Models of mitochondrial DNA transmission genetics and evolution in higher eucaryotes

By ROBERT W. CHAPMAN*, J. CLAIBORNE STEPHENS,
ROBERT A. LANSMAN AND JOHN C. AVISE

*Department of Molecular and Population Genetics, University of Georgia, Athens,
Georgia 30602*

SUMMARY

The future value of mitochondrial DNA (mtDNA) sequence information to studies in population biology will depend in part on understanding of mtDNA transmission genetics both within cell lineages and between animal generations. A series of stochastic models has been constructed here based on various possibilities concerning this transmission. Several of the models generate predictions inconsistent with available data and, hence, their assumptions are provisionally rejected. Other models cannot yet be falsified. These latter models include assumptions that (1) mtDNA's are sorted through cellular lineages by random allocation to daughter cells in germ cell lineages; (2) the effective intracellular population sizes (n_M's) of mtDNA's are small; and (3) sperm may (or may not) provide a low level 'gene-flow' bridge between otherwise isolated female lineages. It is hoped that the models have helped to identify and will stimulate further empirical study of various parameters likely to strongly influence mtDNA evolution. In particular, critical experiments or measurements are needed to determine the effective sizes of mtDNA populations in germ (and somatic) cells and to examine possible paternal contributions to zygote mtDNA composition.

1. INTRODUCTION

In recent years restriction endonucleases and other molecular tools have been employed to study the transmission genetics and evolution of mitochondrial DNA (mtDNA) in higher animals (Gillham, 1978). Although the data are not yet extensive, three important generalizations have emerged from these empirical studies: (1) mtDNA appears maternally inherited (Upholt & Dawid, 1977; Dawid & Blackler, 1972; Avise *et al.* 1979*a, b*; Hutchinson *et al.* 1974); (2) the population of mtDNA molecules in all somatic cells of an individual organism appears to be homogeneous in nucleotide sequence. (For example, no within-animal sequence

* Present address: Chesapeake Bay Institute, The Johns Hopkins University, Baltimore, Maryland.

heterogeneity has been observed in mtDNA purified from livers of 87 gophers (Avise *et al.* 1979b), about 30 cows (Laipis *et al.* 1980, and personal communication), or from placentas of more than 80 humans (Brown, 1980, and personal communication). We are aware of only a single report of heterogeneity in mtDNA from a single animal (Coote, Szabados & Work, 1979), and this interpretation is complicated by the possibility that the mtDNA's from the samples examined were not digested to completion); (3) mtDNA sequence heterogeneity is extensive among individuals of a species and sometimes among individuals within a local inter-breeding population (Upholt & Dawid, 1977; Avise *et al.* 1979a, b; Brown & Wright, 1979).

The first two observations are subject to reservation, since existing methodologies are not sufficiently sensitive to detect minor heterogeneity if, in fact, it is present. Thus a paternal contribution of roughly 5% or less to progeny mtDNA, or the presence of a minor mtDNA in less than about 5% of an organism's total mtDNA complement, would normally remain undetected.

The dynamics of mtDNA genetic variation in animal populations are complicated by the fact that cells contain many mitochondria, and each mitochondrion may contain about 2–10 mtDNA molecules (Gillham, 1978; Potter *et al.* 1975). For example, mouse cells in culture contain about 1250 mtDNA molecules (Birky, 1978), and each mature *Xenopus* egg contains an estimated 10^8 mtDNA molecules (Dawid & Blackler, 1972). However, the number of mtDNA molecules in successive germ or somatic cell generations is not accurately known for any species and may be much smaller. Furthermore, packaging of mtDNA molecules, such that blocks of genetically identical molecules remain associated during cell division, would also reduce the effective mtDNA population size (Ohta, 1980). Thus the apparent dynamics of mtDNA evolution in a population of organisms is interrelated with and partially dependent upon the underlying dynamics of populations of mtDNA molecules within cell lineages of each animal.

Several studies have focused on organelle DNA transmission within cell lineages, and results for lower eucaryotes are cogently summarized in Birky (1978). In sharp contrast to the apparent situation in most higher eucaryotes, mtDNA in many lower eucaryotes is characterized by a relatively small number of mtDNA molecules per cell, and by the common formation of biparental zygotes. These conditions in lower eucaryotes have been modelled mathematically by Birky & Skavaril (1976) and Ohta (1980). The purpose of this paper is to formulate a conceptual framework for dealing with the mtDNA's within and among higher animals. The phenotypes whose dynamics are modelled in this paper are the mtDNA sequences, as for example might be assayed by restriction digests. Our primary approach will be to modify and adapt existing stochastic theories of population genetics to the problem. Since the magnitudes of some parameters important to mtDNA dynamics are unknown, even to a first approximation, a major goal is to develop an orientation to the study of mtDNA in higher eucaryotes which will identify important parameters and establish their probable limits.

2. THE MODELS

(i) *Biological Framework*

Fig. 1 diagrams the general framework for the models. The union of sperm and mature oocytes to form zygotes initiates the cycle. Depending on the model, sperm may or may not contribute mtDNA to the zygote. The percentage of mtDNA
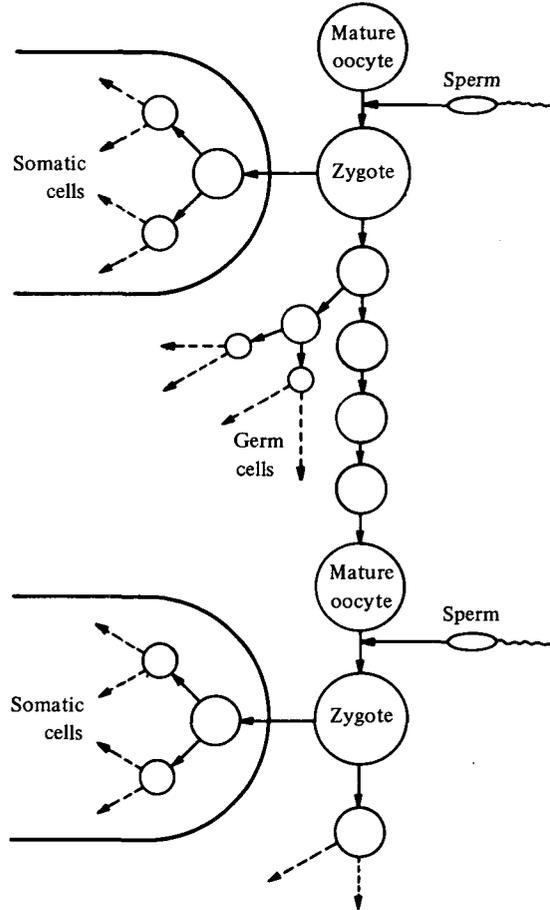

Fig. 1. General framework for mtDNA transmission.

molecules in the zygote which has been contributed by sperm describes the extent of 'paternal leakage', l. Two empirical points bear on the possible magnitude of paternal leakage, if it exists at all: (1) F_1 hybrids of vertebrate parents differing in mtDNA composition have invariably exhibited only the mtDNA of their mothers in tests that would have detected a paternal leakage of 5 % or more (Upholt & Dawid, 1977; Dawid & Blackler, 1972; Avise *et al.* 1979*a*); (2) mature oocytes typically contain thousands or millions of mtDNA molecules, while the mid-piece

of the sperm probably carries only about 100 copies (Gillham, 1978; Dawid & Blackler, 1972; Gresson, 1940). Therefore, the models below consider possible evolutionary consequences of 5% or less paternal leakage per generation.

The zygote gives rise to the somatic and germ cell lineages of each organism. We will focus explicity on mtDNA dynamics in germ cells, since only here will results be relevant to mtDNA evolution. We will make the reasonable assumption that empirical observations on somatic cell mtDNA sequence can be used to infer zygote mtDNA composition. Thus a zygote moderately heterogeneous in mtDNA composition should result either in heterogeneity among somatic tissues, or heterogeneity within somatic cell lineages, depending on whether or not mtDNA is segregated at cell division; and, conversely a homogeneous zygote mtDNA population should give rise to somatic cells homogeneous in mtDNA sequence (barring mutation).

In our models, for simplicity we employ the term 'germ cell generations' when describing mtDNA sequence dynamics. In reality, mtDNA molecules may often replicate independently of cell division (Bogenhagen & Clayton, 1976, 1977). Our models will remain applicable if the number of rounds of mtDNA replication and degradation is used instead of the actual number of cell divisions.

For sake of completeness, we will consider two theoretical possibilities concerning mode of transmission of mtDNA in germ cell lineages. This transmission is defined as 'perfect' (Model I) when random drift in germ cell lineages does not occur. Transmission is 'imperfect' (Model II) if successive germ cells receive some randomly drawn (both numerically and genetically) fraction of the mtDNA sequences present (Table 1).

(ii) *Model I*

It is biologically unrealisitic to assume that mtDNA transmission is 'perfect', but a brief verbal consideration of results of Model I (which neglects random drift at the cellular level) will serve as a useful introduction to more realistic scenarios (Model II).

If mitochondria are strictly maternally inherited, their among-individual population dynamics is the perfect analog of male 'surname transmission' in human societies and can be treated in identical mathematical fashion. Lotka (1931 a, b) used a branching process to investigate the ultimate survival probability (U.S.P.) of male surnames in the United States, and Fisher (1958), Haldane (1927), and Schaffer (1970) have extended and generalized his results. If numbers of surviving female progeny per adult female conform to a Poisson distribution with mean M, the U.S.P. $\simeq 2(M-1)$ (Schaffer, 1970). Exact outcomes for specific offspring distributions other than Poisson can be evaluated by a tedious, iterative branching process, but Bartlett (1966) has provided a useful approximation, U.S.P. $\simeq 2(M-1)/v$, where v is the variance in offspring. In either case, U.S.P. $\simeq 0$ for a non-expanding infinite population (U.S.P. $= 1/N$ for a finite population, where N is the number of females) and remains near zero even in situations where M is

slightly greater than one. The population will, therefore, ultimately be descendants of a single female and contain only her mtDNA complement.

Thus in relatively stable-sized animal populations, between-individual mtDNA heterogeneity will decay through time until eventually only a single lineage remains. If, for instance, we begin with each individual homoplasmic for a different

Table 1. *Possibilities concerning mtDNA transmission in natural populations*

Paternal Leakage l	Germ Cell Transmission Mechanics of mtDNA	
	Perfect	Imperfect
$l = 0$	Model I	Model II A
$0 < l < 0.05$	Model I	Model II B

The following assumptions apply to all models: (*a*) all mtDNA sequences are neutral with respect to fitness: (*b*) somatic cell mtDNA assayed in empirical studies can be used to infer mtDNA composition of zygote.
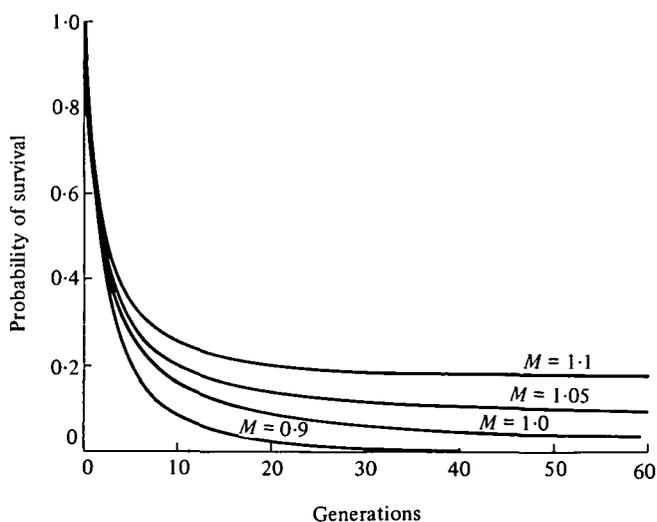

Fig. 2. mtDNA lineage survival probabilities when numbers of surviving progeny per female are Poisson distributed and M assumes various values near one.

sequence, the initial rate of decay can be rapid. With a Poisson distribution of surviving offspring and $M = 1$, about 37 % of mtDNA lineages is lost in the first generation, and the total proportion lost by generation t equals e^{x-1}, where x equals the proportion lost by $t-1$ (Spiess, 1977) (Fig. 2). In later generations the rate of loss slows greatly (Fig. 2) because there are only a few sequences left, in varying proportions.

However, as already noted, animal populations are often heterogenous in

mtDNA sequence, and each animal is apparently homoplasmic. Forces that might be invoked to maintain heterogeneity within a population include paternal leakage, mutation, and interpopulation migration. However, the last of these by itself is only effective for short-term production of heterogeneity, so we turn to the first two possibilities to explain the balance of intrapopulational heterogeneity with its loss over evolutionary time.

The initial effect of paternal leakage would be to shuffle mtDNA sequences already present in the population. In the absence of random drift (or selection) at the cellular level, the ultimate outcome would be an animal population each individual of which is heteroplasmic for the same fraction of those sequences initially present (or their recombinants, if any). Similarly, with 'perfect' transmission, the effect of mutation would be to increase within-individual mtDNA sequence heterogeneity through time. Thus under Model I, neither mutation, paternal leakage, nor a combination of the two can account for the empirical observation of within-individual mtDNA sequence homogeneity. These arguments can readily be given a firm mathematical treatment in terms of Markovian influence matrices (for finite animal populations) or branching processes (for effectively infinite populations) (Chapman, 1980 and unpublished results). However, in view of the lack of consistency with biological observations, it is inappropriate to present these mathematical treatments here, and we turn now to a more realistic approach.

(iii) *Model II*

Model II A. *Germ-line genetic drift, no paternal leakage.* In the foregoing, it was assumed that drift of mtDNA's in germ cells did not occur. No known cellular mechanisms insure that this will be the case. What is needed is an understanding of mtDNA population dynamics within germ cell lineages when drift is allowed.

If we assume that a given zygote is initially highly heterogeneous in mtDNA composition, the problem of the decay of this heterogeneity can be approached by analogy to the branching process in Model I. Now, it is the individual mtDNA molecules within the zygote that are transmitted to germ cells with mean number M and variance v. Thus in a non-expanding, effectively infinite population of mtDNA's within a germ cell lineage ($M \simeq 1$), heterogeneity decays at an initially rapid rate (Fig. 2). The abscissa in Fig. 2 is now interpreted as the number of germ cell generations. Thus if there are 50 germ cell generations per animal generation, roughly 95 % of the different mtDNA sequences initially present in the zygote will be unrepresented in a mature oocyte. For example, if the zygote initially contained 10 000 mtDNA molecules, each of unique sequence, about 500 different sequences would be expected per mature oocyte. Different oocytes stemming from such a zygote would, of course, exhibit very different arrays of mtDNA genotypes. Note, however, that in this treatment we have not allowed recombination of the many types of sequences, which of course would regenerate much if not all of the diversity lost each cell generation. While such a scheme requires an intriguing mathematical analysis, heterogeneous cells or tissues have yet to be observed in nature so we turn to the analysis of less extensive heterogeneity below.

Assuming that the original zygote contained a finite population of just two mtDNA sequences in frequencies p and $1\text{-}p$, and that the sampling variance in transmission from one germ cell generation to the next is binomial, mean times to loss or fixation (barring mutation or recombination) can be estimated by a diffusion approach employed by Watterson (1962) for nuclear genes (see also Crow & Kimura, 1970). Under these assumptions Ohta (1977) demonstrated that the probability of fixation of a mutant mtDNA sequence within a cell lineage is equivalent to that for fixation of a nuclear gene allele in a population. The mean time in cell generations to either fixation or loss is:

$$t_H\,(p) = -\,2n_M\,(p\,\ln\,p + (1-p)\,\ln\,(1-p)) \tag{1a}$$

while the mean time to fixation of an mtDNA sequence, conditional upon its going to fixation, is:

$$t_F\,(p) = -\,(1/p)\,2n_M(1-p)\ln\,(1-p), \tag{1b}$$

where n_M is the effective population size of mtDNA's, closely approximated by the harmonic mean number of mtDNA molecules in successive germ cell generations (Crow & Kimura, 1970).

For a new mutation in an otherwise homoplasmic cell, $1a$ and $1b$ become, approximately:

$$t_H\left(\frac{1}{n}\right) \approx\ 2 + 2\,\ln\,n_M, \tag{2a}$$

$$t_F\left(\frac{1}{n}\right) \approx\ 2n_M, \tag{2b}$$

where n is the actual number of molecules. Thus if $10 < n_M < 10^6$, then $6 < t_H < 30$, while t_F increases linearly with n_M. If we consider that there are roughly 50 germ cell generations per animal generation (Upholt & Dawid, 1977), the vast majority of mutations (Fig. 2) are lost within an animal generation. However, those that are destined to become fixed would be expected to be seen in heteroplasmic condition for several to many animal generations if n_M is much greater than 50. So far only one animal pedigree with segregating progeny is known (Laipis *et al.* 1980). In this case offspring, recently descended from a single female ancestor, are apparently homoplasmic for two types of mtDNA sequence. It would seem that a mutation was fixed in the oocytes leading to the aberrant progeny. Under our model, this is evidence that n_M is small. Clearly more mtDNA pedigrees of families with segregating progeny will be needed to estabish eucaryotic n_M size.

On the other hand, mutations must also be sufficiently frequent in animal populations to balance the continual decay of sequence heterogeneity due to drift at the level of organisms. We have conducted a series of computer simulations which exemplifies how this mutation/organismal-drift balance may be achieved.

Takahata & Maruyama (1981) provide theoretical estimates of the equilibrium levels of homogeneity within an oocyte or zygote as:

$$\hat{H} = 1/(1 + 2n_M\,\mu), \tag{3a}$$

and that within a local population as:

$$\hat{Q} = 1/[1 + (2N_e\, G + 2n_M)\, \mu],\tag{3b}$$

for a model similar to our II A. These estimates may be used to calculate asymptotic levels of:

$$H_1 = 1 - \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K}p_{ij}^2,\tag{4a}$$

and:

$$H_2 = 1 - \sum_{j=1}^{K}\bar{p}_j^2,\tag{4b}$$

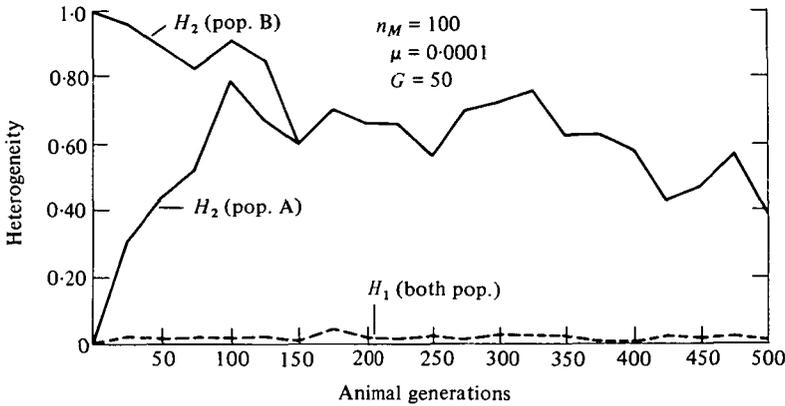the corresponding heterogeneity levels in a population of N females with K

Fig. 3. Time course of mtDNA sequence heterogeneity for population A (initially 500 individuals homoplasmic for the same sequence) and population B (initially 500 individuals each homoplasmic for a different sequence). H_1 and H_2 are calculated as in text, equations (4a) and (4b), respectively. H_1 indexes the intracellular heterogeneity, whereas H_2 indexes that within the animal population. n_M is the number of mtDNA molecules per cell, μ is the probability of mutation per mtDNA molecule per cell generation, and G is the number of germ cell generations per animal generation.

different sequences. In these formulas, p_{ij} is the frequency of the jth sequence in the ith individual, and $\bar{p}_j$ is the frequency of the jth sequence in the population. Thus H_1 measures intracellular heterogeneity in mtDNA sequence, and H_2 indexes inter-individual sequence heterogeneity.

Our simulations were conducted to test the sensitivity of H_1 and H_2 to several potentially important parameters. From Takahata & Maruyama (1981), it would appear that both H_1 and H_2 are sensitive to the effective mtDNA population size (n_M), and to the mutation rate per sequence per cell generation (μ), while H_2 is additionally influenced by the number of cell generations per animal generation (G), and the effective population size of females (N_e).

In our simulations we allow multinomial sampling of mtDNA molecules within germ cell lineages which are heteroplasmic. A cell lineage has a fixed number (n_M)

of mtDNA's each of which can mutate at prescribed rates (μ) each cell generation. After a given number of cell generations (G), each female organism produces female offspring according to a Poisson distribution with mean $= 1$. This cycle is repeated for any number of animal generations.
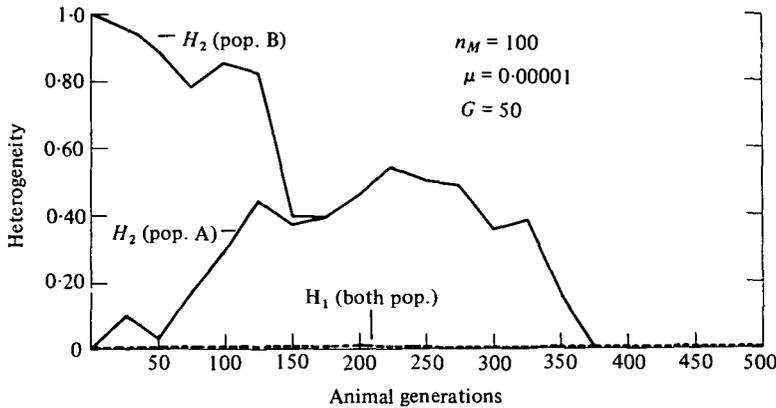

Fig. 4. Time course of mtDNA sequence heterogeneity. n_M, μ, G, H_1, and H_2 are as defined in Fig. 3.

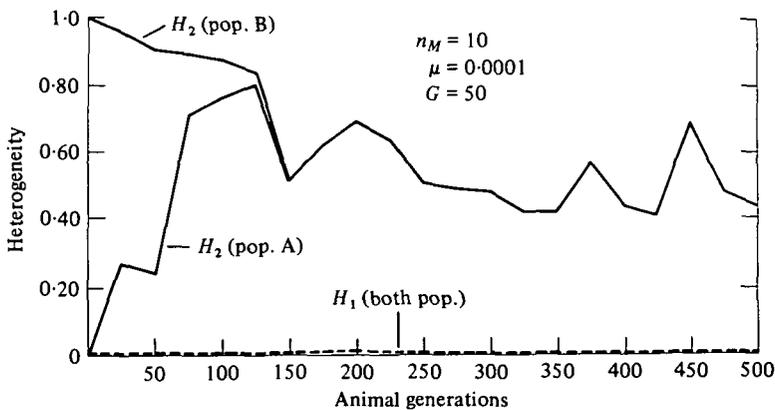

Fig. 5. Time course of mtDNA sequence heterogeneity. n_M, μ, G, H_1, and H_2 are as defined in Fig. 3.

In our models, n_M, μ, and G are fixed for any given simulation, but the actual population size of females proved to be highly variable. Therefore, in order to isolate the effects of changes independent of N_e, we used a single set of random numbers, allowing all stochastic 'decisions' (number of offspring, mutation, sorting of sequences) to be made using the same random number at the same time in each simulation. This has the effect of giving each simulation the same population size trajectory. Finally, for each set of input parameters, two populations started with 500 females were followed: population A was initiated with each

individual homoplasmic for an identical mtDNA sequence ($H_1 = H_2 = 0$); population B was initiated with each individual homoplasmic for an unique mtDNA sequence ($H_1 = 0$; $H_2 = 0.998$).

Several simulations were conducted and H_1 and H_2 were calculated after each animal generation. Fig. 3 plots results for the parameters $n_M = 100$, $\mu = 10^{-4}$, and $G = 50$. Fig. 4–6 plot results of additional simulations in which μ, n_M, and G, respectively, were lowered.
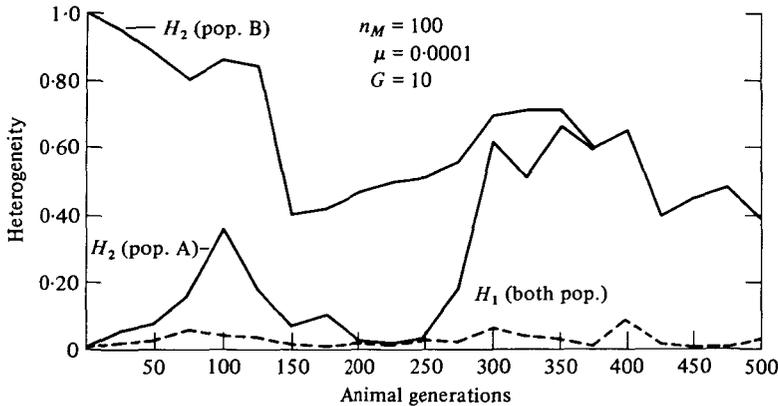

Fig. 6. Time course of mtDNA sequence heterogeneity. n_M, μ, G, H_1, and H_2 are as defined in Fig. 3.

In every case, populations A and B converge to a value of H_2 determined by the collapse of heterogeneity due to drift, balanced by the accumulation of variation by mutation. We can compare representative values of H_2, observed after the point of convergence, against the theoretical equilibrium estimates from equation (3b). Similarly, the equilibrium value of H_1 can be compared to the theoretical estimate from equation (3a). We estimated this value by averaging observed H_1's over at least the last 100 animal generations in the simulation.

For reasons which we do not fully understand, observed and predicted values of H_2 are often quite different (Table 2). One possibility is that N_e, the harmonic mean number of females, is not really appropriate for estimating H_2. This is indicated by the observation that while N_e stabilizes in the latter stages of the simulation, neither N nor H_2 do so. Another possibility is that the theoretical derivation of Takahata and Maruyama is insufficient, in that it appears to neglect the possibility of different oocytes originating from the same mother. By contrast, our simulations do yield H_1's which are in close agreement with the theoretical predictions (Table 2).

Qualitatively, it does appear that lowering μ lowers both H_1 and H_2 (Figs. 3 and 4) as, of course, would be expected. Lowering n_M lowers H_1 without influencing H_2 much (Figs. 3 and 5), while lowering G delays the time until convergence of H_2 in populations A and B with no apparent effect on H_1. Thus, for a given μ (and

Table 2. *Observed and expected levels* of mtDNA sequence heterogeneity within (H_1) and among (H_2) individuals, according to various combinations of effective mtDNA population size (n_M), mutation rate (μ), and germ cell generations (G) per animal generation (see text).*

Simula-tion no.	Simulation conditions		Animal generation....				H_1†	
			200	300	400	500		
		Female census size (N)...	75	276	68	99		
		Harmonic mean (N_e)...	162·57	134·49	139·72	105·00	Observed ±2σ	Expected
1	($n_M = 100, \mu = 10^{-4}, G = 50$)	Observed H_2	0·6592	0·7240	0·5767	0·3914	0·0173 ± 0·0206	0·0196
		Expected H_2	0·6220	0·5771	0·5863	0·5169		
2	($n_M = 100, \mu = 10^{-5}, G = 50$)	Observed H_2	0·4631	0·3556	0·0000	0·0016	0·0017 ± 0·0066	0·0020
		Expected H_2	0·1413	0·1201	0·1241	0·0967		
3	($n_M = 10, \mu = 10^{-4}, G = 50$)	Observed H_2	0·6865	0·7482	0·4433	0·4435	0·0018 ± 0·0058	0·0020
		Expected H_2	0·6194	0·5739	0·5832	0·5127		
4	($n_M = 100, \mu = 10^{-4}, G = 10$)	Observed H_2	—	—	0·6536	0·3939	0·0170 ± 0·0371	0·0196
		Expected H_2	—	—	0·2304	0·1870		

* Observed values are from our computer simulations using the various parameters shown. Expected values were calculated as 1-Ĥ, 1-Q̂, using the formulas (3a and 3b) derived by Takahata & Maruyama (1981).

† Both in our simulations, and in equation (3a) from Takahata & Maruyama (1981), H_1 does not appear to be a function of N_e. Thus each simulation has a single expected H_1.

N_e), the number of cell generations transpired will be the major factor determining times (measured in animal generations) required to attain H_2 equilibrium. What is surprising about Fig. 6 is that there is no apparent drop in equilibrium H_2 levels relative to Fig. 3, as would be predicted by equation (3b). While one expects more mutational events to occur in a cell lineage between mother and daughter when there are more intervening cell generations, it may be that this factor is greatly attenuated by genetic drift.

In any event, in contrast to Model I, the important outcome of these simulations is that biological observations (major inter-individual heterogeneity of mtDNA sequence and very minor intra-individual heterogeneity) can be reproduced with a reasonable model involving genetic drift of mtDNA at the cellular level.

Model IIB. *Germ-line genetic drift with paternal leakage.* For this model each female line may be thought of as a partially isolated mtDNA sub-population (island) which exchanges sequences with other subpopulations through paternal leakage (with effects analogous to migration). This migration is pulsed, occurring only once every animal generation or roughly once per 50 germ cell generations. Thus the effective migration will always be considerably less than l which is already known to be a small number at best.

For an island model of animal population structure, Crow & Kimura (1970) give a probability density function at equilibrium for gene frequencies in a two allele system under migration pressure. The analogous formula corresponding to paternal leakage of an alternative sequence may be written as:

$$\phi(x) = \frac{\Gamma(2n_M m)}{\Gamma(2n_M m x_i)\,\Gamma(2n_M m(1-x_i))}\, x^{2n_M m x_i-1}(1-x)^{2n_M m(1-x_i)-1} \tag{5}$$

where x is the equilibrium frequency of the maternal allele in a cell lineage, n_M is the effective number of molecules in the cell lineage, x_i is the frequency of the alternative allele in fathers, m is the effective migration pressure, and Γ is the gamma function. By analogy to Wright's well known result (1931), this equation predicts that the product of effective mtDNA population size and migration rate between cell lineages determines the tendency for cell lineages to become fixed for alternate mtDNA sequences, or conversely to become homogeneous for all sequences present.

For our purpose, we will consider two equally frequent mtDNA sequences, any female lineage initially fixed for one or the other sequence. Thus $x_i = 0.5$ can be considered the expectation of a father carrying either type. Fig. 7 illustrates the distribution curve at equilibrium, $y = \phi(x)$, for various values of $n_M m$ and $x_i = 0.5$. The shape of the curve changes from U-shape to bell-shape as $n_M m$ changes from smaller ($n_M m$ less than 1) to larger ($n_M m$ greater than 1). For $n_M m$ less than 1, the tendency is for sequences to be fixed within female lineages, and for $n_M m$ greater than 1, the tendency is for all female lineages to approach x_i. In other words, the latter situation leads to within-individual heterogeneity and between-individual homogeneity in mtDNA sequence, a result apparently inconsistent with empirical findings.

Nonetheless, these represent only general tendencies and even with $n_M m$ less than 1, many individuals may be detectably heteroplasmic. When the rarer of two sequences is detectable when present in 0·05 frequency or greater, the expected proportion of heteroplasmic individuals may be found by integrating equation (6) on (0·05, 0·95) and then dividing by the area contained on (0, 1). Table 3 gives some

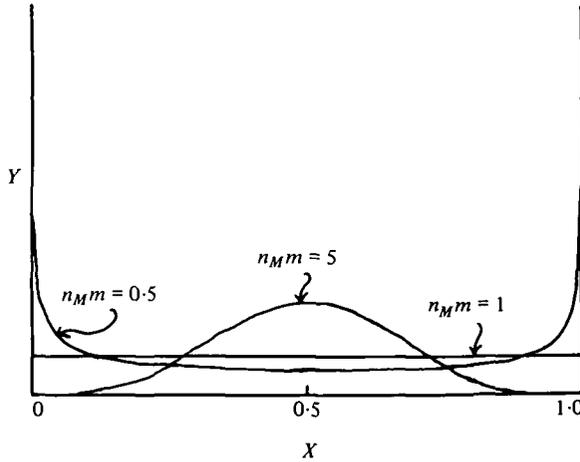Fig. 7. Probability distribution curves for various values of $n_M m$, assuming $x_i = 0.5$ (see text). The figure is adapted from Crow & Kimura (1970).

Table 3. *Expected proportion of female lineages detectably heteroplasmic for mtDNA sequence under Wright's island model (see text).*

$n_M m$	x_i		
	0·5	0·9 or 0·1	0·99 or 0·01
5	1·000	0·630	0·060
2	0·986	0·471	0·052
1	0·900	0·361	0·040
0·5	0·713	0·263	0·029
0·1	0·245	0·088	0·011
0·01	0·029	0·010	0·001
0·001	0·003	0·001	0·000
$k < 0.001$	$(2.94)k$	$(1.06)k$	$(0.117)k$

representative results for various values of $n_M m$ and x_i. Realistically then, paternal leakage, if it exists at all must be such that m is much smaller than the reciprocal of n_M.

So far we have treated the two sequences as though they do not recombine, so no new sequences are produced by paternal leakage. Recombination between parental inputs could indeed create new sequences which might ultimately be fixed, but again, extensive recombination of very different sequences would have a large effect on increasing the number of heteroplasmic individuals. The sorting of such

recombinants is vastly more complicated than the situation treated in Model II A. If recombination between parental mtDNA's is discovered in higher eukaryotes, this situation would then warrant further treatment.

3. RESULTS

The joint empirical observations of within-individual homogeneity and among-individual heterogeneity in mtDNA sequence, recently reported for many higher organisms, are at first thought surprising and counterintuitive. A major stimulus for this report is the need to identify models which can (and those which cannot) account for such empirical observations. We have examined several population genetic approaches which provide useful analogues for study of mtDNA trans-mission and evolution. Specific models which appear readily adaptable to the problem of mtDNA evolution include the 'surname transmission' approach of Lotka (1931 a, b), and the 'island model' of Wright (1943).

Because the predictions of Model I appear incompatible with empirical obser-vations, its tenets are provisionally rejected. On the contrary, various scenarios involving 'imperfect' mtDNA transmission can account for the empirical obser-vations. A primary condition in Model II A is that $n_M \mu$ is small such that either (1) newly arising mutations can go to fixation in a small or moderate number of cell generations or (2) μ is small enough so that the individual cell lineages segregating for two or more sequences will be rare. In this manner, most within-lineage heterogeneity will usually escape detection, while between-lineage heterogeneity accumulates nonetheless. From Model II B, it is also apparent that a moderate level of paternal leakage into the zygote mtDNA pool cannot yet be eliminated as a real possibility, providing $n_M m < 1$ (Fig. 7).

4. DISCUSSION

Little is known about the populations of mtDNA molecules within germ or somatic cell lineages of individuals. Some cells such as mature oocytes typically contain large populations (10^5 or more) of mtDNA molecules, while others such as sperm contain far fewer (10^2 or less). Transmission of mtDNA's between animal generations is also incompletely understood, although the great majority of offspring mtDNA molecules are clearly of the maternal type. Simple but general models have been advanced here which couple intraorganism mtDNA dynamics to problems of longer-term mtDNA evolution. The models consider various possibilities of mtDNA transmission across cell and animal generations (Table 1).

The recent empirical observations on mtDNA sequences in natural populations (within-individual homogeneity and between-individual heterogeneity) are con-sistent only with Model II and then only when $n_M \mu$ is small (or $n_M m < 1$ when $l \neq 0$). Because n_M is approximated by the harmonic mean number of molecules across germ cell generations, its value will be strongly influenced by fluctuations in the total population size of mtDNA and, in particular, by severe reductions or

'bottlenecks' in mtDNA numbers. The vast numbers of mtDNA molecules in mature oocytes may well be descendants of a small number of precursor molecules in earlier germ cell generations of the organism. Thus, it will be important to obtain empirical evidence of mtDNA numbers in intermediate germ cells.

Empirical studies have shown that the number of organelle segregation units seems to be smaller than the total number of cellular mtDNA molecules (c.f. Birky, 1978; Gillham, 1978). If organelles are transmitted largely intact and contain multiple copies of a single mtDNA sequence, then the models constructed here will remain applicable. The value of n_M will, however, be smaller since it applies to the number of segregational units rather than the number of mtDNA molecules. The problem then becomes how to account for homogeneous populations of mtDNA's within organelles. One possibility, raised by Birky & Skavaril (1976), is that of 'gene conversion' in which DNA's within an organelle undergo a process of heteroduplex formation and repair leading rapidly to homogeneity.

For simplicity, we have considered situations in which alternative mtDNA sequences are neutral with respect to fitness. It is certainly conceivable that different mtDNA sequences are strongly selected, but the selection pressure would have to be of a very peculiar nature if it were the only evolutionary force moulding mtDNA heterogeneity: it would somehow have to actively conserve intra-individual nucleotide sequence while still permitting between individual sequence heterogeneity to accumulate rapidly.

Also, we have dealt explicity with mtDNA sorting through germ cell lineages since only here will effects be directly relevant to patterns of mtDNA evolution. Sorting of sequences through somatic cells could also be treated with the models developed here, with the exception that generations will be limited, truncated by death of the individual. Michaelis (1967) has previously examined an alternative model to Model II A, in which plasmid DNA's were exactly duplicated prior to somatic cell division, and daughter somatic cells received equal portions of the parental populations. He concluded that only when n_M was 10 or less could a mixed population be converted to a pure population within 100 cell generations. The diffusion analysis (Model II A) suggests that the sorting may be even faster if there is variance in plasmid DNA replication and population size. This implies that if n_M is small, an initially heterogeneous zygote could give rise to cell lineages or tissues differing in mtDNA composition. Such differences among animal tissues or organs have not been observed, but such data are not extensive.

Finally, the results presented here are relevant to current discussions about whether mtDNA's are strictly maternally inherited. Evidence in favour of strict maternal inheritance comes from experimental crosses in which F_1 progeny exhibit only the maternal mtDNA genotype. However, if l is low, the paternal contribution could have remained undetected. Under a model of 'imperfect' mtDNA transmission between cell generations and $l \neq 0$, within-individual homogeneity and between-individual heterogeneity should be observed, if $n_M m$ is much less than 1. Thus, no available data are incompatible with the possibility of some paternal leakage. Since even small paternal contributions would have important evolu-

tionary consequences as a mechanism insuring gene flow between otherwise isolated female lineages (and as a potential mechanism for producing new sequences through recombination) more stringent experimental conditions to detect low levels of sequence heterogeneity are clearly needed.

REFERENCES

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

AVISE, J. C., GIBLIN-DAVIDSON, C., LAERM, J., PATTON, J. C. & LANSMAN, R. A. (1979b). Mitochondrial DNA clones and matriarchal phylogeny within and among geographic populations of the pocket gopher, *Geomys pinetis*. *Proceedings National Academy of Science U.S.A.* **76**, 6694–6698.

BARTLETT, M. S. (1966). *An Introduction to Stochastic Processes with Special Reference to Methods and Applications*. 2nd edition, 291 pp. Cambridge University Press.

BIRKY, C. W., JR. (1978). Transmission genetics of mitochondria and chloroplasts. *Annual Review Genetics* **12**, 471–512.

BIRKY, C. W. & SKAVARIL, R. V. (1976). Maintenance of genetic homogeneity in systems with multiple genomes. *Genetical Research* **27**, 249–265.

BOGENHAGEN, D. & CLAYTON, D. A. (1976). Thymidylate nucleotide supply for mitochondrial DNA synthesis in mouse L-cells. *Journal of Biological Chemistry* **251**, 2938–2944.

BOGENHAGEN, D. & CLAYTON, D. A. (1977). Mouse L-cell mitochondrial DNA molecules are selected randomly for replication throughout the cell cycle. *Cell* **11**, 719–727.

BROWN, W. M. (1980). Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. *Proceedings National Academy of Science U.S.A.* **77**, 3605–3609.

BROWN, W. M. & WRIGHT, J. W. (1979). Mitochondrial DNA analyses and the origin and relative age of parthenogenetic lizards (genus *Cnemidophorus*). *Science* **203**, 1247–1249.

CHAPMAN, R. W. (1980). *Selected Topics in Evolution Theory*. Ph.D. dissertation, 83 pp. University of Georgia.

COOTE, J. L., SZABADOS, G. & WORK, T. S. (1979). The heterogeneity of mitochondrial DNA in different tissues from the same animal. *FEBS Letters* **99**, 255–260.

CROW, J. F. & KIMURA, M. (1970). *An Introduction to Population Genetics Theory*, 591 pp. New York: Harper and Row.

DAWID, I. B. & BLACKLER, A. W. (1972). Maternal and cytoplasmic inheritance of mtDNA in *Xenopus*. *Developmental Biology* **29**, 152–161.

FISHER, R. A. (1958). *The Genetical Theory of Natural Selection*, 362 pp. New York: Dover Publishers.

GILLHAM, N. W. (1978). *Organelle Heredity*, 602 pp. New York: Raven Press.

GRESSON, R. A. R. (1940). Presence of the sperm middle-piece in the fertilized egg of the mouse (*Mus musculus*). *Nature* **145**, 425.

HALDANE, J. B. S. (1927). A mathematical theory of natural and artifical selection. V. Selection and mutation. *Proceedings of the Cambridge Philosophical Society* **23**, 838–844.

HUTCHINSON, C. A., III, NEWBOLD, J. E., POTTER, S. S. & EDGELL, M. H. (1974). Maternal inheritance of mammalian mitochondrial DNA. *Nature* **251**, 536–538.

LAIPIS, P. J., HAUSWIRTH, W. W., O'BRIEN, T. W. & MICHAELS, G. S. (1980). A physical map of bovine mitochondrial DNA from a single animal. *Biochemica Biophysica Acta* **565**, 22–32.

LOTKA, A. J. (1931a). Population analysis – the extinction of families I. *Journal of the Washington Academy of Science* **21**, 377–380.

LOTKA, A. J. (1931b). Population analysis – the extinction of families II. *Journal of the Washington Academy of Science* **21**, 453–459.

MICHAELIS, P. (1967). The investigation of plasmone segregation by pattern analysis. *The Nucleus* **10**, 1–14.

OHTA, T. (1977). On the gene conversion model as a mechanism for maintenance of homogeneity in systems with multiple genomes. *Genetical Research* **30**, 89–91.

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

POTTER, S. S., NEWBOLD, J. E., HUTCHINSON, C. A., III. & EDGELL, M. H. (1975). Specific cleavage analysis of mammalian mitochondrial DNA. *Proceedings of the National Academy of Science U.S.A.* **72**, 4496–4500.

SCHAFFER, H. (1970). The fate of neutral mutants as a branching process, *in Mathematical Topics in Population Genetics* (ed. K. Kojima), pp. 317–336. New York: Springer-Verlag.

SPIESS, E. B. (1977). *Genes in Populations*, 780 pp. New York: John Wiley.

TAKAHATA, N. & MARUYAMA, T. (1981). A mathematical model of extranuclear genes and the genetic variability maintained in a finite population. *Genetical Research* **37**, 291–302.

UPHOLT, W. B. & DAWID, I. B. (1977). Mapping of mitochondrial DNA of individual sheep and goats: rapid evolution in the *D* loop region. *Cell* **11**, 571–583.

WATTERSON, G. A. (1962). Some theoretical aspects of diffusion theory in population genetics. *Annals of Mathematical Statistics* **33**, 939–957.

WRIGHT, S. (1931). Evolution in Mendelian populations. *Genetics* **16**, 97–159.

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