EXTRANUCLEAR DIFFERENTIATION AND GENE FLOW IN THE FINITE ISLAND MODEL

NAOYUKI TAKAHATA[1] AND STEPHEN R. PALUMBI[2]

Department of Zoology, NJ-15, University of Washington, Seattle, Washington 98195

Manuscript received April 11, 1984
Revised copy accepted October 5, 1984

ABSTRACT

Use of sequence information from extranuclear genomes to examine deme structure in natural populations has been hampered by lack of clear linkage between sequence relatedness and rates of mutation and migration among demes. Here, we approach this problem in two complementary ways. First, we develop a model of extranuclear genomes in a population divided into a finite number of demes. Sex-dependent migration, neutral mutation, unequal genetic contribution of separate sexes and random genetic drift in each deme are incorporated for generality. From this model, we derive the relationship between gene identity probabilities (between and within demes) and migration rate, mutation rate and effective deme size. Second, we show how within- and between-deme identity probabilities may be calculated from restriction maps of mitochondrial (mt) DNA. These results, when coupled with our results on gene flow and genetic differentiation, allow estimation of relative interdeme gene flow when deme sizes are constant and genetic variants are selectively neutral. We illustrate use of our results by reanalyzing published data on mtDNA in mouse populations from around the world and show that their geographic differentiation is consistent with an island model of deme structure.

WRIGHT (1970, 1977, 1978 and references therein) suggested that evolution would act more quickly in species with semi-isolated populations linked by low gene flow and influenced by independent selective regimes. Direct nucleic acid comparisons (either single gene comparisons or larger homologous DNA segments) within or between species have great potential for addressing the problems of population substructure or species divergence. Gene purification and sequencing (BROWN *et al.* 1982; MARTIN, VINCENT and WILSON 1983), thermal renaturation (BRITTEN, CETTA and DAVIDSON 1978; BROWN, GEORGE and WILSON 1979) and restriction endonuclease digestions (AVISE, LANSMAN and SHADE 1979; AVISE *et al.* 1979; FERRIS, WILSON and BROWN 1981; BROWN, GEORGE and WILSON 1979; BROWN 1980; YONEKAWA *et al.* 1981) have all been successfully applied to the study of within- or between-population genetic differentiation. In particular, mitochondrial (mt) DNAs from a wide variety of species have been investigated by use of restriction fragment mapping.

[1] Present address: National Institute of Genetics, Mishima, Shizuoka-ken, 411 Japan.
[2] Present address: Department of Biochemistry and Zoology, University of California, Berkeley, California 94720.

Mitochondrial and chloroplast (extranuclear) genomes are self-replicating units genetically (but not physiologically) independent of the nuclear genome. They are maternally inherited in the majority of species studied (LANSMAN, AVISE and HUETTEL 1983 and references therein). There are multiple copies per cell (GILLHAM 1978; BEALE and KNOWLES 1978), and these are randomly transmitted to daughter cells during mitotic cell divisions (BIRKY 1978, 1983; THRAILKILL *et al.* 1980). Furthermore, mitochondrial genomes have a much higher rate of gene substitution possibly due to deficient repair mechanisms for mtDNA (BROWN, GEORGE and WILSON 1979). As a result, evolutionary patterns exhibited by nuclear and extranuclear genomes are not necessarily congruent and the extensive body of theory on nuclear genetic differentiation and evolution does not apply directly to extranuclear genomes.

The transmission genetics of extranuclear genomes have been modeled recently (OHTA 1980; TAKAHATA and MARUYAMA 1981; CHAPMAN *et al.* 1982; BIRKY, MARUYAMA and FUERS 1983; TAKAHATA 1983a,b; TAKAHATA and SLATKIN 1983, 1984). Others (NEI and LI 1979; NEI and TAJIMA 1981; ENGELS 1981; EWENS, SPIELMAN and HARRIS 1981) have examined the relationship between endonuclease cleavage patterns and within- and between-population genetic variation. However, there is currently little theoretical work allowing a bridge between empirical studies and theoretical models of extranuclear genomes in a species divided into subpopulations.

Here, we study extranuclear genetic variation in a subdivided population based on WRIGHT's (1943) finite island model. Particular attention is paid to the joint effects of both maternal inheritance and sex-dependent gene flow on the local differentiation of extranuclear genomes. First, we develop a model describing the behavior of a single locus on an extranuclear genome. We derive the identity probability (*i.e.*, the probability that two sampled homologous DNA segments will be identical) either in the same or in different demes. This probability is primarily determined by the effective migration rate, which is shown to vary both with paternal contribution to zygotic extranuclear genes and with the relative dispersal of females and males between demes.

Second, we modify existing techniques for calculating genetic diversity based on restriction fragment analysis to allow empirical estimation of the gene identity probabilities we derive in the first section. Third, we analyze published data on mtDNA variation within and between demes to illustrate use of the present results.

MODEL AND ANALYSIS

Population genetics of extranuclear genomes in an island model

We consider the finite island model of L partially isolated demes in which migration may be sex dependent. We denote by m and m^* the per-generation rates of migration of females and males, respectively, between demes. We assume that the number of breeding females (N_f) and of males (N_m) in each deme are the same and are constant between generations. We consider two possible alleles (M and M') at an extranuclear locus (the term "locus" refers

to either a gene or the recognition site of a restriction endonuclease). We assume only neutral mutations between the two alleles with the rate v_1 from M to M' and v_2 for the reverse. This model of mutation allows us to consider the case in which there are K equivalent alleles at a given locus by setting $v_2 = v_1/(K - 1)$ and letting M' represent all alleles other than M (KIMURA 1968). Migration of individuals, random sampling of gemetes and mutation are assumed to occur in this order. Generations are nonoverlapping and discrete.

We allow for paternal contribution to extranuclear genomes in a zygote and for heterogeneity of gametes produced by an individual but assume no heterogeneity within each gamete. Although there are usually multiple copies of the extranuclear genome within a cell, we presume that within-generation drift due to random transmission of extranuclear genomes at mitosis (THRAILKILL *et al.* 1980; BIRKY 1983) guarantees the fixation of either paternal or maternal extranuclear genomes in a cell (TAKAHATA and SLATKIN 1984). We note that this assumption does not necessarily imply homogeneity among different cells within an individual. In other words, we allow for a variety of gametes produced by a heteroplasmic zygote (BROWN 1983; OLIVO *et al.* 1983). Let α be the fraction of genome copies in a zygote at fertilization transmitted from a maternal cell, and $u(\alpha)$ be the fixation probability of such genomes among the cells in an adult individual. If extranuclear inheritance is completely maternal, or if the maternal extranuclear genome preferentially replicates, then $u(\alpha) = 1$. TAKAHATA (1983b) reports criteria for determining the significance of a particular level of paternal contribution to extranuclear genomes. With the assumptions of neutrality and complete homogenization of gametic cells, our model is similar to that of a haploid population. However, to define quantities specific to extranuclear genomes, and to incorporate the effect of recurrent mutation, we need to construct a rigorous mathematical model.

Let $x_t(i)$ and $y_t(i)$ be the frequencies of M in females and males (before migration) in the ith deme in the tth generation. We assume that $x_t(i) = y_t(i)$ at this stage of the life cycle because of random mating in each deme and because there is no differential assignment of different extranuclear genotypes to different sexes. However, if migration is sex dependent ($m \neq m^*$), the frequency $x_t'(i)$ after migration becomes different from $y_t'(i)$. In the Wright finite island model, these postmigration frequencies can be obtained by

$$\begin{aligned} x_t'(i) &= (1 - m)x_t(i) + \frac{m}{L - 1}\sum_{j\neq i}^{L} x_t(j) \\ y_t'(i) &= (1 - m^*)y_t(i) + \frac{m^*}{L - 1}\sum_{j\neq i}^{L} y_t(j) \end{aligned} \tag{1}$$

for $1 \leq i, j \leq L$. The first terms of the right side correspond to the loss of M from the ith deme and the second to the addition of M from the remaining $L - 1$ demes.

Next, we consider the change of the frequency of M due to random sampling of gametes occurring independently in each deme. This process is usually

described by a Markov chain which may be approximated by a diffusion process dependent on both population size and gene frequencies. We describe the dynamics of $x_t(i)$ and $y_t(i)$ by introducing a set of independent random variables, $B_1(i)$ and $B_2(i)$, which satisfy

$$\begin{aligned} E\{B_1(i)\} &= E\{B_2(i)\} = E\{B_1(i)B_2(i)\} = 0 \\ E\{B_1^2(i)\} &= N_f^{-1}x(1-x) \\ E\{B_2^2(i)\} &= N_m^{-1}y(1-y) \end{aligned} \tag{2}$$

where $E\{\ \}$ stands for taking the expectation with respect to the distribution of $B_1(i)$ and $B_2(i)$ conditioned on $x_t'(i) = x$ and $y_t'(i) = y$. $B_1(i)$ and $B_2(i)$ thus defined are related to the inbreeding effective number in each deme, as we will see later. By using $B_1(i)$ and $B_2(i)$, we have

$$\begin{aligned} x_t''(i) &= x_t'(i) + B_1(i) \\ y_t''(i) &= y_t'(i) + B_2(i) \end{aligned} \tag{3}$$

for $1 \leq i \leq L$, where the double primes denote the frequency of M after random sampling of gametes occurring in the gamete pool of each sex.

If maternal inheritance is incomplete, there can be two different homoplasmic and two different heteroplasmic zygotes in the population when gametes of both M and M' combine to form a zygote. If maternal inheritance is complete, however, no heteroplasmic zygotes are produced. In this case, we can ignore the multiplicity of genomes within each cell and model extranuclear genomes as those of a haploid. However, when there are recurrent mutations, heteroplasmic cells are formed even if maternal inheritance is complete. We need to develop a model that adequately incorporates this possibility. We assume that mutation occurs only at the single-cell zygote stage and that the resulting heteroplasmic zygote is subject to within-generation drift, fixing different genotypes in different cells. Table 1 summarizes the possible combinations of different gametes and the input frequencies of M in different genotypes due to the mode of inheritance and mutation. For example, a single-cell zygote fertilized by an egg of M and a sperm of M' is heteroplasmic with relative frequencies of M and M' being α and $\beta = 1 - \alpha$. These frequencies are further changed by mutation to α' and β'. We postulate that they are given by

$$\begin{aligned} \alpha' &= (1 - v_1)\alpha + v_2\beta \\ \beta' &= (1 - v_2)\beta + v_1\alpha \end{aligned} \tag{4}$$

With these frequencies, within-generation drift starts to act, resulting in homoplasmic cells in the adult in the proportion of $u(\alpha')$ to $u(\beta')$. In the case of neutral mutations, we expect that the output frequency of M at the adult stage equals the input frequency, *i.e.*, $u(\alpha') = \alpha'$. Likewise, we obtain the input and output frequencies of M in the other genotypes. Using Table 1, we have the expected frequency of M in females in the next generation

TABLE 1

Relative frequency of genotype M *in single-celled zygotes as a function of gamete genotypes, sex-dependent contribution to zygote extranuclear genes and postfertilization mutation*

	Female gametes	
Male gametes	M $[x]$	M' $[1 - x]$
M $[y]$	$1 - v_1$ $[xy]$	$(1 - v_1)\beta + v_2\alpha$ $[(1 - x)y]$
M' $[1 - y]$	$(1 - v_1)\alpha + v_2\beta$ $[x(1 - y)]$	v_2 $[(1 - x)(1 - y)]$

The relative frequency of occurrence of each gamete or combination of gametes is given in brackets (the symbol to indicate the frequency after random sampling is dropped here). v_1 is the mutation rate from M to M'; v_2 is the reverse rate. The fraction of genome copies maternally transmitted is α and $\beta = 1 - \alpha$. See text for details.

$$\begin{aligned} x_{t+1}(i) &= x_t''(i)y_t''(i)u(1 - v_1) + \{1 - x_t''(i)\}y_t''(i)u(\gamma') \\ &\quad + x_t''(i)\{1 - y_t''(i)\}u(\alpha') + \{1 - x_t''(i)\}\{1 - y_t''(i)\}u(v_2) \\ &= v_2 + \mu\{\alpha x_t''(i) + \beta y_t''(i)\} \end{aligned} \tag{5}$$

for $1 \le i \le L$, where $\gamma' = (1 - v_1)\beta + v_2\alpha$ and $\mu = 1 - v_1 - v_2$. Note that $y_{t+1}(i) = x_{t+1}(i)$ is recovered.

We define the average frequency of $M(X_t)$ and identity probabilities within (I_t) and between (J_t) demes as

$$\begin{aligned} X_t &= \frac{1}{L}\sum_{i=1}^{L} E\{x_t(i)\} \\ I_t &= \frac{1}{L}\sum_{i=1}^{L} E\{x_t^2(i)\} \\ J_t &= \frac{1}{L(L-1)}\sum_{i=1}^{L}\sum_{j\neq i}^{L} E\{x_t(i)x_t(j)\}. \end{aligned} \tag{6}$$

If we assume that all demes are initially homogeneous in their genetic composition or that equilibrium is reached, X_t, I_t and J_t are equal to $E\{x_t(i)\}$, $E\{x_t^2(i)\}$ and $E\{x_t(i)x_t(j)\}$. We will use the same notations to represent these quantities. Recalling that $x_t(i) = y_t(i)$, we first get

$$\begin{aligned} X_t'' &= X_t' = X_t \\ X_{t+1} &= v_2 + \mu X_t \end{aligned} \tag{7}$$

from (5). To calculate the change of I_t and J_t due to migration we first define the identity probabilities of two genes sampled from a male and a female within a deme (Z_t) and between two demes (W_t);

$$Z_t = \frac{1}{L}\sum_{i=1}^{L} E\{x_t(i)y_t(i)\}$$
$$W_t = \frac{1}{L(L-1)}\sum_{i=1}^{L}\sum_{j\neq i}^{L} E\{x_t(i)y_t(j)\}.$$

Making use of (1), we can determine these two probabilities after migration:

$$\begin{aligned} Z_t' &= \left\{1 - m - m^* + \frac{Lmm^*}{L-1}\right\} Z_t + \left\{m + m^* - \frac{Lmm^*}{L-1}\right\} W_t \\ W_t' &= \frac{1}{L-1}\left\{m + m^* - \frac{Lmm^*}{L-1}\right\} Z_t + \left\{1 - \frac{m+m^*}{L-1} + \frac{Lmm^*}{(L-1)^2}\right\} W_t \end{aligned} \tag{8}$$

The equations of I_t and J_t in females can be obtained by replacing m^* by m, whereas the corresponding equations in males, denoted by I_t^* and J_t^*, can be obtained by replacing m by m^*. Since $x_t(i) = y_t(i)$, $Z_t = I_t = I_t^*$ and $W_t = J_t = J_t^*$ hold immediately before migration.

Next, we represent I_{t+1} and J_{t+1} in terms of the identity probabilities after migration. Making use of (5) and then (3), we have

$$\begin{aligned} I_{t+1} &= v_2^2 + \mu^2\left[\alpha^2\left\{I_t' + \frac{1}{N_f}(X_t' - I_t')\right\} + \beta^2\left\{I_t^{*\prime} + \frac{1}{N_m}(X_t' - I_t^{*\prime})\right\}\right. \\ &\qquad \left. + 2\alpha\beta Z_t'\right] + 2v_2\mu X_t' \\ J_{t+1} &= v_2^2 + \mu^2[\alpha^2 J_t' + \beta^2 J_t^{*\prime} + 2\alpha\beta W_t'] + 2v_2\mu X_t'. \end{aligned} \tag{9}$$

Equation (10) suggests defining the effective population number N_e by

$$N_e = \left[\frac{\alpha^2}{N_f} + \frac{\beta^2}{N_m}\right]^{-1} \tag{10}$$

(TAKAHATA and MARUYAMA 1981; *cf.* CROW and KIMURA 1970). Substitution of (8) and the similar equations for I_t', J_t', $I_t^{*\prime}$ and $J_t^{*\prime}$ into (9) yields the change of identity probabilities in a generation. If we assume that m, m^*, v_1, v_2 and N_e^{-1} are all sufficiently small so that the higher order terms of these parameters are negligible, then the recursion equations for I_t and J_t are approximately given by

$$\begin{aligned} I_{t+1} &= \{1 - 2v_1 - 2v_2 - 2m_e - N_e^{-1}\}I_t + 2m_e J_t + \{2v_2 + N_e^{-1}\}X_t \\ J_{t+1} &= \{1 - 2v_1 - 2v_2 - 2m_e(L-1)^{-1}\}J_t + 2m_e(L-1)^{-1}I_t + 2v_2 X_t \end{aligned} \tag{11}$$

where the effective migration rate m_e is defined by

$$m_e = \alpha m + \beta m^*. \tag{12}$$

The effective migration rate is a linearly increasing (decreasing) function of α if $m > m^*(m < m^*)$; it is a constant function of α if $m = m^*$.

These results are formally equivalent to those that should be expected for

nuclear genes in a haploid population if we set $\alpha = \beta = \frac{1}{2}$. Under this condition, $m_e = \frac{1}{2}(m + m^*)$ and $N_e = 4N_fN_m[N_f + N_m]^{-1}$.

Although standard linear algebra yields time-dependent solutions of (7) and (11), we will be concerned only with the equilibrium solutions denoted by the same letters but dropping the subscript for generations. We use a K allele model (KIMURA 1968) and rewrite v_1 simply by v. The equilibrium solutions are readily obtained as

$$I = \frac{1}{K}\frac{Q + (K + KV + LQ)V}{Q + (1 + KV + LQ)KV}$$
$$J = \frac{1}{K}\frac{Q + (1 + KV + LQ)V}{Q + (1 + KV + LQ)KV} \tag{13}$$

since $X = K^{-1}$. In the (13), $Q = 2N_em_e(L - 1)^{-1}$ and $V = 2N_ev(K - 1)^{-1}$. I and J are the identity probabilities for a particular allele. To calculate these probabilities (denoted by $\bar{I}$ and $\bar{J}$) for any allele we need to multiply (13) by K. When K is sufficiently large, $\bar{I}$ and $\bar{J}$ become

$$\bar{I} = \frac{m_e + (L - 1)v}{m_e + \{(L - 1)(1 + 2N_ev) + 2N_em_eL\}v}$$
$$\bar{J} = \frac{m_e}{m_e + \{(L - 1)(1 + 2N_ev) + 2N_em_eL\}v}. \tag{14}$$

Our model explores the consequences of coupled effects of sex-dependent migration and unequal cytoplasmic contribution of both sexes. Although the implications of this are intuitively clear (BIRKY, MARUYAMA and FUERST 1983), the relative importance of sex-dependent migration and the transmission mode is not obvious *a priori*. In addition, the constant generation of heteroplasmic zygotes due to recurrent mutations even in the case of completely maternal inheritance prevents the exact correspondence between the models of an extranuclear genome and a nuclear genome in a haploid organism. Our model gives criteria under which the dynamics of extranuclear genomes in a sexual population are equivalent to those of a haploid population.

Estimation of identity probabilities from restriction maps

NEI and LI (1979), NEI and TAJIMA (1981), EWENS, SPIELMAN and HARRIS (1981) and ENGELS (1981) propose techniques to estimate genetic diversity in a single panmictic population or between two completely isolated populations when data are obtained by the use of restriction endonucleases. Here, we extend these results to estimate identity probabilities and give the relationship between this statistic and the population parameters derived in the previous section. Throughout, we follow the notation and the terminology of ENGELS (1981).

Suppose there are n sequences randomly sampled from a population. Let l be the number of different cleavage sites in the n sequences. We denote by C_i the number of cleavage sites at position i among the n sequences ($1 \leq C_i \leq n$;

$i = 1, 2, \cdots, l$). We denote by Π_i the true frequency of sites at position i in the population. Using restriction map data, we have (ENGELS 1981, p. 6330)

$$\Pi_i^2 = E\left[\frac{C_i(C_i - 1)}{n(n-1)}\right]. \tag{15}$$

The average squared frequency of Π_i taken over all l cleavage sites in the n samples, $\frac{1}{l}\sum_{i=1}^{l}\Pi_i^2$, is the average identity probability for particular "allele" at each cleavage site, given that this site is a recognition sequence. In our model, the probability that r consecutive nucleotides are a recognition sequence is assumed to be $(1/4)^r$, where r is the length of a recognition sequence of a given restriction endonuclease. We will equate 4^r with K. We note that Π_i is conditioned on the presence of at least one cut.

Recalling that I is the identity probability for a particular type of allele, and imposing the condition that the allele must be a recognition sequence, we obtain

$$\bar{I} = I(1/K)^{-1} = \frac{1}{l}\sum_{i=1}^{l}\Pi_i^2. \tag{16}$$

That is, we see that the identity probability, $\bar{I}$, equals that for a particular type of allele conditioned on the probability that that allele is found. From (15) and (16), and replacing expectations by their observed values, we have the estimator

$$\hat{\bar{I}} = \frac{1}{ln(n-1)}\sum_{i=1}^{l} C_i(C_i - 1). \tag{17}$$

When $n = 2$, $\hat{\bar{I}} = (l - k)/l$, where k is the number of polymorphic or unshared sites.

We can extend (17) to consider sampling from two demes. If we sample n and n' sequences from the two demes, respectively, and if l is now the number of sites in the pooled sample, and C_i and C_i' are the numbers of members cut at position i in the n and n' samples, respectively, then the average conditional identity probability per site between two demes is

$$\frac{1}{l}\sum_{i=1}^{l}\Pi_i\Pi_i'. \tag{18}$$

Here, Π_i and Π_i' are the true frequencies of sites at position i in each deme. Since sampling of the n and n' sequences is independent, $\Pi_i\Pi_i' = E\left[\frac{C_iC_i'}{nn'}\right]$.

The average conditional identity probability is then estimated as

$$\hat{\bar{J}} = \frac{1}{lnn'}\sum_{i=1}^{l} C_iC_i' \tag{19}$$

When $n = n' = 1$, or when demes are comprised of individuals with only one genotype, $\hat{\bar{J}} = (l - k)/l$. We use (17) and (19) to estimate $\bar{I}$ and $\bar{J}$, respectively.

As noted by NEI and LI (1979), it is often difficult to obtain unambiguous restriction site maps on the basis of endonuclease cleavage patterns. Lost fragments (ADVISE *et al.* 1979) or insertion and deletion mutations (CANN and WILSON 1983) may prohibit assigning all sites to the sequences that contain them. It is more straightforward to determine the number of DNA fragments shared by digests of two sequences. The total number of restriction sites (l) and the number of unshared sites (k) may be estimated from the total number of fragments in two samples and the number of fragments shared (ENGELS 1981). For pairwise comparisons $\bar{I}$ and $\bar{J}$ can be calculated by $(l - k)/l$, allowing fragment patterns to be used to determine identity probabilities between pairs of individuals or between monomorphic populations. We are currently unable to extend this approach to more than two polymorphic sampled sequences.

Data analysis

FERRIS *et al.* (1983a) present complete restriction maps and complete gel fragment patterns for European mice (*Mus domesticus* and *M. musculus*) from various localities. We apply the analyses to mouse demes pairwise and examine identity probabilities between demes calculated both from restriction site occurrences and from gel fragment patterns. Theoretically, equilibrium $\bar{I}$ and $\bar{J}$ are independent of the demes chosen. However, the values estimated from restriction analyses may vary with demes. A number of causes for this are conceivable; the symmetries of migration pattern and population size in the island model and the equilibrium assumption may not be warranted. In addition, sampling variance would augment heterogeneity. We will discuss the variances of identity probabilities due to random sampling drift.

Estimates of pairwise $\bar{J}$ for Old World mice are given in Table 2. Calculations based on restriction maps were identical with those based on fragment sizes, suggesting that fragment patterns are adequate to estimate $\bar{J}$. Only one mouse was analyzed per site (except for site 3 which was monomorphic for restriction cleavage patterns), so that $\bar{I}$ could not be estimated. The degree of interdeme differentiation measured by $\bar{J}$ is consistent with results presented by FERRIS *et al.* (1983a) based on percentage differences in nucleotide sequences. Strong negative correlation between $\bar{J}$ and percentage difference indicates that $\bar{J}$ is also an appropriate measure for estimating interdeme differentiation. The reason we use identity probabilities is that we could easily relate them to population parameters, particularly to the levels of gene flow.

Geographic patterns in $\bar{J}$ (Table 2) suggest either that there has been substantial gene flow between mouse demes or that these populations have diverged recently. Populations of mice separated by more than 3000 km (deme 11 *vs.* 13) exhibit values of $\bar{J}$ as high as 0.89. Theoretically, $\bar{J}$ declines with declining migration rate (Figure 1) but can remain high even if m_e is on the order of the mutation rate (v). Given the prevalence of trans-Mediterranean commercial traffic (BRAUDEL 1982) which may act as a vector for gene flow, it is not surprising to find high identity probabilities in distant locales. Table 2 suggests, however, that Northern demes (1, 2; 3, 4; 10; 11) exchange genes at a higher rate than do those further apart. This trend is particularly re-

TABLE 2

Estimates of between-deme identity probabilities $\bar{J}$'s for Old World mice

		Deme				
		3, 4	10	11	12	13
M. musculus	1, 2	0.984	0.969	0.861	0.811	0.836
		(0.2)	(0.3)	(1.9)	(3.0)	(2.3)
	3, 4		0.985	0.875	0.825	0.849
			(0.2)	(1.7)	(2.8)	(2.1)
M. domesticus	10			0.863	0.824	0.838
				(2.0)	(3.0)	(2.3)
	11				0.893	0.893
					(1.3)	(1.3)
	12					0.893
						(1.5)

The deme numbers refer to localities where mice were collected by FERRIS *et al.* (1983a). The demes in eastern Europe belonging to authentic *M. musculus* are excluded from the analysis. The demes used here are northern Denmark (1, 2, 3), Sweden (4), Federal Republic of Germany (10), southern Denmark (11), Switzerland (12) and Egypt (13). The numbers in parentheses are estimates of the percentage difference in nucleotide sequence, taken from Table 2 in FERRIS *et al.* (1983a). $\bar{J}$ and P (percentage difference) are strongly inversely correlated (the regression curve: $\bar{J} = 1 - 0.089P + 0.009P^2$).

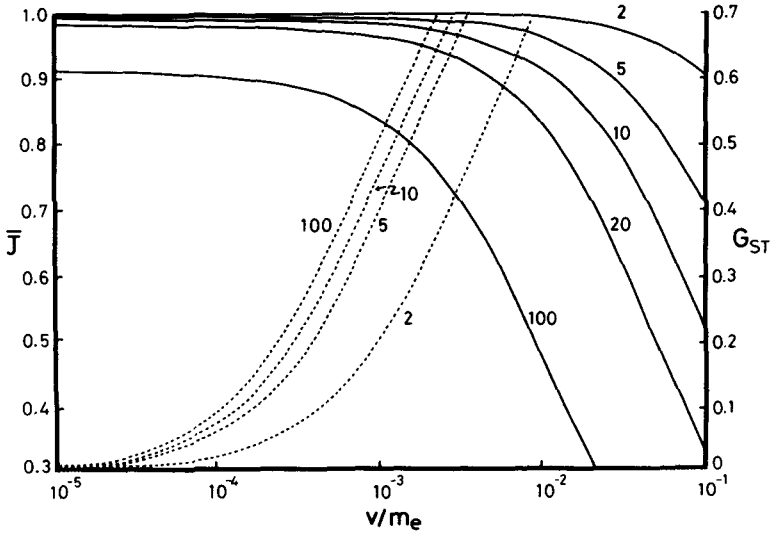

FIGURE 1.—Between-deme identity probability ($\bar{J}$; solid lines) and G_{ST} statistic (dotted lines) as a function of the ratio of mutation rate (v) to effective migration rate (m_e) for populations divided into different numbers of demes (L). The appropriate scales for $\bar{J}$ and G_{ST} are drawn on the left and right margins of the figure, respectively. Here, back mutations is ignored (*i.e.*, the infinite allele model) and $2N_ev$ is kept constant at 0.001. As m_e increases relative to v, genetic similarity between populations increases ($\bar{J}$ increases, G_{ST} decreases). As deme number increases, similarity decreases, especially at low levels of gene flow. The statistic G_{ST} is much less sensitive to L than is $\bar{J}$ as long as N_e and not total population size (N_eL) is constant.

TABLE 3

Within-($\bar{I}$) and interdeme ($\bar{J}$) identity probabilities for ten demes of M. domesticus

	California	Indiana	Maryland	Peru	Israel	Egypt	Morocco	Yugoslavia	Switzerland	Italy
California	0.951	0.950	0.949	0.917	0.906	0.917	0.898	0.915	0.906	0.900
Indiana		1.0	0.946	0.942	0.931	0.961	0.924	0.939	0.947	0.930
Maryland			1.0	0.946	0.936	0.937	0.915	0.937	0.914	0.924
Peru				1.0	0.960	0.925	0.926	0.940	0.910	0.944
Israel					1.0	0.915	0.933	0.937	0.900	0.946
Egypt						0.965	0.901	0.916	0.921	0.916
Morocco							0.910	0.919	0.893	0.912
Yugoslavia								0.953	0.908	0.923
Switzerland									0.956	0.898
Italy										0.953

Restriction site data are from FERRIS *et al.* (1983b). Values on the main diagonal are $\bar{I}$'s calculated with equation (17); values above the main diagonal are $\bar{J}$'s calculated with equation (19) for the pairs of demes listed.

markable since it seems to hold whether comparisons are made intraspecifically (demes 1, 2 *vs.* 3, 4) or interspecifically (deme 10 *vs.* either 1, 2 or 3, 4). Indeed, interspecific mtDNA transmission in Danish mice (FERRIS *et al.* 1983a) signals that mitochondrial genes may be transmitted more extensively than nuclear genes depending on the selective regimes acting on each genome (TAKAHATA and SLATKIN 1984; A. C. WILSON, personal communication). Gene flow appears to play an important role in determining the geographic pattern of mouse mitochondrial genomes. If there has been continuous gene flow, neglect of gene flow would underestimate the rate of mitochondrial divergence in *M. domesticus*.

To illustrate the relationship between $\bar{I}$ and $\bar{J}$, we used data on *M. domesticus* presented by FERRIS *et al.* (1983b). The results reported in their study include the occurrence of all restriction sites from digests with 11 different restriction endonucleases of 29 *M. domesticus* collected worldwide. For illustration, we have arbitrarily grouped mice collected in the same countries (or in the same states from United States mice) and used these to calculate $\bar{I}$'s. We then compare demes pairwise to calculate $\bar{J}$ using equation (19).

$\bar{I}$'s vary from 0.91 to 1.0 (Table 3); $\bar{J}$'s vary from 0.89 (Morocco-Switzerland) to 0.95 (California-Indiana). FERRIS *et al.* (1983b) provide a detailed discussion of how the population biology of mice may affect mtDNA patterns and conclude that mouse demes are remarkably homogeneous in mtDNA even though variation in mtDNA between individuals is high. Ferris *et al.* conclude that this is because mice have long had "commensal ties to the highly mobile human species" (p. 701). The high degree of interdeme similarity between geographic extremes and the lack of a clear relationship between interdeme distance and identity probability in this species (FERRIS *et al.* 1983b) suggests that "isolation by distance" may not be a reasonable explanation for this species. Of course, this may not hold for all species; indeed, other rodents show a correlation between genetic differentiation and geographic distance (AVISE *et al.* 1979).

We may examine Table 3 in a slightly different way by using Wright's F_{ST} or NEI'S (1973) G_{ST} statistic for multiple alleles. G_{ST} represents the fraction of genetic variation within an entire population that is due to interdeme genetic differences and is defined by

$$G_{ST} = \frac{H_T - H_O}{H_T}$$

where $H_O = 1 - \bar{I}$ and $H_T = 1 - \left\{\frac{1}{L}\bar{I} + \left(1 - \frac{1}{L}\right)\bar{J}\right\}$. The equilibrium value of G_{ST} is

$$G_{ST} = \left[1 + 2N_e\left(\frac{L}{L-1}\right)^2\left\{m_e + \left(1 - \frac{1}{L}\right)v\right\}\right]^{-1}$$

Estimates of $\bar{I}$ and $\bar{J}$ are 0.97 and 0.93 and, therefore, $\hat{H}_T \simeq 0.07$ for moderate and large L and $\hat{G}_{ST} \simeq 0.56$. This indicates that more than 50% of the mitochondrial variation per site in the entire *M. domesticus* population can be attributed to interdeme variation. This may in turn suggest that the species is spatially heterogeneous and "microgeographically structured" (Ferris *et al.* 1983b, see DISCUSSION).

DISCUSSION

Restriction fragment patterns are relatively straightforward to obtain from even a large number of individuals (LANSMAN *et al.* 1983). By comparison, restriction maps providing unambiguous identification of endonuclease cleavage sites are much more difficult to construct (CANN, BROWN and WILSON 1982). Our analysis of identity probabilities using both the restriction maps of FERRIS *et al.* (1983a) and the fragment patterns they supply yields identical results. This suggests that fragment patterns, treated as in ENGELS (1981) to estimate the percentage of polymorphic restriction sites, are adequate for pairwise analyses of mtDNA genotypes. However, as mentioned earlier, fragment patterns alone may not allow correct estimation of identity probabilities when multiple samples are taken. In addition, care should be taken to detect missing or altered fragments and insertion or deletion mutations (CANN and WILSON 1983). For instance, several fragments are missing from the patterns presented in AVISE, LANSMAN and SHADE (1979) (*e.g.*, compare *Bam*I patterns *N* and *O*) and from the patterns in FERRIS *et al.* (1983a) (see the cautionary footnote to their Table 1). Adding the minimum number of extra fragments to compensate for those missing will tend to underestimate the identity probabilities.

Equation (12) indicates that m_e varies primarily with the relative contribution of females and males to zygotic extranuclear genomes and with the relative rates of migration of males and females. Even if migration of males is high, m_e is low if the migration of females is low and α is near 1. Yet, in this case, the effective migration rate for nuclear genes may be high since males and females contribute equally to zygotic nuclear content. Thus, for species in

which males migrate more often than do females, genetic differentiation for mitochondrial genomes might be higher than that for nuclear genomes even if mutation or substitution rates are the same. In this context, a comparison of similar species with and without sex-dependent dispersal would be very informative.

Our method is critically based on the assumption of no within-cell heterogeneity in an adult population. TAKAHATA and MARUYAMA (1981) analyzed within- and between-deme identity probabilities in one- and two-dimensional stepping stone models taking into account the possibility of within-cell heterogeneity but assuming sex-independent migration. Within-cell heterogeneity decreases interdeme genetic divergence by the ratio of within-cell heterogeneity to within-deme heterogeneity. With sex-dependent migration the effective migration rate will increase (and genetic divergence will decrease) if there is substantial within-cell heterogeneity since mitochondrial genomes transmitted during any one migration event may be more diverse than if all gametes carry only a single extranuclear genotype. Thus, the present method will overestimate m_e if there is substantial within-cell heterogeneity. At present, within-cell heterogeneity of extranuclear genomes appears of minor importance in higher organisms (AVISE *et al.* 1979; BROWN 1983; BIRKY 1983; LANSMAN, AVISE and HUETTEL 1983; but see Olivo *et al.* 1983).

In our analyses, we have first treated $\bar{I}$ and $\bar{J}$ as independent indices of identity within and between demes. Then, we considered G_{ST}, averaging $\bar{I}$ and $\bar{J}$ over all demes. The ratio of $\bar{J}$ to $\bar{I}$ (or to the geometric mean of $\bar{I}$ values for two demes) may be also used to calculate the genetic distance between extranuclear elements in two populations. NEI (1972) and NEI and FELDMAN (1972) suggested a measure of nuclear genetic distance, D, which might be used as a standard metric for comparing populations. In our terminology, D between two demes is given by

$$D = -\log_e(\bar{J}/\bar{I})$$

and at equilibrium it becomes

$$D = \log_e[1 + (L - 1)v/m_e]. \tag{20}$$

Thus, equilibrium $D \simeq (L - 1)v/m_e$ when v/m_e is small, and it is primarily determined by the ratio of mutation rate to effective migration rate. However, we note that D is fairly sensitive to L. D's estimated pairwise between the ten demes in Table 3 range from 0.013 to 0.087. Lack of information about the deme number in natural populations allows only estimates of the relative levels of gene flow between different pairs of demes. Ratios of D for two pairs of demes vary directly with the ratios of m_e between these pairs, assuming mutation rates do not vary among demes. Effective migration rates vary about seven-fold between pairs of demes (Yugoslavia-Morocco compared to Switzerland-Israel). It should be noted, however, that m_e thus estimated may be inaccurate since D reaches equilibrium only slowly and (20) only holds at equilibrium. In contrast, G_{ST} reaches steady state relatively quickly (J. F. CROW, personal communication). In addition, G_{ST} is rather insensitive to changes in L

(Figure 1). Since $G_{ST} = [1 + 2N_e m_e]^{-1}$ for $m_e \gg v$, and moderate and large L, substitution of $\hat{G}_{ST} = 0.56$ for *M. domesticus* gives $N_e m_e = 0.39$.

So far, we have studied only the mean identity probabilities defined in (6) and their related quantities. For a full discussion of empirically observed genetic heterogeneity, information about the variances is necessary. Although we did not fully study this problem, we can draw the following conclusion about the relationship between the mean and variance of identity probability. Previously, TAKAHATA (1983b) studied this relationship at a single extranuclear locus in a panmictic population. It was clear that the deviation of the variance from STEWART'S (1976) relationship for a nuclear locus is very small. A stochastic force (within-generation drift) characteristic of extranuclear genetic systems does not substantially change the relationship. On the other hand, the population structure (*e.g.*, gene flow and number of demes) may either increase or decrease the variances (Table 2, TAKAHATA 1983c). The variance is large when L is small but tends to be smaller as L increases. From these studies, we anticipate that the variances of identity probabilities at an extranuclear locus take more or less the same values expected from Stewart's relationship when the mean values are given (see BIRKY, MARUYAMA and FUERST 1983). In *M. domesticus*, the expected variance of interdeme identity probability would be approximately 0.015 for the mean 0.93. Thus, the standard deviation becomes 0.12.

Although we are aware that the population parameters we estimate for *M. domesticus* are perhaps greatly dependent on our underlying assumptions, the two main findings by FERRIS *et al.* (1983a,b) are consistent with the predictions from the present study. First, we cannot detect a strong correlation of genetic differentiation with geographic distance between the pairs of demes compared. This "macrogeographic homogeneity" strongly suggests that geographic distance has not been an important barrier in determining the migration pattern of commensal mice. On the other hand, there indeed exists a substantial "microgeographic heterogeneity" despite the commensal mobility of *M. domesticus*. The proportion of interdeme variation to the total may be as high as 50%. In light of the present study, this microgeographic heterogeneity can be attributed either to limited mitochondrial gene flow or to large stochastic errors of statistics we used.

The most powerful use of our results will be in application to mitochondrial data from insular faunas or species inhabiting other discrete habitats. For these species, simultaneous estimates of both within- and between-deme identity probabilities from restriction data will allow estimation of relative interdeme gene flow and the extent to which local populations are isolated demographic (and evolutionary) units.

We thank MONTGOMERY SLATKIN and MASATOSHI NEI for their comments on the earlier version of this paper and anonymous reviewers for their numerous constructive criticisms. This paper is contribution 1577 from the National Institute of Genetics, Mishima, Shizuoka-ken 411, Japan. This work is supported in part by a Grant-in-Aid from the Ministry of Education, Science and Culture of Japan.

LITERATURE CITED

AVISE, J. C., C. GIBLIN-DAVIDSON, J. LAERM, J. C. PATTON and R. A. LANSMAN, 1979 Mitochondrial DNA clones and matriachal phylogeny within and among geographic populations of the pocket gopher, *Geomys pinetis*. Proc. Natl. Acad. Sci. USA **76,** 6694–6698.

AVISE, J. C., R. A. LANSMAN and R. O. SHADE, 1979 The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. I. Population structure and evolution in the genus *Peromyscus*. Genetics **92:** 279–295.

BEALE, G. H. and J. K. C. KNOWLES, 1978 *Extranuclear Genetics*. Edward Arnold, London.

BIRKY, C. W., JR., 1978 Transmission genetics of mitochondria and chloroplasts. Annu. Rev. Genet. **12:** 471–512.

BIRKY, C. W., JR., 1983 Relaxed cellular controls and organelle heredity. Science **222:** 468–475.

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

BRAUDEL, F. 1982 *Civilization and Capitalism 15th–18th Century*, Vol. 2: *The Wheels of Commerce*, English translation by Sian Reynolds. Harper and Row, New York.

BRITTEN, R. J., A. CETTA, and E. H. DAVIDSON, 1978 The single-copy DNA sequence polymorphism of the sea urchin *Strongylocentrotus purpuratus*. Cell **15:** 1175–1186.

BROWN, W. M., 1980 Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. Proc. Natl. Acad. Sci. USA **77:** 3605–3609.

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

BROWN, W. M., M. GEORGE and A. C. WILSON, 1979 Rapid evolution of animal mitochondrial DNA. Proc. Natl. Acad. Sci. USA **76:** 1967–1971.

BROWN, W. M., E. M. PRAGER, A. WANG and A. C. WILSON, 1982 Mitochondrial DNA sequences of primates: tempo and mode of evolution. J. Mol. Evol. **18:** 225–239.

CANN, R. L., W. M. BROWN and A. C. WILSON, 1982 Evolution of human mitochondrial DNA: a preliminary report. pp. 157–164. In: *Human Genetics, Part A: The unfolding genome*, Edited by B. BONNE-TAMIR, P. COHEN and R. N. GOODMAN. Alan R. Liss, New York.

CANN, R. L. and A. C. WILSON, 1983 Length mutations in human mitochondrial DNA. Genetics **104:** 699–711.

CHAPMAN, R. W., J. C. STEPHENS, R. A. LANSMAN and J. C. AVISE, 1982 Models of mitochondrial DNA transmission genetics and evolution in higher eukaryotes. Genet. Res. **40:** 41–57.

CROW, J. F. and M. KIMURA, 1970 *An Introduction to Population Genetics Theory*. Harper and Row, New York.

ENGELS, W. R., 1981 Estimating genetic divergence and genetic variability with restriction endonucleases. Proc. Natl. Acad. Sci. USA **78:** 6329–6333.

EWENS, W. J., R. S. SPIELMAN AND H. HARRIS, 1981 Estimation of genetic variation at the DNA level from restriction endonuclease data. Proc. Natl. Acad Sci. USA **78:** 3748–3750.

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

FERRIS, S. D., R. D. SAGE, E. M. PRAGER, U. RITTE and A. C. WILSON, 1983b Mitochondrial DNA evolution in mice. Genetics **105:** 681–721.

FERRIS, S. D., A. C. WILSON and W. M. BROWN, 1981 Evolutionary tree for apes and humans based on cleavage maps of mitochondrial DNA. Proc. Natl. Acad. Sci. USA **78:** 2432–2436.

GILLHAM, N. W., 1978 *Organelle Heredity*. Raven Press, New York.

KIMURA, M., 1968 Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. Genet. Res. **11:** 247–269.

LANSMAN, R. A., J. C. AVISE, C. F. AQUADRO, J. F. SHAPIRA and S. W. DANIEL, 1983 Extensive genetic variation in mitochondrial DNA's among geographic populations of the deer mouse, *Peromyscus maniculatus.* Evolution **37:** 1–16.

LANSMAN, R. A., J. C. AVISE and M. D. HUETTEL, 1983 Critical experimental test of the possibility of "paternal leakage" of mitochondrial DNA. Proc. Natl. Acad. Sci. USA **80:** 1969–1971.

MARTIN, S. L., K. A. VINCENT and A. C. WILSON, 1983 Rise and fall of the delta globin gene. J. Mol. Biol. **164:** 513–528.

NEI, M., 1972 Genetic distance between populations. Am. Nat. **106:** 283–292.

NEI, M., 1973 Analysis of gene diversity in sub-divided populations. Proc. Natl. Acad. Sci. USA **20:** 3321–3323.

NEI, M. and M. W. FELDMAN, 1972 Identity of genes by descent within and between populations under mutation and migration pressures. Theor. Pop. Biol. **3:** 460–463.

NEI, M. and W.-H. LI, 1979 Mathematical models for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA **76:** 5269–5273.

NEI, M. and F. TAJIMA, 1981 DNA polymorphisms detectable by restriction endonucleases. Genetics **97:** 145–163.

OHTA, T., 1980 Two-locus problems in transmission genetics of mitochondria and chloroplasts. Genetics **96:** 543–555.

OLIVO, P. D., M. J. VAN DE WALLE, P. J. LAIPIS and W. W. HAUSWIRTH, 1983 Nucleotide sequence evidence for rapid genotypic shifts in the bovine mitochondrial DNA D-loop. Nature **306:** 400–402.

STEWART, F. M., 1976 Variability in the amount of heterozygosity maintained by neutral mutations. Theor. Pop. Biol. **9:** 188–201.

TAKAHATA, N., 1983a Linkage disequilibrium of extranuclear genes under neutral mutations and random genetic drift. Theor. Pop. Biol. **24:** 1–21.

TAKAHATA, N., 1983b Population genetics of extranuclear genomes under the neutral mutation hypothesis. Genet. Res. **42:** 235–255.

TAKAHATA, N., 1983c Gene identity and genetic differentiation of populations in the finite island model. Genetics **104:** 497–512.

TAKAHATA, N. and T. MARUYAMA, 1981 A mathematical model of extranuclear genes and the genetic variability maintained in a finite population. Genet. Res. **37:** 291–302.

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

TAKAHATA, N. and M. SLATKIN, 1984 Mitochondrial gene flow. Proc. Natl. Acad. Sci. USA **81:** 1764–1767.

THRAILKILL, K. M., C. W. BIRKY, G. LUCKEMANN and K. WOLF, 1980 Intracellular population genetics: evidence for random drift of mitochondrial allele frequencies in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe.* Genetics **96:** 237–262.

WRIGHT, S., 1943 Isolation by distance. Genetics **28:** 114–138.

WRIGHT, S., 1970 Random drift and the shifting balance theory of evolution. pp. 1–31. In: *Mathematical Topics in Population Genetics,* Edited by K. KOJIMA. Springer-Verlag, Berlin.

WRIGHT, S., 1977 *Evolution and the Genetics of Populations,* Vol. 3: *Experimental Results and Evolutionary Deductions.* University of Chicago Press, Chicago.

WRIGHT, S., 1978 *Evolution and the Genetics of Populations,* Vol. 4: *Variability within and among Natural Populations.* University of Chicago Press, Chicago.

YONEKAWA, H., K. MORIWAKI, O. GOTOH, J. HAYASHI, J. WATANABE, N. MIYASHITA, M. L. PETRAS and Y. TAGASHIRA, 1981 Evolutionary relationships among five sub-species of *Mus musculus* based on restriction enzyme cleavage patterns of mitochondrial DNA. Genetics **98:** 801–816.

Communicating editor: B. S. WEIR