POLYMORPHISM OF MITOCHONDRIAL DNA IN POPULATIONS OF HIGHER ANIMALS

John C. Avise and Robert A. Lansman

INTRODUCTION

The complete nucleotide sequence of the mitochondrial genome has been determined for one individual in each of three species—mouse (*Mus musculus*) (Bibb et al., 1981), cow, and human (Anderson et al., 1981, 1982). Mitochondrial DNA (mtDNA) has also been partially characterized for coding functions, and major features of gene organization are exemplified in Figure 1. Mitochondrial DNA is a closed circular molecule, of length 16,295 base pairs in the mouse. Fully 94% of the molecule encodes functional RNA. Genes for 22 transfer RNAs are interspersed between genes for the 12S and 16S ribosomal RNAs; cytochrome *c* oxidase subunits I, II, and III; ATPase subunit 6; cytochrome *b*; and eight as yet unidentified proteins. Aside from 32 nucleotides scattered about the mouse mtDNA genome, only two significant regions of the molecule lack defined coding function—the 879- and 32-base pair regions around the origins of heavy (D-loop) and light strand replication, respectively. Some adjacent genes exhibit terminal overlap in coding sequence.

Human mtDNA is slightly longer (16,569 base pairs) almost entirely because of an additional 243 nucleotides in the D-loop region. Gene order appears identical to that of the mouse. These results sup-

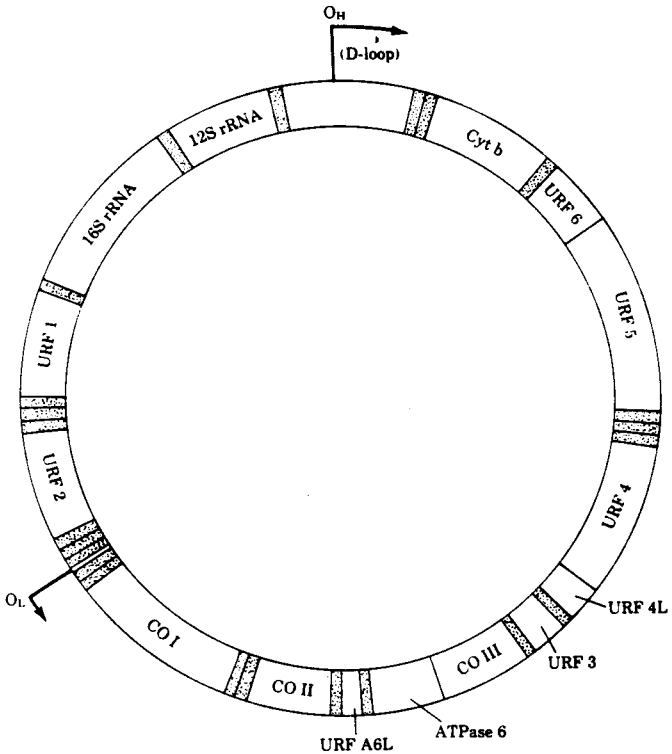

FIGURE 1. Major features of gene organization in mouse mitochondrial DNA. Transfer RNA genes are indicated by stippled regions. CO I, CO II, and CO III are genes for cytochrome oxidase subunits, Cyt b is cytochrome *b*, and URFs are unidentified reading frames. O_H and O_L are origins of heavy and light strand DNA replication. (Modified from Bibb et al., 1981.)

port the general consensus reached for other higher animal mtDNAs not so thoroughly characterized, namely, that mtDNA is remarkably conservative in size, function, and organization (Brown, 1981). The overall sequence divergence between man and mouse mtDNA is approximately 30%, but this is variable across genes. Greatest sequence homology (approximately 70–90%) occurs in the tRNA and rRNA genes, and lowest homologies (approximately 60–70%) are observed at the unidentified protein-coding loci. Portions of the D-loop region also show low homology.

Obtaining these complete mtDNA sequences were monumental efforts. From the perspective of a population geneticist interested in mtDNA sequence polymorphisms among large arrays of individuals, quicker and easier means of sampling nucleotide sites within the mtDNA genome are required. Fortunately, the discovery of Type II

restriction endonucleases has permitted the development of such assays. The purpose of this chapter is to summarize conclusions from research on mtDNA polymorphisms within species of higher animals. For a complementary review of between-species differences and longer-term mtDNA evolution, see Chapter 4 by Brown.

Considering the conserved function of mtDNA in higher animals and the exceptional economy of organization of the molecule, one might expect the primary sequence also to be well conserved. This has not proved to be the case: mtDNA may evolve 5 to 10 times faster than single copy nuclear DNA (Brown et al., 1979; Chapter 4 by Brown). Many of the findings about levels and patterns of mtDNA polymorphism within species are equally surprising and would not likely have been predicted from general knowledge of mitochondrial biology. We will organize this review by presenting, one at a time, the major tentative conclusions reached from early restriction enzyme studies of mtDNA. We will critically reexamine each of these conclusions in the light of more recent data and theory, some of which are still unpublished.

RESTRICTION ENZYME ASSAYS

Approximately 350 restriction endonucleases, involving more than 85 distinct recognition specificities, are now known (Roberts, 1982). A given enzyme recognizes a specific oligonucleotide sequence that is four, five, or six base pairs in length and cleaves two phosphodiester bonds within or near the sequence, one in each strand of the duplex. DNA fragments produced by restriction enzyme cleavage can be separated by molecular weight by gel electrophoresis and can be visualized by any of several staining or autoradiographic techniques (Figure 2). In a typical population survey, mtDNA is purified from tissues of individual organisms and then digested separately by each of a battery of restriction enzymes to obtain "digestion profiles" for the mtDNA molecules. Detailed procedures for isolation and assay of mtDNA have been reviewed (Lansman et al., 1981).

Because mtDNA is a closed circular molecule, the number of linear digestion fragments produced by a restriction enzyme is equal to the number of recognition sites in the mtDNA. For a given enzyme, the approximate anticipated numbers of such fragments can be readily calculated if the size and G + C content of the mtDNA is known and if the frequencies of cleavage sequences in mtDNA are similar to those expected in random sequences of the same base composition (Nei and Li, 1979). As generally predicted, "6-base" enzymes produce 1–8 frag-

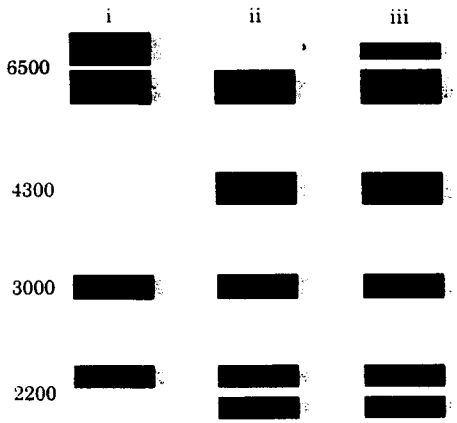

FIGURE 2. Diagrammatic representations of mtDNA "digestion profiles." The pattern in lane iii could result from restriction site heterogeneity or from incomplete digestion. The mtDNAs in lanes i and ii most likely differ by a restriction site that cleaves the 6500-base pair fragment in lane i to the 4300- and 2200-base pair fragments in lane ii.

ments in a typical mtDNA digestion, whereas "4-base" enzymes usually produce as many as 20 or more detectable fragments. Because the potential benefits of increased data from "4-base" enzymes is, in our opinion, offset by increased difficulties of scoring gels and determining fragment homology, for most purposes we prefer to employ 5- and 6-base enzymes (but see Brown, 1980, for a counterargument).

An index to the genetic similarity between mtDNAs from different organisms is the proportion of fragments shared in their digestion profiles. In some cases, further refinement can be obtained by mapping positions of restriction sites relative to one another or to known landmarks of the molecule (Brown and Vinograd, 1974; Nathans and Smith, 1975). Several statistical procedures have been developed to convert raw "fragment" or "site" data to quantitative estimates of nucleotide sequence divergence (p) between mtDNAs (Engels, 1981; Gotoh et al., 1979; Kaplan and Langley, 1979; Kaplan and Risko, 1981; Li, 1981; Nei and Li, 1979; Nei and Tajima, 1981; Upholt, 1977). Qualitative methods of data analysis have also been advocated (Avise et al., 1979a, b; Lansman et al., 1981; Templeton, 1983), as shown later.

BIOLOGY OF MITOCHONDRIAL DNA

Mitochondrial DNA replication and transcription occur within the organelle, autonomously from nuclear DNA. In vertebrate somatic

cells, rate of mtDNA turnover is probably higher than that of nuclear DNA (Rabinowitz and Swift, 1970). A vertebrate cell contains many, often thousands, of mtDNA molecules (Birky, 1978; Gillham, 1978; Potter et al., 1975). Mature oocytes are particularly rich in mtDNA. For example, the exceptionally large *Xenopus* egg carries an estimated 10^8 mtDNA molecules (Dawid and Blackler, 1972). The midpiece of mature sperm also carries small numbers (approximately 100) of mtDNA molecules. In some animals, these mtDNAs disperse into zygote cytoplasm during fertilization (Gresson, 1940; Friedlander, 1980), but in other species sperm mitochondria do not penetrate the egg (Ursprung and Schabtach, 1965). The sheer preponderance of egg mtDNA in zygotes can certainly account for the common observation that most mtDNA is maternally inherited.

Because an individual animal contains many billions of copies of mtDNA, the evolutionary dynamics of mtDNA in an animal population are interrelated with and partially dependent upon the underlying dynamics of the population of mtDNA molecules in each animal. Empirical surveys of mtDNA composition usually employ somatic tissue, so understanding of somatic cell mtDNA segregation is important. However, mtDNA dynamics in germ-cell lineages are of particular interest because only here are results directly relevant to mtDNA evolution. Unfortunately, virtually nothing is known about fundamental characteristics of mtDNA dynamics in germ cells, for example, the number of rounds of mtDNA replication during gametogenesis or the effective population sizes of mtDNA in intermediate germ cells. Recent theoretical models point out the need for such information (Chapman et al., 1982; Birky et al., 1982, 1983; Takahata and Maruyama, 1981).

INTRAINDIVIDUAL MITOCHONDRIAL DNA SEQUENCE HOMOGENEITY

Contrary to some earlier expectations, virtually all studies to date have concluded that the huge populations of somatic cell mtDNAs within an individual organism appear homogeneous in nucleotide sequence. The evidence is as follows. In mtDNAs that are digested to completion with a restriction enzyme, the total molecular weight of the fragments produced typically equals the known genome size of higher animal mtDNA, approximately 16,000 base pairs. Significant sequence heterogeneity in a sample would be evidenced by additional fragments exceeding the total anticipated weight. Furthermore, in gels stained with ethidium bromide, the fluorescence of each band in

the digestion pattern is proportional to the molecular weight of the fragment. Nonstoichiometric bands should appear in digests of tissue heteroplasmic for mtDNA. Studies listed in Table 1 include a total of more than 3000 different single-restriction-enzyme digests of a dozen mammalian species, without report of a case of intra-individual mtDNA sequence heterogeneity. In addition, a few studies have explicitly compared mtDNAs from two or more different tissues (i.e., kidney, heart, liver, platelets, skin fibroblasts) of the same animal (Denaro et al., 1981; Francisco et al., 1979; Potter et al., 1975). No differences among tissues have been documented.

Critique

The above case for homoplasmic intraindividual mtDNA may appear stronger than it really is. A minority mtDNA sequence would have to constitute at least 1–5% of the total mtDNA in a sample to reach limits of normal empirical detectability, even when fragments are assayed by the more sensitive autoradiographic methods. It is thus conceivable that many rare mtDNA sequences commonly coexist with the predominant sequence in a given tissue but remain undetected. Heterogeneity of mtDNA in cultured mammalian cells has been demonstrated (R. Slott, R. O. Shade, and R. A. Lansman, unpublished).

There is another seldom-acknowledged bias operating against reports of possible heteroplasmicity of mtDNA sequence. Not uncommonly, additional faint mtDNA bands are observed on restriction digest gels. These are usually attributed to "incomplete digestions," presumably resulting from partial loss of restriction enzyme activity, low concentration of enzyme relative to mtDNA, too short an incubation period, or to the known several-fold difference in the kinetics of site cleavage (Thomas and Davis, 1975). Such supernumerary bands usually disappear upon more complete redigestions (but see Potter et al., 1975). However, in routine animal surveys not specifically concerned with the issue of individual mtDNA heterogeneity, gel scoring is often done without redigestion, under the assumption that the faint bands represent mtDNA segments not yet fully cleaved. True mtDNA sequence heterogeneity would produce digestion profiles that would not be readily distinguished from those of incomplete digests (Figure 2, lane iii).

Coote et al. (1979) do report heterogeneity of mtDNA in different tissues from a single ox, but their interpretation is also complicated by the possibility of incomplete digestion. The general problem of intraindividual mtDNA sequence variation needs much more empirical study. Because between-individual mtDNA divergence is extensive (see later), the eventual documentation of at least some individual heteroplasmicity seems inevitable. Such a discovery will be especially

TABLE 1. Estimated genetic distances (p, base substitutions per nucleotide) in mtDNAs between conspecific mammals.

Species	Animals surveyed	Geographic source of samples	Restriction enzymes	Approximate $\bar{p}$*	Maximum p	Source‡
Ovis aries	2	Domestics	3	0.02	0.020	1
Capra hircus	3	Domestics	3	0.01	0.010	1
Geomys pinetis	87	Several states, Southeast United States	6	0.02	0.047	2
Peromyscus polionotus	36	Several states, Southeast United States	6	0.01	0.030	3
Peromyscus maniculatus	135	Entire N. Amer. continent	8	0.02	0.070	4
Rattus norvegicus	21	N. Amer.; Puerto Rico; Japan	6	0.01	0.018	5
Rattus rattus	26	N. Amer.; Puerto Rico; East Asia	6	0.04	0.096	5
Mus musculus	≅30	Diverse lab strains; native subspecies worldwide†	10–15	0.03	0.096	6, 7
Pongo pygmaeus	5	Sumatra and Borneo	25	0.02	0.050	8
Pan troglodytes	10	Captives, origin uncertain	25	0.01	0.045	8
Pan paniscus	3	Captives, origin uncertain	25	0.01	0.015	8
Gorilla gorilla	4	Captives, only one subspecies represented	25	0.006	0.009	8
Homo sapiens	21	Three major races, worldwide	18	0.004	<0.01	9, 10

* Usually calculated by the approach of Nei and Li (1979). Mean values are approximate only; actual genetic distances in any given comparison are heavily dependent upon geographic closeness of collection, and individual papers should be consulted for details.
† Taxonomy of *Mus musculus* is controversial, with some workers recognizing two distinct species. Ferris et al. (1982, 1983) also studied mtDNA variation in *M. musculus* and *M. domesticus*, two forms considered subspecies in the work of Yonekawa et al. (1980, 1981).
‡ Source: 1. Upholt and Dawid (1977), 2. Avise et al. (1979b), 3. Avise et al. (1979a), 4. Lansman et al. (1983a), 5. Brown and Simpson (1981), 6. Yonekawa et al. (1981), 7. King et al. (1981), 8. Ferris et al. (1981b), 9. Brown and Goodman (1979), 10. Brown (1980).

significant if it occurs in an organism in which the pattern of mtDNA segregation in somatic tissues, and across animal generations, can be experimentally followed. A promising case of a segregating mtDNA lineage has been reported in a herd of cows descended from a common female parent (Laipis and Hauswirth, 1980). As yet, however, no individual cows have been found that simultaneously exhibit both of the mtDNA sequences. S. Hechtel and W. M. Brown (pers. comm.) have also observed mtDNA sequence heterogeneity in batches of unfertilized eggs from individual sea urchins.

MITOCHONDRIAL DNA POLYMORPHISM WITHIN SPECIES

Table 1 lists the major published studies of mtDNA polymorphism within species. The amount of sequence heterogeneity is striking. Mean values of p (estimated nucleotide sequence divergence) between conspecific individuals commonly range from 0.3 to 4%, and maximum observed p's sometimes approach 10%. Most major studies to date have dealt with mammals, and in the future it will be especially important to survey other groups of animals. Preliminary results from some reptiles, birds, and other groups have been reported (Brown and Wright, 1979; Glaus et al., 1980; Fauron and Wolstenholme, 1980b), and more extensive surveys are underway.

We can briefly summarize results of one typical mammalian study. Nei and Tajima (1981) have suggested the term *nucleomorph* for different restriction-site patterns for a DNA segment such as mtDNA. Among 87 pocket gophers (*Geomys pinetis*) collected from across the species range and assayed by six restriction enzymes, a total of 23 mtDNA nucleomorphs were observed (Avise et al., 1979b). Mean and maximum estimates of sequence divergence between pairs of nucleomorphs were $\bar{p} \simeq 0.020$ and $p = 0.047$, respectively. Within a given local collection, 1–3 nucleomorphs were observed, and estimates of p were less than 1%.

The only reported exception to extensive within-species mtDNA sequence divergence involves humans (gorillas also exhibit limited mtDNA sequence variation, but only a few individuals of one subspecies were assayed; Ferris et al., 1981b). Among 21 humans of diverse ethnic and geographic origin, $\bar{p} \simeq 0.004$. Brown (1980) raises the possibility that all living humans may have evolved from a small, mitochondrially monomorphic population that existed as recently as 200,000 years ago.

Critique

Some of the larger within-species estimates of p in Table 1 are one-fourth to one-third as great as the evaluated sequence divergence of

30% between man and mouse mtDNAs, which have been completely sequenced (Anderson et al., 1981)! Two classes of explanation could account for this surprising result. First, it is conceivable that within-species estimates of p are seriously in error, for some reason biased toward large values. In estimating p from raw fragment or site data, several assumptions are made—that mtDNA polymorphism reflects base substitution differences only; that the frequencies and distributions of cleavage sequences in the mtDNA are similar to those expected in random sequences; and that fragment and site homologies are correctly determined. Justifications for these assumptions have been advanced (Upholt, 1977; Nei and Li, 1979; but see later and also Adams and Rothman, 1982).

Alternatively, if within-species estimates of p are valid, then observed intraspecific polymorphism in mtDNA must not reflect long-term evolutionary rate of accumulation of nucleotide substitution (Upholt and Dawid, 1977). Perhaps some slightly deleterious substitutions contribute to intraspecific polymorphism but do not generally persist over evolutionary time. More likely, certain nucleotides (i.e., in silent positions) may be especially free from evolutionary constraints. Brown and Simpson (1982) report that 94% of the nucleotide substitutions in sequenced portions of the mtDNA genomes from *Rattus norvegicus* and *R. rattus* are silent. Mitochondrial DNA molecules might rapidly become saturated with changes at these positions, whereas the remainder of the molecule evolves much more slowly. Either of these possibilities should result in a curvilinear relationship between p and time. Such empirical curvilinear relationships have been presented by Brown et al. (1979, 1982) and Ferris et al. (1981b).

The issue of the validity of current estimates of p can be finally resolved by use of actual nucleotide sequence data, which for large portions of the mtDNA molecule are now being obtained for conspecific organisms. Complete nucleotide sequences are available for a 900-base pair region surrounding the mtDNA D-loop region in seven humans (Greenberg et al., 1982). Among pairwise comparisons of these seven mtDNAs, evaluated levels of p range from 0.004 to 0.032 ($\bar{p}$ = 0.017). If these same regions of mtDNA had, in fact, been accurately mapped for restriction sites by approximately 25 commercially available restriction enzymes, summary estimates of p for the region would have ranged from 0.006 to 0.034 ($\bar{p}$ = 0.020) (Aquadro and Greenberg, 1983). This strong agreement between restriction site estimates of $\bar{p}$ and sequence evaluation of $\bar{p}$ suggests that current estimates of mtDNA polymorphism within species are not grossly biased toward larger values by the restriction enzyme approach. [Levels of $\bar{p}$ for this 900-base pair region in humans are several-fold higher than estimated

nucleotide differences for the whole human mtDNA genome (Brown, 1980); presumably this is due to the hypervariability of portions of the D-loop region of the molecule (see Aquadro and Greenberg, 1983).]

In any event, mtDNA polymorphism within species is clearly extensive. The joint observations of common individual homoplasmicity and between-individual sequence divergence can be accommodated under a neutral model involving chance sampling drift of mtDNA nucleomorphs in germ-cell lineages. Chapman et al. (1982) provide computer simulations of how the process may occur. In their models, each germ-cell lineage has a fixed number (n_M) of mtDNA molecules, each of which can mutate at a prescribed rate μ each cell generation. Multinomial sampling of mtDNA molecules occurs in heteroplasmic germ cells; and in every animal generation (assumed to be 50 germ-cell generations), each adult female produces female progeny according to a Poisson distribution with mean equal to 1. In the simulations, H_1 (the probability that two randomly drawn mtDNA sequences from the same individual are different) and H_2 (the corresponding probability for the population) are monitored through time: some populations are initiated with all individuals homoplasmic and ide[illegible] (Pop. A); and some populations are initiated with each indiv[illegible] homoplasmic for a unique mtDNA sequence (Pop. B). Thus, H_1 a[illegible] H_2 register within- and between-individual mtDNA sequence heterogeneity, respectively, and their ratio is of particular interest.

Figure 3 shows two typical runs of the simulation. As expected, H_2 in populations A and B converges due to the decay of heterogeneity by drift, balanced by the accumulation of variation by mutation. H_1 is appreciably lower with smaller n_M (Figure 3). Both H_1 and H_2 are lower with lower μ, all else being equal (Chapman et al., 1982). Overall, H_1 remains extremely low and the ratio H_2:H_1 high only when $n_M\mu$ is small. Thus, the models predict that effective population sizes of mtDNA molecules in germ-cell lineages should be small (i.e., 10–100), and they point out the need for empirical data in this area. Census sizes of mtDNAs in mature oocytes are huge, but this may be misleading.

Alternatively, if mtDNA population sizes remain consistently large in germ-cell lineages, some model involving selection might be required. The selection pressure would, however, have to be of a rather peculiar nature because it would require active conservation of within-individual sequence homogeneity while still allowing between-individual differences to accumulate rapidly. Some form of intense intracellular competition among mtDNA nucleomorphs is perhaps conceivable. Another related possibility is that, as in yeast, random choice of a small number of mtDNA molecules for replication may cause cells to move rapidly toward homoplasmicity (Birky et al., 1983).

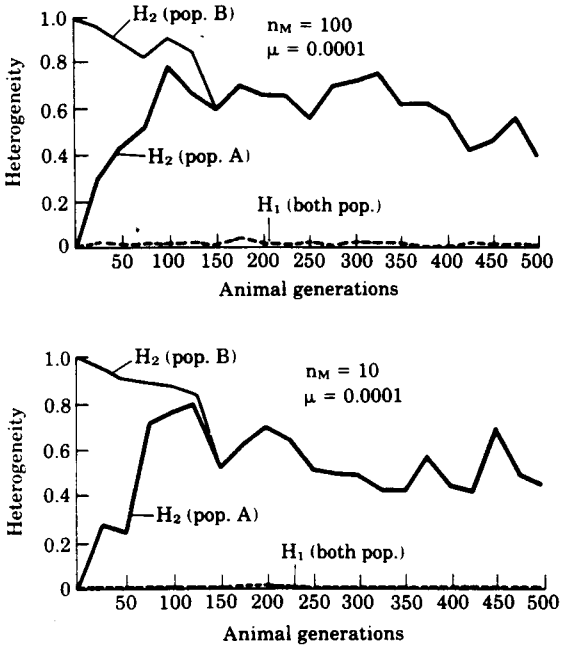

FIGURE 3. Examples of computer-simulated levels of mtDNA sequence heterogeneity for populations A (initiated with all individuals homoplasmic for same sequence) and B (initiated with each individual homoplasmic for unique sequence). H_1 and H_2 register within- and between-individual mtDNA sequence heterogeneity, respectively. n_M is the number of mtDNA molecules per germ cell, and μ is the probability of mutation per mtDNA molecule per germ-cell generation. (From Chapman et al., 1982.)

GENETIC BASIS OF MITOCHONDRIAL DNA POLYMORPHISM

Mitochondrial DNA ranges in size from 15,700 to 17,700 base pairs among assayed species as diverse as sea urchins, amphibians, reptiles, birds, and mammals (Brown, 1981). Among species of *Drosophila,* the size range is even greater (15,600–19,400 base pairs), and as in other animals most of the length differences occur in the region surrounding the origin of heavy strand replication (Fauron and Wolstenholme, 1976; Shah and Langley, 1979b). Within species, some mtDNA size variation has also been reported (Brown, 1980; Brown and Simpson, 1981; Fauron and Wolstenholme, 1980b). For example, one geographic

population of lizards (*Cnemidophorus sexlineatus*) differs from others by an additional 1200 base pairs (Brown, 1981). Genome size variation provides straightforward evidence that additions and/or deletions have played a role in mtDNA evolution.

Nonetheless, the consensus is that the majority of mtDNA polymorphism within species is attributable to gain or loss of particular restriction sites, without detectable alteration of genome size. For example, all differences among a total of 61 mtDNA nucleomorphs observed in geographic samples of the mouse *Peromyscus maniculatus* resulted from changes of individual restriction sites that were mapped on a molecule that appeared constant in size (Lansman et al., 1983a). Particularly within coding regions of genes, restriction site changes are far more likely to arise from base substitutions than from small nucleotide additions or deletions, most of which would alter the reading frame.

Critique

An alternative explanation for apparent gain or loss of restriction sites is differential methylation of cytosine in mtDNA from different animals. Methyl-modified cytosine (5-MeCyt) is the only modified base known in vertebrate DNA, and it protects DNA from cutting by some restriction enzymes (Singer et al., 1979; Bonen et al., 1980). However, several lines of evidence argue that differential methylation does not account for mtDNA polymorphism. First, the maximum amount of base methylation in mouse or hamster mtDNA is very low (<1%), far less than the 21% C content of mouse mtDNA (Nass, 1973; Brown and Goodman, 1979; Singer et al., 1979). Second, restriction enzymes lacking the dinucleotide C-G in their recognition sites uncover at least as much mtDNA restriction site polymorphism as do enzymes that contain it (Brown, 1980). Third, mtDNAs cloned in *Escherichia coli* yield the same fragment patterns as do native mtDNAs when digested with several restriction enzymes (Castora et al., 1980; Chang et al., 1975), despite the fact that cloning has been observed to increase probabilities of altered methylation (Gautier et al., 1977).

Direct confirmation of the common occurrence of base substitution in mtDNA polymorphism has come from recent sequencing studies. From the mtDNAs of each of five humans, Greenberg et al. (1982) cloned and sequenced an approximately 900-base pair noncoding region surrounding the origin of H-strand replication; sequences were further compared to those of three other human mtDNAs previously sequenced (Anderson et al., 1981; Crews et al., 1979; Walberg and Clayton, 1981). Four instances of mtDNA length variation were revealed among the eight samples, each involving additions or deletions of one or two nucleotides. All of the remaining 50+ sequence altera-

tions (>90%) were due to simple base substitutions (Greenberg et al., 1982; Aquadro and Greenberg, 1983). The percentage of mtDNA polymorphism attributable to base substitution may be even higher for the remainder of the mtDNA genome, most of which has active coding function. Limited sequence analysis of four cloned *Peromyscus* mtDNA molecules reveals no additions or deletions outside the D-loop region (J. F. Shapira et al., unpublished).

All available data from closely related populations and species have revealed a great preponderance of transition over transversion substitutions (Aquadro and Greenberg, 1983; Brown et al., 1982; Brown and Simpson, 1982). For example, in the study of Greenberg et al. (1982), 53 of 55 substitutions (96%) were transitions. Furthermore, the distribution of variable sites was highly nonrandom, with the majority of site changes occurring in two regions roughly flanking the D-loop.

MATERNAL INHERITANCE OF MITOCHONDRIAL DNA

In higher animals, mtDNA appears to be inherited maternally—transmitted to progeny through egg cytoplasm. This conclusion stems from examination of mtDNA in progeny of crosses between parents differing in mtDNA restriction fragment pattern. For example, horse and donkey mtDNAs differ in position of *Hae*III restriction sites. Hinnies (progeny of crosses between male horse and female donkey) exhibit the donkey mtDNA pattern, and mules (progeny of the reciprocal cross) exhibit the horse pattern (Hutchison et al., 1974). Similar evidence for maternal inheritance has been reported in experimental crosses of *Peromyscus* (Avise et al., 1979a), *Rattus* (Francisco et al., 1979; Hayashi et al., 1978), *Xenopus* (Dawid and Blackler, 1972), humans (Giles et al., 1980), and others.

Critique

As already mentioned, an mtDNA sequence would have to constitute at least 1–5% of total mtDNA to be noticed in conventional restriction assays. Thus, any "paternal leakage" (P, the proportion of progeny mtDNA derived from the immediate father) less than this amount would normally escape detection. The question of paternal leakage is important, because even very low P can have important evolutionary consequences as a "gene-flow bridge" between female lineages otherwise completely isolated from one another with respect to mtDNA sequence (Chapman et al., 1982).

In an attempt to increase power to detect low-level P, Lansman et al. (1983b) utilized a uniquely favorable backcross strain of tobacco budworms (*Heliothis*). Two budworm species that differ in *Xba*I and *Eco*RI digestion profiles had been hybridized, and the fertile female progeny successively backcrossed to males of one of the species for 91 consecutive generations. In insects, sperm mitochondria have been observed within the zygote, but their fate remains unknown (Friedlander, 1980; Chapman et al., 1982). If effective paternal leakage occurs in *Heliothis* at constant rate P per generation, the total accumulated proportion Q of paternal mtDNA expected in backcross generation i is $Q_i = 1-(1-P)^{i+1}$. Thus, for P as low as 1 molecule in 10,000 per generation, $Q_{91} \cong 1\%$, a level that would be detectable by the autoradiographic assay employed. Nonetheless, absolutely no paternally derived mtDNA was detected in these 91-generation backcross progeny.

Thus, all available information is consistent with strict maternal inheritance of mtDNA. Given the small ratio (S/E) of sperm to egg mtDNAs in zygotes, this result is not too surprising. However, if sperm mtDNAs in the zygote do survive and replicate at rates equal to those of egg mtDNAs, and if random germ-cell lineage drift of mtDNAs does occur as envisioned in the models of Chapman et al. (1982; see also preceding discussion on mtDNA polymorphism within species), then a tiny fraction (S/E) of gametes of the following generation should carry only paternally derived mtDNA. It has also been suggested, however, that sperm mtDNAs are actively altered or degraded and not utilized or transmitted through the zygote. Some evidence for physical alteration of chloroplasts and mitochondria during microsporogenesis in higher plants lends support to this conjecture (Vaughn et al., 1980).

ESTIMATION OF MATRIARCHAL PHYLOGENY

The observed high levels of within-species mtDNA polymorphism provide opportunities for reconstructing evolutionary relationships among conspecifics. If mtDNA is indeed strictly maternally inherited, the phylogenies derived from either mtDNA restriction fragment or site data will represent estimates of matriarchal relationship. Furthermore, because mtDNA genotypes are clonally transmitted (barring mutation) from female to progeny and are not recombined during sexual reproduction, the mtDNA genotype of an individual organism can in theory provide definitive information about the female lineage to which it belongs (Lansman et al., 1981). Attempts to capitalize upon these unique advantages of mtDNA genotypes have only begun (Avise et al., 1979a,b; Brown, 1980; Brown and Simpson, 1981; Lansman et al., 1981, 1983a; Templeton, 1983; Yonekawa et al., 1981).

One general method of data analysis involves manipulation of a quantitative matrix of estimates of nucleotide sequence divergence (p) between mtDNA nucleomorphs (clones). Many different analyses can be performed upon any such matrix, and the method of choice may depend on goals of the analysis. Thus, one could assume that evolutionary rates are roughly constant in all lineages, in which case phenetic clustering procedures might be employed (Sneath and Sokal, 1973). Or one could search for a network whose branch-lengths connecting clones most closely match distances in the matrix from which the network was formed (Farris, 1972; Prager and Wilson, 1978). Some of these approaches have been applied to mtDNA data (Avise et al., 1979a,b; Lansman et al., 1983a; Yonekawa et al., 1981). However, it should be realized that estimates of p involve a large sampling variance when the number of restriction sites assayed is fairly small (Nei and Tajima, 1981; see also Engels, 1981). For example, a typical survey involving five to ten 6-base restriction enzymes and a total of approximately 50 restriction sites might yield an estimate of $p = 0.02$ between two particular mtDNA clones. The standard deviation of this estimate is approximately 0.014 (from formula 19, Nei and Tajima, 1981; Li, 1981). Thus, in population surveys of the size typically conducted to date, many estimates of p between mtDNA clones within a species are not significantly different from one another.

A much stronger method of analysis takes advantage of the qualitative relationships among particular fragment patterns or restriction maps (Lansman et al., 1981). For example, mtDNA restriction maps can often be arranged into a transformation series that reflects the probable steps along which restriction patterns were interconverted during evolution. Figure 4 shows the probable transformations among the assemblage of 11 *Eco*RI restriction maps observed in a study of 135 *Peromyscus maniculatus* collected across North America (Lansman et al., 1983a). In this study, similar networks were likewise developed from the restriction maps for each of seven other endonucleases, and a final composite network was constructed which simultaneously incorporated all data.

For conspecific populations, the composite phylogenetic network (derived solely from the genetic data) can subsequently be superimposed over the geographic sources of collections. In the two cases where this approach has been attempted, the results yielded highly plausible geographic patterns of mtDNA clonal relationships. Thus, both within *Peromyscus maniculatus* (Lansman et al., 1983a) and *Geomys pinetis* (Avise et al., 1979b), mtDNA clones that appear closely related genetically are also usually geographically contiguous. Rare or local

clones are usually related to the widespread clone of that geographic region by one or two assayed site changes. In each species, major genetic "breaks" in mtDNA composition separate clonal assemblages that occupy different portions of the total species range.

Critique

Attempts to infer phylogeny from mtDNA genotypes could be compromised if homoplasy (the evolutionary convergence or reversal of restriction sites) were common. In qualitative networks, homoplasy would have the effect of introducing ambiguity into placement of some clones and would result in total network lengths that exceed the observed minimum mutation distances between clones. By both of these criteria, homoplasy was observed to be fairly common in the extensive mtDNA data for *Peromyscus maniculatus* (Lansman et al., 1983a).

One straightforward approach to document site convergence involves joint examination of restriction maps for pairs of enzyme. Suppose, for example, that *Eco*RI yields restriction map patterns bitrarily labeled A and B, and *Hinc*II produces patterns C and D.

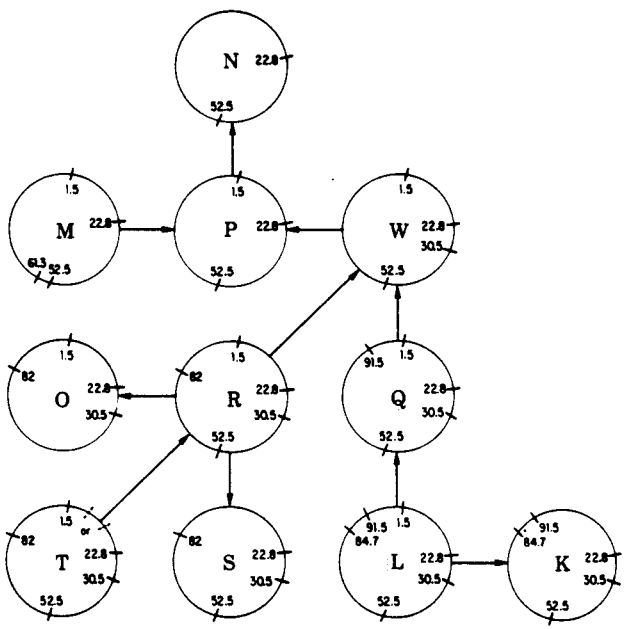

FIGURE 4. Probable evolutionary transformations among the 11 *Eco*RI restriction maps observed in *Peromyscus maniculatus*. Arrows indicate direction of site loss and not necessarily direction of evolution. (From Lansman et al., 1983a.)

Further assume that in a collection of mtDNA clones, all four two-enzyme genotype combinations (AC, AD, BC, BD) were observed. Barring recombination, at least one instance of evolutionary convergence must be invoked to account for this "dilemma," no matter which genotype is considered ancestral or where the four genotypes are placed in a larger phylogenetic network. In the *P. maniculatus* data, several two-enzyme dilemmas were observed. However, each such dilemma involved pairs of restriction maps that were closely related, usually by a single site change. Thus, homoplasy was certainly bounded, and no single-enzyme restriction map appeared to have arisen independently from other restriction maps that were not genetically close to it.

Analysis of the *P. maniculatus* data also revealed a few instances of sites that appeared to "blink" on and off repeatedly during evolution. For example, the presence or absence of one particularly hypervariable *Hinc*II site was all that distinguished members of each of four pairs of *Hinc*II restriction maps that were otherwise genetically very distinct from each other (Lansman et al., 1983a).

Significant homoplasy was further confirmed in recent analyses of nucleotide sequences in human mtDNAs. A minimum-length phylogenetic tree, constructed from seven sequences each approximately 900 nucleotides long, required convergent base substitutions at five of the 45 nucleotide replacement sites (11%) (Aquadro and Greenberg, 1983). This relatively high level of homoplasy is apparently attributable to the hypervariability of particular areas of the mtDNA molecule flanking the D-loop region and to the empirical preponderance of transitional substitutions. The observations that rates of base substitution vary considerably along the mtDNA molecule and that most substitutions are transitions will have to be taken into account in future mathematical or statistical models of mtDNA sequence evolution.

CONCLUSION

Several findings on mtDNA polymorphism within species of higher animals were unanticipated: the high level of polymorphism itself; the apparent rarity of individual heteroplasmicity; the attribution of most polymorphism to base substitution; the preponderance of transitions; and the frequency of convergent site evolution. As stressed in this chapter, by stringent criteria some of these conclusions are inadequately documented. Nonetheless, a general outline for evolution of higher-animal mtDNA is now apparent.

A further recent surprise is the different scenario emerging for plant mtDNA (Sederoff et al., 1981). For example, maize mtDNA is approximately 30 times larger than most animal mtDNA and consists of at least seven classes of molecules that vary in size and abundance. Sederoff et al. (1981) conclude that in maize and teosinte mtDNA sequence homology is generally conserved and most polymorphism is attributable to major reorganizations of sequence. The evolutionary significance of the different pattern of plant mtDNA polymorphism is not known.

Recent studies suggest that the level of mtDNA polymorphism in higher animals is several-fold higher than that of single-copy nuclear DNA (Engels, 1981; Nei and Li, 1979; but see Shah and Langley, 1979b). Some of these conclusions resulted from indirect comparisons of protein-electrophoretic data with data from mtDNA restriction digests. Because restriction enzyme approaches can also be applied to measure polymorphism in the nuclear genome (provided homologous sequences of appropriate size can be isolated and cloned from different animals), direct answers to the question of relative levels of mtDNA versus nuclear DNA polymorphism should be forthcoming (Engels, 1981). Whatever the final outcome, mtDNA polymorphism will continue to be of special interest in its own right. The maternal inheritance and high polymorphism of mtDNA provide unique opportunities for population analysis.