Why Mitochondrial Genes are Most Often Found in Nuclei

Otto G. Berg and C. G. Kurland

Department of Molecular Evolution, Uppsala University Evolutionary Biology Centre, Uppsala, Sweden

A very small fraction of the proteins required for the propagation and function of mitochondria are coded by their genomes, while nuclear genes code the vast majority. We studied the migration of genes between the two genomes when transfer mechanisms mediate this exchange. We could calculate the influence of differential mutation rates, as well as that of biased transfer rates, on the partitioning of genes between the two genomes. We observe no significant difference in partitioning for haploid and diploid cell populations, but the effective size of cell populations is important. For infinitely large effective populations, higher mutation rates in mitochondria than in nuclear genomes are required to drive mitochondrial genes to the nuclear genome. In the more realistic case of finite populations, gene transfer favoring the nucleus and/or higher mutation rates in the mitochondrion will drive mitochondrial genes to the nucleus. We summarize experimental data that identify a gene transfer process mediated by vacuoles that favors the accumulation of mitochondrial genes in the nuclei of modern cells. Finally, we compare the behavior of mitochondrial genes for which transfer to the nucleus is neutral or influenced by purifying selection.

Introduction

It is generally accepted that the genomes of mitochondria are the degraded descendants of free-living α-proteobacteria from which most of the coding sequences have been lost (Attardi 1985; Bulmer 1987; Andersson and Kurland 1991, 1999; Andersson et al. 1998; Martin and Muller 1998; Gray, Burger, and Lang 1999). To our knowledge, the smallest genome of a free-living α-proteobacteria is that of *Bartonella henselae,* with less than 2 million base pairs, and the largest is *Bradyrhizobium japonicum,* with 8.7 million base pairs (Roux and Raoult 1995; Kundig, Henneke, and Gottfert 1993). The Bartonella genome encodes at least 1,600 proteins (Andersson et al., personal communication). Accordingly, we used 1,600 proteins as a conservative estimate for the size of the proteome of the α-proteobacterial ancestor of the mitochondria. In contrast, of the known mitochondrial genomes, the largest is that of *Reclinomonas americana,* with 67 protein-coding genes, and the smallest is that of *Plasmodium falciparum,* with 3 protein-coding genes (Gray, Burger, and Lang 1999). Most common are figures nested between 12 and 24 protein coding genes (Gray, Burger, and Lang 1999).

At least two distinct modes of genetic loss have reduced the coding capacities of mitochondrial genomes from approximately 1,600 proteins in the free-living ancestor to 67 or less in a modern mitochondrion. One reductive mode is the loss of nonessential coding sequences. For example, mitochondria import, rather than synthesize, substrates such as amino acids, nucleosides, and pyruvate (Gray 1992). Consequently, the genes encoding the corresponding biosynthetic activities might be purged from the mitochondrial genome by random mutations that are unopposed by purifying selection.

The second mode of genetic loss involves the transfer of essential genes from the mitochondrial genome to the nuclear genome. Indeed, nearly all of the genes required for the function and propagation of mitochondria normally reside in the host nucleus (Gray 1992). Some of these have originated from the proteobacterial ancestor, but a significant fraction have arisen within the eukaryotic genome (Andersson and Kurland 1999; unpublished data). In the yeast *Saccharomyces cerevisiae,* in which the nucleus encodes more than 400 mitochondrial proteins, at most half of these are of bacterial origin, while the remainder seem to be of eucaryotic origin (Andersson and Kurland 1999; unpublished data). Here, the mitochondrial genome encodes only 17 proteins. Thus, the ancestor's genome has been reduced in yeast to roughly 200 genes, mostly residing in the nucleus.

The reasons for the asymmetrical distribution of mitochondrial genes between the mitochondrial and nuclear genomes are the subject of the present study. We ask why redundant genes are not lost from the nucleus as often as they are from the mitochondrion. Similarly, we ask why the essential genes for the mitochondria are not encoded primarily within the genome of the organelle rather than in the nucleus.

The asexual reproduction of mitochondria might lead to a faster accumulation of deleterious mutations either due to reduced selection efficiency from hitchhiking effects or due to the influence of Müller's ratchet (Felsenstein 1974; Kurland 1992; Lynch 1997; Bergstrom and Pritchard 1998). The details of these two mechanisms are slightly different, but both could, in principle, enhance the mutational load on genes in the mitochondria by reducing both the effective population size and the intensity of purifying selection for mitochondrial genomes. If there was a bias in the partitioning of the mutational load between the mitochondrion and the nucleus, this might drive the transfer of sequences from the high-load to the low-load compartment, with an attendant improvement in cell fitness.

In addition, the cellular processes supporting the transfer of coding sequences from one genome to the other might be much more efficient in one direction than in the other. Transfer of coding sequences from mitochondria and chloroplasts to nuclear genomes has been documented in a variety of organisms (Thorsness and

Key words: mitochondrial genome shrinkage, population dynamic models for gene transfer.

Address for correspondence and reprints: Otto G. Berg, Department of Molecular Evolution, Uppsala University Evolutionary Biology Centre, Norbyvägen 18C, SE-75236 Uppsala, Sweden. E-mail: otto.berg@ebc.uu.se.

Mol. Biol. Evol. 17(6):951–961. 2000

Weber 1996; Martin and Herrmann 1998; Martin et al. 1998). More to the point, an experimental system to study and quantify the transfer of sequences between mitochondrial and nuclear genomes in the yeast *S. cerevisiae* has been developed by Thorsness and Fox (1990, 1993). Thorsness and his collaborators have used this system to study a number of mutant alleles that influence the transfer frequencies of coding sequences to the nucleus (Thorsness, White, and Fox 1993; Hanekamp and Thorsness 1996; Weber, Hanekamp, and Thorsness 1996; Campbell and Thorsness 1998). They suggest that a primary pathway for the uptake of mitochondrial sequences by the nucleus may be provided by autophagy of mitochondria in cellular vacuoles (lysosomes). Here, nucleic acid fragments liberated during the destruction of mitochondria could provide intermediates for sequence transfer to nuclei.

Most important in the present context is that Thorsness and Fox (1990, 1993) were able to estimate frequencies of transfer of intact genes between mitochondrion and nucleus. These frequencies for wild-type cells are roughly one transfer per 10^5 generations to the nucleus from mitochondria and less than one transfer per 10^{10} generations in the opposite direction. Since autophagic vacuoles are ubiquitous in eukaryotes, preferential transfer of genetic sequences from mitochondrion to nucleus is likely to be a normal occurrence. However, we do not know just how polar the transfer processes are in other organisms.

Armed with the results of Thorsness and Fox (1990, 1993), it would be possible, as also suggested by an anonymous referee, to model the transfer to the nucleus from the ancestral symbiont or mitochondria as though it were a one-step process. However, by so doing, relevant biological details of the transfer process would be lost. Therefore, we used somewhat more detailed mathematical models to study both finite and infinite populations with either haploid or diploid gene complements. The consequences of transfer processes that are neutral and selective are discussed. We find that under realistic conditions, the transfer from mitochondria to nuclei is inevitable for genes that can function equally well in the mitochondrion and in the nucleus.

A Neutral Model for Gene Transfer

For simplicity, we consider a population of a unicellular organism. There are two requirements for the functional transfer of a gene from mitochondria to nucleus. One is the possibility of copying coding sequences from the mitochondria to the nucleus. The other is the necessity of directing gene products expressed from nuclear genes to their sites in mitochondria. In addition, the mitochondrial gene copy eventually should be purged from the organelle's genome, since it is observed that most mitochondrial genes reside only in the nucleus.

Again, it is simplest to consider the compound event of copying and addressing together as a transfer event that produces a functional mitochondrial gene in the nucleus. Once a mitochondrial coding sequence is copied (transferred) to the nuclear genome, fixation there can be driven by mutations that inactivate the mitochondrial copy. Thus, at least two distinct processes, transfer and mutation, are required before a gene from the mitochondrial genome has been displaced to the nucleus. The models below were developed to investigate how the rates for the individual steps, their reverse processes, and the effective population size influence this displacement.

Haploid Cells at Infinite Population Size

A mitochondrial gene (or set of genes) that has a functional copy in the mitochondria but not in the nucleus is assigned the *Mn* genotype. The genotype is denoted *mN* when a functional copy of the gene is in the nucleus but not in the mitochondria. Functional copies in both mitochondria and nucleus correspond to the genotype *MN*, while *mn* corresponds to a cell with no functional copies of the gene. The designations *Mn* and *mn* also include the active gene copies in the nucleus but an inability to import the gene product into the mitochondria. Here, we assume that the genotype *mn* is lethal; i.e., we consider only essential genes. Initially, the states corresponding to *Mn*, *MN*, and *mN* are taken as functionally equivalent, i.e., the transitions between these states are neutral. We suspend the neutral constraint below in the section on selection.

The evolutionary scheme for gene transfer and mutation will be

$$\begin{array}{ccc} & \overset{c_1}{\underset{u_1}{\rightleftarrows}} & \\ Mn & & MN \\ \downarrow u_2' & & u_2 \downarrow\uparrow c_2 \\ (mn) & \underset{u_1}{\leftarrow} & mN \end{array} \qquad (1)$$

A mitochondrial gene is transferred as a functional gene into the nucleus with rate c_1 (per cell and per unit time). Based on the data of Thorsness and Fox (1990, 1993), we expect $c_1 = 10^{-5}P_{\text{adr}}$, where P_{adr} is the probability that the gene product receives the appropriate address for reimport into the mitochondria before the gene copy is inactivated. We note that the transfer experiments by Thorsness and Fox (1990, 1993) involved genes that functioned on a plasmid in the nucleus. Thus, P_{adr} should also include the probability of incorporation into the nuclear chromosomes. P_{adr} is not known, but for illustrative purposes we will assume in the calculations below some small number, $P_{\text{adr}} = 1\text{–}10^{-3}$, where the value 1 is assigned to those genes that do not require an addressing sequence to transport their protein products to the mitochondria (Schatz 1996; Neupert 1997).

The nuclear copy is inactivated by mutation with rate u_1, while the mitochondrial allele is inactivated with rate u_2. If the fixation in a cell line is a neutral event in the cellular population, as is assumed for $MN \rightarrow mN$, the effective substitution rate in the cell line is also the same as the rate of appearance of the mutation, u_2 (Kimura 1962). Thus, the rate u_2 implies both the mutational appearance in one mitochondrion and the subse-

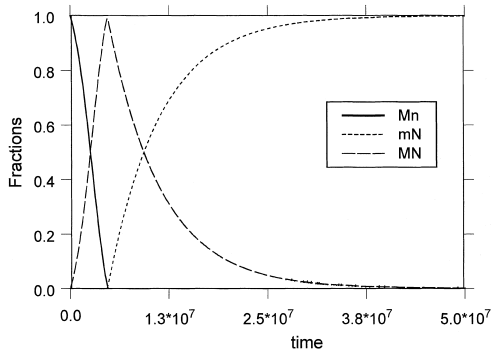

FIG. 1.—Results for the infinite haploid scheme (scheme 1) using $c_1 = 10^{-7}$, $c_2 = 10^{-10}$, $u_1 = 5 \times 10^{-8}$, and $u_2' = u_2 = 2 \times 10^{-7}$.

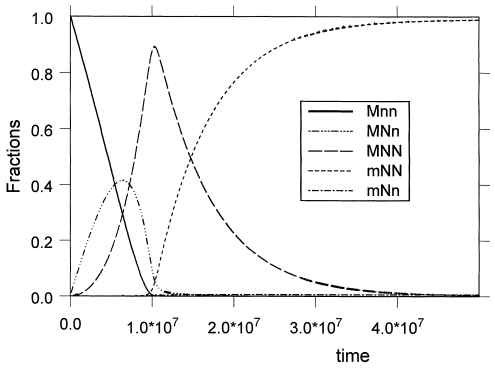

FIG. 2.—Results for the infinite diploid scheme (scheme 3) using $c_1 = 10^{-7}$, $c_2 = 10^{-10}$, $u_1 = 5 \times 10^{-8}$, and $u_2' = u_2 = 2 \times 10^{-7}$.

quent fixation of that mutation within the intracellular mitochondrial population. The process $Mn \rightarrow mn$ is, of course, not a neutral process for essential genes. Here, there is progressive loss of progeny as the number of mutant mitochondria increases until there are no functional mitochondria in that cell line. Thus, the rate u_2' for $Mn \rightarrow mn$ is expected to be somewhat different from u_2 for $MN \rightarrow mN$. Nevertheless, the value of u_2' is not crucial for the results of the model.

The rate of gene inactivation in contemporary bacteria is around 10^{-6}–10^{-8} per generation (Drake 1991). Here, an inactivating mutation is essentially irreversible, since it can be generated at many sites and in many ways, while reversion through back mutations for any one such mutant can occur only at one (or at most a few) sites. For completeness, we consider also the possibility of gene transfer from the nucleus to the mitochondria with rate c_2. Thorsness and Fox (1990, 1993) never observed such an event and estimated $c_2 < 10^{-10}$, which for our purposes is negligible. We consider significantly larger values as well, because such transfers may be more efficient in organisms other than yeast.

In order to be realistic, the scheme also needs to account for the consequences of mating and gene exchange. We assume first that a mating between two host cells implies an exchange of nuclear genes. Then, only mating between the genetic variants *Mn* and *mN* will lead to changes in the phenotypes of the progeny:

$$Mn + mN \rightarrow MN + mn. \tag{2a}$$

Alternatively, we could consider a mating between two host cells as the copying of coding sequences in one nuclear genome by homologs in the other. Then, there are two possible results of a mating:

$$Mn + mN \rightarrow \begin{cases} Mn + (mn) \\ MN + mN. \end{cases} \tag{2b}$$

In effect, there is little difference between these two pictures in the sense that the rates may differ but the consequences are the same.

In an infinite population, all variants are present simultaneously and the state of the population can be described by the fraction of cells that are represented by each genotype. The equations are described in the appendix; the results for $u_2 > u_1$ are shown in figure 1. Due to the ease of generating the lethal *mn* genotype by mating, *mN* and *Mn* cannot be present simultaneously at significant frequencies. This incompatibility is clearly visible and accentuates the sharpness of the transition to takeover by *mN*. Thus, from scheme (1), there are two possible coexistent pairs: *Mn* + *MN* or *MN* + *mN*. Which of these pairs is present is determined largely by the ratio of u_2 to u_1. If $u_2 > u_1$, we find *mN* and *MN* approximately in the ratio $(u_2 - u_1)/c_2$. If, on the other hand, $u_2 < u_1$, there is no takeover, and the end result is *Mn* and *MN* roughly in the ratio $(u_1 - u_2)/c_1$. In this limit, there will always be more *Mn* than *mN* present, and the mating terms (eqs. 2a and b), will remove all newly created *mN*. Thus, a strong polarity in the transfer rates, $c_1 \gg c_2$, is not sufficient by itself to drive the transfer to the nucleus in this case. In an infinite population, effective transfer requires a mutational drive with $u_2 > u_1$. In other words, as long as mutations are more frequent in the mitochondria than in the nucleus, there is a very distinct deterministic behavior leading to the fixation of *mN* independent of the initial state.

In these calculations, we assume that $u_2' = u_2$. Setting $u_2' > u_2$ does not change the shape of the curves but speeds up the takeover somewhat. If the genetic state *mn* is not lethal but strongly counterselected (by $s < 0$), the result looks very much like that of figure 1 as long as $|s| \gg u_1$. Smaller values of $|s|$, however, give incomplete takeover and lead to coexistence of *Mn* and *mN*.

Diploid Cells at Infinite Population Size

Very similar results are obtained for diploid nuclear genomes in an effectively infinite population size. In this case, there are five possible genetic states (*Mnn*, *MNn*, *MNN*, *mNN*, *mNn*) plus one lethal state (*mnn*). The transfer processes can be described for diploids in the following diagram:

$$\begin{array}{ccccc} Mnn & \underset{u_1}{\overset{c_1}{\rightleftarrows}} & MNn & \underset{2u_1}{\overset{c_1}{\rightleftarrows}} & MNN \\ u_2' \downarrow & & u_2 \downarrow\uparrow c_2 & & u_2 \downarrow\uparrow 2c_2 \\ (mnn) & \underset{u_1}{\overset{\leftarrow}{}} & mNn & \underset{2u_1}{\overset{\leftarrow}{}} & mNN \end{array} \tag{3}$$

Here, it is assumed that the mating between cells

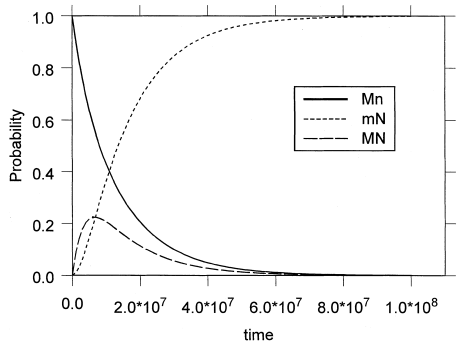

FIG. 3.—The results as described here are the same for a finite population size either using the haploid scheme (scheme 4) with $c_1 = 10^{-7}$, $c_2 = 10^{-10}$, $u_1 = 5 \times 10^{-8}$, and $u_2 = 2 \times 10^{-7}$ or using the diploid scheme (scheme 8) with $c_1 = 2 \times 10^{-7}$, $c_2 = 5 \times 10^{-11}$, $u_1 = 10^{-7}$, and $u_2 = 2 \times 10^{-7}$.

leads to recombination and segregation of nuclear genes from both cells according to the usual rules. In contrast, only the maternal mitochondrial genes are passed on to the progeny. Here, we assume that only one cell within each mating pair is the "mother."

The results of the diploid model in scheme (3) are very similar to those for the haploid case. When $u_2 > u_1$, there is a takeover by *mNN*, and this takeover takes place in two steps with *MNN* as a dominant intermediate (fig. 2). When there are no matings, the takeover on the same timescale is more gradual, and *MNN* is not a dominant intermediate. The timescale for the takeover is also roughly the same as that for the haploid case, with equivalent values of c_1, u_1, and u_2.

Haploid or Diploid Cells in a Finite Population

In the infinite population, no variant is ever totally lost, since even the smallest fraction corresponds to an infinitely large number of individuals. Here, a population will behave as infinite for changes with rate u if $uN_H \gg 1$, where N_H is the population size of the host cells. On the other hand, for transfer processes operating within a finite population, *mN* can be fixed by random drift. In a finite population, transitions between genetic states are rare, but they are rapid when they do occur. Accordingly, we consider the probabilities that the populations are in genetic states *Mn*, *MN*, or *mN*, respectively. Furthermore, since *Mn* and *mN* are not likely to occur simultaneously, there will be little contribution from mating terms. In fact, whenever *Mn* and *mN* do occur together, mating is most likely to destroy both and recreate *MN* (cf. eqs. 2a and b). Thus, any appearance of *mN* before *MN* has taken over completely will only hasten the disappearance of *Mn*, thereby, effectively increasing the rate c_1. Thus, for the finite population, there is little coexistence of different genotypes, except during the short periods of actual takeover. In this case, the haploid scheme (scheme 1) simplifies to

$$Mn \underset{u_1}{\overset{c_1}{\rightleftarrows}} MN \underset{c_2}{\overset{u_2}{\rightleftarrows}} mN \qquad (4)$$

Scheme (4) is a standard two-step transformation that can be solved analytically. If the transfer rates (c_1 and c_2) to *mN* and *Mn*, respectively, are smaller than the mutation rates, *MN* will only be a short-lived transition state on an evolutionary timescale. The probability of finding the population in state *mN* rather than *Mn* is determined by the ratio c_1u_2/c_2u_1. Thus, the probability of eventual takeover by *mN* is given by

$$P_{mN} = c_1u_2/(c_2u_1 + c_1u_2 + c_1c_2). \qquad (5)$$

In this case, a strong polarity in the transfer rates, $c_1 \gg c_2$, is extremely important. This polarity will drive the effective transfer even if mutations cannot do so, which is an important difference between the finite and infinite population schemes. According to this finite scheme, the effective rate of transfer, $Mn \rightarrow mN$, can be written as

$$R_1 = c_1u_2/(u_1 + u_2 + c_1). \qquad (6)$$

If there is a nonvanishing rate of gene transfer to mitochondria as well, $c_2 > 0$, the effective rate of reversal, $mN \rightarrow Mn$, will be

$$R_2 = c_2u_1/(u_1 + u_2 + c_2). \qquad (7)$$

In effect, this scheme suggests that for a finite host population, the transfer is an inevitable consequence of neutral drift, as long as a polarity in transfer, $c_1u_2 \gg c_2u_1$, exists. Since these changes were assumed to be neutral, the rates of fixation (c_1, u_1, and u_2) in the population are the same as the rates of appearance in a single cell. If the host population is small, it takes a larger selective advantage or disadvantage for a given genotype to affect the fixation probability than in it does in large populations. Accordingly, the transitions to *MN* and *mN* in small populations are more likely to be effectively neutral, as assumed in these calculations.

For a diploid in a finite population, the transfer scheme can be reduced in a similar way. If c_1 is the rate of transfer of one gene copy in one cell, the population will begin with one individual in state *MNn* and the rest of the cells in state *Mnn*. For a neutral change, the fixation probability will be $1/2N_H$, so the effective rate of transformation of the population from state *Mnn* to state *MNN* is $c_1/2$. Similarly, an inactivation in a nuclear gene copy would be fixed in the population with rate $u_1/2$. Thus, the scheme will be

$$Mnn \underset{u_{1/2}}{\overset{c_{1/2}}{\rightleftarrows}} MNN \underset{2c_2}{\overset{u_2}{\rightleftarrows}} mNN \qquad (8)$$

and the effective transfer rate from $Mnn \rightarrow mNN$ is

$$R_1 = c_1u_2/(u_1 + 2u_2 + c_1). \qquad (9)$$

Apart from some factors of 2, the scheme for a finite diploid population is identical to that for a finite haploid population.

The results for a finite population size are displayed in figure 3, using the same parameters as in figure 1, and in figure 4 for a range of different parameter values. The calculations describe the probabilities that the population is of one particular genotype, but they do not describe the transition states when the population is

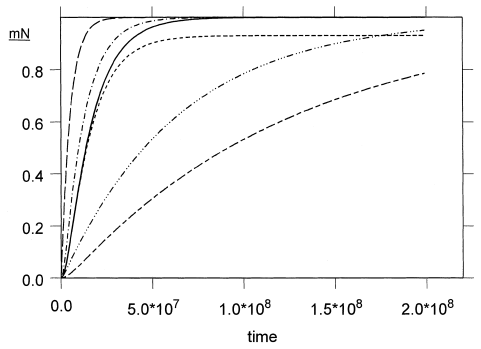

FIG. 4.—Results for the haploid scheme with a finite population size. Only the probability for the takeover by variant *mN* is shown. The solid curve uses the same parameter values as in figures 1–3. The other curves were generated from the solid one by changing one parameter value, from left to right: curve 1 with $c_1 = 10^{-5}$; curve 2 with $u_2 = 5 \times 10^{-7}$; curve 3 (solid) with $c_1 = 10^{-7}$, $c_2 = 10^{-10}$, $u_1 = 5 \times 10^{-8}$, and $u_2 = 2 \times 10^{-7}$; curve 4 with $c_2 = 10^{-8}$; curve 5 with $u_1 = 10^{-6}$; curve 6 with $c_1 = 10^{-8}$.

mixed. The probability for a takeover by *mN* is given by equation (5), and the timescale for the takeover is determined by the inverse of the effective transfer rate given in equation (6). These two equations describe all the results shown in figures 3 and 4. In most of the calculations, we used a strong polarity, $c_1/c_2 = 10^3$. When the polarity is weaker than expected from the data of Thorsness and Fox (1990), the takeovers are less complete (see curve 4 in fig. 4). If the mutation rates u_1 and u_2 are of similar magnitudes and c_1 is smaller, the takeover time is determined primarily by $1/c_1$ (see curve 6 in fig. 4). Assuming that the probability of attaching an addressing sequence for reimport into the mitochondria of the gene product, P_{adr}, is on the order of $1–10^{-3}$, the takeover time may be on the order $10^6–10^8$ generations. If c_1 is much larger than the mutation rates, e.g., if $P_{adr} = 1$, the effective transfer rates from equations (6) and (9) will simply be equal to the mutational inactivation rate, u_2 (see curve 1 in fig. 4). If $u_1 \gg u_2$ while retaining a strong polarity, $c_1u_2 \gg c_2u_1$, the takeover is considerably slowed down (see curve 5 in fig. 4).

If the rate of copying of a gene into the nucleus is very large, one would expect every present-day mitochondrial gene to have a large number of inactive copies in the nucleus. The average number would be given by the ratio of the copying rate and a rate of removal of nuclear genes. We did not find any significant amount of present-day mitochondrial gene sequences in the yeast chromosomes. For this reason, we suggest that the copying rate is smaller than the removal rate in modern yeast.

The Influence of Selection

The results so far were calculated on the basis of two assumptions: (1) that there is no difference in efficiency (fitness) between the variants *Mn*, *MN* and *mN* and the gene under study is essential so that the genotype *mn* is lethal, and (2) that all alleles within a given genotype, such as *M* or *m*, have identical phenotypes. There are, however, some departures from these assumptions that could be relevant to real populations.

First, it is conceivable that a particular gene may be more effective if it is expressed within the mitochondria rather than in the cytoplasm. In other words, the configuration *Mn* may be significantly more fit than *mN*. In such cases, selection will oppose the transfer to the nucleus of the mitochondrial gene. This sort of inequality may in fact be relevant to the relatively small number of genes that have been retained in modern mitochondria (see *Discussion*).

Second, we must also consider the contingencies for which there are selective advantages that provide an additional drive for the transfer. Such advantages may be expressed in the phenotype of the host cell or that of the mutated mitochondrion. Thus, when the mutant mitochondrial phenotype, *m*, leads to an intracellular advantage for mitochondria carrying the mutation, the fixation rate u_2 will be larger than the nominal rate of mutation. For example, mutational inactivation or deletion of a mitochondrial gene might lead to faster replication of the mutant genome. However, this sort of intracellular propagation advantage (or disadvantage) initially will only affect the small number of mitochondria within one cell. Indeed, it is not necessarily so that a cell with more rapidly propagating mitochondria will have a growth advantage (or disadvantage) over other cells. When the rapidly propagating mitochondria do not influence the cell growth rates, cells with this mitochondrial variant will fare within the population as neutral mutants. Thus, to be effective on the cellular level, the mitochondrial variant must confer an advantage on the cell that is expressed as an enhanced selection coefficient, $s' > 1/N_H$.

Indeed, a mitochondrion with a smaller genome might be propagated faster than that with a larger one for at least two reasons (Kurland 1992). First, were a mutant mitochondrion to arise with 1% of its genome deleted, this might be reflected in a ca. 1% replication rate advantage for the mutant. Likewise, a gene that is not expressed, or one that is either deleted or transferred, could provide a growth advantage for the mitochondrion by reducing resources allocated to protein synthesis. In general, we expect the inactivation and deletion of coding sequences that have been transferred to the nucleus to enhance the intracellular growth efficiency of the mitochondria.

Similarly, there might be mitochondrial phenotypes that affect the fitness of the host cell and hence affect its ability to compete at a population level. For example, two gene copies might be better or worse than one, such that *MN* has higher or lower fitness than *mN* and/or *Mn*. A nuclear copy could be more (or less) efficient than a mitochondrial one. However, we do not know of any general tendency toward such selective differences.

Other fitness differences are inherent in the model: the intermediate *MN* is protected to some extent against lethal mutations generated in $Mn \rightarrow mn$. This effectively confers a selective advantage, $s = u_2'$, on *MN* over *Mn*. When *MN* mutates to *mN*, this will lead to a decrease

in the number of *Mn* through the mating $mN + Mn \rightarrow MN + mn$. In other words, as long as *Mn* remains in the population, *mN* will not coexist passively. Rather, *mN* will tend to eliminate *Mn*. Consequently, *MN* also has an advantage, $s = u_2$, over *Mn*. Thus, in the process of fixing *MN* over *Mn*, there is a selective advantage $s = u_2 + u_2'$. This will be important only if the host cell population (N_H) is very large, $u_2 N_H > 1$, which seems unlikely to be a general condition.

The transfer of a mitochondrial gene to its nuclear partner may initiate a selection process at the level of the cell population. Initially, there would be a single *MN* variant in a population of *Mn*. If the mitochondria carry a diversity of alleles for this gene, the transfer will involve just one (*N**) of these alleles. The transferred allele could be better or worse than the majority of the mitochondrial ones. If, by chance, *MN* has received a better allele, *MN**, than the average of *M*, then there would be a selective advantage of *MN** over *Mn*. Since there is no linkage between the nuclear gene and the mitochondrial one, *MN** would, on average, carry an average mitochondrial *M* gene.

There are now two mutually exclusive possibilities: either $MN^* \rightarrow mN^*$ or $MN^* \rightarrow Mn$. If we assume that there are equal expression levels and an averaged fitness for the combination of gene products formed by *MN**, we can deduce the consequences of these two alternative outcomes. Hence, when *N** is a worse-than-average allele in the nucleus, there is a selective disadvantage of *mN** compared with *MN** and a corresponding advantage of *Mn* over *MN**. Such inequalities would favor a return to *Mn* rather than fixation of *mN**. On the other hand, a better-than-average allele *N** would be expected to confer a selective advantage for *mN** over *MN** and a disadvantage for *Mn* over *MN**. This in turn would shift the outcome in favor of fixation of *mN** rather than a return to *Mn*.

Both of these scenarios suggest that mitochondrial genes that are fixed in the nucleus will tend to correspond to alleles with greater fitness on average than those originally distributed among the mitochondria. In effect, the transfer to the nucleus should be attended by a degree of purifying selection. The magnitude of this effect can be calculated (see appendix); the increase in average fitness after transfer is

$$\langle \delta s \rangle = \frac{3u_1 + u_2}{2(u_1 + u_2)} N_H \sigma_s^2 \approx N_H \sigma_s^2, \qquad (10)$$

where σ_s^2 is the variance in the fitness of the alleles of the mitochondrial gene and N_H is the size of the cell population. This can be a significant effect only if the diversity of mitochondrial alleles is large, a condition that could be particularly important where mitochondria are subject to Müller's ratchet. In this case, the majority class of mitochondrial genomes is not a clone, but a mixture of genetically diverse variants, each with a roughly equivalent mutant load that is distributed over the different coding sequences. In this case, there would be significant diversity for the load in individual genes. Hence, the purifying effect in the transfer process could well be large, and the effective transfer rate would increase to a corresponding degree.

Discussion

As expected from the outset, if the mutation rate is higher in the mitochondrion than in the nucleus, a transfer process combined with purifying selection will lead to the preferential accumulation of mitochondrial coding sequences in the nuclei of cells in effectively infinite populations. Here, even with a highly polar transfer process, a higher mutation frequency in mitochondria than in nuclei is required for the disappearance of a gene from the mitochondrion and its fixation in the nucleus. The model for infinite population size is of doubtful relevance to most natural populations, but it does represent an important limiting case with distinct deterministic behavior. In particular, it is useful for demonstrating the incompatibility of some of the mixed genotypes, such as *Mn* and *mN*. Furthermore, the mutation rates for the mitochondria of many organisms are not significantly higher than those of the corresponding nuclei (Wolfe, Li, and Sharp 1987; Palmer and Herbon 1988; Lynch 1996, 1997). Therefore, we need to look to finite populations for a model of gene transfer that is more generally applicable.

A different constraint is required for the fixation of mitochondrial genes in nuclei and their disappearance from the organelle genome in the more realistic situation of finite populations. Thus, finite populations are not deterministic in their behavior. Instead, they fix genes in nuclei and lose genes from mitochondria by genetic drift, i.e., by statistical fluctuations. Accordingly, the relative mutation frequencies will influence the average rates of transfer, but they do not by themselves determine the end point of the process. Here, even if the mutation frequencies in mitochondria are lower than those in nuclei, as long as the cellular processes are biased to a comparable degree to favor transfer from mitochondria to nuclei, mitochondrial genes will be fixed in the nucleus.

Happily, we found in the yeast literature the prototype of a preferential transfer mechanism from mitochondria to nucleus (Thorsness and Fox 1990, 1993). The reasons why we expect this process to be highly polar in all organisms are straightforward. The introduction into mitochondria of exogenous coding sequences such as those on plasmids is notoriously difficult. It has been accomplished routinely only in yeast, but even there it requires ''high velocity microprojectile bombardment'' (Fox et al. 1990). In other words, the one experimental system that permits the transfer of coding sequences into mitochondria requires that the exogenous sequences be shot into the cells to obtain the appropriate physical penetration of some mitochondria. This suggests that there are normally effective physical barriers to the transfer of coding sequences into mitochondrial genomes that would bias transfer in favor of nuclei.

In addition, the vacuoles that are responsible for mediating the transfer of nucleic acid fragments to the

nuclei of yeast cells are ubiquitous (Thorsness, White, and Fox 1993; Hanekamp and Thorsness 1996; Weber, Hanekamp, and Thorsness 1996; Campbell and Thorsness 1998). Apparently, the vacuoles facilitate transfer by degrading mitochondria and thus liberating nucleic acid fragments suitable for incorporation into nuclei. Since the reverse process of nuclear digestion and transfer of the resulting nucleic acid fragments to mitochondria would be lethal to a cell, the vacuole-dependent transfer is necessarily unidirectional. For these reasons, we are inclined to believe that preferential transfer of coding sequences to nuclei is the rule rather than the exception. Nevertheless, the transfer bias is not absolute. An occasional transfer from nucleus to mitochondria may occur as a result of other processes, particularly in cells for which the mitochondrial mutation rates are relatively low, such as those of plants (Thorsness and Weber 1996; Martin and Herrmann 1998). Indeed, it would be useful to obtain estimates of gene transfer rates from mitochondria to nuclei in many more organisms, since it is unlikely that the absolute values of transfer parameters estimated for yeast are relevant to all cells.

In the transfer models based on either finite or infinite populations, the mutational load on mitochondrial coding sequences is expected to hasten the fixation of these genes in the nuclear genome. In both cases, the influence of purifying selection is maintained at the cellular level (Bergstrom and Pritchard 1998). Here, we suggest that the transfer and fixation of mitochondrial sequences in the nucleus will tend to favor the most fit alleles of the mitochondrion. Accordingly, the transfer process may be viewed as a mode of purifying selection (see appendix).

The degree of purifying selection supported by the transfer process will depend very much on the aggregate sequence variation within the populations of mitochondria. In other words, the mutation frequencies of the mitochondrial genomes determine the range of fitness within which the transfer process can winnow the variants (see appendix). We expect that, all else being equal, the sizes of the mitochondrial genomes would be correlated with their total sequence variability. Accordingly, small contemporary mitochondria might not provide particularly fertile ground for this purification effect. However, the putative endosymbiotic ancestors of mitochondria are thought to have had genomes with sequence lengths in the millions (Andersson and Kurland 1999; Gray, Burger, and Lang 1999). For this reason, we suggest that the purifying selection provided by transfer of sequences from mitochondria to nucleus was very much more important earlier in the evolution of the organelle than it is at present.

Transfer to the nucleus is but one of several mechanisms leading to the reduction in the coding capacity of mitochondrial genomes (Andersson and Kurland 1991, 1998; Kurland 1992). Mutational meltdowns of dispensable sequences and deletions by replication slippage, as well as by intrachromosomal recombination, are also likely contributors to the reductive evolution of the mitochondrial genome (reviewed in Andersson and Kurland 1998). Indeed, all of these reductive mechanisms would have been favored by competition within a cell between mitochondria in which those with smaller genomes could propagate faster than the larger ones (Kurland 1992). This mode of selection would tend to reinforce the cellular selection mechanism that is implicit in the nuclear-transfer model.

Purifying selection can also oppose the transfer of genes to the nucleus if the corresponding gene products are less efficient when expressed in the cytosol rather than in the mitochondrion. A simple example of this would be a protein that is properly coded in the nucleus and expressed in the cytosol but inefficiently imported into the mitochondrion. Indeed, a coding sequence transferred to the nucleus is often provided with specific addressing signals to direct its product back to the mitochondrion via a specific transport system (Baker and Schatz 1991; Schatz 1996; Neupert 1997). The great complication here is the variation in the requirements for a more or less defined addressing sequence to guide the protein back into the organelle. Some proteins require no addressing signal at all, others require rather ill defined leader sequences, and still others require a more strictly defined consensus sequence (Baker and Schatz 1991; Kadowski et al. 1996; Schatz 1996; Neupert 1997).

Obtaining a very specific addressing sequence when it is needed might be a very slow process if the agency of chromosome recombination is the sole tagging mechanism. Obviously, the joining of the required signal sequences to newly transferred coding sequences would be greatly facilitated by mechanisms similar to contemporary splicing pathways for exons (Nugent and Palmer 1991; Covello and Gray 1992; Long et al. 1996).

Potential intermediates in the process of gene transfer from mitochondria to nucleus have been identified in recent events in the evolution of legumes (Nugent and Palmer 1991; Covello and Gray 1992). Briefly, the gene for cytochrome oxidase subunit 2 (*cox2*) is most often found in the mitochondrial genomes of plants. Nevertheless, among the legumes, transcriptionally active *cox2* alleles are found in the nucleus, and they may or may not be accompanied by mitochondrial alleles. The exceptional legume here is the pea that seems to have a transcriptionally inactive nuclear allele and an active mitochondrial allele of *cox2*. In addition, the sequence of the nuclear copy seems not to have been copied from the unedited mitochondrial DNA, from which it was inferred that the intermediate for the original transfer event was an edited RNA copy. The finding of an intron between the transit signal sequence and the structural sequence of *cox2* suggests that exon shuffling may have generated the active nuclear version. Finally, the phylogeny of the legumes and the patterns of active *contra* inactive alleles of *cox2* are consistent with a simple transfer scheme (Nugent and Palmer 1991; Covello and Gray 1992).

There is, however, a caveat. Bittner-Eddy, Monroy, and Bramble (1994) have found that a gene present in *Neurospora crassa* both in the nucleus and in the mitochondrion may be transcriptionally inactive or active

in either genome at different stages in the life cycle of the organism. Thus, it is possible that further studies of the expression states of *cox2* in the mitochondria and nuclei of legumes may reveal a more complex expression pattern than that reported initially.

Numerous other examples of genes that have been transferred from mitochondria to nuclei and then recruited for functions outside of the mitochondria can be found in the protein databases. Often, such genes are present in the nucleus as gene duplications with specific cellular targets such as the mitochondrion, the cytosol, and the peroxisome. Most spectacular is the recent discovery in *Arabidopsis thaliana* chromosome 2 of a nearly complete duplication of the entire mitochondrial genome (Lin et al. 1999). Lin et al. (1999) have tentatively concluded that the transfer of this fragment is a relatively recent event.

There is still another dimension to the transfer of genes from bacterial genomes to the nuclear genomes of eukaryotes. Thus, a large number of genes with strong affinities to bacterial homologs and encoding enzymes with metabolic functions in the cytosol of eukaryotes have been identified (Doolittle et al. 1996; Keeling and Doolittle 1997; Feng, Cho, and Doolittle 1997; Rivera et al. 1998). However, in addition to putative α-proteobacterial homologs, there are a majority of other gram-negative and gram-positive bacterial groups that seem to be represented among the nuclear-encoded genes of bacterial origins. It has been suggested that these were transferred to the nuclear genomes from the endosymbiotic bacteria that seeded the evolution of chloroplasts, hydrogenosomes, and mitochondria (Gogarten, Olendzenski, and Hilario 1996; Feng, Cho, and Doolittle 1997; Martin and Schnarrenberger 1997; Martin and Muller 1998). It is also possible that many of these bacterial homologs were transferred to nuclear genomes following the digestion of bacteria by the vacuole-dependent mechanism described above. The point is that cell vacuoles may have facilitated such transfers from diverse bacteria that were not involved in the endosymbiotic events leading to organelle formation.

This brings us to the inevitable question of why there are any genes left in contemporary mitochondrial genomes. It is possible that those genes that remain are also on their way to the nucleus and that sequence technology has discovered the stragglers. Of course, the essential genes that code products that are the most difficult to reroute to the mitochondria from the cytosol might be the last ones to leave the mitochondrial genome. Here, it is worth recalling that there are only three mitochondrial proteins that are found in all mitochondria; these are the three respiratory chain proteins of the *P. falciparum* mitochondrial genome (Gray, Burger, and Lang 1999). All of the other proteins coded by mitochondrial genomes are missing from the genome of at least one other mitochondrion (Gray, Burger, and Lang 1999). Hence, it is possible that aside from these three proteins, the other mitochondrial proteins coded in mitochondrial genomes are simply stragglers.

On the other hand, we have noted that there could be mitochondrial gene products for which the transfer to the nucleus is not a neutral event. Indeed, the identities of the three proteins found in the *P. falciparum* mitochondrial genome are very suggestive. They may represent examples of a selected core of genes in the organelle's genome that must be coded by the mitochondrial genome so that they can regulate gene expression in the organelles (Allen 1993). Thus, there is direct support for the idea that some proteins of the electron transport chain are exploited to provide a regulatory coupling between gene expression and the redox potential of chloroplasts, as well as of mitochondria in plants (Allen 1993; Allen, Hakansson, and Allen 1995; Galvis, Allen, and Hakansson 1998; Pfannschmidt, Nilsson, and Allen 1999; Race, Herrman, and Martin 1999). The *cox1, cox3,* and *cob* genes found in the minimalist genome of *P. falciparum*'s mitochondrion could provide this essential coupling. This is a particularly attractive interpretation, since these are the only proteins that are found in all characterized mitochondria (Gray, Burger, and Lang 1999).

There are other mitochondrial-encoded proteins that could conceivably provide this regulatory coupling. However, they all belong to the group of 64 proteins that are found in the mitochondrial genome of *R. americana* but have been lost from the organelle's genome in one organism or another. At this point, it is not possible to say with confidence why this small, variable core of proteins remains in some mitochondrial genomes but not in others.

Acknowledgments

We thank Marguerite Picard for invaluable advice. We also thank her along with Siv Andersson for critical readings of the manuscript. The Swedish Natural Sciences Research Council supported this work.

LITERATURE CITED

ALLEN, J. F.. 1993. Control of gene expression by redox potential and the requirement for chloroplast and mitochondria genomes. J. Theor. Biol. **165**:609–631.

ALLEN, C. A., G. HAKANSSON, and J. F. ALLEN. 1995. Redox conditions specify the proteins synthesized by isolated chloroplasts and mitochondria. Redox Rep. **1**:119–123.

ANDERSSON, S. G. E., and C. G. KURLAND. 1991. An extreme codon preference strategy: codon reassignment. Bioessays **14**:709–714.

———. 1998. Reductive evolution of resident genomes. Trends Microbiol. **6**:283–268.

———. 1999. Origins of mitochondria and hydrogenosomes. Curr. Opin. Microbiol. **2**:535–541.

ANDERSSON, S. G. E., A. ZOMORODIPOUR, J. O. ANDERSSON, T. SICHERITZ-PONTEN, U. C. M. ALSMARK, R. M. PODOWSKI, A. K. NÄSLUND, A.-S. ERIKSSON, H. H. WINKLER, and C. G. KURLAND. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature **396**:133–143.

ATTARDI, C. 1985. Animal mitochondrial DNA an extreme example of genetic economy. Int. Rev. Cytol. **93**:93–145.

BAKER, K. P., and G. SCHATZ. 1991. Mitochondrial proteins essential for viability mediate protein transport into yeast mitochondria. Nature **349**:205–208.

BERGSTROM, C. T., and J. PRITCHARD. 1998. Germline bottlenecks and evolutionary maintenance of mitochondrial genomes. Genetics **149**:2135–2146.

BITTNER-EDDY, P., A. F. MONROY, and R. BRAMBLE. 1994. Expression of mitochondrial genes in the germinating conidea of *Neurospora crassa.* J. Mol. Biol. **235**:881–897.

BULMER, M. 1987. Coevolution of the genetic code and transfer RNA abundance. Nature **352**:728–730.

CAMPBELL, C. L., and P. E. THORSNESS. 1998. Escape of mitochondrial DNA to the nucleus in *yme1* yeast is mediated by vacuolar-dependent turnover of abnormal mitochondrial compartments. J. Cell Sci. **111**:2455–2464.

COVELLO, P. S., and M. W. GRAY. 1992. Silent mitochondrial and active nuclear genes for subunit 2 of cytochrome c oxidase (cox2) in soybean: evidence for RNA-mediated gene transfer. EMBO J. **11**:3815–3820.

DOOLITTLE, R. F., D.-F. FENG, S. TSANG, G. CHO, and E. LITTLE. 1996. Determining divergence times of the major kingdoms of living organisms with a protein clock. Science **271**: 470–477.

DRAKE, J. W. 1991. A constant rate of spontaneous mutation in DNA-based microbes. Proc. Natl. Acad. Sci. USA **88**: 7160–7164.

FELSENSTEIN, J. 1974. The evolutionary advantages of recombination. Genetics **78**:737–756.

FENG, D.-F., G. CHO, and R. F. DOOLITTLE. 1997. Determining divergence times 3ith a protein clock: update and reevaluation. Proc. Natl. Acad. Sci. USA **94**:13028–13033.

FOX, T. D., L. S. FOLLEY, J. J. MULERO, T. W. MCMULLIN, P. E. THORSNESS, L. O. HEDIN, and M. C. COSTANZO. 1990. Analysis of yeast mitochondrial genes. Methods Enzymol. **194**:149–165.

GALVIS, M. L. E., J. F. ALLEN, and G. HAKANSSON. 1998. Mitochondrial gene expression in Pea IS regulated by the activity of respiration complex II. Pp. 171–174 *in* I. M. MOELLER, P. GARDESTROEM, K. GLIMELIUS, and E. GLASER, eds. Plant mitochondria: from gene to function. Backhuys, Leiden, The Netherlands.

GOGARTEN, J. P., L. OLENDZENSKI, and E. HILARIO. 1996. Dating the cenancestor of organisms. Science **274**:1750–1751.

GRAY, M. W. 1992. The endosymbiont hypothesis revisited. Int. Rev. Cytol. **141**:233–257.

GRAY, M. W., G. BURGER, and B. F. LANG. 1999. Mitochondrial evolution. Science **283**:1476–1481.

HANEKAMP, T., and P. E. THORSNESS. 1996. Inactivation of YME2, an integral inner mitochondrial membrane protein, causes increased escape of DNA from mitochondria to the nucleus in *Saccharomyces cerevisiae.* Mol. Biol. Cell **16**: 2764–2771.

KADOWSKI, K., N. KUBO, K. OZAWA, and A. HIRAI. 1996. Targeting presequence acquisition after mitochonrial gene transfer to the nucleus occurs by duplication of existing targeting signals. EMBO J. **15**:6652–6661.

KEELING, P. J., and W. F. DOOLITTLE. 1997. Evidence that eukaryotic triosephosphate isomerase is of alpha-proteobacterial origin. Proc. Natl. Acad. Sci. USA **94**:1270–1275.

KIMURA, M. 1962. On the probability of fixation of mutant genes in populations. Genetics **47**:713–719.

KUNDIG, C., H. HENNEKE, and M. GOTTFERT. 1993. Correlated physical and genetic map of *Bradyrhizobium japonicum* genome. J. Bacteriol. **175**:613–622.

KURLAND, C. G. 1992. Evolution of mitochondrial genomes and the genetic code. Bioessays **14**:709–714.

LIN, X. Y., S. S. KAUL, S. ROUNSLEY et al. (37 co-authors). 1999. Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana.* Nature **402**:761–768.

LONG, M., S. J. DE SOUZA, C. ROSENBERG, and W. GILBERT. 1996. Exon shuffling and the origin of the mitochondrial targeting function in plant cytochrome c 1 precursor. Proc. Natl. Acad. Sci. USA **93**:7727–7731.

LYNCH, M. 1996. Mutation accumulation in transfer RNAs: molecular evidence for Muller's ratchet in mitochondrial genomes. Mol. Biol. Evol. **13**:209–220.

———. 1997. Mutation accumulation in nuclear, organelle and prokaryotic transfer RNA genes. Mol. Biol. Evol. **14**:914–925.

MARTIN, W., and R. G. HERRMANN. 1998. Gene transfer from organelles to the nucleus: how much, what happens, and why. Plant Physiol. **118**:9–17.

MARTIN, W., and M. MULLER. 1998. The hydrogen hypothesis for the first eukaryote. Nature **392**:37–41.

MARTIN, W., and C. SCHNARRENBERGER. 1997. The evolution of the Calvin Cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis. Curr. Genet. **32**:1–18.

MARTIN, W., B. STOEBE, V. GOREMYKIN, S. HANSMANN, M. HASEGAWA, and K. V. KOWALLIK. 1998. Gene transfer to the nucleus and the evolution of chloroplasts. Nature **393**: 162–165.

NEUPERT, W. 1997. Protein import into mitochondria. Annu. Rev. Biochem. **66**:683–717.

NUGENT, J. M., and J. D. PALMER. 1991. RNA-mediated transfer of the gene coxII from the mitochondrion to the nucleus during flowering plant evolution. Cell **66**:473–481.

PALMER, J. D., and L. A. HERBON. 1988. Plant mitochondrial DNA evolves rapidly in structure but slowly in sequence. J. Mol. Evol. **28**:87–94.

PFANNSCHMIDT, T., A. NILSSON, and J. F. ALLEN. 1999. Photosynthetic control of chloroplast gene expression. Nature **397**:625–628.

RACE, H. L., R. G. HERRMAN, and W. MARTIN. 1999. Why have organelles retained genomes? Trends Genet. **15**:364–370.

RIVERA, M. C., R. JAIN, J. MOORE, and J. A. LAKE. 1998. Genomic evidence for two functionally distinct gene classes. Proc. Natl. Acad. Sci. USA **95**:6239–6244.

ROUX, V., and D. RAOULT. 1995. Inter- and intraspecies identification of Bartonella (Rochalimea) species. J. Clin. Microbiol. **33**:1573–1579.

SCHATZ, G. 1996. The protein import system of mitochondria. J. Biol. Chem. **271**:31763–31766.

THORSNESS, P. E., and T. D. FOX. 1990. Escape of DNA from mitochondria to nucleus in *Saccharomyces cerevisiae.* Nature **346**:376–379.

———. 1993. Nuclear mutations in *Saccharomyces cerevisiae* that affect the escape of DNA from mitochondria to the nucleus. Genetics **134**:21–28.

THORSNESS, P. E., and E. WEBER. 1996. Escape and migration of nucleic acids between chloroplasts, mitochondria, and the nucleus. Int. Rev. Cytol. **165**:207–234.

THORSNESS, P. E., K. H. WHITE, and T. D. FOX. 1993. Inactivation of *YME1,* a gene coding a member of the *SEC18, CDC48* family of putative ATPases, causes increased escape of DNA from mitochondria in *Saccharomyces cerevisiae.* Mol. Cell. Biol. **13**:5418–5426.

WEBER, E. R., T. HANEKAMP, and P. E. THORSNESS. 1996. Biochemical and functional analysis of the *YME1* gene product, an ATP and zinc-dependent mitochondrial protease from S. cerevisiae. Mol. Biol. Cell **7**:307–317.

WOLFE, K. H., W-H. LI, and P. M. SHARP. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. USA **84**:9054–9058.

APPENDIX

Equations for Gene Transfer in an Infinite Haploid Population

Let f_0 denote the fraction of cells that are *MN*, f_1 those that are *Mn*, and f_2 those that are *mN*. Since the genetic state *mn* is assumed to be lethal, each time an event leads to the formation of *mn* it is removed. In order to keep the total population size constant, all variants are increased in proportion to their present fractions. The variants, *Mn*, *MN*, and *mN* are assumed to be selectively neutral variants. We also assume that there are a total of kN_H matings in the population per unit time. Based on schemes (1) and (2a), we obtain the following equations:

$$\begin{aligned}\frac{df_0}{dt} &= k(1 + f_0)f_1f_2 - (u_1 + u_2)f_0 + u_1f_2f_0 \\ &\quad + u_2'f_1f_0 + c_1f_1 + c_2f_2 \\ \frac{df_1}{dt} &= -(1 - f_1)f_1(kf_2 + u_2') + u_1(f_0 + f_1f_2) - c_1f_1 \\ \frac{df_2}{dt} &= -(1 - f_2)f_2(kf_1 + u_1) + u_2f_0 + u_2'f_1f_2 - c_2f_2.\end{aligned} \tag{A1}$$

The terms in the rate equation, for example, for *Mn* with fraction f_1, can be explained in order of appearance as follows: $-kf_1f_2$ is the loss of *Mn* through mating with *mN* (scheme 2a).

$kf_1f_2f_1$ is the creation of the lethal *mn* through mating *Mn* + *mN*, which is subsequently redistributed to *Mn* in the proportion to f_1.

$-u_2'f_1$ is the loss of *Mn* through an inactivating mutation in *M*, which produces the lethal *mn*. This is subsequently redistributed to *Mn* in proportion to f_1 to give the term $u_2'f_1f_1$.

u_1f_0 is the mutation rate for $MN \rightarrow Mn$.

$u_1f_2f_1$ is the mutation rate for $mN \rightarrow mn$ (with rate u_1f_2), which is redistributed to *Mn* in proportion to f_1.

$-c_1f_1$ is the transfer rate for $Mn \rightarrow MN$.

Using scheme (2b) rather than scheme (2a) leads to a replacement of *k* by *k*/2 in equation (A1); this is of little consequence to the final result. The equations can easily be integrated numerically.

Purifying Effect of Gene Transfer

We consider haploid scheme (4) in a finite population. The transfer can be described in two discrete steps. First, there is fixation of a gene in the nucleus (*MN*), and thereafter there is fixation of an inactivated gene in the mitochondria (*mN*) or fixation of an inactivated gene in the nucleus (return to *Mn*). We assume that the transferred gene (*N**) has fitness *s* relative to that of the average mitochondrial gene. If the mixed genotype (*MN**) has equal contributions from the two genes, its fitness relative to the average (*Mn*) may be assumed to be *s*/2. (We also tried the assumption that the mixed genotype only expresses the best allele and obtained essentially the same results as those presented below). The rate of fixation for a variant with selective advantage *s* (or disadvantage if $s < 0$) is $2c_1N_Hs/(1 - e^{-2N_Hs})$. Thus, the fixation probability for a certain *MN** with fitness *s*/2 in a background of average *Mn* is

$$k_1(s) = \frac{c_1N_Hs}{1 - e^{-N_Hs}}. \tag{A2}$$

The gene must be copied many times in individual cells before one of them manages to take over the population. Every variant will arise from a distribution, $\rho(s)$, of alleles with different fitnesses in the mitochondria. Thus, the average rate of fixation of *MN** is

$$\begin{aligned}\langle k_1\rangle &= \int \rho(s)k_1(s)\,ds \\ &\approx c_1[1 + N_H^2\sigma_s^2/12 + O(N_H^4\sigma_s^4)] \approx c_1.\end{aligned} \tag{A3}$$

This result follows from a series expansion of $k_1(s)$ and requires that $\rho(s)$ be symmetric around $s = 0$ with variance σ_s^2. Thus, the expected rate with selection is not much different from that without selection if $N_H^2\sigma_s^2 < 10$. The probability that the gene *N** that is copied and fixed in the population has fitness *s* is

$$p(s) = \frac{k_1(s)\rho(s)}{\int k_1(s)\rho(s)\,ds}. \tag{A4}$$

The average fitness of the transferred gene can be calculated as

$$\langle \delta s\rangle_1 = \int sp(s)\,ds \approx N_H\sigma_s^2/2, \tag{A5}$$

which holds if $N_H^2\sigma_s^2 < 10$. Thus, the expected fitness of the allele that is fixed is $\langle \delta s\rangle_1$ above that of the average.

Once the gene is fixed in the nucleus (*MN**), there are two competing processes of inactivation. If the transferred allele has fitness *s* relative to the average, and an inactivation mutation occurs in the mitochondrial copy (with rate N_Hu_2), the system starts with a single individual of type *mN**. Consequently, this variant has fitness *s*/2 relative to the background *MN** and fixation probability $s/(1 - e^{-N_Hs})$. If the inactivation mutation (with rate N_Hu_1) occurs in the nuclear copy, the system starts with a single individual of type *Mn* which has fitness $-s/2$ relative to *MN**. The probability of fixation will be $s/(e^{-N_Hs} - 1)$. Taken together, the probability that type *mN** rather than *Mn* will be fixed is

$$P_{mN^*} = \frac{u_2}{u_2 + u_1e^{-N_Hs}}. \tag{A6}$$

The overall rate of fixation for $Mn \rightarrow mN^*$ is determined by

$$R_1(s) = k_1(s)P_{\mathrm{mN}^*} = \frac{c_1 N_{\mathrm{H}} s}{1 - e^{-N_{\mathrm{H}} s}} \frac{u_2}{u_2 + u_1 e^{-N_{\mathrm{H}} s}}, \quad \text{(A7)}$$

which reduces to the neutral result, equation (6), when $s = 0$. Thus, the overall probability that a transferred allele has fitness s relative to the average mitochondrial allele of that gene is

$$q(s) = \frac{u_2}{u_2 + u_1 e^{-N_{\mathrm{H}} s}} \frac{p(s)}{Q}. \quad \text{(A8)}$$

The factor $p(s)$ from equation (A3) accounts for the probability that *MN** started with an allele with fitness s, and Q is a normalization factor defined such that $\int q(s)\, ds = 1$. Thus, given a successful gene transfer $Mn \rightarrow mN^*$, the probability that the transferred allele has fitness s relative to the average of the mitochondrial alleles is $q(s)$. The expected improvement in fitness in a gene transfer can be calculated as

$$\langle \delta s \rangle = \int sq(s)\, ds \approx \frac{3u_1 + u_2}{2(u_1 + u_2)} N_{\mathrm{H}} \sigma_s^2 \approx N_{\mathrm{H}} \sigma_s^2. \quad \text{(A9)}$$

WILLIAM MARTIN, reviewing editor

Accepted March 1, 2000