AN APPROACH TO POPULATION AND EVOLUTIONARY GENETIC THEORY FOR GENES IN MITOCHONDRIA AND CHLOROPLASTS, AND SOME RESULTS

C. WILLIAM BIRKY, JR.,* TAKEO MARUYAMA,** AND PAUL FUERST*

*Department of Genetics, The Ohio State University, Columbus, Ohio 43210

**National Institute of Genetics, Mishima 411, Japan

Manuscript received July 1, 1982
Revised copy accepted November 1, 1982

ABSTRACT

We developed population genetic theory for organelle genes, using an infinite alleles model appropriate for molecular genetic data, and considering the effects of mutation and random drift on the frequencies of selectively neutral alleles. The effects of maternal inheritance and vegetative segregation of organelle genes are dealt with by defining new effective gene numbers, and substituting these for $2N_e$ in classical theory of nuclear genes for diploid organisms. We define three different effective gene numbers. The most general is N_λ, defined as a function of population size, number of organelle genomes per cell, and proportions of genes contributed by male and female gametes to the zygote. In many organisms, vegetative segregation of organelle genomes and intracellular random drift of organelle gene frequencies combine to produce a predominance of homoplasmic cells within individuals in the population. Then, the effective number of organelle genes is N_{eo}, a simple function of the numbers of males and females and of the maternal and paternal contributions to the zygote. Finally, when the paternal contribution is very small, N_{eo} is closely approximated by the number of females, N_f. Then if the sex ratio is 1, the mean time to fixation or loss of new mutations is approximately two times longer for nuclear genes than for organelle genes, and gene diversity is approximately four times greater. The difference between nuclear and organelle genes disappears or is reversed in animals in which males have large harems. The differences between nuclear and organelle gene behavior caused by maternal inheritance and vegetative segregation are generally small and may be overshadowed by differences in mutation rates to neutral alleles. For monoecious organisms, the effective number of organelle genes is approximately equal to the total population size N. We also show that a population can be effectively subdivided for organelle genes at migration rates which result in panmixis for nuclear genes, especially if males migrate more than females.

STUDENTS of population genetics and evolution are increasingly turning their attention to genes in mitochondria and chloroplasts. Because the genomes are small, similar in size to those of viruses, it is possible to use restriction endonucleases or base sequencing to survey an entire genome for genetic variability and evolutionary changes in base sequence. In vertebrates the mitochondrial genes evolve at a higher rate than do the nuclear genes, making them more useful for evolutionary studies involving short time spans

(BROWN, GEORGE and WILSON 1979). The organelle genome may also show more polymorphism within a species, revealing details of population structure that are difficult to detect by the usual electrophoretic studies of nuclear-coded enzymes (AVISE *et al.* 1979). Comparative studies of organelle and nuclear genes may be useful for analysis of gene flow via different sexes, for comparisons of sexual and asexual reproduction, and for studies of the role of heterozygosity in evolution.

In spite of the interest in organelle evolution and population genetics, there has been little formal theoretical analysis (TAKAHATA and MARUYAMA 1981; CHAPMAN *et al.* 1982). Classical theory for nuclear genes is not applicable without modification because of the differences in inheritance: organelle genes are usually inherited only from the female parent, and segregate rapidly during mitotic as well as meiotic cell divisions (reviewed by BIRKY 1978). Consequently, heterozygosity for organelle genes is rare. In this paper we present some simple theory and indicate the extent to which maternal inheritance and vegetative segregation affect the stochastic theory of neutral alleles.

Neutral mutations, i.e., selectively equivalent alleles, are fixed or eliminated from finite populations by random drift. For diploid nuclear genes in a population of N individuals, random drift is ordinarily treated by assuming that the gene pool in each generation is a random sample of $2N$ genes from the gene pool of the preceding generation. Theory is often developed assuming perfect random mating, including self-fertilization. To modify equations to fit organisms that are dioecious and thus cannot self, but otherwise mate randomly, the population size N is replaced by an *effective population size* $N_e = 4\ N_m N_f/(N_m + N_f)$ where N_m and N_f are the numbers of males and females, respectively. The effective number of nuclear genes in the population is $2N_e$ for diploids because each zygote receives two independent genes from the parents, and this number determines the rate of random drift. We will derive corresponding effective gene numbers for organelles in dioecious and monoecious sexual organisms, taking into account maternal inheritance and vegetative segregation. We also consider the effect of mutation, using an infinite alleles model (WRIGHT 1949; KIMURA and CROW 1964), one in which the number of possible alleles of a gene is so large that mutation is always to a new allele, and back mutation can be ignored. This model is appropriate when alleles are identified at the molecular level, using techniques such as restriction endonuclease analysis or sequencing of DNA.

Our results verify conjectures (e.g., ENGELS 1981) that the effective number of organelle genes in an organism with strictly maternal inheritance will be equal to the number of females in the population. We show the conditions under which this holds and also give the effective number of genes for organisms with biparental inheritance and varying sex ratios. We also show that population subdivision may be more important for organelle genes than for nuclear genes. Our results have appeared in abstract (BIRKY, MARUYAMA and FUERST 1982).

MODEL OF ORGANELLE GENE HEREDITY

Figure 1 illustrates the basic features of organelle gene heredity. One significant feature is that heteroplasmic cells often produce homoplasmic progeny, so

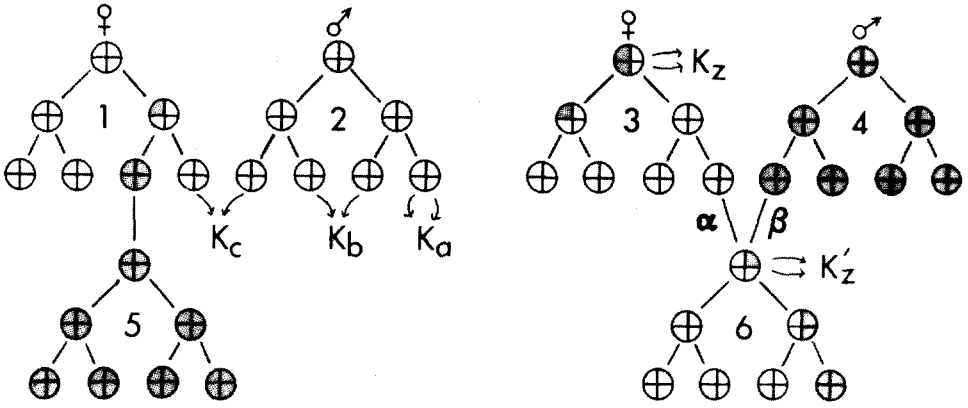

FIGURE 1.—Model of organelle heredity. Six individual organisms are shown, each beginning with a zygote that undergoes cell divisions (two in this example) to produce an adult. Each cell contains multiple copies of an organelle gene, shown as four sectors; open sectors represent wild-type alleles, shaded sectors are mutant alleles. Individual (1) begins as a homoplasmic zygote; a mutation occurs in one daughter cell and by *vegetative segregation* becomes homoplasmic in a gamete. This gamete (egg) gives rise to a homoplasmic mutant individual; the male gamete (not indicated) contributed no organelle genes (*maternal inheritance*). In individual (2) a mutation also occurs but is lost because of *intracellular random drift* taking place within the population of genes inside the cell. Individual (3) begins as a homoplasmic zygote; the mutant allele segregates into one daughter cell but is then lost by intracellular random drift. Individual (6) shows *biparental inheritance*, receiving one mutant allele from the male parent ($\beta = 0.25$) and three wild-type alleles from the female parent ($\alpha = 0.75$). These alleles become homoplasmic by vegetative segregation.

that alleles appear to segregate during mitotic as well as meiotic divisions. This *vegetative segregation* is caused in part by a random component in the partitioning of cytoplasmic organelles or DNA molecules between daughter cells at cell division (BIRKY 1978). Another factor is *intracellular random drift* of gene frequencies, due largely to a random component in the sampling of organelle DNA molecules for replication (BIRKY 1978; BIRKY *et al.* 1982). These processes do not change the mean frequency of neutral alleles in a population of organisms. Consider a population consisting of N individuals, each with n copies of the organelle genomes (and of each gene) per cell; if the mutation rate per gene per generation is u, the number of mutations per generation will be uNn. A new mutant allele arising in a cell will have an initial frequency of $1/n$ in that cell. It can be shown that new mutations will be lost by intracellular random drift with probability $1 - 1/n$. The probability that a mutation will be fixed *in cells* by drift or segregation is $1/n$. The number of cells with fixed new mutations will then be $uNn/n = uN$ per generation. The rate of appearance of homoplasmic mutant cells is thus equal to the mutation rate per gene.

A second important feature of organelle genes is that they are often inherited only from the female parent (BIRKY 1978; BIRKY *et al.* 1982). This is often caused by the paternal gametes transmitting few or no organelle genomes to the egg, which is a large cell with many copies of each organelle gene. Now consider an organism with strictly maternal inheritance; assume further that new mutations occurring in the germ line segregate rapidly so that gametes are nearly always

homoplasmic for the mutant allele or for the wild type allele. Thus only N_f females, carrying an effective number N_f of genes, transmit those genes to future generations. The effective population number and effective gene number are both simply N_f. It should thus be possible to translate existing equations for diploid nuclear genes to the corresponding equations for organelle genes by substituting N_f for $2N_e$ as the effective gene number.

MATHEMATICAL MODEL

This intuitive argument can be verified mathematically, and extended by relaxing the requirements for maternal inheritance and rapid vegetative segregation, as outlined here. We measure genetic variability within a population of organelle DNA molecules in terms of two kinds of parameters, both defined by the operation of drawing two different genes from the population: J = the probability that the two genes are identical, and $K = 1 - J$ = the probability that they are different. K and J will be referred to as "gene diversity" and "gene identity", respectively, following NEI (1973). There are four pairs of these parameters, corresponding to the four kinds of populations involved: (i) J_a and K_a for sampling two distinct molecules from a single cell; (ii) J_b and K_b for sampling two molecules from different cells in an individual organism; (iii) J_c and K_c for sampling two molecules from different individuals in a population of organisms; and (iv) J_d and K_d for sampling two molecules from different populations of a single species. J_z and K_z are used for zygotes; these are special cases of J_a and K_a that require special definition, because in general the measured values of J_a and K_a will be for adult cells or gametes. We use J' and K' to refer to values of J and K in the next generation.

We begin by modelling the cellular and molecular processes that act to reduce genetic variance in the population of organelle genes within a single cell or individual. In the absence of a more precise understanding of the mechanisms of vegetative segregation and intracellular random drift (BIRKY 1978; BIRKY *et al.* 1982), we model these processes by simple binomial sampling of genes at cell division. Each cell has an effective number n_e of organelle DNA molecules; since each molecule is a complete genome, the effective number of genes per cell is also n_e. At cell division each daughter cell receives a random sample, taken with replacement, of n_e genes from the mother cell. Thus n_e is the number of genes per cell that would give the observed rates of vegetative segregation and intracellular drift; normally it will be equal to, or more often less than, the true number of gene copies per cell, n. When we understand the mechanisms of replication and partitioning, we will be able to define n_e explicitly as a function of n and other parameters. With this model K_a changes in one cell generation according to

$$K_a' = \left(1 - \frac{1}{n_e}\right) K_a. \tag{1}$$

Let c be the number of cell generations in a cell lineage leading from a zygote to a cell or gamete in the adult individual. Then K_a in that cell or gamete is

related to K_z by

$$K_a = \left(1 - \frac{1}{n_e}\right)^c K_z. \tag{2}$$

The random sampling process will also reduce the genetic variance within the organism as a whole. K_b in the adult individual is thereby less than K_z in the zygote by a factor F_e which is derived as follows. For any pair of cells in the adult, we can trace their lineages back through 1, 2, or more divisions to their most recent common ancestral cell. If the gene diversity of the ancestral cell was K_a, then K_b for that pair of cells is $(1 - 1/n_e)K_a$. Consider a pair of cells whose lineages diverged from the ancestral cell at the i^{th} division; from equation (2),

$$K_a = (1 - 1/n_e)^{i-1}K_z,$$

so that

$$K_b = (1 - 1/n_e)(1 - 1/n_e)^{i-1}K_z = (1 - 1/n_e)^i K_z.$$

We are interested in K_b for the entire adult organism, which is the mean of K_b for all possible pairs of cells after c divisions from the zygote. There are $\binom{2^c}{2}$ possible pairs; the number of pairs whose lineages diverged at the i^{th} division is 2^{2c-i-1}. It follows that

$$\begin{aligned} K_b = K_z F_e &= K_z \frac{1}{\binom{2^c}{2}} \sum_{i=1}^{c} 2^{2c-i-1}\left(1-\frac{1}{n_e}\right)^i = K_z\left(\frac{2^c}{2^c-1}\right)\sum_{i=1}^{c}\left(\frac{1}{2}\right)^i\left(1-\frac{1}{n_e}\right)^i \\ &= K_z\left(\frac{2^c}{2^c-1}\right)\left(\frac{n_e-1}{n_e+1}\right)\left[1-\left(\frac{1}{2}\right)^c\left(1-\frac{1}{n_e}\right)^c\right] \\ &\approx K_z\left(\frac{n_e-1}{n_e+1}\right) \text{ when } c \geq 10. \end{aligned} \tag{3}$$

Since $K_b = K_z F_e$, F_e is approximately $(n_e - 1)/(n_e + 1)$. This important parameter measures random drift in a whole organism or clone.

K_z' and K_c' are related to K_z, K_c, and K_a in the preceding sexual generation by the following equations:

$$K_z' = \left(1 - \frac{1}{n_e}\right)^c (\alpha^2 + \beta^2)K_z + 2\alpha\beta K_c \text{ and} \tag{4}$$

$$\begin{aligned} K_c' &= \left(1 - \frac{1}{N_{eo}}\right) K_c + \left(\frac{1}{N_{eo}}\right)\left(\frac{n_e - 1}{n_e + 1}\right) K_z \\ &= \left(1 - \frac{1}{N_{eo}}\right) K_c + \left(\frac{1}{N_{eo}}\right)\left(\frac{n_e - 1}{n_e + 1}\right)\left[1 \Big/ \left(1 - \frac{1}{n_e}\right)^c\right] K_a \end{aligned} \tag{5}$$

where α and β are the proportions of organelle genes contributed to the zygote by the female and male gametes, respectively, so that $\alpha + \beta = 1$, and the

effective number of organelle genes in the population is

$$N_{eo} = \frac{N_m N_f}{\alpha^2 N_m + \beta^2 N_f}. \tag{6}$$

Equation 5 can be derived from equation 7 of TAKAHATA AND MARUYAMA (1981).

We wish to describe the stochastic behavior of an allele frequency x as a function of population size. Let δ_x be the infinitesimal change in x between generations t and $t + 1$, so $x_{t+1} = x_t + \delta_x$. The variance of δ_x is $V_{\delta x} = E[(\delta_x)^2]$ where E denotes the expectation. By definition, $K_c = 2x_t(1 - x_t)$ so that $V_{\delta x} = (K_c - K_c')/2$. From this and equation 5 it follows that

$$V_{\delta x} = \frac{x(1-x)}{N_{eo}} - \frac{1}{2N_{eo}}\left(\frac{n_e - 1}{n_e + 1}\right)\left[1 \Big/ \left(1 - \frac{1}{n_e}\right)^c\right] K_a. \tag{7}$$

We have noted above that if vegetative segregation is very rapid, most cells will be homoplasmic; in other words, K_a will be very small. From equation (7) we see that if $K_a \ll K_c$, then

$$V_{\delta x} \approx \frac{x(1-x)}{N_{eo}}. \tag{8}$$

If, in addition, inheritance is primarily maternal, so that $\beta \approx 0$, then $N_{eo} \approx N_f$ and

$$V_{\delta x} \approx \frac{x(1-x)}{N_f}. \tag{9}$$

Equations 8 and 9 are equivalent to the equation $V_{\delta x} = x(1-x)/2N_e$ used for nuclear genes, except that N_{eo} or N_f replace $2N_e$ as the effective number of genes in the population of organisms. Consequently, so long as $K_a \ll K_c$, all equations derived from $V_{\delta x}$ for nuclear genes can be translated to corresponding equations for organelle genes simply by replacing $2N_e$ with N_{eo} or, when inheritance is maternal, by N_f. This verifies the intuitive argument that the rate of change of organelle gene frequencies resulting from random drift is determined by an effective number of genes equal to N_f.

An exact expression for the effective number of organelle genes can be obtained for the important case where a steady rate of decay of gene diversity has been reached. This would apply to mutant alleles that have been present in the population for a few generations. Equations 4 and 5 are recurrence relations of the form $K_z' = AK_z + BK_c$ and $K_c' = CK_z + DK_c$, represented in matrix form

$$\begin{bmatrix} K_z' \\ K_c' \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} K_z \\ K_c \end{bmatrix}.$$

The eigenvalues of these equations are

$$\lambda_1 = \tfrac{1}{2}\{D + A - [(D - A)^2 + 4BC]^{1/2}\} \text{ and} \tag{10}$$

$$\lambda_2 = \tfrac{1}{2}\{D + A + [(D - A)^2 + 4BC]^{1/2}\} \tag{11}$$

where $A = (\alpha^2 + \beta^2)(1 - 1/n_e)^c$; $B = 2\alpha\beta$; $C = (1/N_{eo})(n_e - 1)/(n_e + 1)$; and $D = (1 - 1/N_{eo})$. The eigenvalues λ_1 and λ_2 are similar to λ_2 and λ_1, respectively, in

equation 12 of TAKAHATA and MARUYAMA (1981), except for our omission of mutation and for minor differences between the models. These eigenvalues are interpreted as follows. Let $\Delta K_z = K_z' - K_z$ and $\Delta K_c = K_c' - K_c$. The rates of change of K_z and K_c are $\Delta K_z/K_z$ and $\Delta K_c/K_c$. With time, both rates of change approach to a state of steady decay; the rate of that approach is determined by λ_1. When the steady state of decay is reached it is given by $\Delta K_z/K_z = \Delta K_c/K_c = \lambda_2 - 1$. Thus λ_2 determines the steady rate of decay of gene diversity measured by either K_z or K_c, and λ_1 determines the rate at which that state is approached. The steady rate of decay is reached very quickly when β is large, and more slowly for small β. For strictly maternal inheritance, $\beta = 0$ and $N_{eo} = N_f$. In that case

$$\lambda_1 = \left(1 - \frac{1}{n_e}\right)^c, \text{ and} \tag{12}$$

$$\lambda_2 = \left(1 - \frac{1}{N_f}\right).$$

Here λ_2 still represents the steady rate of decay of K_c, but λ_1 now gives the steady rate of decay of K_z. The exact value of the effective number of organelle genes at the steady rate of decay is thus given by

$$N_\lambda = \frac{1}{(1 - \lambda_2)}. \tag{13}$$

When $\beta = 0$, this reduces to N_f. When $\beta \neq 0$, the effective number of genes is often closely approximated by N_{eo}. N_{eo} usually underestimates the true effective gene number by less than 10% when $\beta \leq 0.01$, $n_e \leq 50$, $c > 10$, and $N_m = N_f \geq 500$. These parameters are quite reasonable for mitochondrial genes in animals and chloroplast genes in the majority of plants (BIRKY 1978), and probably for mitochondrial genes in plants as well. However, for a small multicellular organism with strong paternal transmission of organelle genes, the effective gene number may be larger for organelles than for nuclear genes. For example, if $c = 10$, $N_e = 100$, $\beta = 0.3$, and $N_m = N_f = 500$, the effective number of genes $N_{eo} = 6406$ for organelle genes, whereas for nuclear genes $2N_e = 2000$. The effective number of genes is large for organelles in this case because segregation is slow and the paternal contribution is large. Thus many gametes will be heteroplasmic, rather than being effectively haploid.

CHOICE OF EFFECTIVE GENE NUMBER FOR ORGANELLES

We have described three different effective gene numbers: N_λ, N_{eo}, and N_f. N_λ is exact, but assumes the steady rate of decay is reached. N_{eo} is an excellent approximation, whether or not the system is in the steady state, provided most cells are homoplasmic ($K_a \ll K_c$). This will be true if either one of two conditions are met: 1) Vegetative segregation is very rapid. N_{eo} is then a good approximation even if β is large. Segregation is easily studied in plants, where segregation of wild-type green and mutant white chloroplasts produces variegated plants with sectors of green and white cells. For most plants, this is extremely rapid and is complete within one sexual generation (KIRK and TILNEY-BASSETT 1978; MI-

CHAELIS 1955; for an exception, see VAUGHN 1981). Consequently K_a is approximately zero even if β is large. In animals, K_a for mitochondrial genes cannot be measured directly at present. However, some data suggest that vegetative segregation may take several sexual generations, and hence many cell generations, to complete. HAUSWIRTH and LAPIS (1982) have found a mutant mtDNA genome, with an altered restriction fragment pattern, segregating in a herd of dairy cattle. From their data, if c is between 20 and 50 cell generations in the germ line per sexual generation and $\beta \approx 0$, we estimate that n_e lies between 65 and 163. These values are small enough so that in many cases K_a will be small relative to K_c. 2) If β is nearly zero, K_a may still be small compared to K_c even if vegetative segregation is very slow. This is because only new mutations contribute to gene diversity within a cell (K_a) and the mean time between the occurrence of mutations is very long compared to the mean time required to eliminate the mutation, or to fix it by vegetative segregation. The mean waiting time between mutations in a cell lineage is given by $\tau = c/nu$ where u is the mutation rate per base pair per sexual generation. The mean, $\bar{t}(1/n)$, and the variance of the time to fixation or loss of a new mutation within the individual cells of a lineage are given by equations 15 and 16 below. For any reasonable values of the parameters c, n, n_e, and u, the value of τ will be much larger than $\bar{t}(1/n)$. The difference will be smallest when all parameters are large. Taking $u = 10^{-8}$ (BROWN, GEORGE, and WILSON 1979), $n = 10^4 = n_e$, and $c = 50$, we find $\tau = 5 \times 10^5$ whereas $\bar{t}(1/n) = 18.0$ with standard deviation = 357. Thus the mean K_a taken over a long time will be less than $18.0/(5 \times 10^5)$; for most of the history of a cell lineage, the cells will be homoplasmic. In most animals and plants, the egg will have many more copies of the mtDNA or cpDNA molecules than will the other, smaller cells in the germ line (*cf.* PIKO and MATSUMOTO 1976; BOGENHAGEN and CLAYTON 1974) and we cannot assume that n is constant. We have derived appropriate equations for this case; they actually increase the difference between τ and $\bar{t}(1/n)$.

The assumption that the paternal contribution of organelle genes is negligible is widely applicable (BIRKY 1978). About two-thirds of all plant genera show strictly maternal inheritance of green *vs.* white chloroplasts (KIRK and TILNEY-BASSETT 1978). In some cases the numbers of progeny examined were large enough to allow detection of less than 0.05% paternal chloroplasts ($\beta < 5 \times 10^{-4}$). For mitochondrial genes in animals and plants most of the data are based on restriction fragment analyses that would not detect paternal genotypes less frequent than 1–10% ($\beta < 0.01$–0.1) (references in BIRKY 1978; BIRKY *et al.* 1982). In a few cases the number of mtDNA molecules in an egg is known to be 10^3 to 10^6 times greater than the number in a sperm (DAWID and BLACKLER 1972; M. V. SIMPSON, personal communication). In some plants the pollen contributes no chloroplasts to the zygote (HAGEMANN 1979; SEARS 1980). TAKAHATA and MARUYAMA (1981) derived an equation for the value of $\hat{K}_a$ at equilibrium between mutation and drift (their equation 9, in which their $H = \hat{f}_a = 1 - \hat{K}_a$). The numerical values in their Figure 2 show that $\hat{f}_a \approx 1$ and $K_a \approx 0$ when $\beta < c/n_e$. This same relationship has been derived by R. W. CHAPMAN *et al.* (1982) with the use of a completely different approach. Since c and n_e probably range from

10 to 50 and 10 to 10^3, respectively, c/n_e ranges 0.05 to 1 so $\hat{K}_a \approx 0$ when $\beta < 0.05$.

The effect of maternal inheritance on the behavior of organelle genes is very sensitive to the sex ratio (Figure 2). When the sex ratio $N_m/N_f = 1$, the effective number of nuclear genes, $2N_e$, is four times the effective number of organelle genes, N_f. When the sex ratio is 1/7, $2N_e = N_f$, and when there are more than seven females for every male, $2N_e < N_f$. As the sex ratio increases from 1, the ratio $2N_e/N_f$ slowly and asymptotically approaches 8.

MONOECIOUS ORGANISMS

In monoecious or hermaphroditic species, each individual produces both male and female gametes. This case is especially important for plants. For nuclear genes in a diploid monoecious organism, the effective number of individuals is N and the effective number of genes is $2N$. For organelles, the exact value of the effective gene number at the steady rate of decay is given by N_λ where λ_2 is calculated from equations (4), (5), and (11) as before but replacing N_{eo} with N. This suffices for obligatory outcrossing species. However, if the species is random mating, then $A = (\alpha^2 + \beta^2)(1 - 1/n_e)^c + (2\alpha\beta)(n_e - 1)/N(n_e + 1)$, and $B = 2\alpha\beta(1 - 1/N)$. In either case the effective number of genes reduces to N when $\beta = 0$ and is very closely approximated by N under the same conditions in which it is approximated by N_{eo} for dioecious organisms. For monoecious selfing

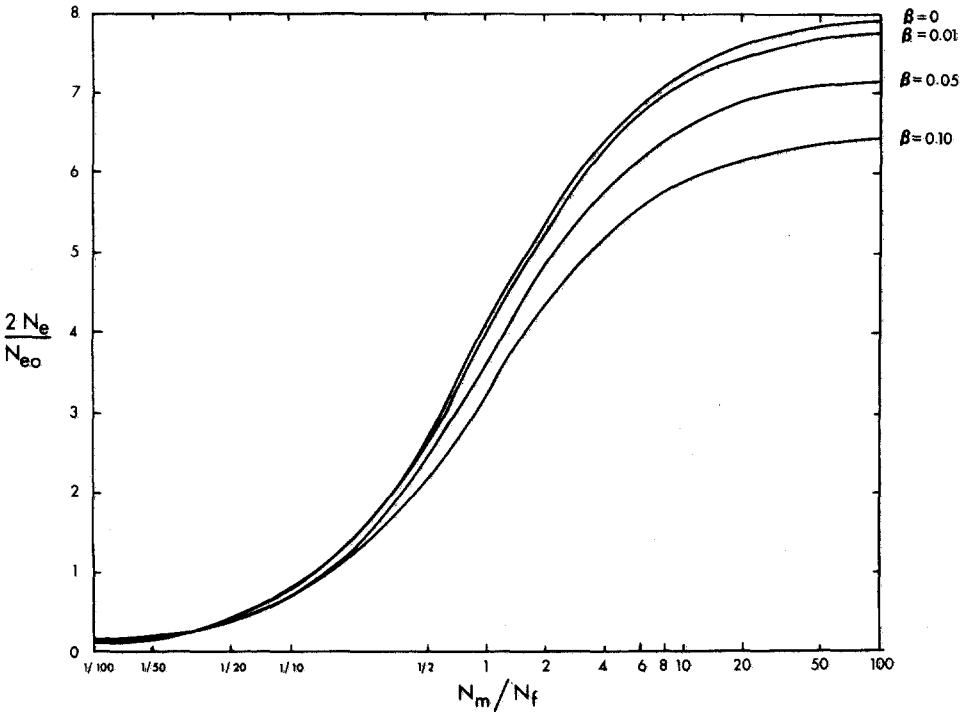

FIGURE 2.—Relationship of effective number of nuclear genes ($2N_e$) to effective number of organelle genes (N_{eo}) as a function of sex ratio (N_m/N_f), with varying degrees of paternal transmission (β).

organisms, $B = 0$ regardless of the value of β and the effective number of genes is always N. Preferential transmission of organelle genes through the egg has no effect in monoecious organisms because every individual produces eggs. However, the fact that rapid vegetative segregation keeps K_a near zero is still important because it makes the effective number of genes approximately equal to the effective number of individuals.

DYNAMICS OF FIXATION OR LOSS OF NEUTRAL ALLELES

Equations for the mean time to fixation or loss of an allele in a cell lineage caused by segregation and intracellular random drift can be obtained from equations for nuclear genes by substituting n_e for $2N_e$ (see also CHAPMAN *et al.* 1982). Time is in *cell* generations, and the initial frequency of the allele is p. When the time to fixation or loss is long enough to include more than one sexual generation, these equations hold only for a lineage of female individuals and require that $\beta = 0$. We give here only the mean time to fixation or loss:

$$\bar{t}(p) = -2n_e[p \ln p + (1 - p)\ln(1 - p)]. \tag{14}$$

For a newly-arisen mutant allele, $p = 1/n$ and the above equation reduces to

$$\bar{t}(1/n) = -2n_e\left[\left(\frac{1}{n}\right)\ln\left(\frac{1}{n}\right) + \left(1 - \frac{1}{n}\right)\ln\left(1 - \frac{1}{n}\right)\right] \approx 2\left(\frac{n_e}{n}\right)[\ln(n + 1)] \tag{15}$$

with variance

$$\sigma_t^2 = \left(\frac{n_e}{n}\right)\left[12.8n_e - 4(\ln n)^2\left(\frac{n_e}{n}\right)\right]. \tag{16}$$

We can also derive equations for the mean time to fixation or loss of an allele from the whole population of organisms by substituting N_λ (or N_{eo} or N_f where appropriate) for $2N_e$ in nuclear gene equations. In this case we assume that the allele is fixed in nearly every individual cell; time is now in *sexual* generations and no assumption about β is required. If p is the initial frequency of the allele in the population, the mean time to loss is

$$t_0(p) = -2N_\lambda[p/(1 - p)]\ln p; \tag{17}$$

the mean time to fixation is

$$t_1(p) = -(1/p)[2N_\lambda(1 - p)\ln(1 - p)]; \text{ and} \tag{18}$$

$$\bar{t}(p) = -2N_\lambda[p \ln p + (1 - p)\ln(1 - p)], \tag{19}$$

(KIMURA and OHTA 1969; MARUYAMA 1977).

From these equations one can easily derive the mean time to fixation or loss of a newly-arisen mutant allele (already fixed in cells) by substituting its initial frequency $p = 1/N$. However, this result is not quite correct for the case of maternal inheritance ($\beta \approx 0$). This is because mutations that occur in males (N_m/N of the total) can never be fixed, and are always lost in one generation: $t_0(1/N_m) = 1$. Mutations that occur in females (N_f/N of all mutations) are fixed

with a probability of $1/N_f$; for these,

$$t_1(1/N_f) = -N_f[2N_f(1 - 1/N_f)\ln(1 - 1/N_f)] \text{ and } t_0(1/N_f) = 2[N_f/(N_f - 1)]\ln N_f.$$

Taking the mean of the equations for mutations occurring in males and females, weighted by the frequencies of mutations in the two sexes, we obtain

$$t_0(1/N) = (N_m/N) + (N_f/N)[2N_f/(N_f - 1)]\ln N_f \approx (N_m/N) + 2(N_f/N)\ln N_f, \quad (20)$$

$$t_1(1/N) = -2N_f(N_f - 1)\ln(1 - 1/N_f) \approx 2N_f, \text{ and} \quad (21)$$

$$\bar{t}(1/N) = (N_m/N) + 2(N_f/N)\ln N_f + 2(1 - 1/N_f)$$

$$\approx (N_m/N) + 2(N_f/N)[\ln(N_f) + 1]. \quad (22)$$

We have verified these equations by Monte Carlo simulations. They give very nearly the same numerical values as the less exact equations, differing by an absolute value of 2 at most when the number of females is very much larger than the number of males. When $\beta \neq 0$, mutations that occur in males will not always be lost immediately. The two approaches then agree even better. In any event, it is important to note that when the effective number of genes is N_f, the time to fixation *or* loss is approximately two times longer for nuclear than for organelle genes.

GENE FREQUENCIES AT EQUILIBRIUM BETWEEN RANDOM DRIFT AND MUTATION

Of special interest in the theory of neutral alleles is the amount of polymorphism in a population in which new alleles are being eliminated by drift as fast as they are introduced by mutation. We can derive the allele frequency distribution. The number of organelle alleles whose frequency is x is given by the following probability density function in which u is the mutation rate (*cf.* KIMURA and CROW 1964):

$$\Phi(x) = 2N_\lambda u x^{-1}(1 - x)^{2N_\lambda u - 1}. \quad (23)$$

We tested this equation by Monte Carlo simulations (Figure 3); the close agreement between theory and simulation verifies that N_λ is a good approximation of the effective number of organelle genes.

Equilibrium values of K_z, K_c and J_c can easily be derived. Making the assumptions that $K_a \approx 0$, $K_c \gg u$, and that terms containing u^2 are negligibly small, the equations are

$$\hat{K}_z \approx 2\alpha\beta\hat{K}_c, \quad (24)$$

$$\hat{K}_c \approx \frac{2N_\lambda u}{2N_\lambda u + 1}, \text{ and} \quad (25)$$

$$\hat{J}_c \approx \frac{1}{2N_\lambda u + 1}. \quad (26)$$

Following STEWART (1976), the variance of $\hat{K}_c$ is

$$V_{K_c} \approx \frac{4N_\lambda u}{(1 + 2N_\lambda u)^2(2 + 2N_\lambda u)(3 + 2N_\lambda u)}. \quad (27)$$

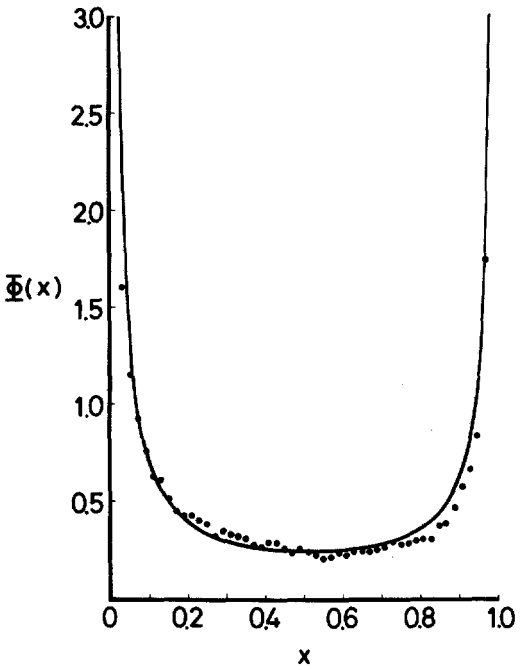

FIGURE 3.—Probability density function of organelle allele frequencies at equilibrium between random drift and mutation. $\Phi(x)$ = number of alleles having frequency x in a population of effective gene number N_λ and mutation rate u. Curve is generated by equation (23); points are from 4000 replicate Monte Carlo simulations with parameters $N_m = N_f = 50$, $c = 10$, $\alpha = 0.95$, $\beta = 0.05$, $n_c = 10$, and $u = 5 \times 10^{-4}$.

Exact solutions for $\hat{K}_z$, $\hat{K}_c$, etc. can be derived and will be published elsewhere. The organelle parameters $\hat{K}_c$ and $\hat{f}_c$ are precisely analogous to nuclear heterozygosity (gene diversity), $H \approx 4N_e u/(4N_e u + 1)$, and homozygosity (gene identity), $F \approx 1/(4N_e u + 1)$. For cases where organelle genes are inherited maternally and vegetative segregation is rapid, $N_\lambda \approx N_f$. Then the ratio of gene diversities in nuclear and organelle genes, $H/\hat{K}_c$ will be approximately $2N_e/N_f$ as long as the mutation rates are the same in the two genomes and also assuming that $N_f u$ is very small. This ratio again depends strongly on the sex ratio. TAKAHATA AND MARUYAMA (1981) derived an equilibrium equation for $\hat{f}_c$ (which they called $\hat{Q}$); their equation (10) gives values similar to our equation (26), noting that their mutation rate ν is per cell generation rather than per sexual generation so that $\nu = u/c$. Since our $\hat{f}_a$ is the same as their $\hat{H}$, their equation for $\hat{Q}/\hat{H}$ reduces to our (26) when their $\hat{H} = 1$, i.e., $\hat{K}_a = 0$. They assume $K_b = K_c$, which is only approximately correct; see equation 3. CHAPMAN *et al.* (1982) used computer simulations to show that, with strictly maternal inheritance, K_c (their H_2) can be large when K_a (their H_1) is small. Their results are in reasonable agreement with calculations from our equation 25.

SUBDIVIDED POPULATIONS

We briefly consider the effects of dividing a population into subpopulations, between which genes are exchanged by migration. We assume that inheritance

of organelle genes is strictly maternal; then only migrating females will carry organelle genes between subpopulations, whereas both males and females will carry nuclear genes. (Migrating males will contribute genes to the subpopulation they enter, but only transiently, to the extent that they form a part of that population until their death. Unless male migration rates are quite high, this effect will be negligible.) MARUYAMA (1971) considered the case of a population arranged on a torus and consisting of s subpopulations, each with N individuals; individuals migrate from one subpopulation to any other at a rate m per generation. The rate of decay of nuclear gene diversity for the whole subdivided population is the same as in a random mating population when $Nm > 1$. When subpopulations are smaller or migration rates are lower ($Nm < 1$), the rate of decay of gene diversity is $1 - \lambda = m/2s$. For organelle genes, where the effective population and gene number is N_f, the transition point occurs at $N_f m_f = 1$, where m_f is the migration rate for females, and the rate of decay of gene diversity below that point is $1 - \lambda = m_f/4s$. When $N_f = N_m = N/2$, the transition point is $Nm_f = 2$. If the migration rate is the same for females and males, the transition point is $Nm = 4$ and $1 - \lambda = m/8s$. In other words, a population will be effectively subdivided for organelle genes at migration rates at which nuclear genes are still panmictic; the difference is fourfold. That difference becomes even greater if males migrate more than females; for example, if the migration rate is fivefold greater for males than for females, then the transition point becomes $Nm = 20$ for organelles but remains at $Nm = 1$ for nuclei. Below the transition, the rate of decay of gene diversity is $1 - \lambda = m/40s$ for organelles and $m/2s$ for nuclei. Subdivision of animal populations with preferential migration of males greatly accentuates the differences in rates of random drift between organelle and nuclear genes. In plants this will be a factor only if pollen dispersal is substantially greater than dispersal by seed or whole plants.

DISCUSSION

Our results show that maternal inheritance and vegetative segregation result in organelle genes having an effective population size approximately one-fourth as large as that of nuclear genes in the same population, with a corresponding increase in rate of gene fixation by drift and a decrease in expected gene diversity at equilibrium between drift and mutation. This fourfold difference is rather small compared to differences in the behavior of nuclear genes in different species. It may disappear or even be reversed in species in which males have harems. Also it may be overshadowed by differences in mutation rates between organelle and nuclear genes. BROWN *et al.* (1979) have shown that animal mitochondria have about a tenfold higher rate of base pair substitution during evolution than do nuclear genes. Maternal inheritance and vegetative segregation do not affect the rate of base pair substitution for neutral organelle alleles. This is equal to the mutation rate per gene just as it is for nuclear genes. Consequently, if most substitutions are neutral in both mitochondria and nuclei, then mitochondria must have a tenfold higher neutral mutation rate. This may be partly a consequence of inefficient proofreading by the mitochondrial DNA polymerase (KUNKEL and LOEB 1981). In addition, the proportion of all mutations that are neutral may be higher for mitochondria (BROWN and SIMPSON 1982).

The data of AVISE *et al.* (1979) for mitochondrial and nuclear genes in the pocket gopher, *Geomys pinetis*, suggest that $\hat{K}_c$ (their p) may be substantially higher than H (calculated from their $\bar{D}$). This result can also be explained if u is higher for organelle genes. However, it is suspected that male pocket gophers maintain harems (J. LAERM, personal communication); the breeding sex ratio may be much less than one and this would also cause $\hat{K}_c$ to be greater than H. The pocket gophers also show substantially more differentiation of local subpopulations for mitochondrial than nuclear genes (AVISE *et al.* 1979). This would be expected if males migrate much more than females.

In this paper we have not considered two-locus theory, where the differences in behavior between nuclear and organelle genes should have an especially important effect. Recombination between two loci on an organelle DNA molecule will obviously be reduced by maternal inheritance and vegetative segregation, which make the heterozygous (heteroplasmic) state rare and transient. This problem is complicated by the existence of multiple genomes per cell; the likelihood that molecules may pair and recombine repeatedly in all cells, not just gametocytes; and the probability that molecules pair for recombination with some degree of randomness (BIRKY 1978).

Supported by National Institutes of Health Grant GM19607 to C. W. BIRKY, JR. and P. S. PERLMAN, by the award of a Distinguished Visiting Professorship by the College of Biological Sciences, OSU, to TAKEO MARUYAMA, and by National Science Foundation grant DEB-8110220 to P. A. FUERST.

LITERATURE CITED

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

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

BIRKY, C. W., JR., A. R. ACTON, R. DIETRICH and M. CARVER, 1982 Mitochondrial transmission genetics: Replication, recombination, and segregation of mitochondrial DNA and its inheritance in crosses. pp. 333–348. In: *Mitochondrial Genes*, Edited by G. ATTARDI, P. BORST AND P. P. SLONIMSKI. Cold Spring Harbor Press, Cold Spring Harbor, New York.

BIRKY, C. W., JR., T. MARUYAMA and P. A. FUERST, 1982 Population genetic theory for genes in mitochondria and chloroplasts. Genetics **100:** s6 (abst.).

BOGENHAGEN, D. and D. A. CLAYTON, 1974 The number of mitochondrial deoxyribonucleic acid genomes in mouse L and human He La cells: Quantitative isolation of mitochondrial deoxyribonucleic acid. J. Biol. Chem. **249:** 7991–7995.

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

BROWN, G. G. and M. V. SIMPSON, 1982 Novel features of animal mtDNA evolution as shown by sequences of two rat cytochrome oxidase subunit II genes. Proc. Natl. Acad. Sci. USA **79:** 3246–3250.

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

DAWID, I. B. and A. W. BLACKLER, 1972 Maternal and cytoplasmic inheritance of mitochondrial DNA in *Xenopus*. Dev. Biol. **29:** 152–161.

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

HAGEMANN, R., 1979 Genetics and molecular biology of plastids of higher plants. pp. 91–115. In: *Stadler Symposium*, Vol. 11, Edited by G. REDEI. Missoui Agricultural Experimental Station.

HAUSWIRTH, W. W. and P. J. LAIPIS, 1982 Rapid variation in mammalian mitochondrial genotypes: Implications for the mechanism of maternal inheritance. pp. 137–141. In: *Mitochondrial Genes*, Edited by G. ATTARDI, P. BORST AND P. P. SLONIMSKI. Cold Spring Harbor Press, Cold Spring Harbor, New York.

KIMURA, M. and T. OHTA, 1969 The average number of generations until fixation of a mutant gene in a finite population. Genetics **61:** 763–771.

KIMURA, M. and J. F. CROW, 1964 The number of alleles that can be maintained in a finite population. Genetics **49:** 725–738.

KIRK, J. T. O. and R. A. E. TILNEY-BASSETT, 1978 *The Plastids*, 2nd ed., Elsevier/North-Holland Biomedical Press, Amsterdam.

KUNKEL, T. A. and L. A. LOEB, 1981 Fidelity of mammalian DNA polymerases. Science **213:** 765–767.

MARUYAMA, T., 1971 Analysis of population structure. II. Two-dimensional stepping stone models of finite length and other geographically structured populations. Ann. Hum. Genet. **35:** 179–196.

MARUYAMA, T., 1977 *Stochastic Problems in Population Genetics. Lecture Notes in Biomathematics*, Vol. 17. Springer-Verlag, New York.

MICHAELIS, M., 1955 Über Gestzmässigkeiten der Plasmon-Umkombination und über eine Methode zur Trennung einer Plastiden-, Chondriosomen-, resp. Sphaerosomen-, (Mikrosomen)- und einer Zytoplasmavererbung. Cytologia **20:** 315–338.

NEI, M., 1973 Analysis of gene diversity in subdivided population. Proc. Natl. Acad. Sci. USA **70:** 3321–3323.

PIKO, L. and L. MATSUMOTO, 1976 Number of mitochondria and some properties of mitochondrial DNA in the mouse egg. Dev. Biol. **49:** 1–10.

SEARS, B. B., 1980 Elimination of plastids during spermatogenesis and fertilization in the plant kingdom. Plasmid **4:** 233–255.

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

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.

VAUGHN, K. C., 1981 Plastid fusion as an agent to arrest sorting out. Curr. Genet. **3:** 243–245.

WRIGHT, S., 1949 Genetics of populations. Encyclopaedia Britannica, 14th ed., **10:** 111–112.

Corresponding editor: B. S. WEIR