Organelle Gene Diversity Under Migration, Mutation, and Drift: Equilibrium Expectations, Approach to Equilibrium, Effects of Heteroplasmic Cells, and Comparison to Nuclear Genes

C. William Birky, Jr.,* Paul Fuerst* and Takeo Maruyama†

*Department of Molecular Genetics, The Ohio State University, Columbus, Ohio 43210, and †National Institute of Genetics, Mishima 411, Japan

Manuscript received February 16, 1988
Accepted for publication December 3, 1988

ABSTRACT

We developed stochastic population genetic theory for mitochondrial and chloroplast genes, using an infinite alleles model appropriate for molecular genetic data. We considered the effects of mutation, random drift, and migration in a finite island model on selectively neutral alleles. Recurrence equations were obtained for the expectation of gene diversities within zygotes, within colonies, and between colonies. The variables are number and sizes of colonies, migration rates, sex ratios, degree of paternal transmission, number of germ line cell divisions, effective number of segregating organelle genomes, and mutation rate. Computer solutions of the recurrence equations were used to study the approach to equilibrium. Gene diversities equilibrate slowly, while $G_{ST}$, used to measure population subdivision, equilibrates rapidly. Approximate equilibrium equations for gene diversities and $G_{ST}$ can be obtained by substituting $N_a$ and $m_n$, simple functions of the numbers of breeding or migrating males and females and of the degree of paternal transmission, for the effective numbers of genes and migration rates in the corresponding equations for nuclear genes. The approximate equations are not valid when the diversity within individuals is large compared to that between individuals, as is often true for the D-loop of animal mtDNA. We used the exact equations to verify that organelle genes often show more subdivision than nuclear genes; however, we also identified the range of breeding and migrating sex ratios for which population subdivision is greater for nuclear genes. Finally, we show that gene diversities are higher for nuclei than for organelles over a larger range of sex ratios in a subdivided population than in a panmictic population.

The last few years have been marked by a strong and growing interest in the evolution and population genetics of genes in mitochondria and chloroplasts. These genes are intrinsically important to the organism and its evolution because they play essential roles in oxidative respiration and photosynthesis. They are potentially useful for crop improvement, and they must be considered along with nuclear genes in the design of conservation programs. Thus it is important to determine the amount and kinds of intraspecific organelle genetic diversity present in wild and domesticated plants and animals. It is also essential to develop a theoretical basis for understanding the diversity we observe. Organelle genes are inherited differently from nuclear genes, so that different models must be used, and different theory must be developed, for their population genetics. In this paper we investigate the diversity of organelle genes and how it differs from that of nuclear genes.

The most striking and general differences between organelle and nuclear gene heredity, and the ones that are relevant in the present context, are that organelle genes show uniparental inheritance and vegetative segregation (reviewed by Birky 1983). Uniparental inheritance simply means that the progeny of a mating often contain organelle genes from only one parent. This is usually the female parent in sexually dimorphic species, in which case we refer to maternal inheritance. Vegetative segregation refers to the segregation of different alleles of organelle genes into different cells during mitotic (vegetative) cell divisions, as well as during the meiotic divisions when nuclear genes segregate. The process is often rapid and complete within one sexual generation, but may take substantially longer.

A general consequence of uniparental inheritance and rapid vegetative segregation is that heteroplasmic cells (those that are heterozygous for cytoplasmic genes) are rare, in contrast to nuclear genes where heterozygosity is common. When this is true, the effective number of organelle genes in a population will be approximately equal to the number of females, rather than being twice the effective population size.
including males and females as it is for nuclear genes. However, many organisms show substantial biparental transmission of organelle genes; this is true for instance of chloroplast genes in about one third of all plant genera (Kirk and Tilney-Basset 1978). Moreover, vegetative segregation is sometimes quite slow, as for mitochondrial genes in animals (Solignac et al. 1984; Rand and Harrison 1986). Furthermore, animal mitochondria may show extremely high mutation rates at some sites, so that it is not uncommon for animals to be heteroplasmic at those sites (e.g., Brown and Desrosiers 1983; Olivo et al. 1983, Solignac, Monnerot and Mounolou 1983; Densmore, Wright and Brown 1985; Harrison, Rand and Wheeler 1985; Bermbingham, Lamb and Avise 1986; Wallis 1987; Bentzen, Legett and Brown 1988). It is therefore of vital importance to have theory that is sufficiently flexible to handle these exceptional cases, which may be of intrinsic interest and especially useful for studying population structure.

We therefore set out to develop new theory for the population and evolutionary genetics of organelles. In our first paper (Birky, Maruyama and Fuerst 1983) we proposed a simple but biologically realistic model of organelle gene heredity. For the case of infinite neutral alleles, we obtained equations for organelle gene diversities within zygotes, gametes, individuals, and populations. We defined effective gene numbers which could be used to convert existing theory for nuclear genes into theory for organelle genes. As expected, for the common situation where there is nearly complete uniparental inheritance, or where the rate of vegetative segregation is high compared to the mutation rate, the effective number of organelle genes is approximately equal to the number of females. Unless the breeding population is heavily biased in favor of females, the gene diversity at equilibrium between mutation and drift will be much lower for organelle than for nuclear genes (assuming mutation rates are similar). Others have reached the same basic conclusions by different approaches which did not include all levels of gene diversity (Takahata and Maruyama 1981; Chapman et al. 1982; Takahata 1985).

In that paper we also argued that populations will tend to be more subdivided for organelle than for nuclear genes. This is because organelle genes will have lower effective migration rates: assuming maternal inheritance and extremely low frequencies of heteroplasmic individuals, migrating males will not transmit any organelle genes to the next generation of their new colony, while migrating females will transmit the equivalent of only one gene. In the present paper we develop rigorous theory for organelle genes in a subdivided population. The results verify and extend our previous conjectures and the work of Takahata and Palumbi (1985), which assumed that gametes are never heteroplasmic. Equally important, we drop that assumption, which is now known to be wrong for certain kinds of variation in animal mitochondrial genomes. We identify situations in which the anticipation of increased subdivision for organelle genes is not met. Further, we compare and evaluate several different measures of population subdivision, and we estimate the rates at which various parameters of gene diversity and population subdivision approach equilibrium. Finally, we compare the expected equilibrium gene diversities of organelle and nuclear genes, identifying the combinations of parameters for which diversity is greater for one type of genome or the other.

MODEL OF ORGANELLE GENE HEREDITY

The model has been described and diagrammed in our earlier paper (Birky, Maruyama and Fuerst 1983). Assume that each cell contains a number \( n \) of complete copies of the organelle genome, and hence \( n \) copies of each gene. In order to model the stochastic replication and partitioning of genes at cell division, we assume that each daughter cell receives a sample, without replacement, of \( n \) genes from the mother cell at each cell division. Then \( n_0 \) is the number of genes which would give the observed rates of vegetative segregation and intracellular random drift for the particular organism. In general it will be much smaller than the real number \( n \) of organelle genes. During sexual reproduction each zygote receives a fraction \( \alpha \) of its organelle genes from the female parent and a fraction of \( \beta = 1 - \alpha \) from the male parent. If inheritance is strictly maternal, \( \beta = 0 \). We assume \( \alpha > \beta \), as is usually the case. However, for the less frequent situation where paternal transmission predominates (such as that of chloroplast genes in gymnosperms: Ohba et al. 1971; Neale, Wheeler and Allard 1986), the definitions of \( \alpha \) and \( \beta \) can simply be reversed.

We consider only alleles that are neutral, i.e., selectively equivalent. Since most of the experimental data on organelle population and evolutionary genetics are sequence data obtained by restriction analysis or sequencing, it is most appropriate to use the infinite alleles model (Wright 1949; Kimura and Crow 1964) in which every mutation is assumed to result in a new allele not already present in the population. Mutations occur with frequency \( u \) per gene copy per sexual generation. Neutral mutant alleles are fixed in cells, i.e., their frequency goes from \( 1/n \) to 1 in a cell, with probability \( 1/n \) by intracellular drift and random partitioning (Backer and Birky 1985). Thus \( u \) is also the frequency of gametes that are homoplasmic for
new mutations, per gene per sexual generation, when heteroplasmic gametes are negligible (Birky, Maruyama and Fuert 1983). For some purposes it is best viewed as the probability that a gene sampled at random from a gamete will have undergone mutation, or be descended from a gene that underwent mutation, during the preceding sexual generation.

FINITE ISLAND MODEL FOR ORGANELLE AND NUCLEAR GENES

The population is assumed to consist of a finite number \( L \) of colonies (subpopulations, demes), within each of which mating is random between separate sexes. Each colony contains \( N_f \) females and \( N_m \) males. The numbers of colonies and of individuals are constant. Each colony exchanges a fraction \( m_f \) of females, and a fraction \( m_m \) of males, for females and males chosen randomly from the entire population (including the same colony). Since female and male migrants will generally transmit organelle genes with different efficiencies to their progeny, and hence to the colonies to which they migrate, we define an effective migration rate \( m_f = \alpha m_f + \beta m_m \), as was also done by Takahata and Palumbi (1985). After migration, the individuals reproduce sexually and die, so that sexual generations are discrete and non-overlapping.

PARAMETERS OF GENE DIVERSITY

As before (Birky, Maruyama and Fuert 1983), genetic variability is measured in terms of the operation of drawing two different copies of a gene from the population. The gene diversity \( K \) is the probability that the two sampled genes will be different alleles. \( K \) can be defined for different levels of the population: (i) \( K_f \) for sampling two distinct genes from a single zygote; (ii) \( K_s \) for sampling two genes from a single adult cell; (iii) \( K_b \) for sampling two genes from different cells in an individual adult organism; (iv) \( K_c \) for sampling two genes from different individuals within a colony; and (v) \( K_d \) for sampling two genes from different colonies within the whole population. In addition, we define new \( K \) values. The first is a version of \( K_f \) that permits genes to be sampled from the same individual as well as from two different individuals:

\[
K^{*}_f = K_f + K_d (N - 1) \quad \text{(1)}
\]

where \( N = N_f + N_m \). The second is \( K_s \), for sampling two genes from anywhere in the population:

\[
K_s = K^{*}_f + K_d (L - 1) \quad \text{(2)}
\]

for organelle genes. For nuclear genes, \( K \) replaces \( K^{*}_f \).

MEASURES OF POPULATION SUBDIVISION

Several different measures of population subdivision have been proposed. We investigated three of these for use with organelle genes. Here we discuss only the most useful one. This is \( G_{ST} \), defined by Nei (1973) as \( (H_T - H_S)/H_T \), where \( H_S \) is the probability that two genes taken from the same colony are different, and \( H_T \) is the probability that two genes taken from anywhere in the population are different. \( G_{ST} \) is a generalization of Wright's \( F_{ST} \), equivalent to \( F_{ST} \) for two alleles, or to the weighted average of \( F_{ST} \) over all alleles Nei (1973). It may be interpreted as the fractional reduction in local gene diversity due to subdivision, or as the ratio of the observed variance in gene frequencies to the maximum variance expected if every colony is fixed for one allele or another (i.e., the maximum possible subdivision of the population). In our terminology, for organelles

\[
G_{ST} = \frac{K_s - K^{*}_s}{K_s}.
\]

For nuclei, \( K_s \) replaces \( K^{*}_s \).

RESULTS

Derivation of recurrence equations for \( K \) values:

We showed previously (Birky, Maruyama and Fuert 1983) that \( K_S \) is related to \( K \) and \( K_i \) by

\[
K_a = (1 - 1/n_f) K_s \quad \text{(4)}
\]

where \( n \) is the number of germline cell divisions in a sexual generation and

\[
K_b = \left( \frac{2^c}{2^c - 1} \right) \left( \frac{n_r - 1}{n_r + 1} \right) \left[ 1 - \left( \frac{1}{2} \right) \left( \frac{1 - \frac{1}{n_r}}{1} \right) \right] K_s \approx \left( \frac{n_r - 1}{n_r + 1} \right) K_s \quad \text{when } c \geq 10.
\]

We can then derive the following recurrence equations. For \( K_i \), the \( K_i \) of the next generation:

\[
K_i' = \alpha^2 K_a + \beta^2 K_s + 2 \alpha \beta (1 - m_f)(1 - m_m) K_i + 2 \alpha \beta m_f \left( \frac{1}{L - 1} \right) K_s + 2 \alpha \beta m_m \left( \frac{L - 2}{L - 1} \right) K_d \quad \text{(6)}
\]

This equation can be understood as follows. Recall that individuals of the first generation (with parameters \( K_a, K_s, K_b, \) and \( K_d \)) migrate, produce gametes, and die. The gametes fuse at random to produce the zygotes of the second generation. \( K_i' \) is the probability that two genes, taken at random from one second-generation zygote, are different alleles. Note that both genes could have come from an egg or both from a sperm, since each egg and each sperm carries many
copies of the organelle genome. Thus $K'_s$ depends first upon whether the two genes both came from the egg (probability $a'^2$), both from the sperm (probability $b'^2$), or one from the egg and one from the sperm (probability $2\alpha\beta$). If both came from the egg or both from the sperm, the probability that they were different in the parent (first generation male or female) is $K_e$. If one sampled gene came from the egg and the other from the sperm, then the two genes came from different parents. These two parents could have both been residents, with probability $(1 - m_j)(1 - m_a)$. The probability that the genes are different would then be $K_e$. Alternatively one parent could have been resident and the other immigrant, with probability $m_f(1 - m_a)$ or $m_a(1 - m_f)$, and the probability that the genes are different is $K_d$. It is also possible that both parents of the zygote were immigrants, with probability $m_f m_a$. In that case they may have both come from the same colony, with probability $1/(L - 1)$ and identity probability $K_e$, or they may have come from different colonies, with probability $(L - 2)/(L - 1)$ and identity probability $K_d$.

The second-generation zygotes mature to produce adults; $K'_s$ is the probability that two genes taken from different second-generation adults are different. The recurrence equation for $K'_s$ is

$$K'_s = \alpha^2 (1 - m_j)^2 \left[ \left( \frac{1}{N_f} \right) K_e + \frac{1}{N_f} K_b \right]$$

$$+ 2m_f(1 - m_j)K_d + m_f^2 \left[ \left( \frac{1}{L - 1} \right) K_e + \left( \frac{L - 2}{L - 1} \right) K_d \right]$$

$$+ \beta^2 (1 - m_a)^2 \left[ \left( \frac{1}{N_m} \right) K_e + \frac{1}{N_m} K_d \right]$$

$$+ 2m_a(1 - m_a)K_d + m_a^2 \left[ \left( \frac{1}{L - 1} \right) K_e + \left( \frac{L - 2}{L - 1} \right) K_d \right]$$

$$+ 2\alpha\beta (1 - m_j)(1 - m_a)K_e$$

$$+ m_f m_a \left( \frac{1}{L - 1} \right) K_e$$

$$+ m_f m_a \left( \frac{L - 2}{L - 1} \right) K_d + [m_f(1 - m_a) + m_a(1 - m_j)]K_d.$$

$K'_s$ depends on the same factors and terms as $K'_s$, with two exceptions. First, the genes are sampled from two different individuals in the second generation, so that they may have come from the same individuals in the first generation or from different individuals. The probabilities are $1/N_f$ or $1/N_m$ vs. $(N_f - 1)/N_f$ or $(N_m - 1)/N_m$. Second, if the sampled genes came from the same parent, the probability that they are different is $K_e$, not $K_a$, since they came from different gametes.

Similar arguments are used to obtain the recurrence equation for $K_d$, the probability of identity for two copies of a gene from different colonies:

$$K'_d = \left[ \alpha^2 m_f(2 - m_f) \left( \frac{1}{L - 1} \right) + \beta^2 m_a(2 - m_a) \left( \frac{1}{L - 1} \right) ight]$$

$$+ 2\alpha\beta (m_f + m_a - m_f m_a) \left( \frac{1}{L - 1} \right) K_e$$

$$+ \left[ \alpha^2 (1 - m_f^2) + m_f(2 - m_f) \left( \frac{L - 2}{L - 1} \right) \right]$$

$$+ \beta^2 (1 - m_a^2) + m_a(2 - m_a) \left( \frac{L - 2}{L - 1} \right)$$

$$+ 2\alpha\beta (1 - m_f(1 - m_a))$$

$$+ (m_f + m_a - m_f m_a) \left( \frac{L - 2}{L - 1} \right) K_d.$$

The recurrence relationship for $K_a$ can be obtained by using Equation 4 to substitute $K_a$ for $K_b$ in Equation 6. $K_a$ can similarly be substituted for $K_c$ in Equation 7 by using (5). The four recurrence equations can then be expressed in matrix form. $K_a$ can be eliminated as an independent variable from the matrix by using Equation 5.

This leaves a system of equations with three unknowns:

$$\begin{pmatrix} K_a \\ K_e \\ K_d \end{pmatrix}' = [P] \begin{pmatrix} K_a \\ K_e \\ K_d \end{pmatrix}$$

where $P$ has elements

$$p_{11} = (1 - 1/N_f)(\alpha^2 + \beta^2)$$

$$p_{12} = 2\alpha\beta \left[ 1 - \left( m_f + m_a - \frac{Lm_f m_a}{L - 1} \right) \right]$$

$$p_{13} = 2\alpha\beta \left[ m_f m_a \left( \frac{L - 2}{L - 1} \right) + m_f(1 - m_a) + m_a(1 - m_f) \right]$$

$$p_{21} = \alpha^2 (1 - m_f)^2 \left[ \frac{1}{N_f} \right] + \beta^2 (1 - m_a)^2 \left[ \frac{1}{N_m} \right]$$

$$+ \left( \frac{2^2 - 1}{2^2 - 1} \right) \left[ \frac{n_r - 1}{n_r + 1} \right] \left[ 1 - \left( \frac{1}{2} \right) \left( \frac{1}{2} - \frac{1}{n_r} \right) \right]$$

$$p_{22} = [\alpha(1 - m_f) + \beta(1 - m_a)]^2 \left[ \frac{1}{L - 1} \right] m_f^2$$

$$- \alpha^2 (1 - m_f) \left[ \frac{1}{N_f} \right] m_f^2$$

$$- \beta^2 (1 - m_a) \left[ \frac{1}{N_m} \right] m_a^2$$

$$p_{23} = \left( \frac{L - 2}{L - 1} \right) m_f^2 + 2m_a(1 - m_f) + \beta(1 - m_a)$$
\[ p_{SS} = \alpha^2 m_f (2 - m_f) \frac{1}{L-1} + \beta^2 m_m (2 - m_m) \frac{1}{L-1} \\
+ 2 \alpha \beta (m_f + m_m - m_f m_m) \frac{1}{L-1} \]
\[ p_{SN} = \alpha^2 \left[ (1 - m_f^2) + m_f (2 - m_f) \frac{L-2}{L-1} \right] \\
+ \beta^2 \left[ (1 - m_m^2) + m_m (2 - m_m) \frac{L-2}{L-1} \right] \\
+ 2 \alpha \beta (1 - m_f) (1 - m_m) \\
+ (m_f + m_m - m_f m_m) \frac{L-2}{L-1} \]

Recurrence Equation 9 does not account for mutation. To do this, we insert terms 
\((1-u)^2\), the probability that there was no mutation in either gene sampled, or 
\(2u - u^2\), the probability that there was a mutation in one or both of the genes sampled, as appropriate:

\[
\begin{pmatrix} K_a \\ K_c \\ K_d \end{pmatrix} = [p_{ij}] \begin{pmatrix} (1-u)^2, 0, 0 \\ 0, (1-u)^2, 0 \\ 0, 0, (1-u)^2 \end{pmatrix} \begin{pmatrix} K_a \\ K_c \\ K_d \end{pmatrix} \\
+ \begin{pmatrix} 2u - u^2 \\ 2u - u^2 \\ 2u - u^2 \end{pmatrix}.
\]

Hence at equilibrium, the vector \((K_a, K_c, K_d)\) satisfies

\[
[I - (1-u)^2(p_{ij})] \begin{pmatrix} K_a \\ K_c \\ K_d \end{pmatrix} = \begin{pmatrix} 2u - u^2 \\ 2u - u^2 \\ 2u - u^2 \end{pmatrix}
\]

or

\[
\begin{pmatrix} K_a \\ K_c \\ K_d \end{pmatrix} = [I - (1-u)^2(p_{ij})]^{-1} \begin{pmatrix} 2u - u^2 \\ 2u - u^2 \\ 2u - u^2 \end{pmatrix}
\]

where \(I\) is an identity matrix and the \(p_{ij}\)s are given in (9).

For nuclear genes there is now drift within individuals so that \(K_a = K_a - K_b\). We have the following recurrence relationships:

\[
K'_i = (1 - m_f)(1 - m_m)K_c + m_f (1 - m_m) \\
+ m_m (1 - m_f)K_d + m_f m_m \left[ \left( \frac{1}{L-1} \right) K_c + \left( \frac{L-2}{L-1} \right) K_d \right].
\]

This differs from the organelle Equation 6 by omitting the first two terms of (6), since the zygote cannot inherit both genes from one parent, and eliminating \(\alpha\) and \(\beta\), since the parents contribute equally to the nuclear genes of their offspring. Other recurrence equations for nuclear genes are

\[
K'_i = \frac{1}{4} \left[ \frac{(1 - m_f)^2}{2N_f} K_c + \left( \frac{N_f - 1}{N_f} \right) (1 - m_f)^2 K_c \\
+ 2m_f (1 - m_f) K_d + m_f^2 \left( \frac{L-2}{L-1} \right) K_d \\
+ \left( \frac{1}{2N_f} \right) K_c \left( \frac{N_f - 1}{N_f} \right) K_c \right] + \frac{1 - m_f}{2N_m} K_c \\
+ \left( \frac{N_m - 1}{N_m} \right) (1 - m_m)^2 K_c + 2m_m (1 - m_m) K_d \right] \right] \right] \\
\]

\[
K'_d = \frac{1}{4} \left[ (1 - m_f)^2 + (1 - m_m)^2 + 2(1 - m_f)(1 - m_m)K_d \\
+ \frac{1}{4} [m_f^2 + m_m^2 + 2m_f m_m + 2m_f (1 - m_f) \\
+ 2m_m (1 - m_m) + 2m_f (1 - m_m) \right] \\
+ \frac{1}{4} \left( 1 - m_f \right) \left( \frac{L-2}{L-1} \right) K_d + \left( \frac{1}{L-1} \right) K_d \right].
\]

These recurrence relationships can be expressed in matrix forms:

\[
\begin{pmatrix} K_a \\ K_c \\ K_d \end{pmatrix}' = [P_L] \begin{pmatrix} K_a \\ K_c \\ K_d \end{pmatrix}
\]

where

\[
p_{11} = 0 \\
p_{12} = (1 - m_m)(1 - m_f) + \frac{m_f m_m}{L-1} \\
p_{13} = m_f (1 - m_m) + m_m (1 - m_f) + m_f m_m \left( \frac{L-2}{L-1} \right) \\
p_{21} = \frac{1}{8} \left[ \frac{1}{N_f} \left( 1 - m_f \right)^2 + \frac{m_f^2}{L-1} \right] \\
+ \frac{1}{N_m} \left( 1 - m_m \right)^2 + \frac{m_m^2}{L-1} \right]
\]

\[
\frac{L-2}{L-1} K_d \\
+ \left( \frac{1}{L-1} \right) K_d \right].
\]
\[
\begin{align*}
 p_{22} &= \frac{1}{4} \left( \frac{N_f - 1}{N_f} \right) (1 - m_f)^2 + \frac{m_f^2}{L - 1} \\
 &+ \frac{1}{4} \left( \frac{N_m - 1}{N_m} \right) (1 - m_m)^2 + \frac{m_m^2}{L - 1} \\
 &+ \frac{1}{2} \left[ (1 - m_f) (1 - m_m) + \frac{m_f m_m}{L - 1} \right], \\
 p_{23} &= \frac{1}{4} \left[ 2 m_f (1 - m_f) + \left( \frac{L - 2}{L - 1} \right) m_f^2 \right] \\
 &+ 2 m_m (1 - m_m) + \left( \frac{L - 2}{L - 1} \right) m_m^2 \\
 &+ \frac{1}{2} \left[ m_f (1 - m_m) + m_m (1 - m_f) + \left( \frac{L - 2}{L - 1} \right) m_f m_m \right]
\end{align*}
\]

\[p_{31} = 0\]
\[p_{52} = \frac{1}{4} \left( \frac{1}{L - 1} \right) (m_f + m_m) (4 - m_f - m_m)\]
\[p_{33} = \frac{1}{4} \left( \frac{L - 2}{L - 1} \right) (m_f + m_m) (4 - m_f - m_m) \]

We can incorporate mutation into (15) by an approach analogous to that for organelle genes, yielding the equilibrium equation

\[
\begin{bmatrix}
K_e \\
K_i \\
K_d
\end{bmatrix} = \left[ I - (1 - u)^{p_0} \right]^{-1} \begin{bmatrix}
2u - u^2 \\
2u - u^2 \\
2u - u^2
\end{bmatrix}.
\]

**Computer evaluation of recurrence equations:** We are interested in evaluating the \(K\) values and measuring population subdivision when there is equilibrium between mutation and random drift. However, it is difficult to obtain an explicit equilibrium solution of Equation 11 without making the simplifying assumption that \(K_e\) is negligibly small. We therefore first solved the recurrence equations by computer, starting from populations with extreme nonequilibrium conditions and following the changes in \(K\) parameters and measures of population subdivision as the populations went to equilibrium. This procedure gave us exact equilibrium values for populations in which \(K_e\) ranged from near zero to very large. These could then be compared to values calculated from explicit equilibrium equations that assumed negligible \(K_e\), in order to determine how robust the assumption is. Our procedure also enabled us to determine how quickly a population goes to equilibrium. This determines the extent to which natural populations, subject to frequent disturbances, are effectively in equilibrium.

Two starting conditions were used: (i) a genetically homogeneous population with only one allele (all \(K\) values = 0), to observe the differentiation of a population under the joint influence of mutation and migration; and (ii) a maximally subdivided population where each colony is fixed for a different allele (\(K\), and \(K_d = 1\)) to study the loss of differentiation (dedifferentiation) when migration is initiated.

**General behavior of differentiating and dedifferentiating populations:** Figures 1 and 2 illustrate the behavior of \(K\) values and measures of differentiation as populations approach equilibrium. Note that for the large number of colonies (50) and the modest number of individuals per colony (20 males and 20
females) in our examples, $K_s$ and $K$, are nearly identical, as are $K_t$ and $K_d$, so only one member of each pair is plotted. Some general aspects of organelle gene diversity, which are the same as in panmictic populations, are unaffected by population subdivision: (i) The total gene diversity between individuals, $K_t$ (approximated closely by $K_d$), is higher when there is no paternal transmission of genes (cf. panels B and C, Figure 1). (ii) The gene diversity within single cells, represented by $K_s$ for zygotes, is generally low. It is higher when there is some paternal transmission of genes (cf: $K_s = 0.0004$ in Figure 1B, 0.01 in Figure 1C). It is also higher when $n_s$ is large so that the vegetative segregation of alleles is slow, or $e$ is small so that segregation cannot be completed before gametes are produced.

The approach to equilibrium: Figures 1 and 2 are examples to illustrate that the organelle gene diversity within zygotes, $K_s$, goes to equilibrium more rapidly than the other $K$ values; equilibration of the diversity within colonies, $K_c$, is much slower; and the between-colony diversity $K_d$ reaches equilibrium last (although when $K_c$ and $K_d$ go to similar equilibrium values, they get there at about the same time). Both $K_c$ and $K_d$ are quite slow to equilibrate, so that a population that has been displaced from equilibrium by changes in population structure (colony number, migration rates) or in gene diversity may require thousands of generations to equilibrate again. The more subdivided the population, the slower the equilibration; in Figure 1B, where $G_{ST} = 0.98$, $K_d$ will equilibrate at about 0.94, but after 22 thousand generations it has only reached 0.65. Even a population with a modest $G_{ST}$ of 0.1 (Figure 1A) takes 5,000 generations to reach equilibrium for within-colony and between-colony diversities.

This slow equilibration of gene diversities in subdivided populations has been noted for nuclear genes by CROW and AOKI (1984), who give approximations of the leading eigenvalues of the transition matrices, which determine the rates of approach to equilibrium. We verified this by determining the leading eigenvalues of the matrix ($p_{ij}$) in equation 9.

Figures 1 and 2 show that $G_{ST}$ and $(1 - G_{ST})$ go to equilibrium very rapidly, even while $K_s$ and $K_d$ are still changing. The structure of the population as reflected in these ratios equilibrates quickly while the gene diversity is still changing. This was previously noted for nuclear genes by CROW and AOKI (1984). If populations are being disturbed frequently in ways that alter gene diversity or structure, it is possible that gene diversities are rarely at their equilibrium values, while $G_{ST}$ usually is.

CROW and AOKI (1984) give a formula for the time required for $G_{ST}$ to go half way to equilibrium:

$$t_{1/2} = \frac{\ln 2}{2m + \frac{1}{2}N_e}.$$  

(17)

This equation can be easily modified for use with organelle genes, following the general approach developed by BIRKY, MARUYAMA and FUERST (1983), for cases where $K_c$ is much smaller than $K_o$. We simply substitute the effective number of organelle genes, $N_o = N_oN_e/(\alpha^2 N_o + \beta^2 N_e)$, for the effective number of nuclear genes ($2N_e$), and substitute the organelle effective migration rate for the nuclear effective migration rate. This yields

$$t_{1/2} = \frac{\ln 2}{2m_e + 1/N_o}.$$  

This figure has the same curves as Figure 1E but on a logarithmic time scale to show details of their early behavior. Note that $K_c$ is also on an expanded vertical scale.
We solved this equation for various combinations of parameter values. The results from (18) are compared to half-times roughly estimated from the solutions of our recurrence equations in Table 1. The agreement is generally good, especially for small values of $\beta$ or when $n_s$ is small or $c$ is large so that $K_d$ is small. Note that these times are quite short, verifying that $G_{ST}$ may be close to equilibrium most of the time.

**TABLE 1**

Generations required for $G_{ST}$ to go halfway to equilibrium, comparison of results from recurrence equations to those calculated from explicit Equation 18

<table>
<thead>
<tr>
<th>$n_s$</th>
<th>$c$</th>
<th>$\beta$</th>
<th>$m_f$</th>
<th>$m_m$</th>
<th>$m_c$</th>
<th>Explicit</th>
<th>Recurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>10</td>
<td>0</td>
<td>0.0003</td>
<td>0.0003</td>
<td>13.7</td>
<td>11-14</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>0</td>
<td>0.0100</td>
<td>0.0100</td>
<td>9</td>
<td>9-10</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>0.02</td>
<td>0.0003</td>
<td>0.1000</td>
<td>13.2</td>
<td>15-16</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>0.40</td>
<td>0</td>
<td>0.0100</td>
<td>20.4</td>
<td>24-35</td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>20</td>
<td>0</td>
<td>0.2000</td>
<td>0.2000</td>
<td>1.54</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>65</td>
<td>20</td>
<td>0</td>
<td>0.0200</td>
<td>0.0200</td>
<td>7.70</td>
<td>7-8</td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0.0003</td>
<td>13.7</td>
<td>11-14</td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>20</td>
<td>0.10</td>
<td>0.0500</td>
<td>0.5000</td>
<td>0.950</td>
<td>3-4</td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>20</td>
<td>0.10</td>
<td>0</td>
<td>0</td>
<td>16.9</td>
<td>23-36</td>
<td></td>
</tr>
</tbody>
</table>

$N_f = N_a = 20$, $L = 50$, $u = 10^{-7}$ throughout. Values from the recurrence equations are given as upper and lower bounds because the computer program did not provide $G_{ST}$ values at every generation after the tenth.

We also studied the behavior of another measure of population subdivision: $L_c = (1 - K^*)/(1 - K)$, equivalent to $f_0/f_1$ of Crow and Maruyama (1971). It equilibrates much more slowly than does $G_{ST}$, and even more slowly than the $K$ values in the case of a highly subdivided population. Consequently it does not seem to be very useful for organelle genes.

**Equilibrium solutions when $K_d$ is negligible:** Takahata and Palumbi (1985) obtained equilibrium expressions for $I = 1 - K^*$ and $f = 1 - K_d$ by assuming that there is no heterogeneity within cells. Although this situation may be unrealistic for cases where there are mutational "hot spots," for instance in the animal mtDNA D-loop, it is clearly a satisfactory approximation for most situations (Birky, Maruyama, and Fuerst 1983). They used a model with a finite number of alleles, but obtained approximations for the case of large numbers of alleles, which should agree closely with an infinite alleles model. Using our symbols, and noting that their $m_s$ is equivalent to our $m_s$, $(L - 1)/L$, their equations are:

$$\hat{K}_{d} \approx \frac{2N_{st}u}{2N_{st}u + u/m_s + 1/L}$$  \hfill (19)

$$\hat{K}_{t} \approx \frac{2N_{st}u}{2N_{st}u + u/m_s + 1/L}.$$  \hfill (20)

**Birky, Maruyama and Fuerst** (1983) showed that equations for organelle gene diversity in an infinite neutral alleles model can be obtained by simply substituting $N_n$ for $2N$, as the effective number of genes in the corresponding equations for nuclear genes, providing $K_d \ll K$. We obtained an equation for the equilibrium expectation of $K$, by making this substitution, and by substituting $m_s$ for $m$, in the equation of Maruyama (1979) for the total nuclear gene diversity in a subdivided population. The resulting equation was the same as (20), as expected because the total diversity $K$, defined in (2), differs from $K_d$ only by the factor $(K_d - K^*)/L$. The corresponding formula for a panmictic population is $\hat{K} = 2N_{st}u / (2N_{st}u + 1)$, where $N_n$ is the total population size, equivalent to $N_nL$ in (19) and (20) (Birky, Maruyama and Fuerst 1983).

From the above equations for $K^*$ and $K_n$,

$$\hat{G}_{ST} \approx \frac{1}{1 + 2N_{st}(L - 1)/m_s + u}.$$  \hfill (21)

In these equations there are two independent parameters, $N_{st}u$ and $N_{st}m_s$, which incorporate a number of variables: the mutation rate, the numbers of males and females, their relative contributions of genes to the zygotes, and their relative migration rates. Thus the results which we present for particular combinations of variables will hold true for many other combinations, so long as the products $N_{st}u$ and $N_{st}m_s$ are constant.

We also obtained an equation for the equilibrium expectation of $G_{ST}$ by making the same substitutions in the nuclear gene equation of Takahata and Nei (1984):

$$\hat{G}_{ST} \approx \frac{1}{1 + N_n(L - 1)/[(1 - m_s^2)(1 - u) - 1]}.$$  \hfill (22)

Equation 21 can be derived from (22) by assuming that $m_s \ll 1$ and $u \ll 1$. As expected, these same equations can be obtained from the equations for $K^*$ and $K_d$ of Takahata and Palumbi (1985). It is encouraging that two independent approaches lead to the same results. With respect to the assumptions about the magnitudes of $m_s$ and $u$, we find that the
approximate equations are in good agreement with computer solutions of the exact equations when \( u \) is as large as \( 10^{-3} \) and \( m \) is as large as 0.1. The approximate equation for \( K_e \) also requires that \( m > u \); we find the approximation to be poor (in error by about 20\% or more) when the ratio \( m/u \) is 5 or less, and good when it is 10 or greater. Finally, the approximate equations for \( K_d \) and \( K_i \) require that \( L(2N_o,m_o+1) \) be 10 or greater.

It is instructive to convert the equation for total organelle gene diversity into an equation for the effective number of alleles in the population: \( 1/j_i = 1/(1-\hat{K}_o) \approx 1/(1-\hat{K}_d) = 1 + 2N_oLu + Lu/m_o. \) This is larger than the corresponding number for a panmictic population by the factor \( Lu/m_o. \) The effective number of nuclear alleles is increased by precisely the same factor in subdivided populations, except that \( m_o \) is replaced by \( m \) in the case of nuclear genes (MARUYAMA 1979).

**When is \( K_o \) negligible?** It is important to know how small the within-individual heterogeneity must be for these approximate equations to be applicable. To investigate this question, we used our matrix equations to determine exact equilibrium values of \( K^p, K_d, K_i, \) and \( G_{ST} \) for several combinations of population sizes, migration rates, paternal transmission frequencies, and mutation rates. For each combination of parameters, we varied the equilibrium values of \( K_o \) and \( K_a \) by varying the number of cell divisions per generation \((c)\) and the effective number of genomes per cell \((n_e)\). These different \( K_a \) values corresponded to different values of \( K^p, \hat{K}, \), and \( G_{ST} \), which were compared to the approximate values calculated from Equations 19–21. The percentage error in the approximate values were plotted against \( \hat{K}_a \). An example is shown in Figure 3.

Figure 3. In all cases we investigated, the approximate equations tend to overestimate \( G_{ST} \) and underestimate gene diversities. For most combinations of parameters, the error is about 5\% or less when \( \hat{K}_a \) is small. When the mutation rate is high and paternal transmission is low, as is likely for those cases where heteroplastic individuals have been found (see DISCUSSION), the error begins to increase sharply between \( \hat{K}_a \) values of 0.01 and 0.1, often exceeding 20\%. The error begins to increase at lower values of \( \hat{K}_a \) when the mutation rate is lower. The error is much greater when there is significant paternal transmission, even if \( \hat{K}_a \) is small; none of the approximate equations seems to be useful for \( \beta > 0.1 \) (as would be the case for chloroplasts in some plants), and the equation for \( G_{ST} \) is not useful for \( \beta = 0.1 \). The error in the approximate equations also depends upon the ratio of the population diversities \( \hat{K}_o, \hat{K}_d \) or \( \hat{K}_i \) to \( \hat{K}_a \), as previously noted by BIRKY, MARUYAMA and FUEST (1983) for panmictic populations. Figure 4 shows that the population diversities need be no more than about tenfold greater than the diversity within cells, when the mutation rate is high. Note that for all of these calculations we kept \( u \leq 10^{-3}, m_o \leq 0.1, \) and \( m_o/u \geq 25, \) since the approximate equations assume that \( u \) is small absolutely and much smaller than \( m_o. \)

**Subdivision for organelle vs. nuclear genes:** Intuitively we expect that populations will generally be more subdivided for organelle than for nuclear genes. This is because the effective migration rates of organelle genes are lower than the actual migration rates, due to maternal inheritance and the scarcity of heteroplasmic individuals. An example is seen in Figure 1, comparing panels C and D. However, as TAKAHATA (1985) noted, this relationship might be overturned if females are in great excess. The relative degree of
differentiation shown by organelle and nuclear genes will also be affected by the relative migration rates of males and females, and by the extent to which males transmit organelle genes. We identified the circumstances under which nuclear genes will show the same or greater amount of population subdivision by using Equation 21 and the corresponding equation for nuclear genes (in which \(2N_o, m\) and \(m\) are replaced by \(4N, m\) and \(m\) to obtain an expression for the ratio of the organelle \(\hat{G}_{ST}\) to the nuclear \(\hat{G}_{ST}\). When the effective migration rates, \(m_0, m_e\), are larger than the mutation rates \(u, L\) is large so that \(L/(L - 1)\) is close to 1, the ratio is closely approximated by

\[
\frac{\hat{G}_{ST/\text{organelle}}}{\hat{G}_{ST/\text{nuclear}}} = \frac{2N_m}{N_o m_e}.
\]

and the organelle and nuclear \(\hat{G}_{ST}\) values are equal when

\[
2N_m = N_o m_e.
\]  \hspace{1cm} (23)

We solved this equation numerically for various combinations of the sex ratio \(R = N_o/N_f\) and the migration ratio \(r = m_o/m_f\), for different values of \(\beta = \text{frequency of paternal transmission}\). Figure 5 shows that for nuclear genes to show more subdivision than organelle genes, the sex ratio must always be less than 1/7, and in fact must be much less than that unless the migration ratio is also very low. Nuclear genes may show more subdivision than organelle genes even if the migration ratio is greater than 1, but this requires that there be little or no paternal transmission and a sex ratio less than 1/15. These conditions might be met by mammals in which males have large harems, even if the males also migrate more than females.

This analysis assumes a diploid organism in which the migrants are diploid, as in many animals, so that the nuclear \(m = m_f + m_o\). When migrants are haploid (e.g., pollen), the results are quite different. The analysis also assumes that organelle and nuclear genes have equal mutation rates. Although this is probably untrue for many or most organisms, it makes little difference for \(\hat{G}_{ST}\) which is very insensitive to mutation rates (Crow and Aoki 1984). In fact, \(u\) can be omitted from Equation 21 with little error so long as the population size is small or \(u/m\) is small.

Heteroplasmy for organelle genes might also have a significant effect on the ratio of nuclear to organelle \(\hat{G}_{ST}\) values. For nuclear genes, each migrating individual carries two alleles. Each migrant could carry as many different alleles as there are copies of the mitochondrial genome in a cell, i.e., hundreds or thousands. When \(K_o\) (or \(K_e\)) is very small, most individuals are homoplasmic and each migrant effectively carries only one organelle allele. But if \(K_e\) is large, so that many migrants are heteroplasmic, then potentially each migrant could carry more than two alleles and effective migration rates could be greater for organelles than for nuclei. We used the recurrence equations to test this possibility, holding \(m\) constant and varying \(K_e\) by varying the rate of segregation, i.e., by varying \(c\) and \(n_e\). The results are shown in Figure 6. Organelle \(\hat{G}_{ST}\) does in fact decrease approximately linearly with increasing \(K_e\), but the effect is very small and is probably negligible for realistic values of \(K_e\).

**Gene diversity for organelle vs. nuclear genes:**

Uniparental inheritance and homoplasmicity reduce the effective number of organelle genes in a popula-
sensitive to mutation rates. Nuclear and organelle genes assume equal mutation rates in organelles and nuclei. Gene diversities are higher for nuclear than for organelle gene diversity only when the breeding ratio of males subdivided, nuclear gene diversity may be greater. However, in contrast to these same features of organelle gene inheritance to nuclear genes. But if the population is subdivided, assuming that complete maternal inheritance, and thereby reduce the gene diversity, relative to nuclear genes. If the population is subdivided, these same features of organelle gene inheritance reduce the effective migration rate and thereby increase the total gene diversity. We asked if there are conditions under which the effects of population subdivision overshadow those of reduced effective gene number so that gene diversity might actually be greater for organelle genes than for nuclear genes.

We used Equation 20 for organelle gene diversity, assuming that \( K_{d} = K_{o} \). For nuclear gene diversity we substituted \( 2N_{e} \) for \( N_{o} \) and \( m \) for \( m_{o} \). Assuming complete maternal inheritance \( (\beta \approx 0) \) and identical mutation rates for organelle and nuclear DNA, organelle and nuclear gene diversities will be equal when

\[
 r = \frac{7R - 1}{1 - 7R + (R + 1)/2N_{e}m_{f}} \tag{24}
\]

where \( r \) and \( R \) are the migration and breeding sex ratios as above.

The results are shown in Figure 7. The vertical line at \( R = 1/7 \) is a reminder that in a panmictic population, nuclear gene diversity is greater than organelle gene diversity only when the breeding ratio of males to females is less the 1/7, which is the area (parameter space) on the left side of the figure (BIRKY, MARUYAMA AND FUERST 1983). But when the population is subdivided, nuclear gene diversity may be greater even for sex ratios greater than 1, providing the migration rates are even moderately biased in favor of males. For example, if \( 2N_{e}m_{f} = 0.1 \), then gene diversities are higher for nuclear than for organelle genes so long as \( r \geq 3 \). We emphasize that these results assume equal mutation rates in organelles and nuclei. However, in contrast to \( G_{SR} \), gene diversities are very sensitive to mutation rates. Nuclear and organelle mutation rates are different in some organisms, and the differences may strongly affect the ratio of organelle to nuclear gene diversities.

**DISCUSSION**

In developing theory for organelle population genetics, we continued to follow the method of our previous paper (BIRKY, MARUYAMA AND FUERST 1983). First we develop recurrence equations that make no a priori assumptions about the magnitudes of the parameters. From these we derive relatively simple equations for equilibrium values of gene diversity (and population subdivision in this paper), making the assumption that the within-cell genetic heterogeneity, \( K_{o} \), is negligible. For effectively neutral alleles whose fate is determined primarily by random drift, these equations are obtained simply by substituting different effective numbers of genes and migration rates in the existing equations for nuclear genes. We are then able to determine the effects of assuming that \( K_{o} \) (or \( K_{d} \)) is negligible, which is usually but not always true. Moreover, we are able to examine the rate at which measures of gene diversity and population subdivision approach equilibrium.

We found that \( G_{SR} \) goes to equilibrium very quickly, while gene diversities within and between populations equilibrate very slowly. This was previously shown for nuclear genes by CROW AND AOKI (1984); we extended it to organelle genes, and also found that measures of
population subdivision other than $G_{ST}$ equilibrate slowly. It is possible that most populations are at or near equilibrium for $G_{ST}$ and far from equilibrium with respect to diversity; this would be the case if they were subjected to major disturbances as infrequently as once in a few thousand generations. There are few hard data on the frequency or magnitude of such disturbances. Moreover, populations may show approximately constant diversity for tens of generations while they are still far from equilibrium (see also VARVIO, CHAKRABORTY and NEI 1986), so that measurements of diversity over the usual period of a field study may give a false impression that the population is in equilibrium. These results should add to the growing recognition among population geneticists of the importance of knowing the history of a population (e.g., SLATKIN 1987).

Our results verify those of TAKAHATA and PALUMBI (1985) and our own less rigorously developed theory (BIRKY, MARUYAMA and FUERST 1983) in demonstrating some important properties of subdivided populations when $K_e$ is small relative to $K_s$. In particular, such populations will usually, but not always, show more subdivision for organelle genes than for nuclear genes. In fact it is possible for a population of organisms to be effectively panmictic for their nuclear genes and still show significant subdivision with respect to their organelle genes. Strong experimental evidence for such a situation has been presented by DESALLE et al. (1987), who found highly significant geographic differentiation for mtDNA in Hawaiian Drosophila mercatorum, but no detectable differentiation for nuclear genes. HALE and SINGH (1987) showed that $G_{ST}$ is also higher for mitochondria than for nuclei in D. melanogaster. Correlated studies of dispersal in D. mercatorum (JOHNSTON and TEMPLETON 1982) show that males and females migrate with the same low frequency and low mean distance. Moreover, most migrating females have been inseminated by 2–4 males, so that 6–8 haploid nuclear genomes are transferred for every 1 mitochondrial genome. However, it is well to remember that some organisms may show the opposite situation, with either the migration or breeding sex ratio biased strongly toward females or female gametes, and consequently populations may show more subdivision for nuclear than for organelle genes. The theory developed here does not apply to plants, in which dispersal may be by means of male gametes (pollen) or by zygotes (seeds or, less often, plants) which may be male, female, or both, all at different rates.

The greater subdivision of organelle genes is expected to affect the measured gene diversity. But the direction of the effect depends upon which $K$ parameter is being measured. This in turn is determined by the mode of sampling of the individuals from the population. If the sampled individuals come mainly or entirely from one localized region within which mating may be random, then one is estimating $K_e$ and this measure of diversity is expected to be lower in a subdivided population than in a panmictic population. As we and others have already shown, organelle genes will usually show less gene diversity than nuclear genes in a panmictic population (assuming equal mutation rates), and this difference will be exacerbated by population subdivision. On the other hand, if one deliberately samples individuals from widely separated localities in a subdivided population, then one is measuring $K_s$. Either of these strategies will give what we call $K_e$ in a panmictic population (BIRKY, MARUYAMA and FUERST 1983). Both measures of gene diversity from a subdivided population will be greater than the corresponding $K_e$ for the panmictic case. We suspect that very often one does not know in advance whether or not there is significant subdivision, and that sampling is either random or more or less uniformly distributed over the population, so that $K_e$ or $K_s$ is measured if the population is in fact subdivided.

There are only a few cases where there are comparable data on nucleotide diversities of nuclear and mitochondrial genes in animals; the data summarized by NEI (1983) suggest that they are similar, in primates as well as in Drosophila. The available plant data (summarized by BIRKY 1988) suggest that intraspecific diversity is lower in chloroplast than in nuclear genomes. The interpretation of these results is becoming more complicated. If organelle and nuclear mutation rates are similar, then diversity will usually be greater for nuclear than for organelle genomes unless: (i) in a subdivided population, $m_n \gg m_o$ so that $m = m_o$; or (ii) in any population, $N_f \gg N_s$ so that $N_m \gg N_s$ (BIRKY, MARUYAMA and FUERST 1983). An examination of Figure 7 shows that the conditions under which nuclear genes may show more diversity than organelle genes are rather restrictive but not impossible for some organisms. On the other hand, mutation rates often differ between nuclei and organelles, judging from comparisons of the rates of silent substitutions between species. For example, in some mammals, mitochondrial genes appear to have a substantially higher mutation rate than nuclear genes (L1, L2O and WU 1985; BROWN 1985; BROWN and SIMPSON 1982; MIYATA et al. 1982). The situation may be reversed in plants, with nuclei having the highest and mitochondria the lowest mutation rates, and chloroplasts in between (WOLFE, LI and SHARP 1987). The effect of different mutation rates may overshadow that of different effective population sizes and migration rates with respect to gene diversities, but not with respect to $G_{ST}$ which is extremely insensitive to muta-
tion rate. A more thorough study of the relative effects of mutation and non-Mendelian inheritance on gene diversities in organelle versus nuclear genes is needed and is underway. And of course a reasonably complete and effective understanding of organelle gene diversity and population structure awaits an evaluation of the roles of selection and linkage.

We tested the validity of the approximate equilibrium solutions for $K$ parameters and $G_{ST}$. In general the approximations are excellent so long as $K_a$ is less than about 0.01, or the ratio of population diversity ($K_a$, $K_d$, or $K_l$) to $K_a$ is greater than about 10. Biologically, this means that the rate of vegetative segregation, measured by the parameter $n_v$, is so great that a new mutation becomes homoplasmic long before another mutation occurs in the same gene and lineage. This is usually true for the most intensively studied mitochondrial and chloroplast genomes (reviewed by Birch 1983), except of course in the presence of mutator genes such as $i o j a p$. An important class of exceptions has recently been discovered: mutational hot spots which produce frequent duplications or deletions, usually if not always in the region of the D-loop (replication origin) of animal mtDNA. Individuals with more than one phenotypic class of mtDNA have been found in the four major vertebrate classes and insects. In the latter, progeny tests of heteroplasmic females show that they produced heteroplasmic germ cells and provided estimates of allele frequencies from which $K_a$ could be calculated for each female. The most complete data are for the crickets Gryllus firmus and G. pennsylvaniae; Rand and Harrison (1989) used restriction analysis of mtDNA from individual animals to show that a collection of 319 individuals included 46% that were heteroplasmic. They calculate an average $K_a$ of 0.1486. Using Equations 4 and 5, they show that $K_a$ is in the range 0.91$K_a$ to 0.98$K_a$, so $K_a = 0.18-0.29$. Also for this sample, $K_a = 0.47$, so the population diversity is only about 3.3 times larger than the within-cell diversity. For a sample of 17 females of Drosophila mauritiana, in which 3 were heteroplasmic (Solignac, Monnerot and Mounolou 1983; Solignac and Monnerot 1986), we calculate $K_a = 0.028$.

In the studies of vertebrates, $K_a$ has not been measured directly; at best, the data include the frequency of heteroplasmic individuals in a collection, together with the numbers of different mitochondrial types within each heteroplasmic individual. Since the frequencies of those types were not given, we assumed that the frequencies were equal (which gives an upper estimate of $K_a$) or that there was one majority allele and the rest were minority types whose frequencies were 0.1 each (which gives a lower estimate of $K_a$). From those frequencies, one can calculate the probability that two molecules taken from anywhere in the individual are different:

$$K_{ab} = 1 - \sum_{i=1}^{n} x_i^2$$

where there are $n$ alleles and $x_i$ is the frequency of the $i$th allele in the organism.

To calculate $K_a$ from $K_{ab}$, we note that if an organism contains $g$ cells, $K_{ab} = K_a + (g - 1)K_a/g$, which is very close to $K_a$ when $g$ is large. Then using equations (4) and (5), we can calculate $K_a$ from $K_{ab}$. For Drosophila we used $c = 10$ and $n_v = 380$, as estimated by Solignac et al. (1984). For a sample of 92 Drosophila melanogaster with 17 heteroplasmic individuals, $K_a = 0.033 - 0.091; K_d = 0.697 so K_d/K_a = 7.7$ (data of Hale and Singh 1986; each female collected from nature was represented by an isofemale line in this study). For vertebrates, we assumed two extreme combinations of $c$ and $n_v$, which probably bracket the true values (Birky, Maruyama and Fuerst 1983): $c = 10, n_v = 500 and c = 50, n_v = 50$. Then we calculated $K_a = 0.020-0.130 and 0.013-0.066$ for the lizards Cnemidophorus tesselatus (18 heteroplasmic/72 individuals) and C. tigris marmoratus (2 heteroplasmic/20 individuals), respectively (Densmore, Wright and Brown 1985; L. D. Densmore, personal communication). For the newt Triturus cristatus, Wallis (1987) found 2 heteroplasmic individuals out of 1185, and $K_a = 0.0007-0.0053$. Bingham, Lamb and Avise (1986) found 4 heteroplasmic/52 bowfin fish (Amia calva), for which $K_a = 0.0052-0.0379; 2$ heteroplasmic/142 treefrogs (Hyla cinerea), for which $K_a = 0.0009-0.0069; and 13 heteroplasmic/163 H. gratiosa, for which $K_a = 0.0055-0.0399$. Heteroplasmic individuals have also been found in domestic cattle (Hauswirth et al. 1984), mice (Mus musculus; Bouricot, Yonekazawa and Bonhomme 1987), frogs (Rana esculenta; Monnerot, Mounolou and Solignac 1984), and bass and perch (R. W. Chapman, personal communication), but in these cases the available data do not permit calculation of $K_a$ for a wild population.

These data suggest that $K_a$ may sometimes lie in the region between $10^{-4}$ and $10^{-2}$ where the error in the approximate equations begins to increase rapidly and may exceed 20%, especially for $G_{ST}$ and $K_a$. This assumes that the heteroplasm is due to a high mutation rate of order $10^{-4}$. If the mutation rate is much lower, so that heteroplasm is due to very slow vegetative segregation and drift (large $n_v$ or $c$), or to substantial paternal contribution (as may be the case for some fish; R. W. Chapman, personal communication), the approximate equations may err by more than 20% even at the lower estimated values of $K_a$.

Our investigations also suggest that mutation rates in these hot spots may be on the order of $10^{-3}$ mutations per gene per sexual generation, since we must
use unrealistically large values of $N_e$ or $c$ to get a large $K_a$ when $u$ is $10^{-4}$ or lower. Rand and Harrison (1989) estimated $u = 1 - 2 	imes 10^{-4}$ from their cricket data. They used the equilibrium equation for $K_a$ in a panmictic population. They note a number of potential sources of error in this calculation; to these we add the possibility that the populations are not in equilibrium for $K_a$. An independent verification that the mutation rates in these cases may be very high comes from Clark (1988), who developed theory for a simple deterministic model of heteroplasmic which includes selection but permits only two alleles. Noting that his parameter $\phi$ is equivalent to our $1/n_a$, one can use his Table 2 to estimate the mutation rate from the frequency of individuals that are heteroplasmic for neutral alleles in a population at equilibrium. For the data given in the preceding paragraphs, the mutation rates range from $10^{-2}$ to $10^{-5}$.

When $K_a$ is large, substantial numbers of germ line cells in migrating animals will be heteroplasmic, containing more than one organelle gene allele; in fact, some may contain more than two alleles. Thus a migrating individual which can carry at most two alleles of a nuclear gene, might carry more than two alleles of an organelle gene. If this happened frequently, migration rates might actually be greater for organelle genes. We suspect that this rarely happens, for two reasons. First, most heteroplasmic animals do not carry more than two alleles of mtDNA. Second, although $G_{ST}$ decreases with increasing heteroplasm, the increase is small for realistic values of $K_a$.

Variation produced by mutational hot-spots may be especially useful for studying population structures. Our investigations suggest that the simple equilibrium equations for the expected gene diversity and population subdivision may not be sufficiently accurate for dealing with this kind of variation. Our precise recurrence equations provide the necessary tools: they make no assumptions about heteroplasmicity and are easily solved by computer. Rand and Harrison (1989) show how our theory may be used to analyze diversity in real populations and how it can be extended to additional levels of organization.

We thank Masatoshi Nei and an anonymous referee for providing helpful comments on the manuscript, and David Rand, Richard Harrison, and Linda Denmmore for providing us with unpublished manuscripts and original data. Supported by National Institutes of Health research grant GM34094 to C. W. B. and P. F.

LITERATURE CITED


Organelle Gene Diversity


Communicating editor: J. AVISE