

Forces maintaining organellar genomes: is any as strong as genetic code disparity or hydrophobicity?

Aubrey de Grey

Department of Genetics, University of Cambridge

Downing Street, Cambridge CB2 3EH, UK

Email ag24@gen.cam.ac.uk

Abstract

It remains controversial why mitochondria and chloroplasts retain the genes encoding a small subset of their constituent proteins, despite the transfer of so many other genes to the nucleus. Two candidate obstacles to gene transfer, suggested long ago, are that the genetic code of some mitochondrial genomes differs from the standard nuclear code, such that a transferred gene would encode an incorrect amino acid sequence, and that the proteins most frequently encoded in mitochondria are generally very hydrophobic, which may impede their import after synthesis in the cytosol. More recently it has been suggested that both these interpretations suffer from serious “false positives” and “false negatives”: genes that they predict should be readily transferred but which have never (or seldom) been, and genes whose transfer has occurred often or early, even though this is predicted to be very difficult. Here I consider the full known range of ostensibly problematic such genes, with particular reference to the sequences of events that could have led to their present location. I show that this detailed analysis of these cases reveals that they are in fact wholly consistent with the hypothesis that code disparity and hydrophobicity are much more powerful barriers to functional gene transfer than any other. The popularity of the contrary view has led to the search for other barriers which might retain genes in organelles even more powerfully than code disparity or hydrophobicity; one proposal, concerning the role of proteins in redox processes, has received widespread support. I conclude that this abandonment of the original explanations for the retention of organellar genomes has been premature. Several other, relatively minor, obstacles to gene transfer certainly exist, contributing to the retention of relatively many organellar genes in most lineages compared to animal mtDNA, but there is no evidence for obstacles as severe as code disparity or hydrophobicity. One corollary of this conclusion is that there is currently no reason to suppose that engineering nuclear versions of the remaining mammalian mitochondrial genes, a feat that may have widespread biomedical relevance, should require anything other than sequence alterations obviating code disparity and causing modest reductions in hydrophobicity without loss of enzymatic function.

Introduction

The number of endosymbiotic events in the history of eukaryotes remains in dispute, but only two “primary” events (ones in which the new endosymbiont did not already contain an endosymbiosis-derived organelle) seem to have occurred, at least in the lineages leading to present-day organisms.⁽¹⁾ Very early in eukaryotic history – indeed, possibly simultaneously with the appearance of eukaryotes – an alpha-proteobacterium became enclosed within the (thereafter, at least) eukaryotic cytoplasm and gave rise to the mitochondrion. Subsequently, an ancestor of plants and algae engulfed a cyanobacterium and thereby acquired a new organelle, the chloroplast.

In both these cases, the bacterium that gave rise to the new organelle possessed a genome of a few thousand genes, as do their free-living bacterial relatives today. This changed rapidly, however, because of the evolution of mechanisms to import proteins from the eukaryotic cytoplasm into the organelle. This development immediately allowed, in principle, a considerable reduction in complexity of the cell’s genetic machinery: any organellar genes that were functionally redundant with ones in the nucleus could simply be lost, with their encoded proteins being replaced by the imported, nuclear-coded counterpart, and genes with no nuclear counterpart could gain one by the insertion of a copy of themselves into the nuclear DNA, after which loss of the organellar gene could occur. Once these processes had run to completion, with all organellar

proteins being nuclear-coded, a large number of (now nuclear) genes or splice variants could be lost entirely: namely, those encoding the organellar DNA replication, DNA maintenance, transcription and translation machinery. The selective pressure for this to occur would not be felt until *all* organellar genes became superfluous, however, so this cannot by itself explain that transfer. Indeed, complete transfer has apparently occurred only quite rarely: the only known examples are hydrogenosomes and mitosomes, genome-free organelles in various eukaryotic lineages that are widely held to be degenerate mitochondria.⁽²⁾ Rather, gene transfer has traditionally been ascribed to the advantage associated with Mendelian inheritance (avoiding Muller's ratchet) and/or the difficulty of maintaining organellar genes in close proximity to electron transport chains that adventitiously generate substantial concentrations of highly mutagenic free radicals:⁽³⁾ these pressures would clearly apply to each gene independently.

There is, however, an unmistakeable (albeit loose) hierarchy among originally organellar genes with regard to tendency to be moved to the nucleus. Eleven subunits of the respiratory chain, for example, are mitochondrial DNA (mtDNA)-encoded in all animals, and two more in almost all, but all others are nuclear-coded in all animals.⁽⁴⁾ Numerous researchers have therefore wondered what characteristics influence the evolutionary pressure to retain a given gene in the organellar genome. Three main candidate barriers to the transfer of a gene to the nucleus are often championed (Table 1).

Summary statement of hypothesis	Name (abbrev.)	Ref.
Divergence of genetic code, especially a codon encoding STOP in the nucleus but not in the organelle, retains genes in the organelle if they contain any such codons. This barrier to functional gene transfer during evolution is greater than any other.	Code disparity hypothesis (CDH)	11
Hydrophobicity (measured as "mesohydrophobicity") of a protein quantitatively impedes its import into the mitochondrial matrix via TIM23, which must unfold its cargo fully so as to translocate it without collapsing the mitochondrial membrane potential. This is the greatest barrier, except code disparity, to functional gene transfer of any gene present in mtDNA as reduced as that of animals. HH does not apply to chloroplasts, because they lack a membrane potential and so can import folded proteins.	Hydrophobicity hypothesis (HH)	32
Proximity of genes to their products' site of action facilitates rapid and microenvironment-specific response of gene expression to metabolic demands. This is especially advantageous for organellar electron transport processes and constitutes a barrier to functional transfer (of either cpDNA or mtDNA genes) stronger than hydrophobicity is for mtDNA and perhaps even stronger than code disparity.	Co-location for redox regulation (CORR)	5, 6

Table 1. Hypotheses for why organelles retail genomes. Note that these are not all mutually exclusive: only CORR and HH are.

The first, disparity of genetic code, is of limited phylogenetic scope (since most organelle genomes use the standard code), but nonetheless merits detailed analysis, as several species might be interpreted as challenges to the conclusion that it is a severe barrier. The second is the encoded protein's hydrophobicity and consequent resistance to import. The hypothesis that the hydrophobicity barrier is the most powerful one in the absence of a code disparity fares quite well for animal and fungal mtDNA-encoded proteins, but many nuclear-coded mitochondrial proteins are highly hydrophobic; in plants, furthermore, various universally chloroplast DNA (cpDNA)-encoded or predominantly mtDNA-encoded ones are hydrophilic. This led Allen, a decade ago, to propose^(5,6) the third currently popular hypothesis: that the involvement of the encoded protein in respiratory or photosynthetic electron transport may impose tight constraints on the sensitivity of gene expression to local redox state. This sensitivity might be hard to maintain when the gene is located far from the encoded protein's site of action, especially bearing in mind that different organelles within a cell might have different requirements at the same instant. Allen's idea is recognised to suffer considerable imperfections

in the form of both organelle-encoded proteins with no known redox function and nuclear-coded ones with pivotal roles in electron transport (including ones in the same complex as, and hence with a requirement for 1:1 stoichiometry with, organelle-encoded ones). It is also open to the general objection that organelles can (and do) autonomously control abundance or activity of nuclear-coded proteins perfectly well by mechanisms such as regulation of protein import rates or allosteric modulation of enzyme activity.⁽⁷⁾ Nonetheless, it is now the most popular hypothesis to explain retention of organellar genes, because most commentators have concluded that the defects of the hydrophobicity and/or code disparity hypotheses are even starker.⁽⁸⁾

Here I critique in detail both classes of candidate counterexample to these hypotheses (hereafter HH and CDH), i.e., genes that appear to have surmounted a code disparity or severe hydrophobicity barrier and genes that are organellar despite the apparent absence of either barrier to their transfer. I argue that:

- disparity of organellar and nuclear genetic codes, or at least the type of disparity seen in animals and fungi, is a virtually insuperable barrier to transfer and fully explains some such candidate counterexamples;
- a variety of subtler features, while less formidable than hydrophobicity or code disparity, play a more substantial role in causing retention of organellar genes than has been supposed;
- the hydrophobic nuclear-coded proteins often cited as challenging HH do not do so because the mechanism whereby hydrophobicity causes some genes to remain organellar does not apply to them

and I review a raft of recent findings which further undermine the view that the taxa with very few mitochondrial genes (animals, fungi, apicomplexans and chlorophyte algae) retain those that they do for any reason other than code disparity or hydrophobicity. I conclude by examining the impact of this conclusion on the plausibility of the functional transgenic introduction of the 13 protein-coding genes of the human mtDNA into the nucleus.

It must be stressed at the outset that there are certainly numerous obstacles to the evolutionary transfer of organellar genes to the nucleus in addition to code disparity and hydrophobicity. This article does not attempt to enumerate all such forces; rather, it focuses on the question of whether any such force compares to (or exceeds) code disparity and hydrophobicity in its severity. I therefore discuss other obstacles to transfer only to the extent that they apply to genes which are retained unusually tenaciously in mtDNA or cpDNA. Hundreds of genes have been found in organellar genomes and we can expect this number to continue to rise rapidly as more such genomes are sequenced. The genes relevant to this article, however, are those present even in unusually reduced mtDNA or cpDNA. All known such genes (except those not encoding proteins) are surveyed here.

The genetic code disparity hypothesis (CDH)

The human mitochondrial genome uses a genetic code in which four of the 64 possible codons have a different meaning than in the nuclear genome.⁽⁹⁾ Thus, a mitochondrial gene transferred to the nucleus would not encode the correct amino acid sequence when translated in the cytosol, so its function would be abrogated even if it were expressed appropriately and successfully imported into mitochondria. All animals and most fungi have at least one such code disparity between the nuclear and mitochondrial genomes, whereas they are much less common in mitochondria of species more distantly related to animals than fungi, nor yet found in chloroplasts of any taxon.⁽¹⁰⁾

Superficially, code disparity would appear to be only a modest barrier to spontaneous gene transfer: while the probability of a transferred gene undergoing just the right point mutations to cancel out the code disparity before undergoing loss-of-function mutations may be very small, most of the differences between mitochondrial and standard genetic codes are conservative, so might not decisively affect gene function. Moreover, it has been noted⁽¹¹⁾ that the appearance of most of the code disparities in various animal lineages greatly postdates the cessation of gene transfer to the nucleus. There are two serious flaws in this logic.⁽¹²⁾ First, even if gene transfer stopped before code disparity emerged, this would not imply that code disparity was a minor impediment to transfer, only that some other impediment (hydrophobicity, for example) was also major. And second, the event of interest is not the bulk of code changes but the first one, which made the

codes different when they had hitherto been identical. This is especially true because it appears⁽¹³⁾ that the first code change was of the UGA codon from STOP to coding, which thus caused subsequently transferred genes to encode truncated proteins – the worst possible prognosis for gene function, far worse than an amino acid change.⁽¹²⁾ And indeed, it seems⁽¹³⁾ that the UGA switch was roughly contemporaneous with the cessation of gene transfer in the animal and fungal lineages.

Rough simultaneity is insufficient to accept the genetic code hypothesis, however: it would in principle be strongly challenged by the discovery of *any* species that has a variant organellar genetic code but yet has some organellar genes that are nuclear in a species whose common ancestor with the former species can confidently be presumed already to have possessed that code disparity. Several potential examples thus merit discussion (Table 2).

Gene	Presumed location in LCAAF	Taxa descended from LCAAF in which location differs from that in LCAAF	Interpretation
ATP synthase subunit 6 (A6)	Mitochondrial	Chaetognaths	<i>Bona fide</i> cross-code (CC) transfer
ATP synthase subunit 8 (A8)	Mitochondrial	Chaetognaths, platyhelminths, some nematodes, some molluscs	Four independent losses (not transfers)
ATP synthase subunit 9 (A9)	Mitochondrial	All animals, filamentous fungi	Two independent non-CC transfers
MutS	Nuclear or absent	<i>Sarcophyton glaucum</i>	Reverse transfer or <i>de novo</i> appearance
AAC04631	Nuclear or absent	<i>Metridium senile</i>	Reverse transfer or <i>de novo</i> appearance
Complex I subunits 1-6, 4L	Mitochondrial	<i>Saccharomyces</i> , <i>Schizosaccharomyces</i>	Two independent losses
Ribosomal small subunit 3 (RpS3)	Mitochondrial	All animals, lower fungi, some <i>Schizosaccharomyces</i> species	3-4 independent non-CC transfers
Other ribosomal proteins mtDNA-encoded in choanoflagellates; mttB	Mitochondrial	All animals, all fungi	Two independent non-CC transfers
Other intronic or unassigned ORFs	Nuclear or absent	Choanoflagellates, fungi	Reverse transfer or <i>de novo</i> appearance (e.g. by gene duplication)

Table 2. Genes present in the mtDNA of some but not all descendants of the last common ancestor (LCA) of animals and fungi (LCAAF).

First we consider the animal kingdom. Four protein-coding genes have been identified that appear in the mtDNA of some animals but not others. The mtDNA of the coral *Sarcophyton glaucum* contains a gene so far found in no other mtDNA, encoding a homologue of the bacterial mismatch repair gene mutS;⁽¹⁴⁾ another cnidarian, the sea anemone *Metridium senile*, contains a putative homing endonuclease within one of its introns.⁽¹⁵⁾ Cnidarians diverged very early from other animals and their genetic code is non-standard only in the translation of UGA. This means that a transfer from the nucleus (where it is found in yeast⁽¹⁴⁾) to the mitochondrion is a plausible origin for *S. glaucum* mutS even though no such transfer has been identified

elsewhere, as the result would have been only a C-terminal extension of the encoded protein, and even that only if the originating nuclear gene had happened to use UGA as its termination codon. Homing endonuclease activity is also seen in fungi and is addressed below.

The third case, reported very recently by two groups, is the absence of ATP synthase subunit 6 (A6) from the mtDNA of two chaetognaths, *Paraspadella gotoi*⁽¹⁶⁾ and *Spadella cephaloptera*.⁽¹⁷⁾ This lineage has been determined to branch close to the protostome clade^(16,17) and has the standard invertebrate mitochondrial genetic code. Thus, the transfer of even one gene to the nucleus seems to weaken CDH considerably. However, the chaetognaths appear to have undergone uniquely intense selection for a compact mtDNA among triploblast animals, as they import all⁽¹⁷⁾ or all but one⁽¹⁶⁾ of their tRNAs from the cytosol, something unknown for *any* tRNAs in other triploblasts, and even more strikingly they also have no non-coding “control region” of the form seen in all other triploblast mtDNAs. This may thus be the extreme “exception that proves the rule” with regard to the genetic code hypothesis.

The fourth and last gene of interest here in the animal kingdom is ATP synthase subunit 8 (hereafter A8), which is absent from the mtDNA not only of chaetognaths but also of platyhelminths,⁽¹⁸⁾ some nematodes⁽¹⁹⁾ and some molluscs,⁽²⁰⁾ clearly as a result of independent events. (The suggestion⁽²¹⁾ that urochordates also lack A8 was recently found to be in error.⁽²²⁾) However, A8 is not present in eubacterial ATP synthases,⁽²³⁾ so the alternative that its gene has simply been lost, not transferred to the nucleus, must be considered. Evidence for this interpretation is mounting. First, no gene for A8 has been found in the complete *C. elegans* genome⁽¹⁶⁾ (though such a small gene might well be missed by inevitably imperfect genome analysis tools). Second and perhaps more compellingly, the mtDNA of the mollusc *Siphonodontalium lobatum* was recently found to contain an A8 gene encoding only a roughly half-sized protein,⁽²⁴⁾ and the second-shortest A8 yet discovered is in a nematode, *Trichinella spiralis*.⁽²⁵⁾ The coincidence of two separate taxa each encompassing some species with exceptionally short A8 and some with “missing” A8 is striking and supports the view that pressure to shorten A8 can lead to its loss rather than transfer.

The protein-coding gene content of fungi is more variable than that of animals, but not by much. There are three categories of such variability that we will not consider in detail. First, it is premature to consider the unassigned ORFs present in such genomes, since lack of data on function (if any such function even exists) makes meaningful speculation difficult. Second, intronic “maturase” ORFs required for the excision of introns are highly variable among fungi and some have been shown to possess homing endonuclease activity, something also identified in the mtDNA of a cnidarian (see above). They seem likely to be the one class (other than the single case of *mutS* in corals, noted above) where genes are added to, rather than removed from, the mtDNA; the homing endonuclease activity would seem to be a prime suspect for why this is, but details are not known at this time.⁽²⁶⁾ However, the arrival of new genes in the mtDNA tells us little about the reasons for retention of ancestrally mitochondrial genes. Thirdly, we can neglect cases where a gene absent from the mtDNA has clearly been simply lost, such as the Complex I subunits in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, species which make do with a nuclear-coded, single-polypeptide enzyme for transferring electrons from intramitochondrial NADH to inner membrane ubiquinone that does not pump protons. Discussion here will therefore focus on the two other protein-coding genes that are mitochondrial in some but not all fungi: ATP synthase subunit 9 (A9) and the ribosomal small subunit protein 3 (RpS3).

In both cases, the genetic code hypothesis appears to predict very accurately the observed distribution of the gene. A9 is nuclear in only one fungal taxon, the filamentous fungi including *Aspergillus*, *Neurospora* and *Podospora*.⁽²⁷⁾ In this group, the only code disparity of the mtDNA is that UGA encodes tryptophan rather than STOP, and it happens that there are no TGA codons in the A9 gene. Additional fungal taxa exist that also possess only this one code disparity but have not transferred A9; however, this cannot be considered contrary to the genetic code hypothesis because A9 is evidently rather hard to transfer for some other reason (such as hydrophobicity, addressed below), being mtDNA-encoded in all species that retain more than 15 assigned, non-endonuclease, protein-coding genes and having resisted functional transfer in *S. cerevisiae* in the laboratory.⁽²⁸⁾ RpS3 was apparently the last mitoribosomal protein-coding gene left in the fungal lineage and was still there when UGA began to code for tryptophan; this trapped it there (according to the genetic code hypothesis) because unlike A9 it does possess a TGA codon. A test for this interpretation is potentially available because RpS3 was transferred without difficulty in the laboratory in both *S. cerevisiae*⁽²⁹⁾ and *S.*

pombe⁽³⁰⁾ after code correction, so should be absent from the mtDNA in any fungal lineage that (for whatever reason) loses this code-disparity barrier. Evidence available to date is consistent with the hypothesis. *S. pombe* mtDNA was once believed to use the universal code, but this was an oversight: none of its respiratory chain genes contains TGA (nor is TGA used as STOP), but its RpS3 homologue indeed uses TGA to encode tryptophan. Two other *Schizosaccharomyces* species' mtDNAs were sequenced recently⁽³¹⁾ and one of them, *S. octosporus*, has retained RpS3 despite losing the TGA codon. In the other (*S. japonicus*), however, RpS3 is absent, presumably having been transferred to the nucleus, indicating that this event can be rapid. Two other fungal taxa have also lost RpS3;⁽³²⁾ conversely, *S. octosporus* is the only example yet discovered of a RpS3 homologue held in the mtDNA in the absence of a code disparity, and the presence of TGA in the *S. pombe* homologue shows that this has not been so for long.

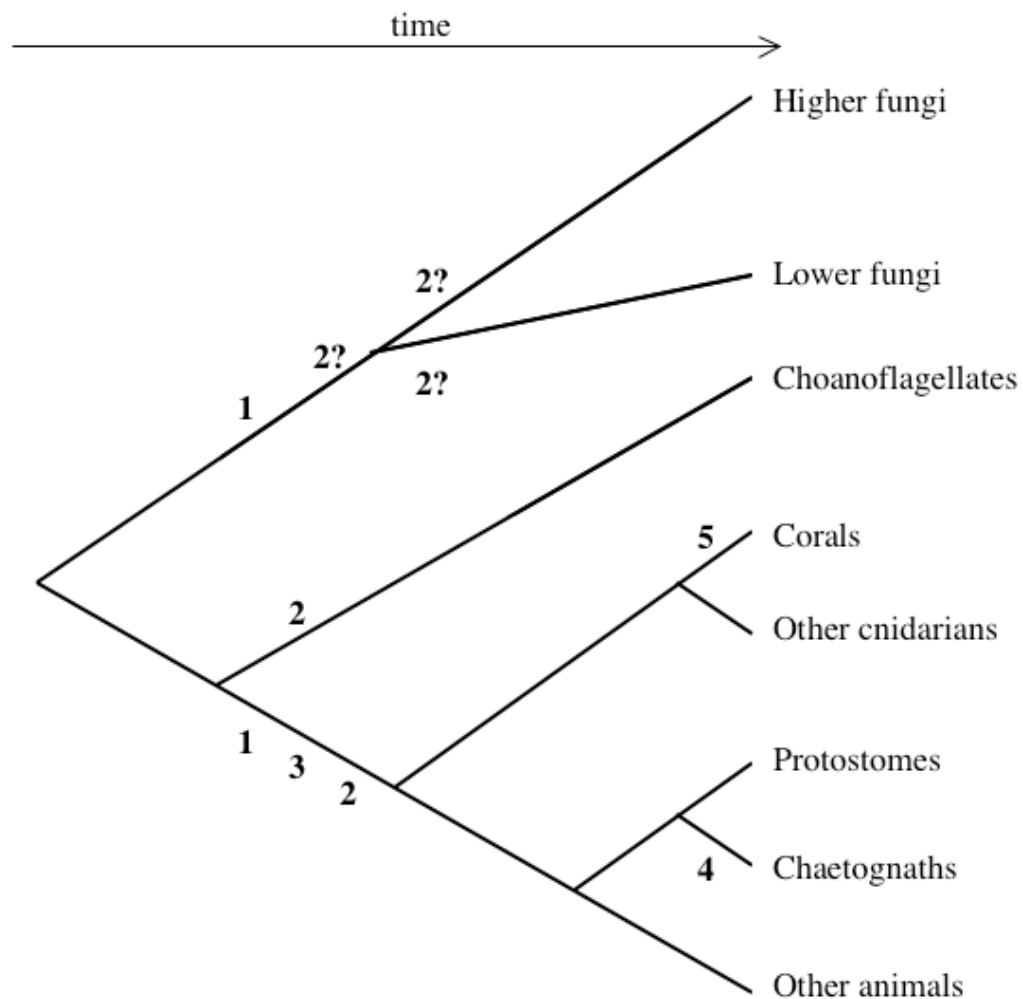


Figure 1. A plausible timing of the key events leading to the present gene content and genetic code in animals, fungi and choanoflagellates, starting from their most recent common ancestor. Numbers denote the following events: 1, transfer of *mttB* and all remaining ribosomal protein genes except RpS3; 2, switch of UGA to encode tryptophan instead of STOP; 3, transfer of RpS3 and A9 in animals; 4, transfer of A6; 5, acquisition of *mutS*. Not shown are: the multiple losses of A8 in animals and RpS3 in fungi; the more recent changes of mtDNA genetic code subsequent to the UGA switch; and endonucleases and unassigned ORFs.

A third taxon must be considered here, however: the choanoflagellates. These organisms have long been thought to be close relatives of animals, and it has now been compellingly shown that they did so after the divergence of fungi.⁽³³⁾ The mtDNA of the choanoflagellate *Monosiga brevicollis* has been sequenced; it encodes the protein translocase gene *mttB* and a slew of mitoribosomal proteins, including most of those encoded by mtDNA in most plants.⁽³³⁾ Intriguingly, its mitochondria decode UGA as tryptophan. This leads to the plausible interpretation (Figure 1) that two key events – the switch of UGA to tryptophan and the hike in gene transfer pressure that moved the genes for *mttB* and the ribosomal proteins – occurred more or less

simultaneously (perhaps causally linked?) but after the divergence of the three relevant lineages (animals, choanoflagellates and fungi), affecting all three independently. (Possibly the lower fungi, whose mtDNAs lack RpS3, had already split from other fungi at this point.) It should be stressed that the near-coincidence of which genes were transferred in fungi and animals in this interpretation is not outlandish, as a third, very distant lineage, the chlorophyte algae, also shows this chronology of transfer (see below). For some reason, choanoflagellates underwent the code change before transferring many mitoribosomal genes (possibly they experienced particular pressure to become AT-rich, since *Monosiga* has very A/T-rich mtDNA⁽³³⁾), whereas fungi moved all but RpS3 before the code change and animals additionally moved both RpS3 and A9. A species also closely related to animals, the ichthyosporean *Amoebidium parasiticum*, may represent an intermediate case because its mtDNA contains only three confirmed ribosomal protein genes;⁽³³⁾ however, sequencing of this mtDNA is incomplete and it also contains many unassigned ORFs, so a definitive conclusion regarding its history is premature.

The hydrophobicity hypothesis (HH)

Historically the most widely-touted hypothesis for the retention of organellar genes in the absence of a code disparity was that they encode proteins whose extreme hydrophobicity prevents their efficient import into the organelle.⁽³⁴⁾ Membrane proteins in general, but some more than others, form globular structures in the aqueous phase. The difficulty that this poses for mitochondrial protein import is that the machinery which translocates proteins through the inner membrane must do so without allowing the free passage of ions, including protons, since the membrane potential across that membrane (which, unlike the pH disparity, would be collapsed by even very slight proton leakage) is intrinsic to many of the mitochondrion's functions, including protein import itself. This means that the import machinery must thread the protein through a small pore, which in the case of hydrophobic proteins entails completely unfolding their preferred globular state. As for CDH, a robust evaluation of HH requires a comprehensive survey of the possible counterexamples so far identified (Table 3).

Gene	Presumed present in primordial organelle?	Taxa in which functionally transferred	Mesohydrophobicity of encoded protein	Interpretation
Cytochrome c oxidase subunit 1, cytochrome b, NADH dehydrogenase subunits 1, 2, 4, 6	Yes	None	Extremely high	Unimportable by TIM23; not assemblable into enzyme complex if imported by TIM22
Cytochrome c oxidase subunits 2 and 3, NADH dehydrogenase subunits 3 and 4L, ATP synthase subunits 6, 8, 9	Yes	Chlamydomonads (all), assorted others (one or two)	Very high	Importable via TIM23 only with much assistance by chaperones etc.; as above re TIM22
Mitoribosomal small subunit protein 3 (RpS3)	Yes	Algae and animals in evolution; <i>S. pombe</i> and <i>S. cerevisiae</i> in lab	Low	Possibly, amino-terminal extensions impede subunit assembly
Several other mitoribosomal proteins	Yes	Chlamydomonads, animals, fungi	Low	Same as RpS3

Mitochondrial anion carriers	No	n/a	Very/extremely high	TIM22 evolved to import these but cannot translocate proteins to matrix
Rubisco large subunit (rbcL)	Yes	None in evolution; tobacco in lab	Low	Post-transfer gene duplication needed
Light-harvesting complex (LHC) core proteins	Maybe	n/a	Extremely high	Chloroplasts can import them even when folded
clpP protease	Yes	None	Low	Cytosolic toxicity
accD	Yes	None	Low	Genetic code pseudo-disparity

Table 3. Genes found unusually often in the “wrong” genome for their hydrophobicity.

All 13 proteins that are mtDNA-coded in animals are very hydrophobic, and those that have never been found encoded in the nucleus of any organism are the most hydrophobic of all, by a measure termed “mesohydrophobicity” developed by Claros.⁽³³⁾ ATP synthase subunit 9, the respiratory chain subunit that was evidently last to be transferred (and, in particular, the only one necessarily remaining in the common ancestor of animals, choanoflagellates and fungi), fits this model well, as its hydrophobicity approaches that of the 13 that humans still retain.⁽³⁵⁾ Further supporting this model, the identical hierarchy of resistance to transfer is seen in the chlorophyte algae, which (if we accept the hypothesis that the apicoplast of apicomplexans was originally a chlamydomonad, whose pre-existing ingenuity in this regard they appropriated⁽³⁶⁾) are the only other group in which as much success in gene transfer has been achieved as in animals. Specifically, the mitochondrial genome of *Scenedesmus obliquus* contains the 12 protein-coding genes mtDNA-encoded in all animals except chaetognaths, A9, and no others (except for seven unassigned ORFs, which we can neglect for reasons already given).⁽³⁷⁾ Moreover, its close relative *Chlamydomonas reinhardtii* has transferred, of those 13, precisely the six that contain the fewest hydrophobic domains and have the lowest mesohydrophobicity.⁽³⁸⁾ Ciliates ostensibly constitute a challenge to the strictness of this hierarchy: *Paramecium aurelia* mtDNA, for example, contains genes for several ribosomal proteins but has been reported to lack three of the eleven pan-animal genes as well as A6.⁽³⁹⁾ However, two of those four have recently been identified in the *P. aurelia* mtDNA, having originally been overlooked on account of extreme divergence from their sequence in other taxa;⁽⁴⁰⁾ the other two might also be present but still unrecognised, especially given the possibility of genes splitting into two as has occurred with, for example, *C. reinhardtii* cytochrome oxidase subunit 2.⁽³⁶⁾

Are there meaningful nuclear-coded counterexamples to HH?

The hydrophobicity hypothesis has been challenged, however, on the basis of a number of nuclear-coded hydrophobic proteins present in mitochondria or chloroplasts. The viability of the hypothesis thus depends on robust arguments that these exceptions “do not count” – that they do not contradict the actual hypothesis, only oversimplifications of it.

One class is the mitochondrial anion carriers, a family of closely-related proteins which have six or twelve transmembrane domains. They are (by some measures) as hydrophobic as animals’ mtDNA-encoded proteins, so HH might be thought to predict that they would pose great difficulties for the import system; the fact that they are all nuclear-coded, even in species with dozens of mtDNA-encoded proteins, has thus been raised as a challenge to the idea that hydrophobicity is much of a problem. This does not follow, however, as the anion carriers are not of eubacterial origin.⁽⁴¹⁾ (An ostensible exception is the phosphate carrier, which has homologues in prokaryotes and is normally imported via the TIM23 complex, the same system used by originally mtDNA-encoded proteins. However, this import mechanism is probably a recent adaptation, since deletion of the N-terminal cleavable signal sequence necessary for TIM23-mediated import only halves import

efficiency.⁽⁴¹⁾ Hence, the anion carrier genes are irrelevant to any hypothesis proposing to explain a hierarchy of resistance to gene transfer, since they have never themselves been transferred.

Eukaryotes have clearly surmounted the hydrophobicity of the anion carriers, however; hence, we might conclude that a solution for hydrophobic proteins of eubacterial origin must also be within evolution's capacity to identify, so there must be something else (something that really *is* too hard for evolution) holding the remaining ones in place. But this oversimplifies the evolutionary situation. The evolutionary pressure to move mitochondrial genes to the nucleus is often described as strong, but by most standards it has surely been rather slight, as no overt improvement in any metabolic function is conferred by gene transfer. The evolution of controlled anion exchange across the inner mitochondrial membrane, on the other hand, presumably led to a dramatic improvement in the ability of the mitochondrion to supply the cell's bioenergetic needs. And indeed, it is now known that a system entirely distinct from the TIM23 complex, termed the TIM22 complex, exists in the inner membrane for the sole purpose of inserting the anion carriers⁽⁴²⁾ – something that would be expected to arise only in response to the genuinely strong selective pressure imposed by a clear effect on metabolic function. Now: even accepting that only the anion carriers, and not the mtDNA-encoded genes, drove the development of the TIM22 complex, one might think that the mtDNA-encoded proteins could surely then have used it, thus making their gene transfer to the nucleus possible after all even if they were too hydrophobic for the TIM23 complex. But this neglects the fact that all the proteins whose hydrophobicity is invoked to explain their resistance to gene transfer are subunits of large multi-component enzyme complexes, whose assembly is a highly intricate process. We should not expect that the orientation of such a protein's final insertion into the membrane could be abruptly altered without disrupting the assembly process. And indeed, the relatively hydrophilic subunits of these complexes, whose genes have been successfully transferred to the nucleus, are imported all the way into the mitochondrial matrix (where, in the primordial mitochondrion, they used to be synthesised) and then re-exported.⁽⁴²⁾ The TIM22 complex, on the other hand, inserts its cargo directly into the membrane, a task that intuitively should be much easier to achieve without making the membrane proton-permeant than translocation of a protein all the way into the matrix. Hence, the arrival of the TIM22 complex on the mitochondrial scene did not greatly lessen the difficulty of transferring mitochondrial genes encoding very hydrophobic proteins to the nucleus.

The above logic highlights the importance of analysing evolutionary scenarios for what they are – sequences of events. Though it may be tempting to presume that hydrophobicity is hydrophobicity, and thus that any extremely hydrophobic, nuclear-coded, mitochondrial protein challenges HH, careful analysis of how such proteins arose shows that this is oversimplistic. In assorting mitochondrial proteins between TIM23 and TIM22, the mitochondrion effectively interrogates their evolutionary origin.

A number of ostensible counter-examples to the hydrophobicity hypothesis are found in chloroplasts, where hydrophobic nuclear-coded proteins indisputably of cyanobacterial origin are found in the thylakoid and must thus be translocated through the inner membrane. Perhaps the examples most directly comparable to mitochondria are the subunits of the thylakoid homologue of mitochondrial Complex I, eleven of which are still cpDNA-encoded in *Arabidopsis thaliana*⁽⁴³⁾ but not in *Pinus thunbergii*,⁽⁴⁴⁾ even though at least three mitochondrial Complex I genes have apparently never been transferred to the nucleus. (This example is in fact provisional, as these genes have not yet been identified in the *P. thunbergii* nucleus.) It has also been noted that the extremely hydrophobic light-harvesting complex (LHC) proteins of Photosystems I and II are always nuclear-coded.⁽⁶⁾ But to compare the inner membranes of chloroplasts and mitochondria for this purpose is invalid, because the hydrophobicity hypothesis concerns the transport of proteins across a membrane that is supporting a membrane potential (and, in particular, a membrane potential on which protein import relies), something that the inner membrane of chloroplasts does not do. It is of note that cyanobacteria, the ancestors of chloroplasts, possess thylakoid-like lamellar structures but perform proton-pumping both there and at their plasma membrane.⁽⁴⁶⁾ Plausibly, the nascent chloroplast surmounted the hydrophobicity barrier to import of advantageous but hydrophobic proteins by abandoning proton-pumping at its inner membrane (which had been the bacterium's plasma membrane) and restricting it to the thylakoid – at which a protein's membrane insertion/translocation requirements are unaltered by its gene transfer to the nucleus, since the cpDNA is in the stroma, not the thylakoid lumen. Direct support for the illegitimacy of generalising across chloroplasts and mitochondria in respect of HH is also available: it has been shown⁽⁴⁵⁾ that proteins can be imported into the

chloroplast stroma in a folded state. There is no problem for the chloroplast if this is accompanied by modest proton leak, because the proton gradient across the chloroplast inner membrane consists only of a pH disparity, whose dissipation requires vastly more proton transfer than is needed to collapse a membrane potential.⁽⁴⁷⁾ On the other hand, import of folded proteins into chloroplasts has not always succeeded⁽⁴⁸⁾ and it has been proposed⁽⁴⁵⁾ that an additional barrier applies to proteins over a certain size. This would help to explain why many proteins located in the thylakoid lumen remain cpDNA-encoded even though the LHC proteins do not.

Are there meaningful organelle-encoded counterexamples to HH?

The foregoing goes only part of the way to demonstrating the compatibility of HH with available data, however. We must also explain why many proteins lacking so much as a hint of hydrophobicity are nonetheless organelle-encoded in most or all species. (Indeed, if code disparity and hydrophobicity in the context of transport across energy-transducing membranes were the only appreciable obstacles to transfer, chloroplast DNA should have disappeared long ago, for the reasons just outlined.) We therefore consider in this section all genes that have been retained even in heavily reduced organellar genomes despite the absence of the code disparity and hydrophobicity barriers. As noted in the Introduction, genes found only in relatively gene-rich organelle genomes are not relevant to this article, as their loss elsewhere shows that the forces holding them in some organelles are relatively modest and our topic is whether there are *severe* obstacles to transfer in addition to code disparity and hydrophobicity.

Possibly the most frequently cited such gene is *rbcL*, the large subunit of Rubisco. Rubisco is an exceptionally lethargic enzyme.⁽⁴⁹⁾ It is an essential component of the Calvin cycle, however, so its activity per cell needs to be high. Consequently, it is the most abundant enzyme on Earth.⁽⁴⁹⁾ This seems to suggest a simple explanation for why it has never been transferred to the nucleus: such transfer would entail a reduction in gene copy number per cell from the number of cpDNA genomes per cell to just two. As others have observed,⁽³⁾ this might not suffice to support the required synthesis rate. Powerful support for this interpretation comes from the observation that the gene for *rbcS*, the small subunit of Rubisco, is indeed present in between two and at least 22 copies per haploid genome in all taxa in which it is nuclear.⁽⁵⁰⁾ The implication is that the transfer of these genes can only succeed if substantial duplication of the nuclear gene follows gene transfer soon enough to provide an adequate protein synthesis rate before the nuclear gene degenerates by random mutation. (Clearly other alternatives can be imagined, but they can be ignored, because the ubiquity of *rbcS* duplication demonstrates that all such alternatives are even less frequent.) This evidently occurred for *rbcS*, but probably only once (since *rbcS* has never been found in the cpDNA of plants or green algae and never in the nucleus of red algae⁽⁵⁰⁾). Additional support for this hypothesis derives from the successful insertion of *rbcL* into the nucleus, where it supported only about 10% of the natural synthesis rate per cell.⁽⁵¹⁾

A very general potential obstacle to transfer is the toxicity of an organellar protein to cytosolic constituents. This seems not to hold for ATP synthase subunit 8, as was once proposed,⁽¹¹⁾ but it may be the reason why *clpP* is apparently never nuclear-coded (though the belief that it cannot simply be lost is now challenged⁽⁵²⁾): being a protease, it must be sequestered from any proteins that it might inappropriately degrade. Putting this less teleologically: the restriction of *clpP* to the chloroplast stroma has potentially allowed the evolution of cytosolic proteins that it would rapidly degrade. The toxicity associated with expression of *clpP* from the nucleus might thus kill such cells before there was time to co-evolve *clpP* and these proteins to prevent this degradation.

The only other nearly always cpDNA-encoded chloroplast protein of known function,⁽⁵³⁾ other than ribosomal proteins (which are addressed below), is *accD*, a component of the fatty acid biosynthesis pathway.⁽⁵⁴⁾ (Rice and other Gramineae lack *accD* in their cpDNA but appear to have lost it rather than transferred it, since they lack activity for the enzyme of which *accD* is a subunit, the herbicide-insensitive acetyl CoA carboxylase.⁽⁵⁴⁾) Recently it was noted that *accD* may be retained in the cpDNA by the requirement for editing of its mRNA.⁽⁵⁵⁾ In cells in which RNA editing is absent from the nucleus or cytosol and only occurs in organelles, a requirement for editing is functionally equivalent to a variant genetic code: a transfer of the gene to the nucleus would result in an incorrect amino acid sequence being synthesised in the cytosol, and if this severely affected gene function then the transfer could thus succeed only if that error were corrected by mutation

before other mutations had eliminated function. Other genes may be hard to transfer because of the cpDNA's use of non-standard initiation codons in some genes, which would result in amino-terminal truncation (or, more unusually, extension) of encoded proteins after gene transfer.

Perhaps the above obstacles all seem rather feeble and thus unlikely to have sufficed to maintain organellar DNA in the face of evidently strong selective pressure to transfer genes. However, this interpretation overconfidently conflates the situations of mitochondria and chloroplasts. A measure of the relative selective pressure on the two organellar genomes is circumstantially provided by the organellar ribosomal proteins, all of whose genes have been transferred in the animal lineage but many of which remain organelle-encoded in all plant mitochondria and chloroplasts except the mtDNA of green algae. The reason why they have been slow to transfer is unclear, but perhaps it is that the ribosome's subunits are both large globular structures, with a low surface-to-volume ratio. This means that many ribosomal proteins are embedded within the structure; ribosomal assembly may therefore be intolerant of the addition of a few amino-terminal amino acids to such proteins. Amino-terminal extension is a likely adjunct to gene transfer, as the sequence recognised by the mitochondrial processing peptidase is a good deal more specific than the remarkably loosely constrained sequences that can target genes to mitochondria.⁽⁵⁶⁾

Additionally, one of the most plausible proposed mechanisms underlying gene transfer to the nucleus implies that the pressure to do so would be less in chloroplasts than in mitochondria: namely, relative mutation rate. On account of its proximity to the respiratory chain, mtDNA is exposed to relatively high concentrations of mutagenic free radicals, which will tend to disable mitochondrial genes more rapidly than nuclear ones.⁽⁵⁷⁾ Hence, whenever this loss of function can be tolerated because a nuclear copy is being successfully expressed and targeted, that nuclear copy will tend to encounter selective pressure for retention (due to loss of the mitochondrial gene) more often than the converse. This effect should be more pronounced in species with higher free radical production; it would be intriguing to determine whether animals and fungi generally have higher basal mitochondrial free radical production than plants and protists. But it should also be more pronounced in mitochondria than in chloroplasts, for the same reason.

In summary, it thus appears plausible that there is relatively weak selective pressure to transfer chloroplast genes to the nucleus, which can be outweighed by a wide variety of only modestly severe obstacles to transfer, each affecting a minority of genes. No all-embracing explanation for the retention of cpDNA is required.

Conclusion: no evidence for a barrier as great as code disparity or hydrophobicity

The survey presented here has covered all genes so far identified that might be thought to imply the existence of a barrier to organellar gene transfer comparable in severity to code disparity and hydrophobicity, and has shown that none of them actually implies this. While future data on other species may of course unearth contrary evidence, this survey thus demonstrates that there is presently no case for invoking such a barrier. In particular, no functional property of organellar proteins, such as involvement in redox processes, appears to need to be invoked as a more severe (or even comparably severe) barrier. Since the only such hypothesis currently entertained, CORR, is challenged by numerous examples (most starkly, the many apparently early transfers of genes encoding subunits of enzyme complexes other of whose subunits remain universally organelle-encoded), the combination of CDH and HH must at present be considered unequivocally the most economical explanation of available data so far suggested.

This conclusion is relevant to the development of modified mitochondrial genes which could be inserted into the nuclear genome to complement mutations in the mitochondrial copy, a technology termed "allotopic expression".⁽⁵⁸⁾ If a mitochondrial location for the gene were inherently necessary for an encoded protein's function, such efforts might be doomed to fail: moving the gene to the nucleus might have only slight functional consequences, relevant in evolution but not biomedically, but it is equally plausible that the problem would be much more severe (just as hydrophobicity is). Such reservations have been voiced by very senior figures in the field,^(59,60) possibly contributing to the lack of manpower that has hitherto been applied to developing allotopic expression in mammals. The obstacles to functional allotopic expression are anyway highly challenging – modulating these proteins' hydrophobicity without impairing function will require much

experimentation – but evolution does not, after all, appear to be telling us that recent progress in this area^(61,62) is a false dawn. The biomedical potential of successful, comprehensive allotopic expression is considerable: it would provide a cure for many presently incurable diseases caused by mtDNA mutations.⁽⁶³⁾ it might prove to play a major role in postponing aging and age-related disease (though this remains controversial⁽⁶⁴⁾), and it would also open the way for the use of non-human oocytes for therapeutic cloning, thereby eliminating the oocyte supply problem that is such a major barrier to that technology at present.⁽⁶⁵⁾ Hence, it is important to recognise that one frequent argument against attempting to develop allotopic expression – namely, that these genes will probably not work from the nucleus even if the code disparity and hydrophobicity barriers are removed, or else more genes would have been transferred during evolution – is not presently supported by the evidence available.

Acknowledgements

I am indebted to two anonymous referees and the Editor for highly constructive comments on the initial draft of this article.

References

1. Cavalier-Smith T. 1992. The number of symbiotic origins of organelles. *Biosystems* 28:91-106.
2. Embley TM, van der Giezen M, Horner DS, Dyal PL, Bell S, Foster PG. 2003. Hydrogenosomes, mitochondria and early eukaryotic evolution. *IUBMB Life* 55:387-395.
3. Blanchard JL, Lynch M. 2003. Organellar genes: why do they end up in the nucleus? *Trends Genet* 16:315-320.
4. Lang BF, Gray MW, Burger G. 1999. Mitochondrial genome evolution and the origin of eukaryotes. *Annu Rev Genet* 33:351-397.
5. Allen JF. 1993. Control of gene expression by redox potential and the requirement for chloroplast and mitochondrial genomes. *J Theor Biol* 165:609-631.
6. Allen JF. 2003. The function of genomes in bioenergetic organelles. *Philos Trans R Soc Lond B Biol Sci* 358:19-37.
7. Bender E, Kadenbach B. 2000. The allosteric ATP-inhibition of cytochrome c oxidase activity is reversibly switched on by cAMP-dependent phosphorylation. *FEBS Lett* 466:130-134.
8. Race HL, Herrmann RG, Martin W. 1999. Why have organelles retained genomes? *Trends Genet* 15:364-370.
9. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG. 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290:457-465.
10. Andersson GE, Kurland CG. 1991. An extreme codon preference strategy: codon reassignment. *Mol Biol Evol* 8:530-544.
11. Jacobs HT. 1991. Structural similarities between a mitochondrially encoded polypeptide and a family of prokaryotic respiratory toxins involved in plasmid maintenance suggest a novel mechanism for the evolutionary maintenance of mitochondrial DNA. *J Mol Evol* 32:333-339.
12. de Grey ADNJ. 1999. Are those 13 proteins really unimportable? In: *From Symbiosis to Eukaryotism - Endocytobiology VII* (E. Wagner et al., eds.), Geneva University Press, pp. 489-502.
13. Leblanc C, Richard O, Kloareg B, Viehmann S, Zetsche K, Boyen C. 1997. Origin and evolution of mitochondria: what have we learnt from red algae? *Curr Genet* 31:193-207.
14. Pont-Kingdon G, Okada NA, Macfarlane JL, Beagley CT, Watkins-Sims CD, Cavalier-Smith T, Clark-Walker GD, Wolstenholme DR. 1998. Mitochondrial DNA of the coral *Sarcophyton glaucum* contains a gene for a homologue of bacterial MutS: a possible case of gene transfer from the nucleus to the mitochondrion. *J Mol Evol* 46:419-431.
15. Beagley CT, Okimoto R, Wolstenholme DR. 1998. The mitochondrial genome of the sea anemone *Metridium senile* (Cnidaria): introns, a paucity of tRNA genes, and a near-standard genetic code. *Genetics* 148:1091-1108.
16. Helfenbein KG, Fourcade HM, Vanjani RG, Boore JL. 2004. The mitochondrial genome of *Paraspadella gotoi* is highly reduced and reveals that chaetognaths are a sister group to protostomes. *Proc Natl Acad*

Sci USA 101:10639-10643.

17. Papillon D, Perez Y, Caubit X, Le Parco Y. 2004. Identification of chaetognaths as protostomes is supported by the analysis of their mitochondrial genome. *Mol Biol Evol*, in press.
18. McManus DP, Le TH, Blair D. 2004. Genomics of parasitic flatworms. *Int J Parasitol* 34:153-158.
19. Okimoto R, Macfarlane JL, Clary DO, Wolstenholme DR. 1992. The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. *Genetics* 130:471-498.
20. Hoffmann RJ, Boore JL, Brown WM. 1992. A novel mitochondrial genome organization for the blue mussel, *Mytilus edulis*. *Genetics* 131:397-412.
21. Yokobori S, Ueda T, Feldmaier-Fuchs G, Paabo S, Ueshima R, Kondow A, Nishikawa K, Watanabe K. 1999. Complete DNA sequence of the mitochondrial genome of the ascidian *Halocynthia roretzi* (Chordata, Urochordata). *Genetics* 153:1851-1862.
22. Gissi C, Iannelli F, Pesole G. 2004. Complete mtDNA of *Ciona intestinalis* reveals extensive gene rearrangement and the presence of an atp8 and an extra trnM gene in ascidians. *J Mol Evol* 58:376-389.
23. Altendorf K, Stalz W, Greie J, Deckers-Hebestreit G. 2000. Structure and function of the F(o) complex of the ATP synthase from *Escherichia coli*. *J Exp Biol* 203:19-28.
24. Dreyer H, Steiner G. 2004. The complete sequence and gene organization of the mitochondrial genome of the gadilid scaphopod *Siphonodontalium lobatum* (Mollusca). *Mol Phylogenet Evol* 31:605-617.
25. Lavrov DV, Brown WM. 2001. *Trichinella spiralis* mtDNA: a nematode mitochondrial genome that encodes a putative ATP8 and normally structured tRNAs and has a gene arrangement relatable to those of coelomate metazoans. *Genetics* 157:621-637.
26. Chatterjee P, Brady KL, Solem A, Ho Y, Caprara MG. 2003. Functionally distinct nucleic acid binding sites for a group I intron encoded RNA maturase/DNA homing endonuclease. *J Mol Biol* 329:239-251.
27. Ridder R, Kunkele KP, Osiewacz HD. 1991. Sequence of the nuclear ATP synthase subunit 9 gene of *Podospira anserina*: lack of similarity to the mitochondrial genome. *Curr Genet* 20:349-351.
28. Gray RE, Law RH, Devenish RJ, Nagley P. 1996. Allotopic expression of mitochondrial ATP synthase genes in nucleus of *Saccharomyces cerevisiae*. *Methods Enzymol* 264:369-389.
29. Sanchirico M, Tzellas A, Fox TD, Conrad-Webb H, Periman PS, Mason TL. 1995. Relocation of the unusual *VARI* gene from the mitochondrion to the nucleus. *Biochem Cell Biol* 73:987-995.
30. Neu R, Goffart S, Wolf K, Schafer B. 1998. Relocation of *urf a* from the mitochondrion to the nucleus cures the mitochondrial mutator phenotype in the fission yeast *Schizosaccharomyces pombe*. *Mol Gen Genet* 258:389-396.
31. Bullerwell CE, Leigh J, Forget L, Lang BF. 2003. A comparison of three fission yeast mitochondrial genomes. *Nucleic Acids Res* 31:759-768.
32. Bullerwell CE, Burger G, Lang BF. 2000. A novel motif for identifying *rps3* homologs in fungal mitochondrial genomes. *Trends Biochem Sci* 25:363-365.
33. Burger G, Forget L, Zhu Y, Gray MW, Lang BF. 2003. Unique mitochondrial genome architecture in unicellular relatives of animals. *Proc Natl Acad Sci USA* 100:892-897.
34. Popot JL, de Vitry C. 1990. On the microassembly of integral membrane proteins. *Annu Rev Biophys Chem* 19:369-403.
35. Claros MG, Perea J, Shu Y, Samatey FA, Popot JL, Jacq C. 1995. Limitations to *in vivo* import of hydrophobic proteins into yeast mitochondria. The case of a cytoplasmically synthesized apocytochrome b. *Eur J Biochem* 228:762-771.
36. Funes S, Davidson E, Reyes-Prieto A, Magallon S, Herion P, King MP, Gonzalez-Halphen D. 2002. A green algal apicoplast ancestor. *Science* 298:2155.
37. Nedelcu AM, Lee RW, Lemieux C, Gray MW, Burger G. 2000. The complete mitochondrial DNA sequence of *Scenedesmus obliquus* reflects an intermediate stage in the evolution of the green algal mitochondrial genome. *Genome Res* 10:819-831.
38. Michaelis G, Vahrenholz C, Pratje E. 1990. Mitochondrial DNA of *Chlamydomonas reinhardtii*: the gene for apocytochrome b and the complete functional map of the 15.8 kb DNA. *Mol Gen Genet* 223:211-216.
39. Pritchard AE, Seilhamer JJ, Mahalingam R, Sable CL, Venuti SE, Cummings DJ. 1990. Nucleotide sequence of the mitochondrial genome of *Paramecium*. *Nucleic Acids Res* 18:173-180.
40. Brunk CF, Lee LC, Tran AB, Li J. 2003. Complete sequence of the mitochondrial genome of *Tetrahymena thermophila* and comparative methods for identifying highly divergent genes. *Nucleic*

Acids Res 31:1673-1682.

41. Zara V, Palmieri F, Mahlke K, Pfanner N. 1992. The cleavable presequence is not essential for import and assembly of the phosphate carrier of mammalian mitochondria but enhances the specificity and efficiency of import. *J Biol Chem* 267:12077-12081.
42. Rehling P, Pfanner N, Meisinger C. 2003. Insertion of hydrophobic membrane proteins into the inner mitochondrial membrane – a guided tour. *J Mol Biol* 326:639-657.
43. Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S. 1999. Complete structure of the chloroplast genome of *Arabidopsis thaliana*. *DNA Res* 6:283-290.
44. Wakasugi T, Tsudzuki J, Ito S, Nakashima K, Tsudzuki T, Sugiura M. 1994. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. *Proc Natl Acad Sci USA* 91:9794-9798.
45. Clark SA, Theg SM. 1997. A folded protein can be transported across the chloroplast envelope and thylakoid membranes. *Mol Biol Cell* 8:923-934.
46. Padan E, Schuldiner S. 1978. Energy transduction in the photosynthetic membranes of the cyanobacterium (blue-green alga) *Plectonema boryanum*. *J Biol Chem* 253:3281-3286.
47. Nicholls DG, Ferguson SJ. 2002. *Bioenergetics 3*. Academic Press, London.
48. Wu C, Seibert FS, Ko K. 1994. Identification of chloroplast envelope proteins in close physical proximity to a partially translocated chimeric precursor protein. *J Biol Chem* 269:32264-32271.
49. Ellis RJ. 1979. The most abundant protein in the world. *Trends Biochem Sci* 4:241-244.
50. Spreitzer RJ. 2003. Role of the small subunit in ribulose-1,5-bisphosphate carboxylase/oxygenase. *Arch Biochem Biophys* 414:141-149.
51. Kanevski I, Maliga P. 1994. Relocation of the plastid *rbcL* gene to the nucleus yields functional ribulose-1,5-bisphosphate carboxylase in tobacco chloroplasts. *Proc Natl Acad Sci USA* 91:1969-1973.
52. Cahoon AB, Cunningham KA, Stern DB. 2003. The plastid *clpP* gene may not be essential for plant cell viability. *Plant Cell Physiol* 44:93-95.
53. Wolfe KH, Morden CW, Palmer JD. 1992. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proc Natl Acad Sci U S A* 89:10648-10652.
54. Konishi T, Shinohara K, Yamada K, Sasaki Y. 1996. Acetyl-CoA carboxylase in higher plants: most plants other than Gramineae have both the prokaryotic and the eukaryotic forms of this enzyme. *Plant Cell Physiol* 37:117-122.
55. Bungard RA. 2004. Photosynthetic evolution in parasitic plants: insight from the chloroplast genome. *BioEssays* 26:235-247.
56. Lemire BD, Fankhauser C, Baker A, Schatz G. 1989. The mitochondrial targeting function of randomly generated peptide sequences correlates with predicted helical amphiphilicity. *J Biol Chem* 264:20206-20215.
57. Wallace DC, Ye JH, Neckelmann SN, Singh G, Webster KA, Greenberg BD. 1987. Sequence analysis of cDNAs for the human and bovine ATP synthase beta subunit: mitochondrial DNA genes sustain seventeen times more mutations. *Curr Genet* 12:81-90.
58. de Grey ADNJ. 2000. Mitochondrial gene therapy: an arena for the biomedical use of inteins. *Trends Biotechnol* 18:394-399.
59. von Heijne G. 1986. Why mitochondria need a genome. *FEBS Lett* 198:1-4.
60. Hartl FU, Neupert W. 1990. Protein sorting to mitochondria: evolutionary conservations of folding and assembly. *Science* 247:930-938.
61. Manfredi G, Fu J, Ojaimi J, Sadlock JE, Kwong JQ, Guy J, Schon EA. 2002. Rescue of a deficiency in ATP synthesis by transfer of MTATP6, a mitochondrial DNA-encoded gene, to the nucleus. *Nat Genet* 30:394-399.
62. Zullo SJ. 2001. Gene therapy of mitochondrial DNA mutations: a brief, biased history of allotopic expression in mammalian cells. *Semin Neurol* 21:327-335.
63. Finsterer J. 2004. Mitochondriopathies. *Eur J Neurol* 11:163-186.
64. de Grey ADNJ. 2004. Mitochondrial mutations in mammalian aging: an over-hasty about-turn? *Rejuvenation Res* 7:171-174.
65. de Grey ADNJ. 2004. Inter-species therapeutic cloning: the looming problem of mitochondrial DNA and two possible solutions. *Rejuvenation Res* 7:95-98.