J Mol Evol (1984) 20:99–105

Journal of Molecular Evolution

Demographic Influences on Mitochondrial DNA Lineage Survivorship in Animal Populations

John C. Avise, Joseph E. Neigel, and Jonathan Arnold

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

Summary. Probability models of branching processes and computer simulations of these models are used to examine stochastic survivorship of female lineages under a variety of demographic scenarios. A parameter Π, defined as the probability of survival of two or more independent lineages over G generations, is monitored as a function of founding size of a population, population size at carrying capacity, and the frequency distributions of surviving progeny.

Stochastic lineage extinction can be very rapid under certain biologically plausible demographic conditions. For stable-sized populations initiated by n females and/or regulated about carrying capacity $k = n$, it is highly probable that within about 4n generations all descendants will trace their ancestries to a single founder female. For a given mean family size, increased variance decreases lineage survivorship. In expanding populations, however, lineage extinction is dramatically slowed, and the final k value is a far more important determinant of Π than is the size of the population at founding. The results are discussed in the context of recent empirical observations of low mitochondrial DNA (mtDNA) sequence heterogeneity in humans and expected distributions of asexually transmitted traits among sexually reproducing species.

Key words: Branching process — Mitochondrial DNA evolution — Lineage extinction — Human populations

Offprint requests to: J.C. Avise

Introduction

Polymorphism in mitochondrial DNA (mtDNA) is being increasingly exploited for the study of genetic relationships among closely related organisms (Brown et al. 1979; Ferris et al. 1981; Adams and Rothman 1982; Aquadro and Greenberg 1983; Avise et al. 1983; Brown 1983; Lansman et al. 1983a; Powell 1983; Templeton 1983). Because mitochondria appear to be strictly maternally inherited in many higher animals (Avise and Lansman 1983; Lansman et al. 1983b), mtDNA-generated evolutionary trees are interpreted as estimates of matriarchal phylogeny (Avise et al. 1979). If the survivals and extinctions of mtDNA lineages during evolution do indeed represent a sorting of asexually transmitted traits in otherwise sexually reproducing species, they can be formally modeled in analogous fashion to "male surname evolution" in many human societies (Lotka 1931a, b; Chapman et al. 1982). In general, the particular demographies of female populations should significantly influence the evolutionary dynamics of mtDNA lineages.

This brief article addresses the question, "How far back in time might pairs of extant organisms have last shared a common female parent?" We employ probability models (such as those in Harris, 1963) and computer simulations to examine stochastic lineage survivorship under various demographic scenarios. Although the approach and results should have general applicability, we will focus our discussion on a previously published conclusion that specifically motivated this report. From an analysis of mtDNA-sequence diversity among humans of diverse racial and geographic origin, Brown

(1980) concluded that "the amount of sequence heterogeneity observed, 0.18%, could have been generated from a single mating pair that existed 180–360 × 10^3 years ago, suggesting the possibility that present-day humans evolved from a small mitochondrially monomorphic population that existed at that time." Is this a plausible or necessary scenario? How small need that population have been to be compatible with the empirical observations on human mtDNA?

Methods and Results

Density-Unregulated Model

Our models involve only a slight modification of the general "branching process" approach used to study surname dynamics in human populations and evolutionary fates of newly arising mutants (Haldane 1927; Fisher 1930; Lotka 1931a, b; Ross 1970; Schaffer 1970; Li and Nei 1977; Spiess 1977). For a given branching process, a specified distribution of family sizes determines probabilities of loss or survival of a given lineage. For example, when adult females produce daughters according to a Poisson distribution with mean μ, the probability of loss of a given female lineage after one nonoverlapping generation (or the proportion of lineages lost from the population) is $e^{-\mu}$ (Spiess 1977, chapter 13). Through the use of generating functions, probabilities of loss after G generations (p_G) can also be determined (Li 1955; Crow and Kimura 1970; Spiess 1977). In the Poisson case, $p_G = e^{\mu(x-1)}$, where x equals the probability of loss in the previous generation (p_{G-1}).

If survival and extinction of different lineages occur independently of one another, as would be true for a density-unregulated population, the probability of various numbers of lineages surviving through G generations can be obtained from the approach exemplified in Table 1. Suppose that a population is initiated with n unrelated females. The first and second terms of the binomial expansion are the probabilities that all except zero or one lineage, respectively, will be extinct after G generations. The sum of all other terms of the expansion (Π, the probability of survival of two or more lineages) is of special interest. If Π is near one, it is very likely that a population founded G generations earlier will still carry descendants of two or more original founders. Conversely, if Π is near zero, all mtDNA-sequence diversity in the population or species will almost certainly have arisen less than

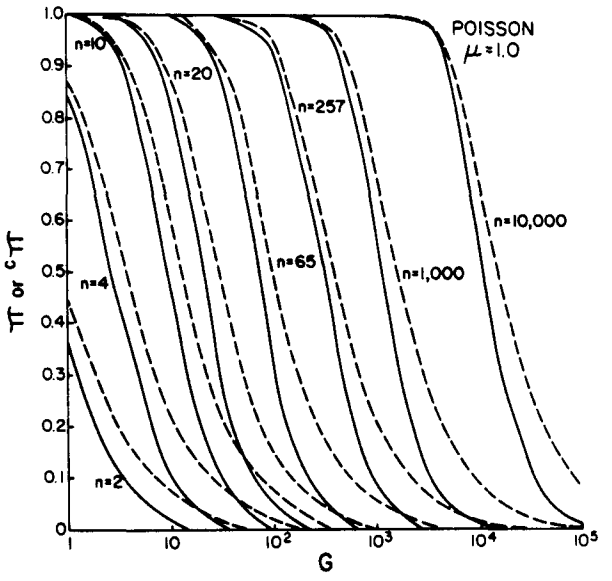

Fig. 1. Probabilities of survival of two or more founding lineages through G generations in populations initiated with n females producing daughters according to a Poisson distribution with mean $\mu = 1.0$. Solid lines, Π; dashed lines, ᶜΠ

Table 1. Probabilities of lineage survival through G generations in a population initiated with n = 4 females, each producing offspring according to a Poisson distribution with mean $\mu = 1$

			Binomial expansion			
G	p	q	p^4	$4p^3q$	Sum of other terms (Π)	ᶜΠ
1	0.3679	0.6321	0.0183	0.1259	0.8558	0.8718
2	0.5315	0.4685	0.0798	0.2814	0.6338	0.6925
5	0.7319	0.2681	0.2869	0.4205	0.2926	0.4103
10	0.8417	0.1583	0.5019	0.3775	0.1206	0.2421
20	0.9125	0.0875	0.6933	0.2659	0.0408	0.1330
100	0.9807	0.0193	0.9250	0.0728	0.0022	0.0293
General	$p_G = e^{\mu(p_{G-1}-1)}$	$1-p$	p^n (extinct)	$np^{n-1}q$	$1-p^n-np^{n-1}q$ (extant)	$1-\left(\frac{np^{n-1}q}{1-p^n}\right)$

p and q are probabilities of extinction and survival, respectively, of single lineages. The first and second terms of the binomial expansion are the probabilities that zero or one lineage, respectively, survive to generation G

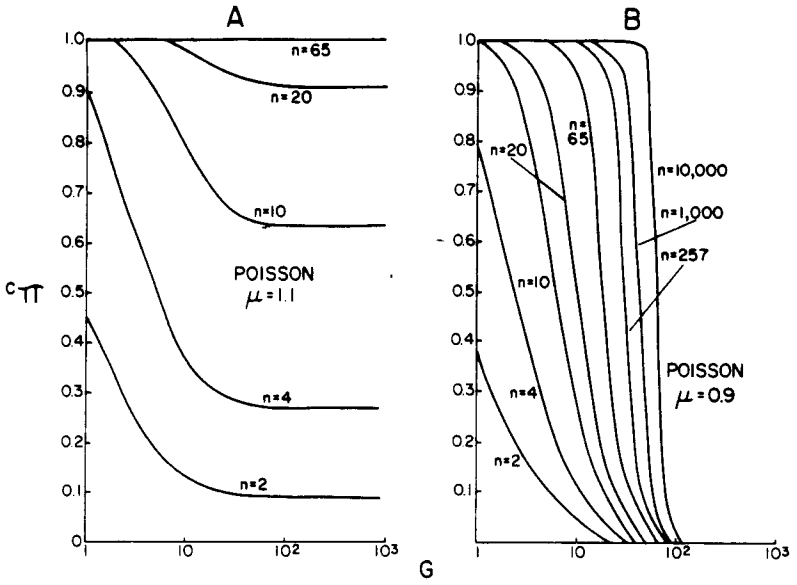

Fig. 2A and B. Probabilities of survival of two or more founding lineages through G generations in populations initiated with n females producing daughters according to a Poisson distribution with means **A** 1.1 and **B** 0.9

Table 2. Examples of generating functions $p_G(x)$ for various family-size distributions[a]

Parametric family of distributions	Mean (μ)	Variance (v)	Generating function $[p_G(x)]$
Poisson	μ	μ	$e^{\mu(x-1)}$
Binomial	np	npq	$(q + px)^n$
Geometric	$\frac{q}{p}$	$\frac{q}{p^2}$	$\frac{p}{1 - qx}$
Negative binomial	$\frac{rq}{p}$	$\frac{rq}{p^2}$	$\left(\frac{p}{1 - qx}\right)^r$
Logarithmic $\left(\alpha = \frac{-1}{\log(1 - q)}\right)$	$\frac{-\alpha q}{p}$	$\frac{-\alpha q(1 + \alpha q)}{p^2}$	$1 + \alpha \log\left[1 - \frac{q}{p}(x - 1)\right]$

[a] The generating function is used to calculate recursively the probability of loss in each generation. Let x equal the probability of loss in the previous generation (G − 1). Then $p_G(x)$ is the probability of loss in generation G. The probability of loss in the first generation is $p_1(0)$

G generations prior. We generated values of Π for large n on a Digital PDP-11/34A computer, using the procedure shown in Table 1. Several weeks of computer time were required for the following models and simulations.

Figure 1 plots values of Π against G for populations founded by n = 2–10,000 females and a Poisson distribution of offspring with $\mu = 1$. Under these conditions stochastic lineage sorting can be very rapid. Consider, for example, the case for n = 65. If G > 200, Π is less than 0.1, and it is thus very likely that this population would now contain only the descendants of a single founder female. On the other hand, if G < 40, then Π > 0.9, and it is very likely that the descendants of at least two founding females are still represented among the surviving lineages. In general, for the range of n examined, Π > 0.9 only for G < ~0.5n, and Π < 0.1 only when G > ~4n.

We can also define a conditional probability $^c\Pi$ (last column, Table 1) that two or more lineages survive through G generations given that the population remains extant. Curves for Π and $^c\Pi$ are generally very similar at higher values (Π and $^c\Pi$ > 0.5), but diverge increasingly at lower values (Fig. 1).

In this density-independent model with Poisson offspring distribution of $\mu = 1$, most populations founded by small numbers of females will go extinct within a few generations (column of p^4 values, Table 1). For example, about 92% of populations with founder size n = 4 are likely to expire within 100 generations (Table 1), whereas only about 1% of populations with n = 250 will be lost within the same time span. However, the probabilities of population extinction and the Π curves will of course also be strongly influenced by the family-size distributions, which determine the population-size trajectories. Figures 2A and 2B plot $^c\Pi$ curves for Poisson offspring distributions with means $\mu = 1.1$ and

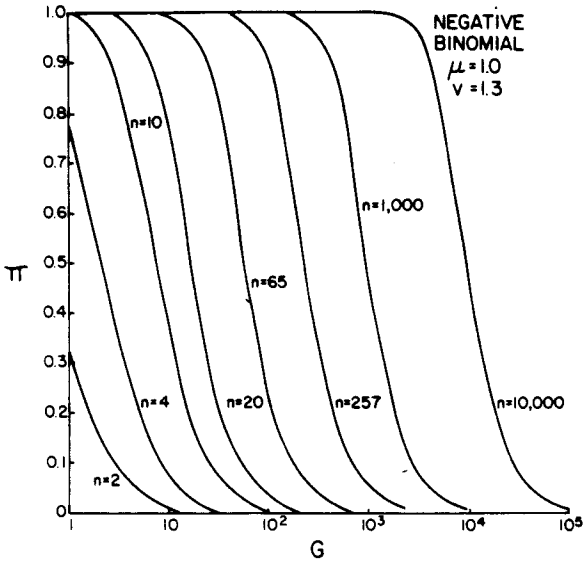

Fig. 3. Probabilities of survival of two or more founding lineages through G generations in populations initiated with n females producing daughters according to a negative binomial distribution with $\mu = 1.0$ and v = 1.3

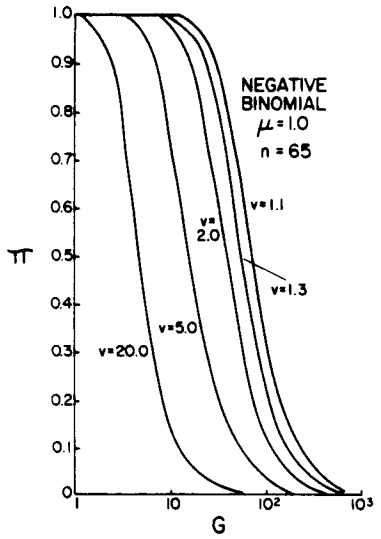

Fig. 4. Probabilities of survival of two or more founding lineages through G generations in populations initiated with 65 females producing daughters according to a negative binomial distribution with $\mu = 1.0$ and variances ranging from 1.1 to 20.0

0.9, respectively. In the biologically unrealistic situation in which females continually produce excess surviving progeny ($\mu > 1.0$, indefinitely expanding population), $^{c}\Pi$ values remain positive for any n value (Fig. 2A). When females fail to fully replace themselves with daughters ($\mu < 1.0$, declining population), $^{c}\Pi$ values quickly go to zero even for large founding n value (Fig. 2B).

The Poisson distribution is a model for counts of rare events. Some organisms (perhaps many fishes, amphibians, etc.) may maintain relatively constant population size by producing large numbers of offspring, of which only a small fraction survives. Considering survival as a rare event, progeny numbers in such populations may follow a Poisson distribution with $\mu = 1.0$. Other progeny distributions may be more realistic elsewhere. For example, Kojima and Kelleher (1962) reported that the 1950 census data for the growing U.S. population conforms well to a negative binomial distribution. The approach exemplified in Table 1 and Fig. 1 may be generalized to any offspring distribution, provided appropriate generating functions are available for recursively calculating p_G (Table 2). Figure 3 shows curves of Π for a negative binomial offspring distribution with $\mu = 1.0$ and variance v = 1.3. These curves are very similar in form to those for the Poisson situation (Fig. 1).

One would expect the variance as well as the mean of the progeny distribution to influence lineage survivorship (Schaffer 1970), and Fig. 4 shows that this is indeed true. For a negative binomial distribution with a given μ and n (in this case 1.0 and 65, respectively), probabilities of lineage survival through time decrease as v increases, with overall effects on Π equivalent to those of lowering n. For example, the Π curve for v = 5.0 and n = 65 is nearly identical to that for v = 1.3 and n = 20 (compare Figs. 4 and 3).

Density-Regulated Models

In the foregoing models, because the dynamics of different lineages are assumed to be independent, most populations initiated with small or moderate n values rapidly go extinct, and the only regulation of population size is through the constancy of the mean μ. In the following computer simulations (by J.E.N.), population size is explicitly regulated through time about some specified carrying capacity k; the mean μ changes each generation and depends on n logistically. The simulations were done with a DEC system random-number generator tested by several criteria.

In these simulations each population is founded by n = k unrelated females. In the first generation these females are allowed to produce progeny according to a Poisson distribution with $\mu = 1.0$. By chance, total numbers of daughters will often differ from k, but in subsequent generations population size is buffered as follows: For any monitored n, the mean (and variance) in number of offspring per female for the following generation is calculated by $\mu = e^{(k-n)/k}$. Thus for $n < k$, $\mu > 1.0$, and the population grows logistically until $n > k$ ($\mu < 1.0$), at which point the population temporarily declines. Throughout the simulations the female lineages initiated by the founders are labeled, their fates are

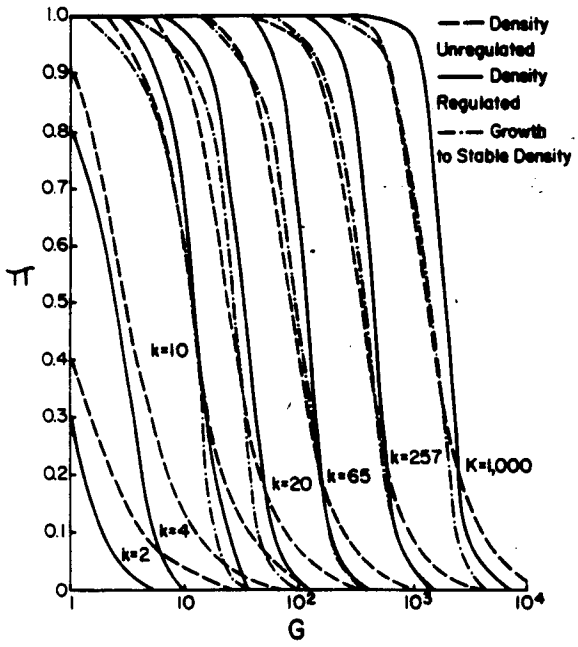

Fig. 5. Solid lines, Π curves for populations regulated at carrying capacities k; dashed lines, Π curves for density-unregulated populations initiated with n female founders reproducing according to a Poisson distribution with $\mu = 1.0$; dashed/dotted lines, Π curves for populations that have grown logistically to stable carrying capacity k after founding by n = 4 females

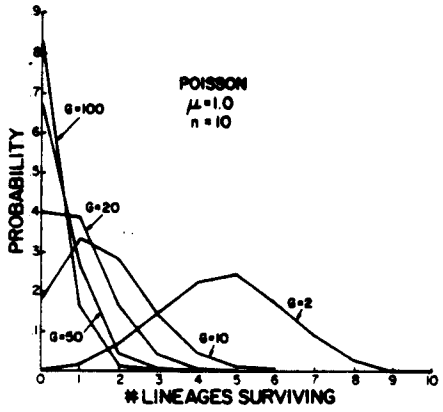

Fig. 6. Frequency distributions of probabilities of lineage survival through G generations for populations initiated with 10 females producing daughters according to a Poisson distribution with $\mu = 1.0$

followed, and the time course of Π is determined. For each specified k value a total of 100 simulations was run; the Π values represent the proportion of times among these runs that two or more lineages remained extant through G generations.

Results for computer-simulated populations with k = 2–1000 are plotted in Fig. 5. For comparison the Π curves from the previous density-unregulated models are also shown. (In the density-regulated case, curves of $^{c}\Pi$ are not shown because they are very similar to those of Π, particularly for large k values.) For a given n = k, the two sets of curves are very similar in form and magnitude. This suggests that the results for density-unregulated populations founded by n females are applicable, to a first approximation, to the more realistic situation in which population size is explicitly regulated around k = n.

However, regulation by density may be important in a species composed of many spatially isolated populations each of which is buffered against extinction through density effects on population growth. Although the diversity of lineages *within* each population may decay rapidly (depending on the demographic parameters, as already discussed), at least one lineage per extant population will be retained indefinitely. The maximum age of separation of lineages within a species composed of isolated populations could be no less than the age of the populations themselves, and could be much older. The maximum age would be strongly dependent on the lineages introduced into the populations at their time of founding, which in turn would be a function of the sorting of lineages in the ancestral stock (as already modeled). In such cases the Π value for the species as a whole could remain larger than a consideration based solely on composite census size would suggest.

Extensions of the Models and Simulations

Further information can be extracted from the models. For example, the probability that any number of founding lineages remain extant through G generations can readily be calculated from additional terms of the binomial expansion (Table 1). Figure 6 shows such an example for n = 10 and G values ranging from 2 to 100.

The simulations can also be modified to accommodate temporally changing demographic conditions. One situation that is probably common in nature involves a small founding population that subsequently grows to a much higher carrying capacity. Results of a number of computer simulations of such a situation are shown in Fig. 5. Artificial populations were founded by n = 4 females and allowed to grow logistically ($\mu = e^{(k - n)/k}$) to carrying capacities ranging from 10–1000. (Mean times to k, measured in generations, ranged from 3.6 for k = 10 to 10.2 for k = 1000.) The Π curves for these conditions appear nearly identical to the respective curves for the previously discussed density-unregulated models and density-regulated simulations with the same n and k values. Evidently, for a population that expands rapidly after establishment, the final k value is a far more important determinant of Π than is the founding n value. This makes intuitive sense, because during the expansion phase

most mtDNA lineages survive (see Fig. 2), and once k is reached, the rate of loss of lineages becomes lower as the absolute number of surviving lineages declines.

Discussion

The major conclusion to be derived from these models and simulations is that stochastic extinction of mtDNA lineages can be very rapid under certain plausible demographic conditions. For nonexpanding populations initiated by n females and/or regulated about a carrying capacity k = n, it is highly probable that within about 4n generations all descendants will trace their ancestries to a single female founder. The probability that two or more founding lineages remain extant (Π) is very high only when the number of generations (G) since founding is less than about 0.5n. However, the plots of Π versus time are strongly influenced by the particular means and variances of family-size distributions and by the attendant population-size fluctuations. During the expansion phase of a population, the probability Π consistently remains much greater than zero; during a contraction phase lineage extinction is especially rapid, and Π quickly diminishes.

Although treatments using branching processes and generating functions have been used previously to examine related population problems (such as mean times to loss or fixation of newly arising mutants), we have introduced the probability Π because of its clearer relevance to two issues recently raised about mtDNA evolution: the rate of sorting of female lineages in ancestors of modern humans; and the distribution of mtDNA lineages among related species.

Human mtDNA Evolution

The observation of exceptionally low mtDNA-sequence diversity within and among human races led Brown (1980) to speculate that all living humans may have descended from a single mating pair that lived between 180,000 and 360,000 years ago (about 9000–18,000 generations). Brown went on to postulate a severe bottleneck in human population numbers—a "small mitochondrially monomorphic population." This may have occurred, but the results of our models indicate that a dramatic population reduction does not necessarily have to be invoked to account for the human mtDNA data, even if all living humans have indeed descended from a recent female parent.

The branching process approach shows that with a Poisson offspring distribution of mean 1.0, a population founded by n = 15,000 *unrelated* females will yield a value of $\Pi \approx 0.5$ within 18,000 generations. (For a negative binomial offspring distribution with mean 1.0 and variance 3.0, a population of n = 45,000 yields $\Pi \approx 0.5$ in the same time.) In other words, under these circumstances our "Eve" could have belonged to a population of many thousands or tens of thousands of females, the remainder of whom left no descendants to the present day, due simply to the stochastic lineage extinction associated with reproduction. The absolute population size could have been even larger than this if the variance in numbers of surviving progeny was greater.

Clearly the human population has expanded dramatically since the Pleistocene. However, prior to the agricultural revolution that began about 10,000 years ago, the human population was much smaller, and conceivably relatively stable for long periods of time (Ehrlich and Ehrlich 1970; Cavalli-Sforza and Bodmer 1971). Deevey (1960) and Ehrlich and Ehrlich (1970) suggest a mean population size of about 125,000 (perhaps 62,000 females) for the Lower Paleolithic (roughly 1 million to 300,000 years ago). The mtDNA data and results of the branching process models appear compatible with a population of this general order of magnitude for much of the Middle Paleolithic as well. Probably it was only after the more recent expansion of the human population that the rate of loss of mtDNA lineages decreased dramatically.

mtDNA Lineages in Related Species

The demographically influenced rates of mtDNA-lineage sorting also have important consequences for expected patterns of mtDNA relationships among closely related species. We plan a fuller treatment of this topic elsewhere, and will limit comments here to a few general thoughts.

If more than about 4n generations have elapsed since a given *stable-sized* species last split from its nearest relatives, all mtDNA-sequence divergence within it will likely postdate the speciation event, and the species will be monophyletic in its matriarchal ancestry. On the other hand, in a derivative species less than 0.5n generations old (or for a species whose population has expanded since its origin), some mtDNA lineages will almost certainly predate the separation from ancestral stock. In comparison with other extant taxa, the possibility exists for individuals or populations within this species to be polyphyletic or paraphyletic (Farris 1974; Wiley 1981). Avise et al. (1983) presented empirical evidence for just such a situation. The rodent *Peromyscus polionotus* inhibits the southeastern United States, while its close relative *P. maniculatus* occurs throughout most of North America. Restriction enzyme digestion of mtDNA showed that several high-

ly differentiated lineages were present in *P. maniculatus*, and importantly, with respect to *P. polionotus*, *P. maniculatus* appeared paraphyletic in matriarchal genealogy.

Because mtDNA evolves rapidly, it has proven to be most useful for assessing relationships among conspecific populations and recently diverged species. It is in these recently separated species that patterns of polyphyly or paraphyly in mtDNA are most likely. In conventional phylogenetic scenarios, speciations are usually pictured as dichotomous splits leading to monophyletic species. Considerations of the demographic events associated with speciations suggest that with respect to uniparentally inherited traits such as mtDNA, these scenarios may often be incorrect.

Acknowledgments. This work was supported by NSF grant BSR-8217291, by the Biomedical Research Support Grant Program BRSG S07 RR07025-17 of the NIH, and by the U.S. Army Research Office under training grant DAAG29-83-6-0111. The computer was made available to our department as a gift from the Digital Equipment Corporation.

References

Adams J, Rothman ED (1982) The estimation of phylogenetic relationships from DNA restriction patterns and selection of endonuclease cleavage sites. Proc Natl Acad Sci USA 79: 3560–3564

Aquadro CF, Greenberg BD (1983) Human mitochondrial DNA variation and evolution. Analysis of nucleotide sequences from seven individuals. Genetics 103:287–312

Avise JC, Lansman RA (1983) Polymorphism of mitochondrial DNA in populations of higher animals. In: Nei M, Koehn RK (eds) Evolution of genes and proteins. Sinauer, Sunderland, Massachusetts, p. 147

Avise JC, Giblin-Davidson C, Laerm J, Patton JC, Lansman RA (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

Avise JC, Shapira JF, Daniel SW, Aquadro CF, Lansman RA (1983) Mitochondrial DNA differentiation during the speciation process in *Peromyscus*. Mol Biol Evol 1:38–56

Brown WM (1980) Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. Proc Natl Acad Sci USA 77:3605–3609

Brown WM (1983) Evolution of animal mitochondrial DNA. In: Nei M, Koehn RK (eds) Evolution of genes and proteins. Sinauer, Sunderland, Massachusetts, p 62

Brown WM, George M Jr, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. Proc Natl Acad Sci USA 76: 1967–1971

Cavalli-Sforza LL, Bodmer WF (1971) The genetics of human populations. WH Freeman, San Francisco

Chapman RW, Stephens JC, Lansman RA, Avise JC (1982) Models of mitochondrial DNA transmission genetics and evolution in higher eucaryotes. Genet Res 40:41–57

Crow JF, Kimura M (1970) An introduction to population genetics theory. Harper & Row, New York

Deevey ES Jr (1960) The human population. Sci Am 203:194–204

Ehrlich PR, Ehrlich AH (1970) Population, resources, environment. WH Freeman, San Francisco

Farris JS (1974) Formal definitions of paraphyly and polyphyly. Syst Zool 23:548–554

Ferris SD, Wilson AC, Brown WM (1981) Evolutionary tree for apes and humans based on cleavage maps of mitochondrial DNA. Proc Natl Acad Sci USA 78:2432–2436

Fisher RA (1930) The genetical theory of natural selection. Clarendon Press, Oxford

Haldane JBS (1927) A mathematical theory of natural and artificial selection V. Selection and mutation. Proc Cambridge Philos Soc 23:838–844

Harris T (1963) The theory of branching processes. Springer-Verlag, Berlin

Kojima K-I, Kelleher TM (1962) Survival of mutant genes. Am Nat 96:329–346

Lansman RA, Avise JC, Huettel MD (1983a) Critical experimental test of the possibility of "paternal leakage" of mitochondrial DNA. Proc Natl Acad Sci USA 80:1969–1971

Lansman RA, Avise JC, Aquadro CF, Shapira JF, Daniel SW (1983b) Extensive genetic variation in mitochondrial DNAs among geographic populations of the deer mouse, *Peromyscus maniculatus*. Evolution 37:1–16

Li CC (1955) Population genetics. University of Chicago Press, Chicago

Li W-H, Nei M (1977) Persistence of common alleles in two related populations or species. Genetics 86:901–914

Lotka AJ (1931a) Population analysis—the extinction of families. I. J Wash Acad Sci 21:377–380

Lotka AJ (1931b) Population analysis—the extinction of families. II. J Wash Acad Sci 21:453–459

Powell JR (1983) Interspecific cytoplasmic gene flow in the absence of nuclear gene flow: evidence from *Drosophila*. Proc Natl Acad Sci USA 80:492–495

Ross SM (1970) Applied probability models with optimization applications. Holden-Day, San Francisco

Schaffer H (1970) The fate of neutral mutants as a branching process. In: Kojima K (ed) Mathematical topics in population genetics. Springer-Verlag, New York, p 317

Spiess EB (1977) Genes in populations. John Wiley & Sons, New York

Templeton AR (1983) Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. Evolution 37:221–244

Wiley EO (1981) Phylogenetics. John Wiley & Sons, New York

Received September 9, 1983/Revised December 16, 1983