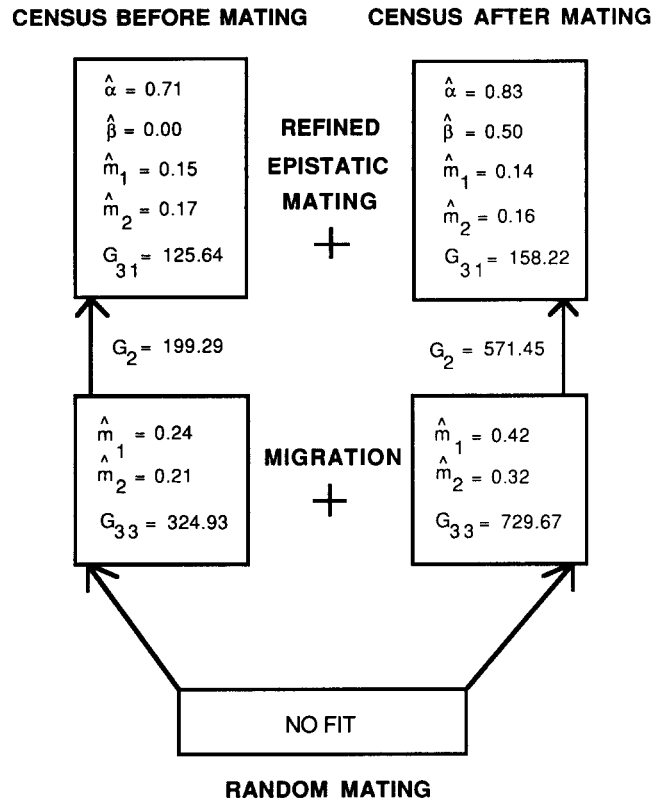

FIGURE 4.—Best fits of five models to the Hyla data. The bottom (random mating) model assumes random mating and no migration, and corresponds to Equations 7–8, or equivalently Equations B1–B4, with $m_1 = m_2 = 0$ and $\alpha = \beta = 0$. Models in the left branch correspond to Equations 7–8; those in the right branch correspond to Equations B1–B4. Models in the middle tier assume migration alone, with random mating ($\alpha = \beta = 0$). Those on the top tier assume both migration and refined epistatic mating. The goodness-of-fit is measured by the usual G-statistic, with degrees of freedom denoted by the subscript.

migrants consist solely of the two genetically distinct parental species, with the values specified in (5). The parameter vector $\boldsymbol{\beta}$ is the same across allozyme loci (and independent of k). We used a derivative-free method of maximizing the loglikelihood (Press *et al.* 1986, p. 299). An exhaustive search of the parameter space on a 0.01 grid verified that the local maximum found by the derivative-free method was indeed a global maximum. Standard errors on the parameter estimates were obtained by resampling 1000 times through the bootstrap method (Efron 1983) based on the observed proportions of each cytonuclear genotype $u_{1s}, \ldots, w_{2s}$ for each allozyme locus in combination with the mitochondrial locus. This same procedure was repeated for each of the five models. In each case, G-tests (as in Asmussen, Arnold and Avise 1987) were used to assess goodness-of-fit, and differences among G-statistics were used to compare models.

Figure 4 shows the fit of the five models to the Hyla data. The bottom tier in the figure summarizes results for the simplest model—that with no migration into

TABLE 2

Maximum likelihood estimates for the assortative mating rates (α, β) and the migration rates (m_1, m_2) for the Hyla data under the model of migration and refined epistatic mating

Locus	$\hat{\alpha}$	$\hat{\beta}$	$\hat{m}_1$	$\hat{m}_2$	G_3
Alb	0.77 (0.06)	0.13 (0.21)	0.13 (0.03)	0.15 (0.03)	11.01
Pgi	0.70 (0.05)	0.00 (0.01)	0.15 (0.01)	0.18 (0.01)	45.29
Ldh	0.70 (0.07)	0.00 (0.01)	0.15 (0.01)	0.18 (0.01)	22.31
Pep	0.71 (0.04)	0.00 (0.03)	0.15 (0.01)	0.17 (0.01)	29.50
Mdh	0.72 (0.05)	0.00 (0.02)	0.15 (0.01)	0.17 (0.01)	17.53
Overall	0.71 (0.005)	0.00 (0.009)	0.15 (0.001)	0.17 (0.001)	1.07[a]

Censusing is assumed to take place before mating. G_3 is the goodness-of-fit statistic with 3 degrees of freedom.
[a] Homogeneity test (df = 16).

the hybrid population and random mating of individuals in and around the pond; the census timing is irrelevant in this case. The presence of the two parental types has a likelihood of zero under this simplest hypothesis, and thus precludes random mating alone (or refined epistatic mating alone) as a tenable model for the Hyla population. Models in the middle tier of Figure 4 assume random mating coupled with migration, either with censusing before mating (left side of figure) or after mating (right side). Both can be fit to the data, unlike the model without migration. However, models in the top tier of Figure 4, which assume refined epistatic mating coupled with migration, offer significant improvement over the nested models with migration alone. The first conclusion is that both migration and positive assortative mating between conspecifics may play a role in generating the observed data.

The goodness-of-fit statistic in the upper migration models is smaller for the case in which censusing takes place before rather than after refined epistatic mating (125.64 vs. 158.22). Since these models are not nested, the usual test based on a difference of G-test statistics cannot be made, but a bootstrap method may nonetheless be employed (WAHRENDORF, BECHER and BROWN 1987). In 1000 bootstrap samples from the original data, the G-statistic always showed a lower value in the model with censusing before mating, so we conclude that the improvement in fit is indeed statistically significant. This model also coincides with the presumed census timing for this population.

Table 2 reports the resulting estimates of the mating fidelities (α for H. cinerea and β for H. gratiosa) and the gene flow parameters (m_1 and m_2 for the two respective species) for each allozyme locus, as well as for the combined estimates across loci. [Estimates for individual loci were obtained by maximizing (20).] A G-test of heterogeneity (G = 1.07, 16 df) between allozyme loci is not significant, indicating that all assayed genes provide consistent estimates of rates of migration and assortative mating in this hybrid population. Under the best-fitting model, an estimated

TABLE 3

Comparison of observed disequilibria in a hybrid zone of Hyla, and best-fit equilibrium values expected under refined epistatic mating and migration ($\hat{\alpha}$ = 0.71, $\hat{\beta}$ = 0.00, $\hat{m}_1$ = 0.15, $\hat{m}_2$ = 0.17)

Nuclear locus	Cytonuclear disequilibria			
	D	D_1	D_2	D_3
Alb	0.16	0.19	-0.06	-0.13
Pgi	0.19	0.22	-0.07	-0.15
Ldh	0.18	0.20	-0.05	-0.15
Pep	0.18	0.21	-0.06	-0.15
Mdh	0.17	0.21	-0.07	-0.14
Expected	0.15	0.17	-0.05	-0.12

Censusing is assumed to take place before mating.

32% of the inhabitants of the hybrid zone are immigrants each generation, and both parental species appear to make a nearly equal contribution to the hybrid zone (m_1 = 0.15 and m_2 = 0.17).

The current results, coupled with our earlier mating models (ARNOLD, ASMUSSEN and AVISE 1988), demonstrate that in principle, an asymmetrical contribution of conspecifics to the hybrid zone ($m_1 \neq m_2$), and/or an asymmetry in the mating behavior ($\alpha \neq \beta$), can produce a significant D_2. Since m_1 and m_2 are very close in the best-fitting model (Table 2), the significant D_2 in the Hyla data are probably best explained by an asymmetry in mating behavior—H. cinerea females display a considerably stronger fidelity to conspecific males than do H. gratiosa females (Table 2). This is consistent with a previous behavioral prediction that H. cinerea males may intercept and mate H. gratiosa females, while H. gratiosa males, because of their distinct calling-site positions, have little opportunity to encounter H. cinerea females (LAMB and AVISE 1986).

Overall, a model of assortative mating coupled with migration can, given appropriate choice of parameter values ($\hat{\alpha}$, $\hat{\beta}$, $\hat{m}_1$, $\hat{m}_2$), lead to equilibrium expectations for cytonuclear associations ($\hat{D}$, $\hat{D}_1$, $\hat{D}_2$, $\hat{D}_3$) that closely match the observed values in the hybrid Hyla population (Table 3). [There is, however, a small but

TABLE 4

Observed (O) and expected (E) counts of the eight cytonuclear genotypes for the Albumin locus in the Hyla hybrid population

	AA (pure)	AA (hybrid)	Aa	aa (hybrid)	aa (pure)	
M	103	23	11	5	—	O
	102	15	21	5	—	E
	(0)	(2)	(−2)	(0)	—	R
m	—	20	54	29	60	O
	—	22	59	25	58	E
	—	(0)	(−1)	(1)	(0)	R

$$G_3 = 11.01$$
$$P = 0.011$$

The expected counts were generated from the best-fit equilibrium values for the Albumin locus (Table 2) under a model of refined epistatic mating and migration ($\hat{\alpha} = 0.77$, $\hat{\beta} = 0.13$, $\hat{m}_1 = 0.13$, and $\hat{m}_2 = 0.15$). The residuals (R) are a measure of the discrepancies between O and E (see ANSCOMBE 1982, ch. 12). Residuals are squared and summed to yield the goodness-of-fit statistic, $G_3 = 11.01$.

significant discrepancy between the observed and expected genotypic counts (Table 4).] Of course, these findings do not necessarily indicate that the models reflect the biological reality of the hybrid *Hyla* situation, since other (unexplored) models might also "fit" the genetic patterns. In particular, the hybrid population may not be at equilibrium with respect to cytonuclear associations, various hybrid classes (F_1's, backcrosses, later generation hybrids) may mate nonrandomly, and particular hybrid classes may be selected against. The effects of selection on cytonuclear disequilibria (CLARK 1984) will be taken up elsewhere.

In conclusion, cytonuclear systems offer a novel perspective on the study of gene flow and mating pattern in a hybrid zone. The uniparental transmission of a cytoplasmic genome, and its lack of linkage with nuclear genes, produces a cytoplasmic frequency trajectory that is particularly sensitive to gene flow, but relatively unconstrained by the mating system, even when there is an epistatic interaction between the cytoplasm and nucleus in determining mating preference. Yet the cytonuclear associations are influenced in predictable ways by both the magnitude of migration into a hybrid zone, and the assortative mating propensities of the species involved. A comparison of the separate versus joint behaviors of cytoplasmic and nuclear genes can help to partition the effects of migration and the mating system on the genetic architecture of hybrid or other admixture zones.

LITERATURE CITED

ANSCOMBE, F. J., 1982 *Computing in Statistical Science through APL.* Springer-Verlag, New York.

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

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.

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

EFRON, B., 1983 The jacknife, the bootstrap, and other resampling plans. Soc. Ind. Appl. Math. CBMS-NSF Monograph 38.

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

PRESS, W. H., B. P. FLANNERY, S. A. TEUKOLSKY and W. T. VETTERLING, 1986 *Numerical Recipes, the Art of Scientific Computing.* Cambridge University Press, New York.

SCHLEFER, E., M. ROMANO, S. I. SUTTMAN and S. RUTH, 1986 Effects of twenty years of hybridization in a disturbed habitat on *Hyla cinerea* and *Hyla gratiosa.* J. Herpetol. **20:** 208–219.

WAHRENDORF, J., H. BECHER and C. C. BROWN, 1987 Bootstrap comparison of non-nested generalized linear models: applications in survival analysis and epidemiology. Appl. Statist. **36:** 72–81.

Communicating editor: B. S. WEIR

APPENDIX A

Dynamical behavior of the parental frequencies, u_{1s} and w_{2s}, under census method 1: The u_{1s} recursion in (7a) shows that its equilibrium value satisfies the quadratic equation

$$f(u_{1s}) = (1 - \alpha)(1 - m)u_{1s}^2$$

$$+ [\alpha(1 - m) - 1]u_{1s} + m\bar{u}_{1s} = 0. \tag{A1}$$

A sketch of the graph of $f(u_{1s})$ based on the sign of $f'(u_{1s})$ shows that whenever $0 < \bar{u}_{1s}$, $m < 1$ and $0 \le \alpha < 1$, the function $f(u_{1s})$ has two positive real roots. One of these roots lies in the interval $(0,1)$ and thus determines a valid equilibrium value, while the other root exceeds 1. There is therefore always exactly one equilibrium frequency, $\hat{u}_{1s}$, for pure species 1, which is the

smaller of the two roots of (A1). From the quadratic formula this is given explicitly by

$$\hat{u}_{1s} = \frac{1 - \alpha(1 - m) - \sqrt{\Delta_1}}{2(1 - \alpha)(1 - m)} \qquad \text{(A2a)}$$

where

$$\Delta_1 = [1 - \alpha(1 - m)]^2 - 4(1 - \alpha)(1 - m)m\bar{u}_{1s}. \qquad \text{(A2b)}$$

Although it is not possible to provide an explicit solution for the frequency of species 1 through time, its qualitative behavior can nonetheless be easily deduced from the fact that the recursion for u_{1s} in Equation 7a is a monotonic transformation (*i.e.*, $du_{1s}'/du_{1s} = (1 - m)[2(1 - \alpha)u_{1s} + \alpha]$, which is positive for all $0 < u_{1s} \leq 1$). The frequency of pure species 1 in the hybrid zone thus either monotonically increases or decreases to the unique equilibrium, $\hat{u}_{1s}$, depending on whether its initial frequency is below or above this limiting value.

The final observation, concerning the relation between $\hat{u}_{1s}$ and $\bar{u}_{1s}$, is evident from further analysis of the graph of the function $f(u_{1s})$ in (A1), which shows that $f(u_{1s}) > 0$ if $0 \leq u_{1s} < \hat{u}_{1s}$ and $f(u_{1s}) < 0$ if $\hat{u}_{1s} < u_{1s} \leq 1$. Since $f(\bar{u}_{1s}) = -(1 - \alpha)(1 - m)\bar{u}_{1s}(1 - \bar{u}_{1s}) < 0$, it follows that $\hat{u}_{1s} < \bar{u}_{1s}$, and thus that the equilibrium frequency of pure species 1 is always both positive and less than that in the overall migrant pool. Analogous steps establish that whenever $0 < \bar{w}_{2s}$, $m < 1$ and $0 \leq \beta < 1$, the frequency of pure species 2 in the hybrid zone monotonically approaches

$$\hat{w}_{2s} = \frac{1 - \beta(1 - m) - \sqrt{\Delta_2}}{2(1 - \beta)(1 - m)} \qquad \text{(A3a)}$$

where

$$\Delta_2 = [1 - \beta(1 - m)]^2 - 4(1 - \beta)(1 - m)m\bar{w}_{2s} \qquad \text{(A3b)}$$

and $0 < \hat{w}_{2s} < \bar{w}_{2s} < 1$. Note that the equilibrium frequency of each parental species is uniquely determined by its assortative mating rate, its frequency in the combined migrant pool, and the total frequency of immigrants (m).

APPENDIX B

Method 2—censusing after mating and before migration: Under this alternative censusing scheme, the frequencies in Table 1 and Equation 4 represent the values in the hybrid zone following mating and reproduction (Figure 1). Each generation cycle begins with an influx of immigrants, and ends after reproduction.

The state of the hybrid zone population at the next census depends on the interim values of five critical variables after migration: the frequencies of the two pure parental species, the nuclear and cytoplasmic gene frequencies, and the allelic disequilibrium. These are readily calculated as the result of admixture between the joint migrant pool and the existing residents of the hybrid zone to be

$$u_{1s}^* = m\bar{u}_{1s} + (1 - m)u_{1s}, \quad w_{2s}^* = m\bar{w}_{2s} + (1 - m)w_{2s} \qquad \text{(B1)}$$
$$p^* = m\bar{p} + (1 - m)p \qquad x^* = m\bar{x} + (1 - m)x$$
$$D^* = m\bar{D} + (1 - m)D + m(1 - m)(x - \bar{x})(p - \bar{p}).$$

The final values at the next hybrid zone census are then obtained by substituting the (starred) interim values from (B1) into the righthand side of the refined epistatic mating recursions (ARNOLD, ASMUSSEN and AVISE 1988). The new hybrid zone frequencies of the two parental species are thus

$$u_{1s}' = u_{1s}^*[\alpha + (1 - \alpha)u_{1s}^*] \quad \text{and} \qquad \text{(B2)}$$
$$w_{2s}' = w_{2s}^*[\beta + (1 - \beta)w_{2s}^*],$$

while the new allelic frequencies are

$$p' = p^* + \tfrac{1}{2}(\alpha u_{1s}^*q^* - \beta w_{2s}^*p^*) \quad \text{and} \qquad \text{(B3)}$$
$$x' = x^* = m\bar{x} + (1 - m)x.$$

The recursions for the six composite cytonuclear genotypes are similarly found to be

$$u_1' = (p^*)^2 x^* + p^*D^* + \alpha u_{1s}^*q^*$$
$$v_1' = 2p^*q^*x^* + (q^* - p^*)D^* - \alpha u_{1s}^*q^*$$
$$w_1' = (q^*)^2 x^* - q^*D^*$$
$$u_2' = (p^*)^2 y^* - p^*D^* \qquad \text{(B4)}$$
$$v_2' = 2p^*q^*y^* - (q^* - p^*)D^* - \beta w_{2s}^*p^*$$
$$w_2' = (q^*)^2 y^* + q^*D^* + \beta w_{2s}^*p^*,$$

which in turn show that the new marginal genotypic frequencies at the nuclear locus are

$$u' = (p^*)^2 + \alpha u_{1s}^*q^*$$
$$v' = 2p^*q^* - \alpha u_{1s}^*q^* - \beta w_{2s}^*p^* \qquad \text{(B5)}$$
$$w' = (q^*)^2 + \beta w_{2s}^*p^*.$$

Finally, Equations B1 and B3–B5 together with the definitions in (1)–(3) produce the disequilibrium recursions

$$D' = \tfrac{1}{2}D^* + \tfrac{1}{2}(\alpha u_{1s}^*y^*q^* + \beta w_{2s}^*x^*p^*)$$
$$D_1' = p^*D^* + \alpha u_{1s}^*y^*q^*$$
$$D_2' = (q^* - p^*)D^* - \alpha u_{1s}^*y^*q^* + \beta w_{2s}^*x^*p^* \qquad \text{(B6)}$$
$$D_3' = -q^*D^* - \beta w_{2s}^*x^*p^*.$$

Only the cytoplasmic frequency has exactly the same trajectory under the two census schemes. The hybrid zone frequencies of the two pure parentals in this case monotonically approach

$$\hat{u}_{1s} = \frac{1 - \alpha(1 - m) - 2(1 - \alpha)(1 - m)m\bar{u}_{1s} - \sqrt{\Delta_1}}{2(1 - \alpha)(1 - m)^2} \qquad \text{(B7a)}$$

and

$$\hat{w}_{2s} = \frac{1 - \beta(1 - m) - 2(1 - \beta)(1 - m)m\bar{w}_{2s} - \sqrt{\Delta_2}}{2(1 - \beta)(1 - m)^2} \qquad \text{(B7b)}$$

where $0 < \hat{u}_{1s} < \bar{u}_{1s} < 1$ if $0 < \bar{u}_{1s}$, $m < 1$ and $0 \leq \alpha < 1$; $0 < \hat{w}_{2s} < \bar{w}_{2s} < 1$ if $0 < \bar{w}_{2s}$, $m < 1$ and $0 \leq \beta < 1$; and the discriminants, Δ_1 and Δ_2, are as in (A2b) and (A3b) for the first census method.

The nuclear gene frequency now has the corresponding polymorphic equilibrium value

$$\hat{p} = \frac{m\bar{p}(2 - \alpha\hat{u}_{1s}^* - \beta\hat{w}_{2s}^*) + \alpha\hat{u}_{1s}^*}{2m + (1 - m)(\alpha\hat{u}_{1s}^* + \beta\hat{w}_{2s}^*)}. \qquad \text{(B8)}$$

where

$$\hat{u}_{1s}^* = m\bar{u}_{1s} + (1 - m)\hat{u}_{1s} \text{ and } \hat{w}_{2s}^* = m\bar{w}_{2s} + (1 - m)\hat{w}_{2s} \qquad \text{(B9)}$$

are the unobserved equilibrium frequencies of the two parental species midway through a census interval, immediately after migration. As with the first census method, the equilibrium nuclear frequency equals the overall nuclear gene frequency in the migrant pool when hybrid zone matings are random.

Under the current census scheme, the four equilibrium associations are, from (B6):

$$\hat{D} = \frac{m\bar{D} + \alpha\hat{u}_{1s}^{*}\bar{y}\hat{q}^{*} + \beta\hat{w}_{2s}^{*}\bar{x}\hat{p}^{*}}{1 + m}$$

$$\hat{D}_1 = \hat{p}^{*}\hat{D}^{*} + \alpha\hat{u}_{1s}^{*}\bar{y}\hat{q}^{*} \tag{B10}$$

$$\hat{D}_2 = (\hat{q}^{*} - \hat{p}^{*})\hat{D}^{*} - \alpha\hat{u}_{1s}^{*}\bar{y}\hat{q}^{*} + \beta\hat{w}_{2s}^{*}\bar{x}\hat{p}^{*}$$

$$\hat{D}_3 = -\hat{q}^{*}\hat{D}^{*} - \beta\hat{w}_{2s}^{*}\bar{x}\hat{p}^{*},$$

where

$$\hat{p}^{*} = m\bar{p} + (1 - m)\hat{p} \quad \text{and} \quad \hat{D}^{*} = m\bar{D} + (1 - m)\hat{D} \tag{B11}$$

are the equilibrium nuclear gene frequency and allelic association midway through the census interval, with $\hat{q}^{*} = 1 - \hat{p}^{*}$. Note that the corresponding (midcensus) equilibrium cytoplasmic frequencies are simply the migrant frequencies $\bar{x}$ and $\bar{y} = 1 - \bar{x}$, since $\hat{x} = \bar{x}$ implies that $\hat{x}^{*} = m\bar{x} + (1 - m)\hat{x} = \bar{x}$. The equilibrium values for the six basic cytonuclear genotypic frequencies can now be obtained by substituting this relation plus the formulas for $\hat{u}_{1s}^{*}$, $\hat{w}_{2s}^{*}$, $\hat{p}^{*}$, and $\hat{D}^{*}$ specified by (B7)–(B11), into the righthand sides of (B4). As for the first census method, the unique joint hybrid zone equilibrium is locally (and we conjecture also globally) stable whenever $0 < \hat{u}_{1s}, \hat{w}_{2s} < 1$, $0 < m < 1$, and $0 \le \alpha, \beta < 1$.

All the relationships found among the equilibrium associations in (13)–(14), for the first census method, also apply to those in (B10) under the current census scheme. In this case, the equilibrium pattern, $\hat{D}_3 < 0 < \hat{D}$, $\hat{D}_1$, holds more generally whenever $\bar{D} > 0$. A slightly modified argument is needed to show from (B10) and (5) that again $\hat{D} < \bar{D}$, whenever migrants consist solely of the two genetically distinct parental species. The key fact for this census scheme is that the equilibrium relations, $\hat{u}_{1s} < \bar{u}_{1s} (= \bar{x})$ and $\hat{w}_{2s} < \bar{w}_{2s} (= \bar{y})$, imply from (B9) that $\hat{u}_{1s}^{*} < \bar{x}$ and $\hat{w}_{2s}^{*} < \bar{y}$.

Special case of census method 2—random mating populations: For random mating hybrid zones, explicit solutions can again be given for the trajectories of all the cytonuclear variables except the frequencies of the two pure parentals. Those for the two gene frequencies are the same as those in (11) and (15) for the first census method. As before, we focus on the solutions for the special case where the hybrid zone is initialized with just the two genetically distinct parental species in the same relative proportions as in the subsequent migrants (*i.e.*, $u_{1s}^{(0)} = \bar{u}_{1s} = \bar{x} = m_1/m$ and $w_{2s}^{(0)} = \bar{w}_{2s} = \bar{y} = m_2/m$).

The allelic and genotypic disequilibria are then given in every generation $t \ge 1$ by

$$D_t = \hat{D} + \frac{\bar{D}}{1 + m}\left(\frac{1 - m}{2}\right)^{t} \rightarrow \hat{D}$$

$$D_1^{(t)} = 2\bar{p}D_t \rightarrow 2\bar{p}\hat{D}$$

$$D_2^{(t)} = 2(\bar{q} - \bar{p})D_t \rightarrow 2(\bar{q} - \bar{p})\hat{D} \tag{B12}$$

$$D_3^{(t)} = -2\bar{q}D_t \rightarrow -2\bar{q}\hat{D}$$

where

$$\hat{D} = m\bar{D}/(1 + m). \tag{B13}$$

(Although their transient behavior depends in general on the initial hybrid zone frequencies, the equilibrium values of all four cytonuclear disequilibria have the same form under all initial and migrant compositions.)

The other qualitative results for the disequilibrium trajectories in (16)–(17), under the first census scheme, apply here as well, with the single exception that under this second census method the ratio $\hat{D}/\bar{D}$ only increases to $\frac{1}{2}$ as the total migration rate increases from 0 to the limiting value of 1. The solutions in (11), (15) and (B12)–(B13) show in conjunction with (B4) that in each generation $t \ge 1$ the six basic cytonuclear frequencies are

$$u_1^{(t)} = \bar{p}^2\bar{x} + 2\bar{p}D_t$$

$$v_1^{(t)} = 2\bar{p}\bar{q}\bar{x} + 2(\bar{q} - \bar{p})D_t$$

$$w_1^{(t)} = \bar{q}^2\bar{x} - 2\bar{q}D_t$$

$$u_2^{(t)} = \bar{p}^2\bar{y} - 2\bar{p}D_t \tag{B14}$$

$$v_2^{(t)} = 2\bar{p}\bar{q}\bar{y} - 2(\bar{q} - \bar{p})D_t$$

$$w_2^{(t)} = \bar{q}^2\bar{y} + 2\bar{q}D_t.$$

While different in form, these have the same qualitative behavior as the corresponding solutions in (18) for the first census method.

APPENDIX C

Derivation of Equation 19: The equilibrium relation in (19) can be easily and directly verified for the frequencies of the pure parentals (u_{1s} and w_{2s}) using (A2) with (B7a) and (A3) with (B7b), the nuclear gene frequency (p) using (12) with (B8), and the allelic disequilibrium (D) using (13) with (B10)–(B11). Since these four variables plus the migrants' cytoplasmic frequency suffice to describe the remaining equilibrium values in the hybrid zone, it follows from (8) plus (B4), (9) plus (B5), and (14) plus (B10)–(B11) that (19) also holds for the three genotypic disequilibria (D_1, D_2, D_3), as well as for the frequencies of the six basic cytonuclear types ($u_1, v_1, w_1, u_2, v_2, w_2$) and the marginal frequencies of the three nuclear genotypes (u, v, w).