Published in final edited form as:

Nat Rev Genet. 2018 October ; 19(10): 635–648. doi:10.1038/s41576-018-0035-9.

Cytonuclear integration and coevolution

Daniel B. Sloan[1,*], **Jessica M. Warren**[1], **Alissa M. Williams**[1], **Zhiqiang Wu**[1], **Salah E. Abdel-Ghany**[1], **Adam J. Chicco**[2], and **Justin C. Havird**[1]

[1]Department of Biology, Colorado State University, Fort Collins, CO 80523, USA

[2]Department of Biomedical Sciences, Colorado State University, Fort Collins, CO 80523, USA

Abstract

The partitioning of genetic material between the nucleus and cytoplasmic (i.e., mitochondrial and plastid) genomes within eukaryotic cells necessitates coordinated integration between these genomic compartments, with important evolutionary and biomedical implications. Classic questions persist about the pervasive reduction of cytoplasmic genomes via a combination of gene loss, transfer, and functional replacement — and yet why they are almost always retained in some minimal form. One striking consequence of cytonuclear integration is the existence of 'chimeric' enzyme complexes composed of subunits encoded in two different genomes. Advances in structural biology and comparative genomics are yielding important insights into the evolution of such complexes, including correlated sequence changes and recruitment of novel subunits. Thus, chimeric cytonuclear complexes provide a powerful window into the mechanisms of molecular coevolution.

Abstract

Eukaryotic cells must regulate not just their nuclear genomes, but also the intimate functional interplay with organelle genomes, such as in mitochondria and plastids. This Review discusses the functional and evolutionary implications of cellular genomes being partitioned between the nucleus and organelles, including the translocation of genes and gene products, chimeric enzyme complexes, and insights into molecular coevolution.

* dbsloan@rams.colostate.edu.

Author contributions:

All authors contributed to researching data for the article, discussion of content, and reviewing and editing the manuscript for submission. D.B.S. was primarily responsible for writing the manuscript.

Competing interests

The authors declare no competing interests.

Subject categories

Biological sciences / Evolution / Evolutionary genetics [URI /631/181/2474]

Biological sciences / Genetics / Genomics / Genome evolution [URI /631/208/212/2304]

Biological sciences / Structural biology [URI /631/535]

Biological sciences / Genetics / Genetic interaction [URI /631/208/2490]

Biological sciences / Cell biology / Organelles / Mitochondria [URI /631/80/642/333]

Biological sciences / Cell biology / Organelles / Chloroplasts [URI /631/80/642/1374]

Biological sciences / Genetics / Genome / Mitochondrial genome [URI /631/208/726/2129]

Introduction

The genetic control of eukaryotic cells is divided between the nucleus and one or more cytoplasmic genomes. In one sense, this genomic division of labour is an obvious outcome of the endosymbiotic origins of mitochondria and plastids, as these organelles retain remnants of ancestral genomes from their respective alphaproteobacterial and cyanobacterial progenitors[1]. Nevertheless, it is remarkable that this complex arrangement of multiple genomic compartments has persisted through billions of years of evolution. The genetics of nuclear and cytoplasmic genomes differ in nearly every respect, including genome copy number, mutation rates, modes of inheritance, and mechanisms of replication and expression (Table 1). But they still function in an integrated fashion to maintain arguably the most intimate and important symbioses in the history of life.

The functional interactions between nuclear and cytoplasmic genomes have important evolutionary consequences. The creation of ‘mismatched’ nuclear and cytoplasmic genotypes via hybridization often has notable phenotypic effects, which has sparked recent debate as to whether cytonuclear interactions contribute disproportionately to speciation[2-5]. Cytonuclear interactions also have implications for human health. There are numerous examples of mitonuclear epistasis, such that the penetrance of disease-causing mitochondrial mutations depends upon the nuclear background on which they occur — and vice versa[6]. Here too, however, there has been controversy over the magnitude and biomedical relevance of such effects, which is currently playing out in debates about the risks of performing mitochondrial replacement therapy as an assisted reproductive technology in humans[3,7-9]. Thus, open questions in multiple fields have created a pressing need to better understand the molecular-genetic basis of cytonuclear integration and interactions.

The ultimate source of cytonuclear interactions is the long-term evolutionary process by which mitochondria and plastids have transitioned from free-living bacteria to endosymbionts to organelles. One of the dominant themes in eukaryotic genome evolution is the extreme reduction of cytoplasmic genomes. However, it is rare for mitochondrial genomes (mitogenomes) and plastid genomes (plastomes) to be lost entirely despite their ancient origins ~1-2 billion years ago. These observations raise two classic questions asked from opposite perspectives. First, why has eukaryotic genome evolution asymmetrically favoured reduction of cytoplasmic genomes and transfer of genetic control to the nucleus? Second, why has this process generally stopped short of complete elimination of cytoplasmic genomes?

Because a small number of key genes have been retained in mitogenomes and plastomes, one defining outcome of cytonuclear integration is the maintenance of enzyme complexes composed of interacting subunits from both cytoplasmic and nuclear genomes[10]. These ‘chimeric’ complexes include enzymes involved in cellular respiration and photosynthesis, which are central to eukaryotic bioenergetics. Because these enzymes are endosymbiotically derived, we can compare them to their bacterial counterparts and dissect their history of diversification across eukaryotes.

Here, we review recent advances and future prospects for understanding cytonuclear integration. First, we describe the shift of genetic control from organelle genomes to the nucleus. We also consider how recent studies of younger endosymbionts have provided insights into the process and why it rarely results in complete loss of organelle genomes. Then, we review recent evidence that the chimeric enzyme complexes that now exist within organelles often represent arenas for rapid molecular coevolution. We note that having interacting gene products from two different genomes enhances our ability to detect and dissect general coevolutionary mechanisms that are likely to be playing out in other multisubunit complexes throughout the cell. In general, we focus on the molecular-genetic basis of these interactions and direct readers to recent reviews on related subjects pertaining to cell function, including mechanisms of protein trafficking and import, and intracellular (i.e., anterograde or retrograde) signalling[11-14].

Transfer of genetic control to the nucleus

Cytoplasmic genome reduction.

The reduction in gene content that occurs when organisms evolve to live exclusively inside the cells of other species is a remarkably repeatable phenomenon. In addition to mitochondria and plastids, this pattern of reductive genome evolution has occurred in several single-celled eukaryotes that have transitioned to an obligately intracellular lifestyle[15] and in a tremendous diversity of bacterial and archaeal endosymbionts that are independently derived from divergent phylogenetic lineages and found in a wide range of hosts[16,17]. There are at least three mechanisms responsible for this pattern: first, outright gene loss owing to relaxed selection on genes that are redundant with host/nuclear genes or unnecessary in an intracellular context; second, endosymbiotic gene transfer (EGT) to the nucleus (also known as intracellular gene transfer (IGT)); and third, recruitment of other nuclear genes to replace endosymbiont/organellar functions[18-22].

By comparing mitogenome content across the extant diversity of eukaryotes, researchers have been able to infer the evolutionary timing of mitochondrial gene loss events in different lineages [21-24]. Such analyses have divided the history of mitochondrial gene loss into two general phases[24]. The first is a period of extensive reduction preceding the last eukaryotic common ancestor (LECA). As such, all known mitogenomes have some subset of a small pool of 69 ancestral protein-coding genes (Figure 1), which represents a tiny fraction of the >1,000 genes estimated to have been present in the bacterial progenitor of mitochondria[25]. The second phase of mitogenome reduction is a heterogeneous one with variation both through time and across phylogenetic lineages, such that extant eukaryotes differ greatly in organelle gene content[23,24]. Sequenced mitogenomes contain anywhere from just two protein-coding genes (*cox1* and *cox3*) in the alveolate *Chromera velicia*[26,27] to as many as 66 of the 69 ancestral proteins in the jakobid *Andalucia godoyi*[28] (Figure 1). Some mitogenomes encode a full complement of tRNAs, whereas other species have lost all of these genes[29,30]. For individual lineages, current mitochondrial gene content has been shaped by long periods of stasis punctuated by episodes of accelerated gene loss. For example, prior to the divergence of bilaterian animals, lineage-specific losses and transfers led to a highly reduced set of only 13 protein-coding genes in the mitogenome, but this set

has remained nearly unchanged for the ensuing hundreds of millions of years. By contrast, active loss and EGT are currently occurring in many angiosperm plant lineages[31]. Plastome evolution has exhibited a similar pattern of temporal and phylogenetic variation, with plastomes of photosynthetic organisms containing anywhere from 21 protein-coding genes in Cladophorales green algae to over 200 in red algae, and anywhere from a complete set of tRNA genes to none at all[32,33]. Even though there is substantial variation in the extent of mitochondrial and plastid gene loss among lineages, there is striking parallelism between mitogenomes and plastomes with respect to the types of genes that are more or less likely to be lost, which even extends down to the level of the specific ribosomal protein subunits that have been retained[34]. Importantly, this parallel and directional history of gene loss may lead us to misinterpret some independent losses as resulting from a single, older loss event[24]. Regardless, in extant mitogenomes and plastomes, there is a highly reduced gene content, with the retained genes being dominated by the core subunits of cellular-respiration and photosynthetic enzyme complexes as well as key components of translational machinery.

Asymmetrical and opposite movement of genes versus gene-products between cytoplasmic genomes and the nucleus.

There are two striking asymmetries that have defined the long-term evolution and integration of nuclear and cytoplasmic genomes (Figure 2). First, the history of EGT has been almost unidirectional from cytoplasmic genomes to the nucleus[18,19]. Second, the trafficking of gene products (i.e., RNAs and proteins) occurs overwhelmingly in the opposite direction — from the nucleus and cytosol into the organelles[11,14]. Neither of these asymmetries is absolute. Many mitogenomes have acquired genes from the nucleus or other foreign sources[35-39], and horizontal gene transfer (HGT) has also introduced novel genes into some plastomes[40,41]. In addition, there is growing recognition that organelle-derived RNAs (including rRNAs[42-44], tRNAs[45,46], other noncoding RNAs[47-49], and RNAs encoding small peptides[50,51]) can be released to the cytosol, with likely roles in signalling and other pathways. Nevertheless, it is clear that the overall tide of gene movement has been towards the nucleus, with gene products generally trafficked in the reverse direction. These patterns often lead to the false assumption that all transferred genes encode proteins that are targeted back to the organelles or that all imported proteins are encoded by nuclear genes formerly found in cytoplasmic genomes. Phylogenetic assessments have strongly rejected this oversimplified view. Many of the products of nuclear genes derived from EGT are not targeted to the organelles[25,52], and major fractions of the mitochondrial and plastid proteomes are derived from pre-existing nuclear genes that have been recruited into novel functional roles[53-55] (Figure 2).

Doolittle[56] argued that the propensity of organellar (and other endosymbiotic) genes to be transferred to the nucleus and replace existing functions is not unexpected. Instead, it can be viewed as a natural byproduct of asymmetrical opportunities for DNA movement between genomes[57], likely owing to the high copy number of cytoplasmic genomes and frequent release of their DNA into the cytosol[58]. The logic follows that, even though the vast majority of cytoplasmic DNA insertions into the nucleus have no effect and are eventually lost, the perpetual transfer of DNA creates the opportunity for occasional insertions to become expressed and duplicate the function of an existing nuclear gene. In some of these cases, the original (but now redundant) nuclear gene might then be lost, resulting in a functional

replacement. Even if replacement events are rare, this model predicts a directional process in which transferred genes increasingly take over pre-existing nuclear-encoded functions because the paucity of DNA transfer in the opposite direction precludes reversing this trend. As such, each functional replacement would amount to the irreversible "clicking" of a "gene-transfer ratchet"[56].

We suggest that the early evolution of specialized machinery for import of nuclear gene products across organelle membranes (i.e., the TIM/TOM[59] and TIC/TOC[60] translocases in mitochondria and plastids, respectively) created an analogous asymmetry that allowed for widespread recruitment of existing nuclear gene products into novel functional roles in the organelles (Figure 2). In a sense, the establishment of these import mechanisms may have created a 'protein-import ratchet'. Specifically, their presence may result in the persistent opportunity for promiscuous or leaky import of nuclear-encoded proteins into the organelles. Analogous to Doolittle's proposed gene-transfer ratchet, background levels of promiscuous import of nuclear-encoded proteins may lead to functional redundancy with genes in the mitogenome or plastome. In occasional cases where the redundant cytoplasmic genes are then lost, the end result would be a 'locked in' functional replacement by a pre-existing nuclear gene and selection for more specific and sustained import of that nuclear-encoded protein. Under a gene-transfer ratchet, the functional replacement process is asymmetrical because the raw movement of DNA is biased to go from cytoplasmic genomes to the nucleus. Under a protein-import ratchet, the functional replacement process is asymmetrical because the early evolution of mitochondrial and plastid import machinery led to greater opportunities for non-adaptive import of nuclear-encoded proteins into the organelles than for non-adaptive export of cytoplasmically encoded gene products to other parts of the cell.

These proposed 'ratchets' are unlikely to fully explain the asymmetrical and opposite movement of genes and gene products. But they establish null models for the direction of these asymmetries, and the effects of other evolutionary forces can be cast as either accelerating or slowing/halting the ratcheting.

Mechanisms and causes of gene transfer to the nucleus.

The initial step in a functional EGT event is the physical insertion of cytoplasmic DNA into the nuclear genome. First detected decades ago, mitochondrial and plastid DNA insertions in the nucleus are now referred to as nuclear mitochondrial DNAs (NUMTs) and nuclear plastid DNAs (NUPTs), respectively[61-63]. Key advances came with the development of methods that used reporter constructs that could only be expressed if they were relocated to the nucleus to directly observe the physical movement of mitochondrial and plastid DNA to the nuclear genomes. These approaches revealed that DNA transfer is surprisingly frequent[57,64] and that genomic DNA (rather than cDNA or RNA) is likely to be the most common vehicle[65,66]. These experimental methods have been complemented with comparative approaches based on the proliferation of genomic sequence data[58], and recent bioinformatic advances have helped overcome technical hurdles arising from the filtering of organellar sequences during nuclear genome assembly and the duplication and/or fragmentation that occurs after initial NUMT or NUPT insertion[67]. Insertions of cytoplasmic DNA can be large and even encompass entire genome copies, such as the 620 kb NUMT on

chromosome 2 of *Arabidopsis thaliana*[68]. They can also be structurally complex, with multiple fragments from different regions of the mitogenome and/or plastome often merged into the same location in the nuclear genome[65,69,70]. Comparative studies have revealed enormous variation in the amount of NUMT or NUPT sequence across species, including positive associations with nuclear-genome size[58] and the number of organelles per cell[71]. The latter observation is consistent with the hypothesis that the lack of organelle breakdown in species with only a single organelle per cell creates a barrier to DNA transfer[72]. Together, these observations support a model wherein breakdown of mitochondria and plastids leads to release of free organellar DNA into the cytosol, which can then passively enter the nucleus, with incorporation into the nuclear genome most likely via mechanisms that repair double-stranded breaks in nuclear DNA (although alternative mechanisms may also play a role[58]).

Insertion of cytoplasmic DNA into the nuclear genome is necessary but not sufficient for functional EGT. Indeed, the vast majority of NUMT and NUPT insertions are likely to be nonfunctional, with either deleterious or neutral consequences, and are eventually deleted from the genome. A true functional ‘transfer’ is a multi-step process that involves: first, insertion of a copy into the nucleus; second, acquisition of the necessary functional elements for expression/targeting; and third, subsequent loss of the original cytoplasmic copy. This process can play out over tens of millions of years and include long transitional stages in which expressed gene copies persist in both the nucleus and the cytoplasmic genome[18,73,74]. The second step in this process is complex. Nuclear and cytoplasmic genomes rely on different machinery for gene expression, so functional EGT requires acquisition of nuclear-specific promoters and regulatory elements. Moreover, although some proteins have innate features that enable mitochondrial or plastid import[75,76], the majority of targeting is based on an amino-terminal presequence that must also be acquired if transferred genes are to be targeted back to the organelle[11,14]. This process can depend on fortuitous nuclear insertion sites and take advantage of existing regulatory or targeting elements in the nuclear genome[75]. In addition, changes that occur post-insertion, including exon shuffling and evolution of alternative splicing, can facilitate the acquisition of functional elements[75,77].

Compartment-specific features of gene expression also create barriers to functional EGT. Many mitogenomes and some plastomes now employ modified genetic codes, and cytoplasmic gene expression often involves complex forms of intron splicing and RNA editing that do not occur in the nucleus[78]. In such cases, directly transferred cytoplasmic genes would not yield the same protein if expressed in the nucleus, and functional EGT often necessitates removal of introns and ‘hardcoding’ of RNA-editing events at the DNA level. As such, there has been a long-running debate[79] as to whether functional EGT may be mediated by movement of cDNAs (i.e., reverse-transcribed mRNAs that have already been spliced and edited). This hypothesis has been supported by the fact that many transferred genes resemble mature mRNAs, but it is at odds with the fact that cDNA-based movement is undetectably rare in experimental studies, whereas movement of genomic DNA is prolific[79]. Experimental techniques have shown that the presence of a single intron or RNA-editing site is not prohibitive for functional EGT and can be overcome by promiscuous splicing activity or use of alternative start codons[66,80], but it is difficult to envision how the highly fragmented and/or edited genes found in some cytoplasmic genomes could become functional in the nucleus after direct insertion of genomic DNA. A possible reconciliation

was proposed almost two decades ago by Henze and Martin[81]. These authors hypothesized that ‘retroprocessing ’ may first occur within mitochondria and/or plastids, resulting in mature coding sequences that are incorporated back into cytoplasmic genomes, which are then more amenable to functional EGT via conventional mechanisms of genomic-DNA transfer. The first empirical support for this hypothesis was provided by two recent phylogenetic studies that reconstructed the evolutionary timing of events and showed that specific cases of EGT in angiosperms were preceded by retroprocessing within the mitogenome[73,82]. It remains to be seen how widespread this mechanism is and whether the derived features of gene expression in many cytoplasmic genomes create a long-term barrier to EGT and contribute to the retention of cytoplasmic genomes (Box 1).

Although substantial progress has been made in understanding EGT mechanisms, the role of natural selection in driving this process is less clear. It has long been hypothesized that relocation of genes to the nucleus confers a benefit by enabling them to escape the high mutation rates and/or reduced efficacy of selection found in predominantly asexual organelle genomes[83-85]. However, assessing this hypothesis has proven challenging, and the classic assumption that organelle genomes are subject to deleterious mutation accumulation has become a source of renewed controversy[5,86,87]. Therefore, it remains difficult to reject the null model that the extensive history of functional EGT is largely the product of a non-adaptive gene-transfer ratchet[56], making this an important area for new empirical and theoretical research.

‘Interchangeable parts’: replacement of cytoplasmic genes by nuclear counterparts.

In addition to EGT, the history of cytonuclear integration has also involved extensive repurposing of existing nuclear genes to replace cytoplasmically encoded functions. Examples include early events in eukaryotic evolution, such as the recruitment of nuclear genes to function in mitochondrial protein translocation[21] and mitochondrial division[88]. More recent, lineage-specific events have also occurred including the replacement of a mitochondrial-encoded subunit of the mitochondrial ribosome in the common ancestor of seed plants with its nuclear-encoded counterpart from the cytosolic ribosome[89]. The acquisition of plastids in some eukaryotic lineages has enabled further functional replacement via dual-targeting of individual proteins to both mitochondria and plastids[90,91]. HGT from other foreign sources has also led to functional replacement, such as the use of phage-like proteins now found in the nucleus for replication and transcription of cytoplasmic genomes[21]. Indeed, recent investigations into the establishment of other obligate bacterial endosymbionts has stimulated an active debate about the role of HGT from diverse sources in the origin of mitochondria and plastids (Box 2).

Although EGT and functional replacement have both played major roles in the history of cytonuclear integration, their relative impacts have differed across functional classes of genes. For example, to our knowledge, there are no documented cases of functional transfer of a cytoplasmic tRNA gene to the nucleus. Instead, tRNA-gene losses from the mitogenome generally coincide with novel import of existing nuclear-encoded, cytosolic tRNAs[29,90]. Moreover, many aminoacyl-tRNA synthetases (aaRSs) have a history of sharing or replacement across cytosolic and organellar compartments[90]. Conversely, EGT has been the

dominant mode of cytoplasmic gene loss for subunits in oxidative phosphorylation (OXPHOS) and photosynthetic enzyme complexes. One obvious explanation is that some molecular systems (e.g., protein translation) have homologous or analogous counterparts across cellular compartments to draw from, whereas others (e.g., OXPHOS and photosynthesis) do not. This view is consistent with the argument of Woese et al.[92] that proteins such as aaRSs with relatively modular functions that are shared across different systems are most amenable to HGT (and by extension to EGT). Nevertheless, it is striking that functional replacement is still feasible and so widespread given the ancient timescales of divergence between the mitochondrial (alphaproteobacterial), plastid (cyanobacterial), and nuclear (archaeal) compartments — especially when juxtaposed with the observation that even single-nucleotide substitutions can be sufficient in some cases to produce major cytonuclear incompatibilities among close relatives[93,94].

Cytonuclear enzyme complex coevolution

Enzyme complexes composed of subunits from both nuclear and cytoplasmic genomes (e.g., organellar ribosomes and many OXPHOS and photosynthetic enzymes) are hallmarks of cytonuclear integration[10]. These complexes are arenas for intimate interactions between genes that are subject to different mutation rates, effective population sizes, and modes of inheritance. One general hypothesis is that higher mutation rates and/or inefficient selection against weakly deleterious mutations lead to rapid evolution in cytoplasmic genomes, creating selection for subsequent changes in nuclear-encoded protein sequence and the overall structure and composition of enzyme complexes[95,96]. From this perspective, cytoplasmic genomes are viewed as the primary ‘drivers’ of cytonuclear coevolution, whereas nuclear genomes are the primary ‘responders’. There is growing evidence for components of this model, but many of its key tenets await rigorous testing.

Rates of sequence evolution.

If cytonuclear enzyme complexes are indeed arenas for molecular coevolution, one expectation is that nuclear-encoded subunits in these complexes will exhibit faster rates of evolution than other nuclear-encoded proteins. A number of studies have found support for this prediction[97-99]. However, alternative interpretations have also been proposed. For example, nuclear-encoded subunits may not be subject to strong functional constraint because they occupy peripheral positions in the complexes[100] or because selection for efficiency in some mitochondrial processes such as translation is not very intense compared to corresponding nuclear processes[29,101,102].

To disentangle these alternatives, recent studies have used eukaryotic lineages with large variation in rates of mutation and/or sequence evolution in cytoplasmic genomes[101,103-109]. These studies have found that lineages with rapidly evolving cytoplasmic genomes also show elevated rates of amino-acid substitution in interacting nuclear-encoded proteins (but not in other nuclear proteins) compared to related lineages with slower rates of cytoplasmic evolution. They have also often detected signatures of positive selection. Therefore, cytonuclear coevolution appears to be a major determinant of rates of sequence evolution in interacting nuclear-encoded proteins. Comparative studies have also taken advantage of

classic ‘built-in’ negative controls, such as OXPHOS Complex II (succinate dehydrogenase), to show that complexes that are involved in the same functional pathways but are composed entirely of nuclear-encoded subunits do not exhibit the same signatures of cytonuclear coevolution[104,110]. Examining patterns of amino-acid substitution in the context of protein structures has provided additional support, suggesting that changes occur in a spatially correlated fashion at interfaces between cytoplasmic and nuclear proteins[95,103,104,109]. Using a phylogenetic framework to assess timing of sequence changes, Osada and Akashi[95] further showed that substitutions in mitochondrial-encoded subunits in OXPHOS Complex IV (cytochrome c oxidase) often precede changes at interacting sites in nuclear-encoded subunits, suggesting that nuclear genomes are responding to cytoplasmic changes.

Collectively, this recent body of work has highlighted the advantages of studying molecular coevolution in systems where interacting gene products are encoded in different genomes. The above findings have often been interpreted as evidence that coevolutionary changes in the nucleus act to compensate for genomic deterioration in mitochondria and plastids. However, there is little direct evidence for this specific mode of compensatory coevolution, and debates about the efficiency (or inefficiency) of selection on cytoplasmic genomes have cast doubt on the generalities of nuclear compensation[86,87]. Thus, a clear challenge in the field is to distinguish cases in which nuclear changes are counteracting deleterious cytoplasmic mutations from those in which coevolutionary responses in the nucleus are spurred by initially beneficial or neutral cytoplasmic mutations[5].

More generally, research in this area often suffers from inconsistent use of the term ‘coevolution’. First described by Ehrlich and Raven[111] in the context of interactions between plants and butterflies, the most rigorous definition of coevolution involves reciprocal effects of selection on interacting populations (or interacting molecules in the cases we discuss). As such, both parties are expected to change in response to selective pressures exhibited by the other. However, the term coevolution is often applied in looser ways and sometimes assumed based only on co-occurrence and interactions without direct evidence for a history of changes in response to selection[112]. For example, in cases of correlated changes in rates of evolution, it is always possible that new selection pressures on the overall function of a complex result in correlated responses across all subunits, but such patterns do not necessarily mean that the subunits are coevolving in response to the changes in other subunits. Therefore, the recent research efforts described above that are examining the relative evolutionary timing of changes and their effects on structural interactions are key for testing predictions about effects of coevolutionary pressures[10] and whether such pressures are truly reciprocal or largely unidirectional among interacting subunits. In addition, comparisons between chimeric cytonuclear complexes and those encoded solely by nuclear genes are an informative way to identify effects that are specific to cytonuclear interactions[104,110].

Recruitment of novel subunits.

Cytonuclear interactions and the potential for coevolutionary dynamics are also evident at a structural level in chimeric multisubunit complexes. Mitochondrial ribosomes and OXPHOS complexes have been reshaped and substantially enlarged relative to their ancestral,

bacterial-like forms via the acquisition of nuclear-encoded supernumerary subunits (i.e., eukaryotic-specific subunits gained since endosymbiosis)[21,113,114] (Figure 3). Plastid enzymes have also recruited novel subunits, although they have generally done so to a much lesser extent and more closely resemble their bacterial counterparts[96,115,116] (Figure 3). The evolution and function of the novel subunits within cytonuclear complexes have been a source of intrigue, and recent progress in comparative genomics and structural biology is yielding answers to longstanding questions.

Broad phylogenetic sampling and application of sensitive methods to detect sequence homology have revealed that the majority of supernumerary mitochondrial subunits were incorporated pre-LECA, coinciding with a period of extensive EGT and functional gene loss and replacement in cytoplasmic genomes[21,52]. For example, mitochondrial OXPHOS Complex I (NADH:ubiquinone oxidoreductase) has an especially rich and well-studied history of subunit acquisition[117,118]. The entirety of this complex in many extant bacteria consists of only 14 subunits, but it has expanded to a total of 45 and 49 subunits in the mitochondria of mammals and angiosperms, respectively (Figure 3). In addition to the 14 ancestral subunits, at least 26 eukaryotic-specific subunits in this complex appear to have been present prior to the radiation of extant eukaryotic lineages[117]. Nevertheless, a more limited rate of recruitment of novel subunits has continued post-LECA, with a variable number of gains across eukaryotic lineages[21]. Many supernumerary subunits are members of larger gene families, and it appears that they are often opportunistically recruited from the pool of nuclear-encoded proteins that are already targeted to the mitochondria. Indeed, some supernumerary subunits can be traced back to duplications from other mitochondrial complexes or even other subunits in the same enzyme complex[114,118]. However, the specific origins of many supernumerary subunits remain uncertain. For example, 21 of the acquired subunits in mammalian Complex I do not have any identified homologues outside of eukaryotes[118].

Ever since the discovery of supernumerary subunits in mitochondrial OXPHOS and ribosomal complexes, there has been extensive interest and speculation about their functional role. These subunits tend to ‘coat’ the periphery of enzyme complexes (Figure 3), and one early idea was that they serve as ‘molecular prostheses’. This hypothesis was motivated by the observation that rRNAs in mitochondrial ribosomes in metazoans are shortened[119], and mitochondrial-encoded OXPHOS subunits are also unusually short in metazoans[96]. Thus, it seemed plausible that nuclear-encoded supernumerary subunits filled the structural gaps resulting from reductions in these mitochondrial-encoded subunits. However, this hypothesis has been discounted based on subsequent structural data, because supernumerary subunits do not tend to occupy the spaces left by shortened mitochondrial-encoded subunits[96,120]. It is also inconsistent with the timing of subunit gains, as the pre-LECA incorporation of most supernumerary subunits is well in advance of metazoan-specific reductions in the size of mitochondrial-encoded subunits[21,52,96].

Alternative hypotheses have focused on the possibility that supernumerary subunits may confer entirely novel functions to their respective enzyme complexes. For example, the inner surface of the membrane arm of OXPHOS Complex I often contains lineage-specific supernumerary subunits, including a novel carbonic anhydrase in plants, which has led to the

suggestion that this region serves as a ‘workbench’ for attaching proteins with novel functions[117].

Recent studies have increasingly shifted focus to the role of supernumerary subunits in assembly and stability of their respective complexes. For example, Stroud and colleagues[121] individually knocked out 31 supernumerary subunits in human OXPHOS Complex I. They found that 25 of these subunits (81%) are required for proper assembly of the complex and that subunit loss generally affected the stability of neighbouring subunits. In addition, van der Sluis *et al.*[96] modelled structural deficiency in OXPHOS complexes and found that models including supernumerary subunits were more stable. These efforts have been augmented by publication of the first high-resolution structures of Complex I and related ‘supercomplexes’, which are higher-order associations involving multiple different OXPHOS complexes[122-126] and have some assembly features that are widely conserved across eukaryotes[127].

Complex I is the primary site of electron entry to the respiratory system of eukaryotes, reflecting the metabolic simplification of the more linearized mitochondrial electron transport system relative to bacteria with many branched, alternative metabolic pathways[115]. Thus, enhanced interactions with downstream complexes may improve the efficiency of electron transfer and reduce the formation of reactive oxygen species (ROS), with supernumerary subunits playing an important role in supercomplex assembly and stability. Formation of these respiratory supercomplexes appears to specifically enhance Complex I stability, and may be dynamically regulated to facilitate cellular metabolic adjustments and ROS-mediated cell signalling[128]. A ‘scaffold’ of supernumerary subunits linking the peripheral arm and transmembrane domain of complex I[129] is thought to confer additional complex stability and reduce formation of ROS[130], and perhaps regulate the coupling of ubiquinone reduction to proton translocation[123]. At least four supernumerary subunits of Complex I appear to facilitate Complex III (cytochrome *bc*1 complex) dimerization and association with Complex I through direct protein-protein interactions that involve mitochondrial-encoded subunits[124]. These interactions might improve the efficiency of electron transfer within and between these complexes, and facilitate allosteric regulation of Complex I by matrix components[131]. In addition, a nuclear-encoded Complex IV subunit (COX7A2L) is thought to stabilize the interaction of Complex III and Complex IV, as well as Complex IV dimerization[132].

As the assembly and stability of complexes and supercomplexes emerge as the predominant (though not exclusive) roles for supernumerary subunits, questions remain about the evolutionary mechanisms that yielded these functions and why the subunits are absent in bacteria. It is possible that unique demands of organelle function in eukaryotes have created novel selection pressures for altered and augmented mechanisms to assemble and stabilize these enzyme complexes. For example, there is evidence for inducible subunits that facilitate cellular stress responses, such as regulation of Complex IV activity in hypoxia[133]. In addition, adaptive evolutionary shifts in respiratory complex oligomerization across and within taxa[134] may enable selective interactions between respiratory chain redox centres and electron carriers to facilitate electron entry and channelling from specific substrate oxidation pathways[135].

It has also been proposed that recruitment of novel subunits serves as another form of compensatory cytonuclear coevolution that offsets the effects of weakly deleterious changes in cytoplasmic genes[96]. Under this model, after weakly deleterious and destabilizing changes occur in cytoplasmically encoded subunits, there is selection to recruit entirely new subunits to the complex to provide structural stabilization. A third possibility is that the expanded subunit composition is a form of 'constructive neutral evolution'[21,136]. This alternative model proposes that subunit recruitment would have initially been neutral, but it would have also altered the fitness landscape for subsequent mutations in core mitochondrial-encoded subunits. As such, some mutations that would otherwise have disrupted complex stability could then occur and spread neutrally by genetic drift without harmful consequences, thereby 'locking in' the functional importance of formerly non-adaptive supernumerary subunits. Such non-adaptive thinking parallels neutral 'ratchet' arguments that establish null models for the evolutionary history of functional EGT and protein import (see "Asymmetrical and opposite movement of genes versus gene-products between cytoplasmic genomes and the nucleus"). However, little effort has yet been made to distinguish between these various models. Regardless of the initial evolutionary pressures that led to recruitment, the presence of nuclear-encoded supernumerary subunits in mitochondrial enzyme complexes now provides much of the raw genetic material for mitonuclear coevolution.

Gene duplication and tissue-specific paralogues.

Gene duplication represents another key mechanism that has shaped the evolution of cytonuclear enzyme complexes. As noted in the preceding section, paralogues have been one source of the novel subunits recruited to these complexes. For example, the metazoan duplication of the nuclear *MRPS10* gene, which encodes a component of the mitochondrial small ribosomal subunit, produced a paralogue (*MRPL48*) that has been incorporated into the large ribosomal subunit[114]. The caseinolytic protease (Clp) complex in plastids offers another striking case of gene duplication. In bacteria, the core of this complex typically comprises 14 copies of the same subunit encoded by a single gene (*clpP*). In cyanobacteria, the lineage from which plastids were endosymbiotically derived, this ancestral gene has expanded into a family of four paralogues. Since the origin of plastids, there has been further gene duplication such that the core Clp complex in land plants is encoded by a single plastid gene and eight or more related nuclear genes[137], which collectively exhibit some of the most dramatic examples of correlated evolutionary rates in cytonuclear interactions[103].

Another outcome of gene duplication is the generation of tissue-specific paralogues — a phenomenon that has occurred repeatedly in metazoans, especially in OXPHOS Complex IV. In this complex, related nuclear gene copies encode interchangeable subunits that are differentially expressed across tissues, such as heart, skeletal muscle, liver, and testis[138]. Many of these paralogues can be highly divergent in amino-acid sequence. The flexibility of OXPHOS complexes to 'swap' such divergent subunits across tissue types paints a somewhat contrasting view from studies that have focused on the disruptive effects of individual amino-acid substitutions on the molecular interactions between mitochondrial and nuclear gene products[93-95,104]. The role of these paralogous subunits is generally thought to provide a match to the metabolic demands of their respective tissues[138]. However, the

existence of testis-specific paralogues also raises the intriguing possibility that gene duplication may provide a mechanism for the nucleus to compensate for the inherent conflict associated with the activity of maternally inherited cytoplasmic genes in male-specific functions (i.e., 'mother's curse')[139-142]. More generally, gene duplication in the nucleus creates an additional dimension to cytonuclear coevolution that is still largely underappreciated.

Conclusions and future directions

The history of cytonuclear integration has been characterized by two striking asymmetries: first, the movement of genes from cytoplasmic genomes to the nucleus, many of which have evolved functions outside of mitochondria and plastids; and second, the recruitment of nuclear-encoded gene products into mitochondria and plastids, many of which did not originate from transferred cytoplasmic genes. These patterns may be partially attributed to basic asymmetries between the nuclear and cytoplasmic genomes, including differences in genome copy number, which determines relative rates of DNA movement, and the early evolution of mitochondrial and plastid protein import (but not export) machinery. The outcome of this long-term evolutionary process is an essential set of chimeric enzyme complexes composed of both nuclear- and cytoplasmically-encoded subunits. These complexes represent active sites of molecular coevolution, exhibiting complementary amino-acid substitutions and recruitment of novel subunits. Research in this field is taking advantage of the opportunity to better separate cause and effect in molecular coevolution when genes residing in entirely different genomic contexts must intimately interact.

Our understanding of the historical process of cytonuclear integration and coevolution has advanced in recent years thanks to the proliferation in comparative genomic data and key achievements in structural biology. Nevertheless, a host of unanswered questions remain, offering exciting challenges to shape ongoing research efforts. For example, the history of cytonuclear integration has involved remarkable cases of functional replacement between anciently divergent genes. Although the evidence for functional replacement is often clear, our understanding of the evolutionary process is far more limited. Notably, when mitochondrial tRNAs have been lost in favour of importing nuclear-encoded cytosolic tRNAs, whole suites of enzymes responsible for maturation and charging of tRNAs have also apparently been replaced or retargeted[29], raising fascinating questions about the order of such changes and how they occurred to entire interacting systems without harmful disruptions. Lineages such as angiosperms, in which this turnover of mitochondrial tRNAs is actively occurring[143], offer attractive models to understand these dynamics.

Another open question pertains to the apparently conservative nature of plastid evolution. Smith and Keeling[78] have recently argued that, relative to their mitogenome counterparts, plastomes generally have not evolved nearly as extreme and unusual genome architectures. Plastomes also retain more of their ancestral gene content. This pattern seems to extend to the structure of cytonuclear enzyme complexes and recruitment of supernumerary subunits, as plastid photosynthetic enzymes and ribosomes have remained far more 'bacterial-like' (Figure 3). An immediately obvious explanation is that plastids are evolutionarily younger organelles and have thus had less time to diverge. However, this argument appears weak

when confronted with evidence that most of the radical changes in mitochondrial gene content and enzyme composition arose in the earliest stages of mitonuclear integration[21,24,144]. Researchers are thus left to ponder features of plastid biology, such as the extreme levels of protein expression and intense selection on photosynthetic efficiency, that might explain the imperfectly parallel trajectories of mitochondrial and plastid evolution.

Our focus in this Review has been on chimeric cytonuclear enzyme complexes. They are obvious arenas for cytonuclear interactions, and they are an exquisite model for investigating general processes of molecular coevolution. But that does not mean they are the sole or even the most important source of epistatic interactions in cytonuclear genetics. Cytonuclear incompatibilities may involve nuclear-encoded proteins that function outside of these complexes or even outside of mitochondria and plastids entirely. Therefore, the interactions within chimeric cytonuclear enzyme complexes may only be the tip of the iceberg when it comes to the functional, biomedical, and evolutionary consequences of divided genetic control in the eukaryotic cell. We eagerly await efforts to measure the relative importance of these chimeric complexes within the full scope of cytonuclear genetic interactions.

Acknowledgements

We thank the reviewers for their insightful comments on an earlier version of this manuscript. This work was supported by a National Institutes of Health (NIH) postdoctoral Fellowship (F32 GM116361 to J.C.H.), a National Science Foundation (NSF) Graduate Research Fellowship (DGE-1321845 to A.M.W.), an NSF-funded GAUSSI Graduate Fellowship (DGE-1450032 to J.M.W.) and grants from NSF (MCB-1412260 and MCB-1733227 to D.B.S.), NIH (NIGMS R01 GM118046 to D.B.S.), and the U.S. Department of Agriculture (2015-67017-23143 to A.J.C.).

Glossary

Plastids
Organelles that were endosymbiotically derived from cyanobacteria and can differentiate into multiple functional types, the most well-known of which is the chloroplast.

Endosymbiotic gene transfer (EGT).
The process by which genes are functionally transferred from cytoplasmic genomes to the nucleus (also known as intracellular gene transfer).

Last eukaryotic common ancestor (LECA).
The most recent common ancestor of all extant eukaryotes — an organism that is thought to have already acquired mitochondria and undergone substantial cytonuclear integration.

TIM/TOM
The translocase of the inner mitochondrial membrane (TIM) and translocase of the outer mitochondrial membrane (TOM) mediate import of nuclear-encoded proteins into the mitochondria.

TIC/TOC
The translocase of the inner chloroplast membrane (TIC) and translocase of the outer chloroplast membrane (TOC) mediate import of nuclear-encoded proteins into the plastids.

Nuclear mitochondrial DNAs (NUMTs).
Insertions of mitochondrial DNA into the nucleus (usually non-functional).

Nuclear plastid DNAs (NUPTs).
Insertions of plastid DNA into the nucleus (usually non-functional).

Double-stranded breaks
Breaks in DNA molecules, that when subsequently repaired by processes such as non-homologous end-joining can result in incorporation of other DNA sequences.

Retroprocessing
The process by which a mature RNA transcript is reverse transcribed and recombined back into the genome.

Oxidative phosphorylation
A biochemical process that occurs in the mitochondria and is mediated by a set of cytonuclear enzyme complexes, in which energy generated by electron transfer results in the synthesis of adenosine triphosphate (ATP).

Supernumerary subunits
Protein subunits within cytonuclear enzyme complexes that were not present in the bacterial progenitors of mitochondria or plastids and have been recruited to these complexes during eukaryotic evolution.

Paralogues
Genes that are related to each other as the result of an earlier gene duplication event within a genome.

Mother's curse
The concept articulated by Frank and Hurst[139] and later named by Gemmell *et al.*[140]that cytoplasmic alleles that are harmful to male reproduction may persist in populations because strict maternal inheritance shields these effects from selection.

References

1. Gray MW & Archibald JM in Genomics of Chloroplasts and Mitochondria, Advances in Photosynthesis and Respiration (eds Bock R & Knoop V) 1>–30 (Springer, 2012).
2. Burton RS & Barreto FS A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? Molecular Ecology 21,4942–4957, doi:10.1111/mec.12006; 10.1111/mec.12006 (2012). [PubMed: 22994153] This review argues for a special role of cytonuclear interactions in the process of speciation and summarizes extensive work in Tigriopus copepods, a classic system for mitonuclear biology.
3. Eyre-Walker A Mitochondrial replacement therapy: are mito-nuclear interactions likely to be a problem? Genetics 205, 1365–1372 (2017). [PubMed: 28360127] The author of this paper argues on both theoretical and empirical grounds that the effects of mitonuclear incompatibilities on reproductive isolation and the risks associated with mitochondrial replacement therapy have been exaggerated.
4. Hill GE The mitonuclear compatibility species concept. Auk 134, 393–409 (2017).
5. Sloan DB, Havird JC & Sharbrough J The on-again, off-again relationship between mitochondrial genomes and species boundaries. Molecular Ecology 26, 2212–2236 (2017). [PubMed: 27997046]

6. Dowling DK Evolutionary perspectives on the links between mitochondrial genotype and disease phenotype. Biochimica et Biophysica Acta 1840, 1393–1403, doi:10.1016/j.bbagen.2013.11.013 [doi] (2014). [PubMed: 24246955]
7. Reinhardt K, Dowling DK & Morrow EH Mitochondrial replacement, evolution, and the clinic. Science 341, 1345–1346, doi:10.1126/science.1237146 [doi] (2013). [PubMed: 24052294]
8. Chinnery PF et al. The challenges of mitochondrial replacement. PLoS Genetics 10, e1004315, doi: 10.1371/journal.pgen.1004315 [doi] (2014). [PubMed: 24762741]
9. Dobler R, Dowling DK, Morrow EH & Reinhardt K A systematic review and metaanalysis reveals pervasive effects of germline mitochondrial replacement on components of health. Human Reproduction Update, doi:10.1093/humupd/dmy018 (2018).This metanalysis reports evidence that mitonuclear 'mismatches' can have detrimental consequences, and the authors contend that this is an important consideration for applications of mitochondrial replacement therapy in humans.
10. Rand DM, Haney RA & Fry AJ Cytonuclear coevolution: the genomics of cooperation. Trends in Ecology & Evolution 19, 645–653, doi:10.1016/j.tree.2004.10.003 (2004). [PubMed: 16701327]
11. Paila YD, Richardson LGL & Schnell DJ New insights into the mechanism of chloroplast protein import and its integration with protein quality control, organelle biogenesis and development. Journal of Molecular Biology 427, 1038–1060 (2015). [PubMed: 25174336]
12. Chan KX, Phua SY, Crisp P, McQuinn R & Pogson BJ Learning the languages of the chloroplast: retrograde signaling and beyond. Annual Review of Plant Biology 67, 25–53 (2016).
13. Quirós PM, Mottis A & Auwerx J Mitonuclear communication in homeostasis and stress. Nature Reviews Molecular Cell Biology 17, 213–226 (2016). [PubMed: 26956194]
14. Wiedemann N & Pfanner N Mitochondrial machineries for protein import and assembly. Annual Review of Biochemistry 86 (2017).
15. Curtis BA et al. Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. Nature 492, 59–65 (2012). [PubMed: 23201678]
16. Nowack EC & Melkonian M Endosymbiotic associations within protists. Philosophical Transactions of the Royal Society of London B: Biological Sciences 365, 699–712 (2010). [PubMed: 20124339]
17. McCutcheon JP & Moran NA Extreme genome reduction in symbiotic bacteria. Nature Reviews Microbiology 10, 13–26, doi:10.1038/nrmicro2670; 10.1038/nrmicro2670 (2012).
18. Adams KL & Palmer JD Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Molecular phylogenetics and evolution 29, 380–395 (2003). [PubMed: 14615181]
19. Timmis JN, Ayliffe MA, Huang CY & Martin W Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. Nature Review Genetics 5, 123–135 (2004).
20. Gray MW Mitochondrial evolution. Cold Spring Harbor Perspectives in Biology 4, a011403, doi: 10.1101/cshperspect.a011403 [doi] (2012). [PubMed: 22952398]
21. Huynen MA, Duarte I & Szklarczyk R Loss, replacement and gain of proteins at the origin of the mitochondria. Biochimica et Biophysica Acta-Bioenergetics 1827, 224–231 (2013).
22. Johnston IG & Williams BP Evolutionary inference across eukaryotes identifies specific pressures favoring mitochondrial gene retention. Cell Systems 2, 101–111 (2016). [PubMed: 27135164] This analysis infers the relative timing of gene losses from mitogenomes in different eukaryotic lineages and indentifies the features that favour retention of specific genes in the mitogenome.
23. Kannan S, Rogozin IB & Koonin EV MitoCOGs: clusters of orthologous genes from mitochondria and implications for the evolution of eukaryotes. BMC Evolutionary Biology 14, 237 (2014). [PubMed: 25421434]
24. Janouškovec J et al. A new lineage of eukaryotes illuminates early mitochondrial genome reduction. Current Biology 27, 3717–3724 (2017). [PubMed: 29174886] This study included a reannotation of mitochondrial gene content across diverse eukaryotic lineages, allowing for reconstruction of heterogeneous rates of gene loss through time and across lineages.
25. Wang Z & Wu M Phyogenomic reconstruction indicates mitochondrial ancestor was an energy parasite. PloS One 9, e110685, doi:10.1371/journal.pone.0110685 [doi] (2014). [PubMed: 25333787]

26. Petersen J et al. Chromera velia, endosymbioses and the rhodoplex hypothesis--plastid evolution in cryptophytes, alveolates, stramenopiles, and haptophytes (CASH lineages). Genome Biology and Evolution 6, 666–684 (2014). [PubMed: 24572015]
27. Flegontov P et al. Divergent mitochondrial respiratory chains in phototrophic relatives of apicomplexan parasites. Molecular Biology and Evolution 32, 1115–1131 (2015). [PubMed: 25660376]
28. Burger G, Gray MW, Forget L & Lang BF Strikingly bacteria-like and gene-rich mitochondrial genomes throughout jakobid protists. Genome Biology and Evolution 5, 418–438, doi: 10.1093/gbe/evt008 [doi] (2013). [PubMed: 23335123]
29. Pett W & Lavrov DV Cytonuclear interactions in the evolution of animal mitochondrial tRNA metabolism. Genome Biology and Evolution 7, 2089–2101 (2015). [PubMed: 26116918]
30. Salinas-Giegé T, Giegé R & Giegé P tRNA biology in mitochondria. International Journal of Molecular Sciences 16, 4518–4559 (2015). [PubMed: 25734984]
31. Adams KL, Qiu YL, Stoutemyer M & Palmer JD Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. Proceedings of the National Academy of Sciences of the United States of America 99, 9905–9912, doi:10.1073/pnas.042694899 [doi]; 042694899 [pii] (2002). [PubMed: 12119382] This classic study provided extensive documentation of the variation in mitochondrial gene content across angiosperms, revealing a rapid and ongoing process of EGT.
32. Barbrook AC, Voolstra CR & Howe CJ The chloroplast genome of a Symbiodinium sp. clade C3 isolate. Protist 165, 1–13 (2014). [PubMed: 24316380]
33. Del Cortona A et al. The plastid genome in Cladophorales green algae is encoded by hairpin chromosomes. Current Biology 27, 3771–3782 (2017). [PubMed: 29199074]
34. Maier U-G et al. Massively convergent evolution for ribosomal protein gene content in plastid and mitochondrial genomes. Genome biology and evolution 5, 2318–2329 (2013). [PubMed: 24259312]
35. Pont-Kingdon G et al. 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. Journal of Molecular Evolution 46, 419–431 (1998). [PubMed: 9541536]
36. Knie N, Polsakiewicz M & Knoop V Horizontal gene transfer of chlamydial-like tRNA genes into early vascular plant mitochondria. Molecular Biology and Evolution 32, 629–634 (2014). [PubMed: 25415968]
37. Milani L, Ghiselli F, Maurizii MG, Nuzhdin SV & Passamonti M Paternally transmitted mitochondria express a new gene of potential viral origin. Genome Biology and Evolution 6, 391–405 (2014). [PubMed: 24500970]
38. Qiu Y, Filipenko SJ, Darracq A & Adams KL Expression of a transferred nuclear gene in a mitochondrial genome. Current Plant Biology 1, 68–72 (2014).
39. Korovesi AG, Ntertilis M & Kouvelis VN Mt-*rps3* is an ancient gene which provides insight into the evolution of fungal mitochondrial genomes. Molecular Phylogenetics and Evolution, doi: 10.1016/j.ympev.2018.04.037 (2018).
40. Ruck EC, Nakov T, Jansen RK, Theriot EC & Alverson AJ Serial gene losses and foreign DNA underlie size and sequence variation in the plastid genomes of diatoms. Genome Biology and Evolution 6, 644–654 (2014). [PubMed: 24567305]
41. Yurchenko T, Ševčíková T, Strnad H, Butenko A & Eliáš M The plastid genome of some eustigmatophyte algae harbours a bacteria-derived six-gene cluster for biosynthesis of a novel secondary metabolite. Open Biology 6, 160249 (2016). [PubMed: 27906133]
42. Amikura R, Kashikawa M, Nakamura A & Kobayashi S Presence of mitochondria-type ribosomes outside mitochondria in germ plasm of *Drosophila* embryos. Proceedings of the National Academy of Sciences 98, 9133–9138 (2001).
43. Villegas J, Araya P, Bustos-Obregon E & Burzio LO Localization of the 16S mitochondrial rRNA in the nucleus of mammalian spermatogenic cells. Molecular Human Reproduction 8, 977–983 (2002). [PubMed: 12397209]

44. Zheng Z, Li H, Zhang Q, Yang L & Qi H Unequal distribution of 16S mtrRNA at the 2-cell stage regulates cell lineage allocations in mouse embryos. Reproduction 151, 351–367 (2016). [PubMed: 26762401]
45. Maniataki E & Mourelatos Z Human mitochondrial tRNAMet is exported to the cytoplasm and associates with the Argonaute 2 protein. RNA 11,849–852 (2005). [PubMed: 15872185]
46. Cognat V et al. The nuclear and organellar tRNA-derived RNA fragment population in *Arabidopsis thaliana* is highly dynamic. Nucleic Acids Research 45, 3460–3472 (2017). [PubMed: 27899576]
47. Landerer E et al. Nuclear localization of the mitochondrial ncRNAs in normal and cancer cells. Cellular Oncology 34, 297–305 (2011).
48. Dietrich A, Wallet C, Iqbal RK, Gualberto JM & Lotfi F Organellar non-coding RNAs: emerging regulation mechanisms. Biochimie 117, 48–62 (2015). [PubMed: 26143009]
49. Pozzi A, Plazzi F, Milani L, Ghiselli F & Passamonti M SmithRNAs: Could Mitochondria “Bend” Nuclear Regulation? Molecular Biology and Evolution 34, 1960–1973 (2017). [PubMed: 28444389]
50. Lee C, Yen K & Cohen P Humanin: a harbinger of mitochondrial-derived peptides? Trends in Endocrinology & Metabolism 24, 222–228 (2013). [PubMed: 23402768]
51. Lee C et al. The mitochondrial-derived peptide MOTS-c promotes metabolic homeostasis and reduces obesity and insulin resistance. Cell Metabolism 21,443–454 (2015). [PubMed: 25738459]
52. Gray MW Mosaic nature of the mitochondrial proteome: Implications for the origin and evolution of mitochondria. Proceedings of the National Academy of Sciences 112, 10133–10138 (2015).This paper provides a broad review of mitochondrial evolution, arguing for diverse phylogenetic sources contributing to the origins of the mitochondrial proteome in eukaryotes.
53. Karlberg O, Canback B, Kurland CG & Andersson SG The dual origin of the yeast mitochondrial proteome. Yeast (Chichester, England) 17, 170–187, doi:2-V (2000).
54. Gabaldón T & Huynen MA From endosymbiont to host-controlled organelle: the hijacking of mitochondrial protein synthesis and metabolism. PLoS Computational Biology 3, e219 (2007). [PubMed: 17983265]
55. Szklarczyk R & Huynen MA Mosaic origin of the mitochondrial proteome. Proteomics 10, 4012–4024 (2010). [PubMed: 21080492]
56. Doolittle WF You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. Trends in Genetics 14, 307–311 (1998). [PubMed: 9724962] This paper lays out the hypothesis for a non-adaptive gene-transfer ratchet as an explanation for frequent replacement of pre-existing host functions by endosymbiotic genes.
57. Thorsness PE & Fox TD Escape of DNA from mitochondria to the nucleus in Saccharomyces cerevisiae. Nature 346, 376–379 (1990). [PubMed: 2165219]
58. Hazkani-Covo E, Zeller RM & Martin W Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLoS Genetics 6, e1000834, doi:10.1371/journal.pgen.1000834 [doi] (2010). [PubMed: 20168995]
59. Dolezal P, Likic V, Tachezy J & Lithgow T Evolution of the molecular machines for protein import into mitochondria. Science 313, 314–318 (2006). [PubMed: 16857931]
60. Shi L-X & Theg SM The chloroplast protein import system: from algae to trees. Biochimica et Biophysica Acta-Molecular Cell Research 1833, 314–331 (2013).
61. Farrelly F & Butow RA Rearranged mitochondrial genes in the yeast nuclear genome. Nature 301, 296–301 (1983). [PubMed: 6296698]
62. Timmis JN & Scott NS Sequence homology between spinach nuclear and chloroplast genomes. Nature 305, 65–67 (1983).
63. Lopez JV, Yuhki N, Masuda R, Modi W & O'Brien SJ Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. Journal of Molecular Evolution 39, 174–190 (1994). [PubMed: 7932781]
64. Huang CY, Ayliffe MA & Timmis JN Direct measurement of the transfer rate of chloroplast DNA into the nucleus. Nature 422, 72–76, doi:10.1038/nature01435 (2003). [PubMed: 12594458]
65. Huang CY, Ayliffe MA & Timmis JN Simple and complex nuclear loci created by newly transferred chloroplast DNA in tobacco. Proceedings of the National Academy of Sciences 101,9710–9715 (2004).

66. Fuentes I, Karcher D & Bock R Experimental reconstruction of the functional transfer of intron-containing plastid genes to the nucleus. Current Biology 22, 763–771 (2012). [PubMed: 22503505]

67. Hazkani-Covo E & Martin WF Quantifying the number of independent organelle DNA insertions in genome evolution and human health. Genome Biology and Evolution 9, 1190–1203 (2017). [PubMed: 28444372]

68. Stupar RM et al. Complex mtDNA constitutes an approximate 620-kb insertion on Arabidopsis thaliana chromosome 2: implication of potential sequencing errors caused by large-unit repeats. Proceedings of the National Academy of Sciences of the United States of America 98, 5099–5103, doi:10.1073/pnas.091110398 (2001). [PubMed: 11309509]

69. Ricchetti M, Fairhead C & Dujon B Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. Nature 402, 96–100 (1999). [PubMed: 10573425]

70. Richly E & Leister D NUPTs in sequenced eukaryotes and their genomic organization in relation to NUMTs. Molecular Biology and Evolution 21, 1972–1980, doi: 10.1093/molbev/msh210 [doi]: (2004). [PubMed: 15254258]

71. Smith DR, Crosby K & Lee RW Correlation between nuclear plastid DNA abundance and plastid number supports the limited transfer window hypothesis. Genome Biology and Evolution 3, 365–371 (2011). [PubMed: 21292629]

72. Barbrook AC, Howe CJ & Purton S Why are plastid genomes retained in non-photosynthetic organisms? Trends in plant science 11, 101–108, doi: 10.1016/j.tplants.2005.12.004 (2006). [PubMed: 16406301]

73. Wu Z et al. Mitochondrial retroprocessing promoted functional transfers of rpl5 to the nucleus in grasses. Molecular Biology and Evolution 34, 2340–2354 (2017). [PubMed: 28541477]

74. Adams KL et al. Intracellular gene transfer in action: dual transcription and multiple silencings of nuclear and mitochondrial *cox2* genes in legumes. Proceedings of the National Academy of Sciences 96, 13863–13868 (1999).

75. Liu SL, Zhuang Y, Zhang P & Adams KL Comparative analysis of structural diversity and sequence evolution in plant mitochondrial genes transferred to the nucleus. Molecular Biology and Evolution 26, 875–891, doi:10.1093/molbev/msp011 (2009). [PubMed: 19168566]

76. Bonen L & Calixte S Comparative analysis of bacterial-origin genes for plant mitochondrial ribosomal proteins. Molecular Biology and Evolution 23, 701–712, doi:10.1093/molbev/msj080 (2006). [PubMed: 16368778]

77. Long M, De Souza SJ, Rosenberg C & Gilbert W Exon shuffling and the origin of the mitochondrial targeting function in plant cytochrome c1 precursor. Proceedings of the National Academy of Sciences 93, 7727–7731 (1996).

78. Smith DR & Keeling PJ Mitochondrial and plastid genome architecture: Reoccurring themes, but significant differences at the extremes. Proceedings of the National Academy of Sciences of the United States of America 112, 10177–10184, doi:10.1073/pnas.1422049112 [doi] (2015). [PubMed: 25814499]

79. Timmis JN Endosymbiotic evolution: RNA intermediates in endosymbiotic gene transfer. Current Biology 22, R296–R298 (2012). [PubMed: 22575464]

80. Sheppard AE et al. Introducing an RNA editing requirement into a plastid-localised transgene reduces but does not eliminate functional gene transfer to the nucleus. Plant Molecular Biology 76, 299–309 (2011). [PubMed: 21404088]

81. Henze K & Martin W How do mitochondrial genes get into the nucleus? Trends in Genetics 17, 383–387 (2001). [PubMed: 11418217]

82. Grewe F, Zhu A & Mower JP Loss of a *trans*-splicing *nad1* intron from Geraniaceae and transfer of the maturase gene *matR* to the nucleus in *Pelargonium*. Genome Biology and Evolution 8, 3193–3201 (2016). [PubMed: 27664178]

83. Allen JF & Raven JA Free-radical-induced mutation vs redox regulation: costs and benefits of genes in organelles. Journal of Molecular Evolution 42, 482–492 (1996). [PubMed: 8662000]

84. Blanchard JL & Lynch M Organellar genes - why do they end up in the nucleus? Trends in Genetics 16, 315–320 (2000). [PubMed: 10858662]

85. Brandvain Y & Wade MJ The functional transfer of genes from the mitochondria to the nucleus: the effects of selection, mutation, population size and rate of self-fertilization. Genetics 182, 1129–1139 (2009). [PubMed: 19448273]

86. Cooper BS, Burrus CR, Ji C, Hahn MW & Montooth KL Similar efficacies of selection shape mitochondrial and nuclear genes in both Drosophila melanogaster and Homo sapiens. G3: Genes| Genomes| Genetics 5, 2165–2176 (2015). [PubMed: 26297726]

87. Christie JR & Beekman M Uniparental inheritance promotes adaptive evolution in cytoplasmic genomes. Molecular Biology and Evolution 34, 677–691 (2016).

88. Osteryoung KW & Nunnari J The division of endosymbiotic organelles. Science 302, 1698–1704 (2003). [PubMed: 14657485]

89. Adams KL, Daley DO, Whelan J & Palmer JD Genes for two mitochondrial ribosomal proteins in flowering plants are derived from their chloroplast or cytosolic counterparts. The Plant Cell 14, 931–943 (2002). [PubMed: 11971146]

90. Duchêne A-M, Pujol C & Maréchal-Drouard L Import of tRNAs and aminoacyl-tRNA synthetases into mitochondria. Current Genetics 55, 1–18 (2009). [PubMed: 19083240]

91. Carrie C & Small I A reevaluation of dual-targeting of proteins to mitochondria and chloroplasts. Biochimica et Biophysica Acta-Molecular Cell Research 1833, 253–259 (2013).

92. Woese CR, Olsen GJ, Ibba M & Söll D Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. Microbiology and Molecular Biology Reviews 64, 202–236 (2000). [PubMed: 10704480]

93. Meiklejohn CD et al. An Incompatibility between a mitochondrial tRNA and its nuclear-encoded tRNA synthetase compromises development and fitness in Drosophila. PLoS Genetics 9, e1003238, doi:10.1371/journal.pgen.1003238; 10.1371/journal.pgen.1003238 (2013). [PubMed: 23382693] This study provides a detailed characterization of incompatibilities arising from individual subsiuttions in a nuclear-encoded aaRS and a mitochondrial-encoded tRNA.

94. Harrison JS & Burton RS Tracing hybrid incompatibilities to single amino acid substitutions. Molecular Biology and Evolution 23, 559–564, doi:10.1093/molbev/msj058 (2006). [PubMed: 16280539]

95. Osada N & Akashi H Mitochondrial-nuclear interactions and accelerated compensatory evolution: evidence from the primate cytochrome C oxidase complex. Molecular Biology and Evolution 29, 337 (2012). [PubMed: 21890478] This study analyzes both the structural positions and relative timing of amino acid substitutions in a mitonuclear enzyme complex to infer the coevolutionary effects of mutation pressure in the mitochondrial genome.

96. van der Sluis EO et al. Parallel structural evolution of mitochondrial ribosomes and OXPHOS complexes. Genome Biology and Evolution 7, 1235–1251 (2015). [PubMed: 25861818] This study models structural interactions involving supernumerary subunits in mitnuclear enzyme complexes and advances the hypothesis that these subunits were recruited to compensate for destabilizing changes in mitochondrial-encoded subunits.

97. Barreto FS & Burton RS Evidence for compensatory evolution of ribosomal proteins in response to rapid divergence of mitochondrial rRNA. Molecular Biology and Evolution 30, 310–314, doi: 10.1093/molbev/mss228; 10.1093/molbev/mss228 (2013). [PubMed: 22993236]

98. Willett CS & Burton RS Evolution of interacting proteins in the mitochondrial electron transport system in a marine copepod. Molecular Biology and Evolution 21,443–453, doi: 10.1093/molbev/msh031 [doi] (2004). [PubMed: 14660687]

99. Grossman LI, Wildman DE, Schmidt TR & Goodman M Accelerated evolution of the electron transport chain in anthropoid primates. Trends in genetics : TIG 20, 578–585, doi: 10.1016/j.tig.2004.09.002 (2004). [PubMed: 15475118]

100. Zhang F & Broughton RE Mitochondrial-nuclear interactions: compensatory evolution or variable functional constraint among vertebrate oxidative phosphorylation genes? Genome Biology and Evolution 5, 1781–1791 (2013). [PubMed: 23995460]

101. Sloan DB, Triant DA, Wu M & Taylor DR Cytonuclear interactions and relaxed selection accelerate sequence evolution in organelle ribosomes. Molecular Biology and Evolution 31,673–682, doi:10.1093/molbev/mst259; 10.1093/molbev/mst259 (2014). [PubMed: 24336923] This paper was one of many recent studies to compare closely related species with dramatically

different rates of cytoplasmic genome evolution in order to infer coevolutionary consequences in the nucleus.

102. Adrion JR, White PS & Montooth KL The roles of compensatory evolution and constraint in aminoacyl tRNA synthetase evolution. Molecular Biology and Evolution 33, 152 (2016). [PubMed: 26416980]
103. Rockenbach KD et al. Positive selection in rapidly evolving plastid-nuclear enzyme complexes. Genetics 204, 1507–1522 (2016). [PubMed: 27707788]
104. Havird JC, Whitehill NS, Snow CD & Sloan DB Conservative and compensatory evolution in oxidative phosphorylation complexes of angiosperms with highly divergent rates of mitochondrial genome evolution. Evolution 69, 3069–3081 (2015). [PubMed: 26514987]
105. Weng ML, Ruhlman TA & Jansen RK Plastid-nuclear interaction and accelerated coevolution in plastid ribosomal genes in Geraniaceae. Genome Biology and Evolution 8, 1824–1838 (2016). [PubMed: 27190001]
106. Zhang J, Ruhlman TA, Sabir J, Blazier JC & Jansen RK Coordinated rates of evolution between interacting plastid and nuclear genes in Geraniaceae. The Plant cell 27, 563–573 (2015). [PubMed: 25724640]
107. Li Y et al. The molecular evolutionary dynamics of oxidative phosphorylation (OXPHOS) genes in Hymenoptera. BMC Evolutionary Biology 17, 269 (2017). [PubMed: 29281964]
108. Havird JC, Trapp P, Miller C, Bazos I & Sloan DB Causes and consequences of rapidly evolving mtDNA in a plant lineage. Genome Biology and Evolution 9, 323–336 (2017).
109. Yan Z, Ye G.-y. & Werren J Evolutionary rate coevolution between mitochondria and mitochondria-associated nuclear-encoded proteins in insects. bioRxiv, 288456 (2018).
110. Ellison CK & Burton RS Disruption of mitochondrial function in interpopulation hybrids of Tigriopus californicus. Evolution 60, 1382–1391 (2006). [PubMed: 16929655]
111. Ehrlich PR & Raven PH Butterflies and plants: a study in coevolution. Evolution 18, 586–608 (1964).
112. Janzen DH When is it coevolution? Evolution 34, 611–612 (1980). [PubMed: 28568694]
113. Desmond E, Brochier-Armanet C, Forterre P & Gribaldo S On the last common ancestor and early evolution of eukaryotes: reconstructing the history of mitochondrial ribosomes. Research in Microbiology 162, 53–70 (2011). [PubMed: 21034815]
114. Smits P, Smeitink JA, van den Heuvel LP, Huynen MA & Ettema TJ Reconstructing the evolution of the mitochondrial ribosomal proteome. Nucleic Acids Research 35, 4686–4703 (2007). [PubMed: 17604309]
115. Berry S Endosymbiosis and the design of eukaryotic electron transport. Biochimica et Biophysica Acta-Bioenergetics 1606, 57–72 (2003).
116. Sharma MR et al. Cryo-EM study of the spinach chloroplast ribosome reveals the structural and functional roles of plastid-specific ribosomal proteins. Proceedings of the National Academy of Sciences 104, 19315–19320 (2007).
117. Senkler J, Senkler M & Braun HP Structure and function of complex I in animals and plants-a comparative view. Physiologia Plantarum 161, 6–15 (2017). [PubMed: 28261805]
118. Elurbe DM & Huynen MA The origin of the supernumerary subunits and assembly factors of complex I: a treasure trove of pathway evolution. Biochimica et Biophysica Acta- Bioenergetics 1857, 971–979 (2016).This review provides a detailed documentation of the history of supernumerary subunit acquistion in mitochondrial OXPHOS Complex I.
119. O'Brien TW Evolution of a protein-rich mitochondrial ribosome: implications for human genetic disease. Gene 286, 73–79 (2002). [PubMed: 11943462]
120. Sharma MR et al. Structure of the mammalian mitochondrial ribosome reveals an expanded functional role for its component proteins. Cell 115, 97–108 (2003). [PubMed: 14532006]
121. Stroud DA et al. Accessory subunits are integral for assembly and function of human mitochondrial complex I. Nature 538, 123 (2016). [PubMed: 27626371] This study included a systematic analysis of the effects of knocking out individual supernumerary subunits in mitochondrial Complex I, revealing common roles in complex assembly and stability.
122. Fiedorczuk K et al. Atomic structure of the entire mammalian mitochondrial complex I. Nature 538, 406–410 (2016). [PubMed: 27595392]

123. Zhu J, Vinothkumar KR & Hirst J Structure of mammalian respiratory complex I. Nature 536, 354 (2016). [PubMed: 27509854]

124. Wu M, Gu J, Guo R, Huang Y & Yang M Structure of mammalian respiratory supercomplex I1III2IV1. Cell 167, 1598–1609. e1510 (2016). [PubMed: 27912063]

125. Guo R, Zong S, Wu M, Gu J & Yang M Architecture of human mitochondrial respiratory megacomplex I2III2IV2. Cell 170, 1247–1257. e1212 (2017). [PubMed: 28844695]

126. Gu J et al. The architecture of the mammalian respirasome. Nature 537, 639 (2016). [PubMed: 27654917]

127. Davies KM, Blum TB & Kühlbrandt W Conserved in situ arrangement of complex I and III2 in mitochondrial respiratory chain supercomplexes of mammals, yeast, and plants. Proceedings of the National Academy of Sciences 115, 3024–3029 (2018).This structural study compared mitochondrial OXPHOS supercomplex assemblies across diverse eukaryotes, revealing widely conserved features of these higher-order structures.

128. Genova ML & Lenaz G Functional role of mitochondrial respiratory supercomplexes. Biochimica et Biophysica Acta-Bioenergetics 1837, 427–443 (2014).

129. Angerer H et al. A scaffold of accessory subunits links the peripheral arm and the distal proton-pumping module of mitochondrial complex I. Biochemical Journal 437, 279–288 (2011). [PubMed: 21545356]

130. Hirst J Why does mitochondrial complex I have so many subunits? Biochemical Journal 437, e1–e3 (2011). [PubMed: 21711245]

131. Melber A & Winge DR Inner secrets of the respirasome. Cell 167, 1450–1452 (2016). [PubMed: 27912054]

132. Cogliati S et al. Mechanism of super-assembly of respiratory complexes III and IV. Nature 539, 579–582 (2016). [PubMed: 27775717]

133. Hayashi T et al. Higdla is a positive regulator of cytochrome c oxidase. Proceedings of the National Academy of Sciences 112, 1553–1558 (2015).

134. Chaban Y, Boekema EJ & Dudkina NV Structures of mitochondrial oxidative phosphorylation supercomplexes and mechanisms for their stabilisation. Biochimica et Biophysica Acta-Bioenergetics 1837, 418–426 (2014).

135. Lapuente-Brun E et al. Supercomplex assembly determines electron flux in the mitochondrial electron transport chain. Science 340, 1567–1570 (2013). [PubMed: 23812712]

136. Lukeš J, Archibald JM, Keeling PJ, Doolittle WF & Gray MW How a neutral evolutionary ratchet can build cellular complexity. IUBMB life 63, 528–537, doi:10.1002/iub.489 [doi] (2011). [PubMed: 21698757]

137. Nishimura K & van Wijk KJ Organization, function and substrates of the essential Clp protease system in plastids. Biochimica et Biophysica Acta-Bioenergetics 1847, 915–930 (2015).

138. Sinkler CA et al. Tissue-and condition-specific isoforms of mammalian cytochrome c oxidase subunits: from function to human disease. Oxidative Medicine and Cellular Longevity 2017, 1534056 (2017). [PubMed: 28593021]

139. Frank SA & Hurst LD Mitochondria and male disease. Nature 383, 224 (1996). [PubMed: 8805695]

140. Gemmell NJ, Metcalf VJ & Allendorf FW Mother's curse: the effect of mtDNA on individual fitness and population viability. Trends in Ecology & Evolution 19, 238–244 (2004). [PubMed: 16701262]

141. Gallach M, Chandrasekaran C & Betran E Analyses of nuclearly encoded mitochondrial genes suggest gene duplication as a mechanism for resolving intralocus sexually antagonistic conflict in Drosophila. Genome Biology and Evolution 2, 835–850 (2010). [PubMed: 21037198]

142. Patel MR et al. A mitochondrial DNA hypomorph of cytochrome oxidase specifically impairs male fertility in Drosophila melanogaster. eLife 5, e16923 (2016). [PubMed: 27481326]

143. Small I et al. The strange evolutionary history of plant mitochondrial tRNAs and their aminoacyl-tRNA synthetases. Journal of Heredity 90, 333–337 (1999).

144. Roger AJ, Muñoz-Gómez SA & Kamikawa R The origin and diversification of mitochondria. Current Biology 27, R1177–R1192 (2017). [PubMed: 29112874]

145. Daley DO & Whelan J Why genes persist in organelle genomes. Genome Biology 6, 110 (2005). [PubMed: 15892877]

146. Björkholm P, Harish A, Hagström E, Ernst AM & Andersson SG Mitochondrial genomes are retained by selective constraints on protein targeting. Proceedings of the National Academy of Sciences 112, 10154–10161 (2015).

147. Martin W & Schnarrenberger C The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis. Current Genetics 32, 1–18 (1997). [PubMed: 9309164]

148. Oca-Cossio J, Kenyon L, Hao H & Moraes CT Limitations of allotopic expression of mitochondrial genes in mammalian cells. Genetics 165, 707–720 (2003). [PubMed: 14573482]

149. Allen JF The function of genomes in bioenergetic organelles. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 358, 19–37 (2003).

150. Allen JF Why chloroplasts and mitochondria retain their own genomes and genetic systems: colocation for redox regulation of gene expression. Proceedings of the National Academy of Sciences 112, 10231–10238 (2015).This review provides a recent summary of the classic CoRR hypothesis for why mitochondria and plastids still retain their own genomes after billions of years of evolution.

151. Kanevski I & Maliga P Relocation of the plastid rbcL gene to the nucleus yields functional ribulose-1,5-bisphosphate carboxylase in tobacco chloroplasts. Proceedings of the National Academy of Sciences 91, 1969–1973 (1994).

152. Tovar J et al. Mitochondrial remnant organelles of Giardia function in iron-sulphur protein maturation. Nature 426, 172–176 (2003). [PubMed: 14614504]

153. Freibert S-A et al. Evolutionary conservation and in vitro reconstitution of microsporidian iron-sulfur cluster biosynthesis. Nature Communications 8, 13932 (2017).

154. Karnkowska A et al. A eukaryote without a mitochondrial organelle. Current Biology 26, 1274–1284 (2016). [PubMed: 27185558]

155. Smith DR & Lee RW A plastid without a genome: evidence from the nonphotosynthetic green algal genus Polytomella. Plant Physiology 164, 1812–1819 (2014). [PubMed: 24563281]

156. Molina J et al. Possible loss of the chloroplast genome in the parasitic flowering plant Rafflesia lagascae (Rafflesiaceae). Molecular Biology and Evolution 31,793–803 (2014). [PubMed: 24458431]

157. Bennett GM & Moran NA Small, smaller, smallest: the origins and evolution of ancient dual-obligate symbioses in a phloem-feeding insect. Genome Biology and Evolution 5, 1675–1688 (2013). [PubMed: 23918810]

158. Sloan DB et al. Parallel histories of horizontal gene transfer facilitated extreme reduction of endosymbiont genomes in sap-feeding insects. Molecular Biology and Evolution 31, 857–871 (2014). [PubMed: 24398322]

159. Nowack EC et al. Endosymbiotic gene transfer and transcriptional regulation of transferred genes in Paulinella chromatophora. Molecular biology and evolution 28, 407–422 (2011). [PubMed: 20702568]

160. Singer A et al. Massive protein import into the early-evolutionary-stage photosynthetic organelle of the amoeba Paulinella chromatophora. Current Biology 27, 2763–2773 (2017). [PubMed: 28889978]

161. Nakabachi A, Ishida K, Hongoh Y, Ohkuma M & Miyagishima SY Aphid gene of bacterial origin encodes a protein transported to an obligate endosymbiont. Current Biology 24, R640–R641 (2014). [PubMed: 25050957]

162. Nowack ECM et al. Gene transfers from diverse bacteria compensate for reductive genome evolution in the chromatophore of Paulinella chromatophora. Proceedings of the National Academy of Sciences In Press (2016).

163. Husnik F et al. Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. Cell 153, 1567–1578 (2013). [PubMed: 23791183] This comparative genomic analysis reveals the role that HGT from multiple sources in the more recent establishment of obligate bacterial endosymbionts in sap-feeding insects.

164. Bennett GM & Moran NA Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole. Proceedings of the National Academy of Sciences 112, 10169–10176 (2015).

165. Husnik F & McCutcheon JP Repeated replacement of an intrabacterial symbiont in the tripartite nested mealybug symbiosis. Proceedings of the National Academy of Sciences 113, E5416–5424 (2016).

166. Ball SG et al. Toward an understanding of the function of Chlamydiales in plastid endosymbiosis. Biochimica et Biophysica Acta-Bioenergetics 1847, 495–504 (2015).

167. Domman D, Horn M, Embley TM & Williams TA Plastid establishment did not require a chlamydial partner. Nature Communications 6, 6421 (2015).

168. Larkum AW, Lockhart PJ & Howe CJ Shopping for plastids. Trends in Plant Science 12, 189–195 (2007). [PubMed: 17416546]

169. Ku C et al. Endosymbiotic gene transfer from prokaryotic pangenomes: Inherited chimerism in eukaryotes. Proceedings of the National Academy of Sciences 112, 10139–10146 (2015).The authors of this study argue that inferences of diverse phylogenetic donors in establishment of endosymbtionts and organelles are often overestimated because they fail to account for errors in phylogenetic reconstruction and for the history of HGT in bacteria that preceded endosymbiosis.

170. Kapust N et al. Failure to recover major events of gene flux in real biological data due to method misapplication. Genome Biology and Evolution 10, 1198–1209 (2018). [PubMed: 29718211]

Box 1: The stubborn retention of cytoplasmic genomes: constraint or adaptation?

Given the prolific rates of cytoplasmic gene loss, replacement, and transfer during eukaryotic evolution, why have cytoplasmic genomes been retained at all? Proposed answers to this classic question fall into two broad categories. The first focuses on constraints, including the genetic barriers that impede functional expression of cytoplasmic DNA in the nucleus (see “Mechanisms and causes of gene transfer to the nucleus” in the main text). Other constraint-based hypotheses invoke gene-specific effects of biochemical and targeting mechanisms. The ‘hydrophobicity hypothesis’ is based on the observation that retained cytoplasmic genes preferentially encode highly hydrophobic, membrane-bound subunits within oxidative phosphorylation (OXPHOS) and photosynthetic enzymes, which may be prohibitive to transport across organelle membranes[18,145]. Related arguments contend that hydrophobic proteins are likely to be mistargeted to the endoplasmic reticulum[146] or have toxic effects in the cytosol[147,148]. The second broad group of hypotheses to explain cytoplasmic genome retention focuses on beneficial consequences rather than constraints. The most prominent of these is the ‘co-location for redox regulation’ (CoRR) hypothesis[149,150], which argues that locally expressing the core components of OXPHOS and photosynthetic enzyme complexes allows for fine-tuned response to redox state in individual organelles — a necessity when there are numerous organelles per cell. As such, the CoRR hypothesis provides an alternative explanation for the preferential retention of ‘core’ subunits.

An extensive comparative analysis of mitochondrial gene loss versus retention recently arrived at the pluralistic conclusion that GC nucleotide content, protein hydrophobicity, and ‘energetic centrality’ all have explanatory power when predicting gene retention[22]. Experimental strategies to relocate cytoplasmic genes to the nucleus[146,148,151] have also been valuable in testing predictions associated with the above hypotheses. To date, however, such studies have been limited to only one or a few genes at a time. Given improvements in genetic-engineering throughput, there are tantalizing prospects to comprehensively catalogue the effects of relocating each individual mitochondrial and plastid gene on targeting fidelity, complex assembly, organelle function, and organismal phenotypes.

Of course, there are always exceptions to the rule, and some cytoplasmic genomes have been lost entirely, such as those in mitochondrial-derived organelles that have lost the capacity for cellular respiration (e.g., hydrogenosomes and mitosomes)[144]. In many mitogenomes, all remaining protein-coding genes are for OXPHOS subunits, so it is unsurprising that loss of cellular respiration eliminates any selection to retain a mitogenome. However, mitochondrial-like organelles are generally still retained in these cases, which may be largely explained by their essential role in iron-sulfur cluster assembly[144,152,153]. This interpretation is supported by the recent report of a eukaryote (the *oxymonad Monocercomonoides* sp. PA 203) that appears to lack any mitochondrial-like organelle and is likely to have replaced the mitochondrial-based iron-sulfur cluster system with a cytosolic sulfur mobilization system acquired via horizontal gene transfer (HGT) from bacteria[154]. In plastids, loss of photosynthesis does not appear to be

sufficient for complete plastome loss. Most parasitic plants retain a plastome, probably because of the roles of just a few genes in essential biosynthetic pathways[72]. However, none of these key non-photosynthetic genes are universally retained within plastomes, and the first likely examples of complete plastome loss have been identified in a parasitic alga[155] and an angiosperm[156].

Box 2: What can be learned from younger endosymbiotic events?

Genomic and functional analyses of obligate bacterial endosymbionts found in eukaryotic hosts[16,17] have blurred, if not entirely obliterated, the boundary between ‘organelles’ and ‘endosymbionts’. Endosymbionts have been identified that are smaller in genome size and/or gene content than some mitochondria and plastids[157]. Direct gene transfer has occurred from endosymbionts to their hosts[158,159], and endosymbionts can import host-encoded proteins[160,161]. Many intimate endosymbiotic relationships have evolved over more recent timescales (tens to hundreds of millions of years) compared to mitochondria and plastids (billions of years), allowing for more accurate reconstruction of evolutionary processes that likely apply more broadly to cytonuclear integration.

One emerging theme is the role of horizontal gene transfer (HGT) to the host nucleus from multiple bacterial sources during endosymbiont establishment. In organisms such as sap-feeding insects and the photosynthetic protist *Paulinella chromatophora*, many of the host nuclear genes that play a role in maintaining relationships with obligate bacterial endosymbionts are of bacterial origin themselves, but they derive from a heterogeneous pool of donors[158,162,163]. Additionally, many lineages have experienced a history of serial endosymbiotic replacement, in which a long-term bacterial endosymbiont is lost and functionally replaced by an independently derived partner[164,165]. These observations of endosymbiont evolution have fuelled a debate about the origins of mitochondrial proteomes. Phylogenetic analyses have long struggled to trace the majority of mitochondrial proteins back to their presumed origins within the alphaproteobacteria. Such studies have often pointed to diverse clades of bacteria outside the alphaproteobacteria. Observations from younger endosymbionts, in which phylogenetic analyses are more statistically robust, may be seen as support for arguments that the establishment of mitochondria involved multiple genetic contributors and perhaps multiple endosymbiotic partners[52]. There have been similar (albeit controversial) arguments that the establishment of plastids involved an additional symbiotic partner from within the Chlamydiales[166,167]. The so-called ‘shopping bag’ model[168] posits that host cells were able to acquire many of the key genetic pieces that eventually led to the stable, long-term incorporation of an organelle through a series of endosymbionts (loosely analogous to purchasing items from a series of stores that all wind up in the same shopping bag). This model offers an explanation for the heterogeneous phylogenetic signals within organellar proteomes and for how eukaryotes may have overcome major early barriers to cytonuclear integration. However, these interpretations have been vehemently disputed by other researchers[169,170], who argue that the phylogenetic patterns can be explained by: first, the statistical artefacts associated with reconstructing trees for individual genes across such deep timescales, and second, a rampant history of HGT among diverse lineages of bacteria that may have affected the mitochondrial ancestor prior to eukaryogenesis. This ongoing debate highlights one of the central uncertainties about the earliest stages of cytonuclear integration.

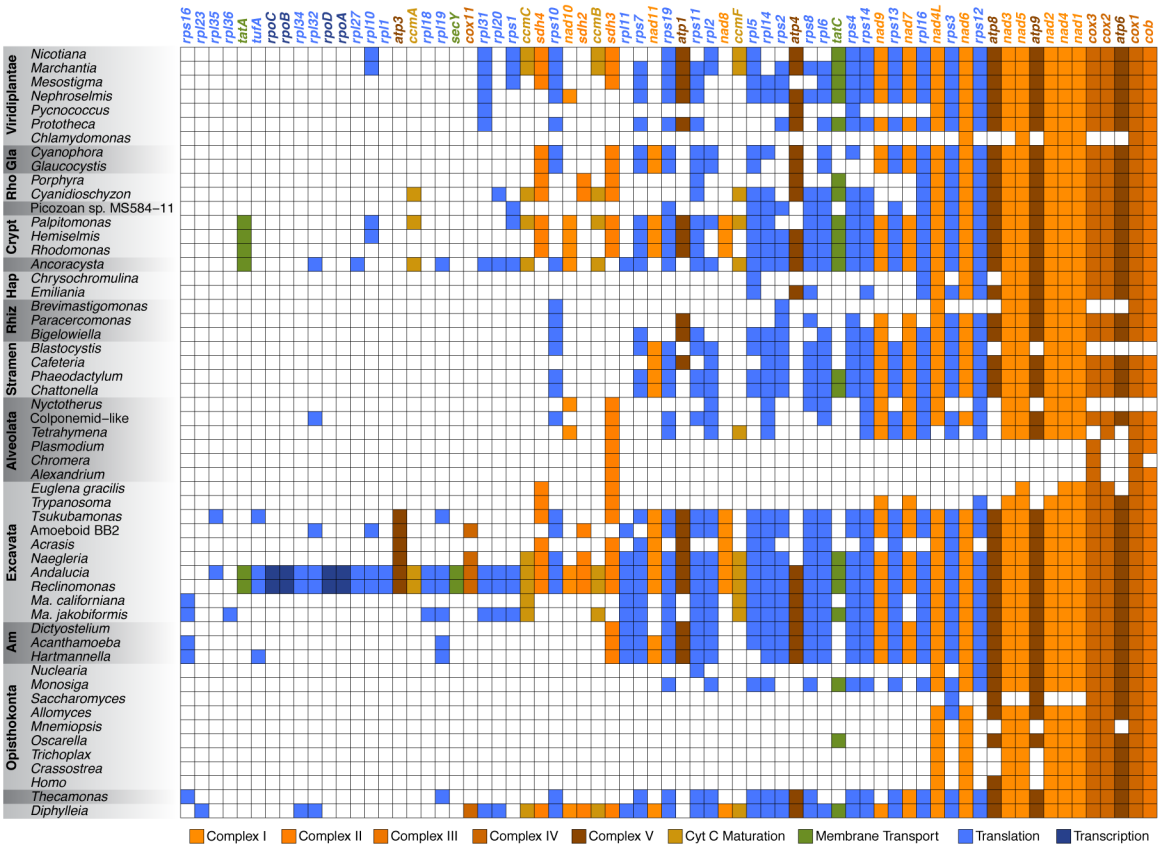

Figure 1.

Variation in mitochondrial gene content across eukaryotic lineages. Filled squares indicate the presence of the gene in the mitogenome of the corresponding species. Gene ordering reflects analysis of gene retention rates by Johnston and Williams[22] (with the exception of the four left-most genes, which were not included in that analysis). Genes are colour-coded by functional category. Gene content is based on summary by Roger *et al.*[144], incorporating additional sampling and reannotations by Janouškovec et al.[24]. In all known eukaryotes, the large complement of proteincoding genes ancestrally found in the mitochondrial progenitor has been reduced to some subset of these 69 genes, reflecting massive gene loss early in the evolution of eukaryotes. The extent of the reduction in gene content varies considerably across eukaryotic lineages, but genes encoding certain key subunits of OXPHOS complexes are almost always retained in the mitogenome. Taxonomic abbreviations are as follows. Am, Amoebozoa; Crypt, Cryptophyta; Gla, Glaucophyta; Hap, Haptophyta; *Ma, Malawimonas;* Rhiz, Rhizaria; Rho, Rhodophyta; Stramen, Stramenopiles. OXPHOS, oxidative phosphorylation.

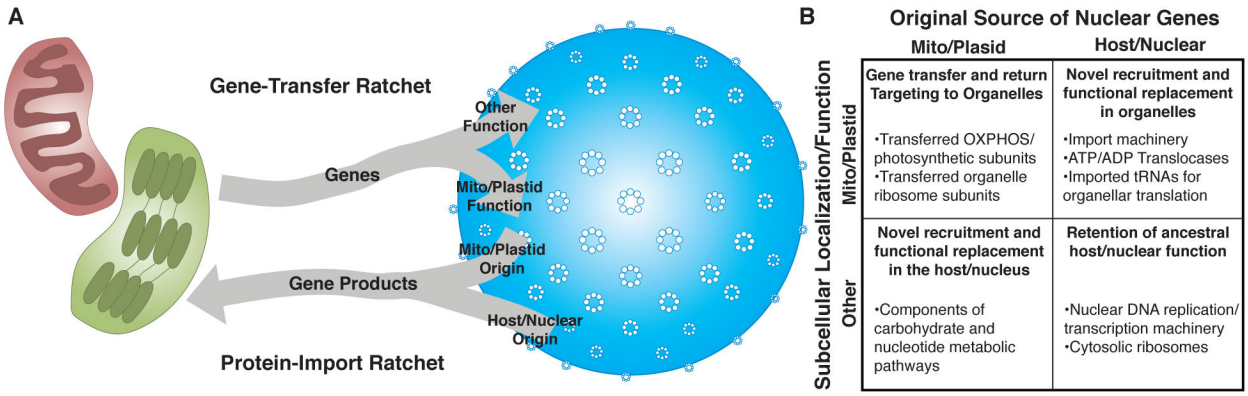

Figure 2.

Opposite trends in the cytonuclear movement of genes versus gene products. **a** | Schematic summary of the asymmetrical and opposite movement of genes versus gene products between cytoplasmic organelles and the nucleus during the history of cytonuclear integration. As described in the main text, non-adaptive ratchet-like processes related to functional gene-transfer and protein-import may contribute to these two asymmetries. **b** | Categorization of functional gene classes in the nucleus, highlighting the fact that not all cytoplasmically derived genes are targeted back to the mitochondria and plastids, and that not all nuclear genes with mitochondrial or plastid function originated from cytoplasmic genomes. Representative examples are provided for each category. OXPHOS, oxidative phosphorylation.

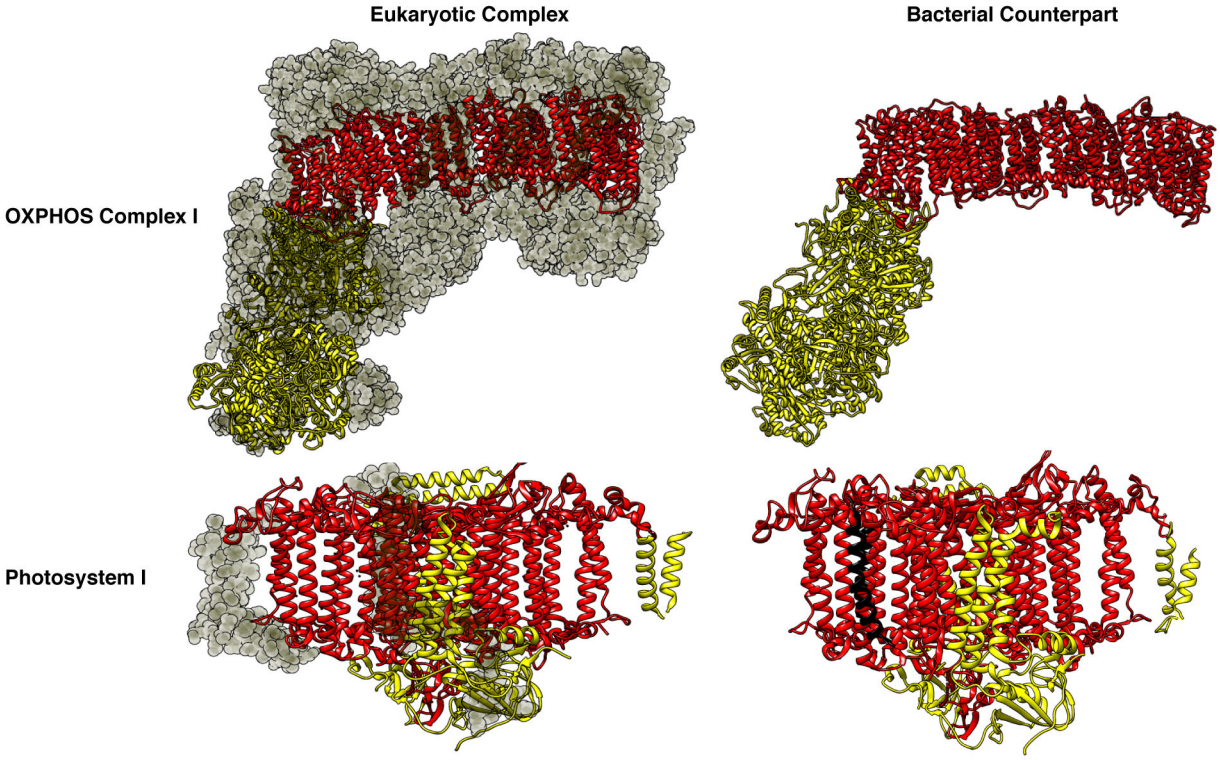

Figure 3.

Acquisition of supernumerary subunits in chimeric cytonuclear enzyme complexes. Subunits drawn with grey spheres are eukaryotic-specific (i.e., nuclear-encoded supernumerary subunits). Structures drawn with ribbons are 'core' subunits that were ancestrally present in bacteria. Red subunits correspond to genes that are still retained in cytoplasmic genomes, whereas yellow subunits correspond to genes that have been transferred to the nucleus. The same coloration is applied to homologous subunits in bacteria for the sake of comparison, even though there is no division into genomic compartments in bacteria. Subunits drawn with black ribbons in the cyanobacterial structure (PsaM and PsaX) have been lost from the eukaryotic structure. Complexes from both mitochondria and plastids were selected to illustrate the much more extensive recruitment of supernumerary subunits in mitochondria. Structures are based on the following Protein Data Bank (PDB) accessions. OXPHOS Complex I: *Ovis aries* (5LNK) and *Thermus thermophilus* (4HEA); Photosystem 1: *Pisum sativum* (4XK8) and *Synechococcus elongatus* (1JB0).

Table 1.

Contrasting features that differentiate nuclear and cytoplasmic genomes

	Nuclear	Mitochondrial or plastid
Genome copy number	Typically two copies (diploidy)	Highly multicopy
Genome size	Typically >10 Mb and often >1 Gb	Typically <500 kb
Genome structure	Linear chromosomes	Highly variable (linear, circular, branched, singlechromosome versus multichromosomal, etc.)
Inheritance	Typically biparental	Often uniparental (typically maternal)
Sex/recombination	Typically sexual	Often effectively asexual
Independent DNA replication machineries	Host (archaeal) origins	Bacterial (endosymbiotic) origins and functional replacement with phage-derived machinery
Independent transcription machineries	Host (archaeal) origins	Bacterial (endosymbiotic) origins and functional replacement with phage-derived machinery
Independent translation machineries	Host (archaeal) origins	Mostly bacterial (endosymbiotic) origins with occasional functional replacement and supernumerary subunits
Mutation rates	Highly variable in both nuclear and cytoplasmic genomes. Cytoplasmic mutation rates can be much higher than in the nucleus (e.g., in bilaterian animals) or much lower than in the nucleus (e.g., in land plants).	