Review

Dinoflagellate chloroplasts as a model for extreme genome reduction and fragmentation in organelles – the COCOA principle for gene retention

Christopher J. Howe [a,b]

Adrian C. Barbrook [a]

[a]*Department of Biochemistry, University of Cambridge, Downing Site, Tennis Court Road, Cambridge, CB2 1QW, UK*

[b]*Stellenbosch Institute for Advanced Study, (STIAS), Wallenberg Research Centre at Stellenbosch University, Stellenbosch 7600, South Africa*

ch26@cam.ac.uk, acb18@cam.ac.uk

Abstract

The genomes of peridinin-containing dinoflagellate chloroplasts have a very unusual organisation. These genomes are highly fragmented and greatly reduced, with most of the usual complement of chloroplast genes relocated to the nucleus. Dinoflagellate chloroplasts highlight evolutionary changes that are found to varying extents in a number of other organelle genomes. These include the chloroplast genome of the green alga *Boodlea* and other Cladophorales, and the mitochondrial genomes of blood-sucking and chewing lice, the parasitic plant *Rhopalocnemis phalloides*, the red alga *Rhodosorus marinus* and other members of the Stylonematophyceae, diplonemid flagellates, and some Cnidaria. Consideration of the coding content of the remnant chloroplast genomes indicates that organelles may preferentially retain genes for proteins important in initiating assembly of complexes, and the same is largely true for mitochondria. We propose a new principle, of CO-location for COntrol of Assembly (COCOA), indicating the importance of retaining these genes in the organelle. This adds to, but does not invalidate, the existing hypotheses of the multisubunit completion principle, CO-location for Redox Regulation (CORR) and Control by Epistasy of Synthesis (CES).

Keywords

Minicircle; organelle genome; photosystem; electron transfer; endosymbiont gene transfer; endosymbiosis

Introduction

Dinoflagellates are a widespread group of aquatic protists of great importance in their own right and as symbiotic partners of a wide range of organisms including corals. About 50% of dinoflagellate taxa are photosynthetic (Gómez, 2012). Most of the photosynthetic species contain a chloroplast with chlorophylls *a* and *c*, and various carotenoids, most notably peridinin (Schnepf and Elbrächter, 1999). The latter has a light harvesting function (Hofmann et al., 1996). The chloroplast is believed to be a secondary one, acquired from a red alga, and present ancestrally in

the dinoflagellates. In some lineages, the peridinin chloroplast has been replaced, for example by one characterised by the carotenoid fucoxanthin (from a haptophyte lineage) or chlorophyll *b* (from a green algal lineage), reviewed by Dorrell and Howe (2015). Dinoflagellates have many unusual features, such as the large-scale replacement of conventional nuclear histones, primarily with a protein apparently of viral origin (Gornik et al., 2012). We focus here on the chloroplast genome of the peridinin-containing dinoflagellates. This is another unusual feature of dinoflagellates, as it is highly fragmented and reduced (reviewed by Barbrook et al., 2019). We compare organelle genome organisation in other groups where extreme fragmentation, and in some cases reduction, have also been identified, and argue that dinoflagellates provide a model for what has happened. We focus on instances where individual subgenomic fragments carry only a few protein-coding genes, typically no more than three. Except where specific comparisons are helpful, we will not consider examples where genomes have fragmented into a small number of subgenomic molecules carrying many genes, as in the green alga *Koshicola spirodelophila* (whose chloroplast genome is in three large chromosomes that all carry many genes (Watanabe et al., 2016)), or in mitochondria of plants, algae or many other organisms (e.g. Fan and Lee, 2002; Palmer and Shields, 1984). We exclude dinoflagellate mitochondria, whose genome organization remains unclear (Nash et al., 2007, 2008; Slamovits et al., 2007; Shoguchi et al., 2015). We also exclude the trypanosomatids, whose mitochondrial genomes comprise maxicircles and minicircles, with the latter encoding guide RNAs for extreme editing of transcripts from the former (reviewed by Callejas-Hernandez et al., 2021). We present a new model for why certain genes are retained in organelles, COCOA (CO-location for Control of Assembly), which integrates important existing models.

The dinoflagellate chloroplast genome

Zhang et al. (1999) were the first to show that in the peridinin dinoflagellate *Heterocapsa triquetra* chloroplast genes reside on small, plasmid-like, topologically circular molecules of 2.0 – 3.0 kbp, each carrying a single gene. They reported such molecules carrying genes for the photosystem I components PsaA and PsaB, the photosystem II components PsbA, PsbB and PsbC, the PetB component of the

cytochrome b_6f complex that links the two photosystems in the photosynthetic electron transfer chain, the AtpA component of the chloroplast ATP synthase, and 23S and 16S rRNA. They also showed a similar arrangement might occur in some other dinoflagellates. In 2000, Barbrook and Howe reported that this was the case for the dinoflagellate *Amphidinium carterae* (referred to at the time as *A. operculatum*) and added *atpB* and *petD* to the list of genes identified (Barbrook and Howe, 2000). Since then, a range of other peridinin dinoflagellates have been shown to have a similar chloroplast gene organisation, including *Adenoides eludens*, *Ceratium horridum* (now *Tripos horridus*), *Pyrocystis lunula*, *Protoceratium reticulatum, Symbiodinium sp.*, and *Fugacium kawagutii* (Barbrook et al., 2019; He et al., 2024). Among dinoflagellates, this unusual chloroplast genome organisation seems to be confined to the peridinin-containing taxa, although there are suggestions that the replacement chloroplast genomes in the fucoxanthin dinoflagellates may be evolving in the same direction (Dorrell and Howe, 2012).

The general features of dinoflagellate chloroplast DNA minicircles have been reviewed in detail elsewhere (Barbrook et al., 2019 and refs therein). In general, the genes are located on minicircles of 2-5 kbp. Each minicircle contains a 'core' which is similar (but not identical) between minicircles carrying different genes. The core may contain repeated sequences. The genes on different minicircles are all in the same orientation with respect to the core. It has been suggested that the core is involved in initiation of transcription and replication (arguably resembling the non-coding region ('D-loop') of mammalian mitochondrial DNA (Tan et al., 2024)). There are also a range of anomalous minicircles. These include molecules that are recognisably minicircles by virtue of having a core, but have no (as yet) identified protein coding sequences or RNA genes. Some of them are only a few hundred bp in size and are referred to as 'microcircles'. There are also 'jumbled' minicircles that contain fragments of recognized genes (Zhang et al., 2001). These anomalous minicircles, and the microcircles, may be derived by recombination across repeated sequences. In some organisms, such as *A. carterae* (Barbrook et al., 2001) and *Adenoides eludens* (Nelson and Green, 2005), there are instances of minicircles with more than one gene, such as the *petB* and *atpA* combination in *A. carterae*. They

are combinations that do not appear together in other chloroplast or cyanobacterial genomes, suggesting they arose by fusion between minicircles. There appear to be incomplete sets of tRNA genes, apparently limited to *trnM*, *trnP* and *trnW* (Barbrook et al, 2006b; Nelson et al. 2007). In some taxa transcripts have been shown to receive 3' polyU tails of uncertain function. Minicircles also contain a small number of lineage-specific ORFs (open reading frames) that have not been identified as homologous to sequences in other chloroplasts or cyanobacteria, including a small number of ORFs that are suggested to have been transferred horizontally from bacteria (Mackiewicz et al., 2013). The latter suggestion is controversial, as some have suggested the sequences may actually be from bacterial contaminants in DNA preparations. Setting these ORFs aside, it is striking that the dinoflagellate chloroplast retains no more than 13 identified protein-coding genes, compared to the 100 or so found in typical chloroplast genomes or the few dozen retained in essentially all chloroplast genomes from other photosynthetic organisms (Martin et al., 2002). These 13 genes all encode subunits of the complexes that carry out the light reactions of photosynthesis. They also show a high level of divergence from their homologues in other organisms (He et al., 2024 and refs therein). This divergence is reflected in those structures of components of the photosynthetic machinery that are available (Li et al., 2024). In general, the remaining genes that are typically found in chloroplasts of other lineages have been relocated to the dinoflagellate nucleus, rather than lost, consistent with the retention of chloroplast function (Hackett et al., 2004). The functions of the products of the retained genes are discussed below. There is also evidence in some lineages of editing, and the use of non-canonical translation initiation codons, although we will not discuss the details here.

The chloroplast genome of the green alga *Boodlea composita*

The photosynthetic green alga *Boodlea composita* also has a reduced and fragmented genome, although the fragmentation is somewhat different from that in dinoflagellates (Del Cortona et al., 2017). In *Boodlea* (and probably other members of the Cladophorales), the chloroplast genes are located on multiple hairpin molecules, corresponding to separate sequencing contigs. Each hairpin contains

only one gene, though individual elements may contain additional fragments of that gene, somewhat resembling the arrangement found in dinoflagellate mitochondria (Nash et al., 2008). In total, 21 protein-coding genes have been retained in the *Boodlea* chloroplast, together with a 16S rRNA gene. As with the dinoflagellate chloroplasts, all 21 proteins are components of the photosynthetic electron transfer chain (or the large subunit of ribulose *bis*-phosphate carboxylase), and there is a high degree of convergence with dinoflagellates as to which components remain chloroplast-encoded (Table 1). Essentially all the genes retained in dinoflagellates are also retained in *Boodlea*. The retention of these genes (and loss of others, such as the ribosomal protein genes) contrasts with the situation in plastids of organisms that have lost photosynthesis. In plastids of non-photosynthetic organisms, genes for proteins involved in photosynthesis have generally been lost or turned into pseudogenes, and genes for proteins involved in protein synthesis retained (e.g. Wilson, 2005). The *Boodlea* chloroplast genes are highly divergent at the sequence level from conventional chloroplast homologues, as with dinoflagellates (Del Cortona et al., 2017). Strikingly, another member of the Cladophorales with a similar chloroplast genome organization to *Boodlea, Pithophora roettleri*, shows 3' polyU addition to transcripts from a number of the genes, as with dinoflagellates (Meade et al., 2020). The chloroplast genome of the Cladophorales therefore resembles that of the peridinin dinoflagellates in showing fragmentation, with a single, highly divergent, gene per genetic element, 3' polyuridylylation of some transcripts, and convergent retention of a small number of genes.

The mitochondrial genome of lice

A remarkably similar genome organization to dinoflagellate chloroplasts was found by Shao et al. (2009) in the mitochondrial genome of the human body louse. They reported that the genome had become fragmented into at least 18 circular minichromosomes, 3-4 kbp in size, most containing one or two protein coding genes. Some minichromosomes contain tRNA genes as well, and some contain tRNA genes only. There is also a non-coding region of several hundred nucleotides on each minichromosome. This non-coding region is very similar (though not identical) among minichromosomes carrying the same gene, and also similar among

minichromosomes with different genes. The non-coding region was proposed to act as a transcription initiation site. With the exception of the *nad1* minicircle in a number of taxa, the genes are in the same orientation with respect to the non-coding region (Shao et al., 2012). A similar gene organization has been reported in a wide range of other sucking and chewing lice. The same set of protein-coding genes is present, although there is wide variation in which genes are present on the same minichromosomes, probably reflecting active recombination among the minichromosomes (Fu et al., 2023), and reminiscent of the recombination inferred among dinoflagellate chloroplast minicircles. Within the genus *Liposcelis*, which includes booklice, a less dramatic fragmentation into two minichromosomes has happened on multiple occasions (Wei et al., 2012; Feng et al., 2022), and there is evidence of increased sequence divergence and rearrangement of gene order both in taxa showing fragmentation and in closely related taxa that do not. This may be indicative of an intermediate stage to fragmentation, as discussed below.

It is difficult to comment with certainty on the retention of the tRNA genes, given the possibility of change in usage by anticodon mutations, but the fragmented lice mitogenomes generally have the same protein coding content as other arthropods (Hassanin, 2006 and Table 2), with the possible exception of *nad4*, which may be absent from human pubic lice although this is not certain (Shao et al., 2012).

The mitochondrial genome of the plant *Rhopalocnemis phalloides*

Rhopalocnemis phalloides is a non-photosynthetic, parasitic plant. It has a highly-reduced, AT-rich chloroplast genome (Schelkunov et al., 2019). However, the mitochondrial genome is particularly striking. It comprises 21 minichromosomes, 4.9 to 7.9 kbp in size (Yu et al., 2022), most containing 3-6 genes or exons. (A number of proteins are encoded in exons on separate minichromosomes.) The chromosomes all contain a highly conserved core region, flanked by less conserved regions, which is proposed to act as an origin of replication. In most, but not all, cases the genes on a given minichromosome are all on the same strand. Yu et al. (2022) reported a high level of variation between individual sequence reads, indicating (as the DNA preparation came from a single plant) a high level of

heteroplasmy. The gene content is broadly similar to the typical angiosperm mitochondrial content, although detailed analysis is difficult, given the lack of mitochondrial genome sequences from close relatives. Comparison with the mitochondrial genome of *Malania oleifera*, within the same order (Santales) as *Rhopalocnemis*, suggests that some genes for proteins in the small subunit of the ribosome may have been lost from the *R. phalloides* mitochondrion, as *rps7* and *rps14* are absent from the latter, but present in the mitogenome of *M. oleifera* (Luo et al., 2020).

The mitochondrial genome of the red alga *Rhodosorus marinus*

The mitochondrial genomes of red algae do not generally show evidence of fragmentation, although some, such as *Galdieria* species, show high levels of gene sequence divergence from other mitochondrial homologues (Cho et al., 2020). However, the mitochondrial genomes of the unicellular red alga *Rhodosorus marinus* and relatives within the Stylonematophyceae show a very different organisation from other red algae (Lee at al., 2023). In *Rh. marinus*, the genome was found to be organised into minicircles of 2.9 – 6.3 kbp. Somewhat larger minicircles of 7.2 – 17.0 kbp were found in another species in the Stylonematophyceae, *Chroodactylon ornatum*. Minicircles contain at most two genes, with one gene (*nad5*) split between two minicircles. Most of the minicircles include a repeat-containing non-coding region that is essentially identical between the minicircles in a given species, but not between species. Coding regions are generally in the same orientation with respect to the non-coding region. Only one tRNA gene (*trnM*) was identified in *Rh. marinus*, with two (*trnW* and *trnM*) in *C. ornatum* (Lee et al., 2003, and ACB, unpublished analysis of the data therein). Interestingly, both of these tRNA genes are also present in the very limited tRNA gene complement of dinoflagellate chloroplasts. The protein coding sequences in *Rh. marinus* and *C. ornatum* mitogenomes are highly divergent, and the gene complements are significantly reduced in both taxa. Genes that are usually retained in the mitochondrion in the Bangiophycean and Florideophycean red algae (sisters to the group containing the stylonematophycean *Rh. marinus* and *C. ornatum*), but have been lost to the nucleus in both *Rh. marinus*

and *C. ornatum* include *atp4, atp8*, *nad3*, *nad4L, nad6*, *sdhC, sdhD,* and genes for a few proteins of the small and large subunits of the ribosome.

The mitochondrial genome of diplonemids

Diplonemids are heterotrophic marine flagellates within the Euglenozoa, with a particularly complex mitochondrial genome structure (Valach et al., 2016), best characterised in *Diplonema papillatum*. The mitogenome comprises over 80 minicircular chromosomes in two size classes, 6.0 - 6.3 kbp or 7.0 – 7.3 kbp. Each minicircle generally carries only part of a gene, around 40 – 450 bp, with full transcripts assembled by *trans*-splicing. The rest of the minicircle is similar between minicircles. Part of this non-coding region is essentially identical among minicircles of the same size class, and part of it is essentially identical across all minicircles (Marande et al., 2005; Marande and Burger, 2007). As well as *trans*-splicing, transcripts undergo editing and addition of 3' tails which can include polyU and polyA,U (i.e. heteropolymers). These tails can be added to transcripts of internal exons, and are therefore lost during *trans*-splicing. They can also be added to the 3' terminal exons, in which case they remain on the transcripts after splicing (Kaur et al., 2020). Diplonemid protein sequences show high levels of divergence, although this is also true of other members of the Euglenozoa. This divergence hampered the identification of a number of diplonemid mitochondrial genes, but the gene complement is now believed to include *atp6*, *cob*, *cox1, cox2, cox3*, *nad1, nad2, nad3, nad4, nad4L, nad5, nad6, nad7, nad8* and *nad9* (as well as *rrnL* and *rrnS*) and an unidentified open reading frame, but no tRNA genes (Kaur et al., 2020). It is difficult to compare this complement with that of the diplonemids' close relatives the trypanosomatids, because of the challenge of identifying some of the open reading frames in the latter. Nevertheless, although the *rps12* gene is present in the trypanosomatids (Callejas-Hernandez et al., 2021) it has not yet been reported in the diplonemids. By contrast, *Euglena gracilis*, which is less closely related to the diplonemids than the trypanosomatids are, but whose mitochondrial genome is not fragmented into mincircles, has a more reduced mitochondrial genome than the diplonemids (Dobákova et al., 2015). It therefore seems that there has not been

significant mitochondrial genome reduction specifically in the fragmented diplonemid line.

Mitochondrial genomes in Cnidaria

The cnidarian parasite *Enteromyxum leei* has a large mitochondrial genome fragmented into at least eight circular molecules of 22-24 kbp (Yahalomi et al., 2017). They include genes for rRNAs, cytochrome *b*, COX1, COX2, NAD1, NAD5 and some unidentified open reading frames. The coding regions identified are highly divergent, and tRNA genes are probably lacking. A large part of each circle (in most of the molecules approximately 22 kbp) is highly conserved between circles with different genes. Less extreme fragmentation has also occurred in the mitochondrial genome of another cnidarian, the box jellyfish *Alatina moseri* (Smith et al., 2011). This has a gene content typical of other Cnidaria (Kayal et al., 2012), including a gene for a putative DNA polymerase subunit, and an unidentified ORF. The genes are on eight molecules. These molecules are linear, rather than circular, but with telomeric regions that are identical at both ends and on all the linear molecules. The jellyfish *Morbakka sp.* also has a partially fragmented mitogenome and the same protein gene content as *Alatina moseri* (Ling et al., 2023). Not all Cnidaria or even all Medusozoa (jellyfish) have fragmented genomes, however (Yahalomi et al., 2017; Ling et al., 2023). Whether the lack of genes in *Enteromyxum* compared to other Cnidaria (Table 2A) is due to a failure to detect them (perhaps given their divergence) or a genuine absence, is not fully clear.

Fragmented linear mitochondrial genomes

Large-scale genome fragmentation has also taken place in some cases, but with fragmentation into linear, rather than circular, molecules (reviewed by Lavrov and Pett, 2016; Formaggioni et al., 2021). In the calcaronean sponges, the mitochondrial genes (*atp6*, *atp9*, *cob*, *cox1*, *cox2*, *cox3*, *nad1*, *nad2*, *nad3*, *nad4*, *nad5* and a number of tRNA genes) are highly divergent and located on individual, probably linear, molecules 1-2 kbp in size and with similar telomeric sequences (Lavrov et al., 2016). In another group of sponges, the Calcinea (e.g., *Clathrina clathrus*), there is also some fragmentation, but with more genes per subgenomic molecule than in the

Calcaronea (Lavrov et al., 2012). The mitogenomes of Calcinea and a group with non-fragmented mitogenomes (hexactinellid sponges) also lack *atp8*, but contain *nad4L* and *nad6* (Rosengarten et al., 2008), suggesting loss of the last two in the Calcaronea lineage (Table 2A).

The mitochondrial genome of the ichthyosporean parasite *Amoebidium parasiticum* comprises several hundred linear molecules of a few kbp, which may contain no, one, or many genes. The molecules have repeated telomeric sequences, which are identical on a given chromosome, but not between chromosomes. In addition to the genes that are normally found in animal mitochondria, there are *atp9*, *rps3*, *rps4* and *rps13* genes (Burger et al., 2003). The non-fragmented mitochondrial genome of the ichthyosporean relative *Sphaerothecum destruens* (Sana et al., 2020) contains in addition *ccmC* and *ccmF* (whose products are involved in cytochrome maturation), and *tatC* (whose product is involved in protein import). It also contains *rpl2, rpl16, rps13* and *rps14*, but lacks the *rps3* and *rps4 that A. parasiticum* has. This suggests there may have been some gene loss from the *A. parasiticum* lineage (Table 2B).

Three mitochondrial genes, *COX1*, *COX2* and *COX3*, along with two putative tRNA genes (*trnR* and *trnG*) have been reported in the poorly characterised parasite *Dicyema*, on circles of 1.6 – 1.7 kbp (Watanabe et al., 1999). However, it seems likely that additional genes remain to be identified.

General features of fragmented genomes

The features of the highly fragmented genomes described are summarised in Table 3. The dinoflagellate chloroplast genome epitomises the changes seen in these genomes. The genomes are fragmented, usually into minicircles with a region conserved among minicircles in the same species, and probably functioning in replication and transcription. For some linearised genomes, the telomeres are similar between different molecules. The coding sequences are highly divergent. Transcripts may acquire 3' polyU tails and be edited. The gene content may be reduced, with genes relocated to the nucleus, although this has happened to a greater extent in the chloroplasts than the mitochondria. It seems that most mitogenomes had already

undergone more reduction prior to fragmentation than have chloroplast genomes, and were already at (or close to) the minimal gene content as discussed below. Why genome reduction should have happened particularly in mitochondria prior to any fragmentation is not clear. As a result, there is generally less opportunity for genome reduction in mitochondria than in chloroplasts. Nevertheless, where a genome has been fragmented, movement of genes from the organelle to the nucleus (and therefore organelle genome reduction) may offer a selective advantage in reducing the risk of genes being lost during division, as discussed below.

Which genes are retained? Chloroplasts and the COCOA hypothesis.

The peridinin dinoflagellate chloroplasts present the most remarkable example of genome reduction. This chloroplast is of secondary origin from a red alga, so the gene content has been reduced by an order of magnitude since that secondary origin (Dorrell and Howe, 2015). Most of the genes that are located in the chloroplast in essentially all other photosynthetic lineages (Martin et al., 2002) have been lost to the nucleus. This raises the question of why certain genes have been retained. It is striking that genes for the proteins that are particularly important in the early steps of assembly of the complexes of the photosynthetic electron transfer chain have all been retained (Barbrook et al., 2019). Assembly of Photosystem I begins with the insertion of PsaA and PsaB into the thylakoid and formation of a heterodimer. The other subunits are then assembled around this (Yang et al., 2015). In Photosystem II assembly, four modules are initially assembled separately, and then assembled together in sequence. These are the D1 module (built on PsbA), the D2 module (built on PsbD), the CP47 module (built on PsbB) and the CP43 module (built on PsbC) (Komenda et al., 2024). PsbE assembles with PsbD, and PsbI assembles with PsbA prior to assembly of the D2 and D1 modules and the addition of the CP47 and CP43 modules. The assembly of the cytochrome b_6f complex begins with the association of cytochrome b_6 (PetB) with subunit IV (PetD) (Hassan et al., 2013). The assembly of the CF1 portion of the ATP synthase probably begins with the association of AtpA and AtpB (Drapier et al., 2007). These eleven early-assembly proteins (PsaA, PsaB, PsbA, PsbB, PsbC, PsbD, PsbE, PsbI, PetB, PetD, AtpA and AtpB) are all chloroplast-encoded in dinoflagellates.

The Boodlea chloroplast genome is not as streamlined as the dinoflagellate one, but the same set of early-assembled proteins remains organelle encoded, with the possible exception of PsbI. [The role and importance of PsbI are not fully clear, and may differ between taxa (Dobáková et al. 2007)]. RbcL is also organelle-encoded in *Boodlea.* Assembly of the Rubisco (ribulose *bis*-phosphate carboxylase) complex begins with RbcL, eventually forming a complex including the assembly factor RAF1 (and possibly other chaperones). RbcS is then incorporated into this complex, displacing RAF1. RbcL therefore plays a key role in initiating the assembly of the $RbcL_8S_8$ complex (Wietrzynski et al., 2021 and refs therein) and the retention of its gene in the chloroplast follows the same principle as for genes encoding components of the electron transfer chain. It is worth noting that carbon fixation provides a sink for photosynthetically generated electrons, so the fact that Rubisco's genes follow a similar pattern to those for electron transfer chain components is not surprising. In peridinin dinoflagellates and their close relatives *Chromera velia* and *Vitrella brassicaformis*, the Rubisco typical of oxygenic photosynthetic organisms has been replaced with a Type II enzyme typical of anoxygenic bacteria, presumably by lateral transfer (Morse et al., 1995; Janouškovec et al., 2010). The gene for this is nuclear in dinoflagellates, *Chromera* and *Vitrella,* and it seems likely this would be the case when it was first acquired, as it is very rare for the chloroplast to acquire foreign DNA, unlike the nucleus and mitochondrion (Cui et al., 2021 and refs therein). The nuclear location of *rbcL* in dinoflagellates would not be a problem for the principle of early-assembled subunits being chloroplast-encoded, as the Type II Rubisco comprises dimers of a L subunit, without an S subunit (Morse et al., 1995).

There appears to be a clear correlation between a gene being retained in these highly reduced genomes and its product being involved in early stages of assembly of a protein complex involved in electron transfer or a closely associated function. We suggest that retaining the genes for proteins that initiate complex assembly gives the chloroplast greater control over the overall biogenesis of the complexes. Assembly of the complexes cannot begin without the organelle-encoded subunits. This may help to avoid the accumulation of incomplete complexes, which might

generate reactive oxygen species. We call the requirement for the retention in the organelle of genes for proteins initiating assembly of electron transfer complexes (and other closely associated complexes, such as Rubisco) the CO-location for Control of Assembly, COCOA, principle.

Mitochondrial genomes

Fragmented mitochondrial genomes show less reduction compared to their unfragmented relatives than fragmented chloroplast genomes, as they were already highly reduced prior to fragmentation (Table 2). Nevertheless, the COCOA principle still applies, as there is also an apparent association between proteins being mitochondrially encoded and being involved in early steps of electron transfer complex assembly. For complex I, the ND1, ND2, ND4 and ND5 subunits are involved early in the assembly of modules (ND-1, ND-2, ND-4 and ND-5) that together form complex I (Signes and Fernandez-Virra, 2018). Subunits ND3, ND4L and ND6 are involved at later stages in the formation of the ND-2 module. Most mitochondrial genomes encode ND1, ND2, ND3, ND4, ND4L, ND5 and ND6, as do the lice, *Rhopalocnemis* and diplonemid minicircles. However, ND3, ND4L and ND6 genes (whose products are involved at later assembly stages) have been lost from one or more of the Stylonematophyceae (red algal) mitogenomes, ND4L and ND6 genes have also apparently been lost from the calcaronean sponge mitogenomes (and possibly some or all of ND2, ND3, ND4, ND4L and ND6 lost from the *Enteromyxum* mitogenome).

For Complex II it is not obvious that any one subunit plays the lead role in initiating assembly of the whole (van Kranken et al., 2015), and this seems to be reflected in the range of coding locations. In many organisms, all subunits are nuclear encoded. In many plants (Sloan et al., 2012) including *Rhopalocnemis* and *Malania*, subunits C and D (the membrane-associated ones) are mitochondrially encoded. In the Florideophycean and Bangiophycean red algae subunit B (which carries the electron transfer wire) is mitochondrially encoded in addition to subunits C and D. However, in *Rhodosorus marinus* (Stylonematophyceae) all three genes have been lost from the mitogenome, with the gene for subunit B retained in *Chroodactylon ornatum*.

Complex III assembly begins with cytochrome *b* (Signes and Fernandez-Virra, 2018), which is always mitochondrially encoded. Complex IV assembly requires the separate assembly of three modules, MT-CO1, MT-CO2, and MT-CO3, and these require COX1, COX2 and COX3 respectively for initiating module assembly (Signes and Fernandez-Virra, 2018). These three proteins are mitochondrially encoded in all the examples here, with the possible exception of COX3 in the Cnidarian parasite *Enteromyxum leei.* It may be that the gene for this is present in the *E. leei* mitochondrion, but has not yet been detected, although it is also interesting that the MT-CO3 module is the last to be added, so the COX3 gene would be the strongest candidate for relocation to the nucleus according to the COCOA hypothesis. Complex V requires assembly of an F_1 subcomplex, together with the proton channel, and the stator. ATP6 and ATP8 are added late (He et al., 2018). ATP6, which provides the two proton half-channels through the mitochondrial inner membrane, is mitochondrially encoded in all the organisms considered, in spite of its later incorporation. *ATP8* is mitochondrial in the lice and *Rhopalocnemis*, but not in diplonemids, Stylonematophyceae (and some other related red algae, so it is possible its loss predates fragmentation) and the sponges. It is interesting that, although ATPA and ATPB are important for initiating assembly of the F_1 component, they are not mitochondrially encoded, whereas their chloroplast equivalents are chloroplast encoded. It is therefore not clear that the assembly of ATP synthases accords with the COCOA principle, although it is also worth noting that the ATP synthase does not directly carry out direct electron transfer reactions.

COCOA and other hypotheses

The COCOA principle therefore proposes that organelle genomes retain genes for the subunits of their electron transfer chain complexes, and closely associated complexes, that are important in early steps of assembly of those complexes (or sub-complex modules). Where organelle genomes have become reduced (perhaps as a consequence of fragmentation) those are the genes least likely to be lost. As a corollary of COCOA, genes involved in later stages of assembly of electron transfer chain complexes, or in other functions (such as the ribosomal proteins) will be lost if

organelle genomes reduce. Why *ATP6* is apparently an exception to this is not clear. Other additional genes may be present in organelles in some lineages and not others, but COCOA deals with those genes that must be retained, even in highly reduced genomes.

COCOA adds to, but does not displace other hypotheses on organelle genomes. Ellis proposed the 'Multisubunit completion principle' in 1977, which states that organelle-encoded proteins interact with at least one nuclear encoded protein (Ellis, 1977). COCOA is fully compatible with this, and adds that the proteins that remain organelle encoded in spite of genome reduction not only follow this principle, but are fundamental to initiating assembly of the multisubunit complexes of which they are part.

It has also been proposed that retention of genes by organelles allows control of expression to be linked to the redox state of the organelle, in the 'CORR' or 'CO-location for Redox Regulation' hypothesis (Allen, 2003). This was expanded on by Barbrook et al. (2006a) who suggested that regulation might be in response to other biochemical needs of the organelle, not just the redox state. Another hypothesis is that retention of genes by organelles allows regulation of expression of subunits of a complex in response to levels of unassembled components of the same complex. This is often referred to as 'CES' or 'Control by Epistasy of Synthesis' (e.g., Wietrzynski et al., 2021). CES is embodied in the 'Cytoplasmic control principle' also proposed by Ellis (1977), which argues more specifically that the synthesis of organelle-encoded subunits of complexes is regulated by the same subunits in a negative feedback loop. COCOA does not replace the CORR or CES hypotheses. COCOA explains which genes must be retained in the organelle, although other genes may additionally be present in some lineages. CORR explains why the retained genes may need to be redox regulated. For example, the COCOA hypothesis 'requires' that the *psaA, psaB* and *psbA* genes be retained in the chloroplast, and the CORR hypothesis 'requires' that they be redox regulated (Pfannschmidt et al., 1999). This does not exclude the possibility that, according to CORR, other genes may also be retained in many lineages, and be redox regulated.

Similarly, we argue that the COCOA principle requires that the Type I *rbcL* gene be retained in the chloroplast. Synthesis of RbcL is inhibited by unassembled RbcL, so it is regulated by CES (Wietrzynski et al., 2021). COCOA does not require the cytochrome *f* gene (*petA*) to be retained by the chloroplast. However, *petA* is present in the chloroplast in the green alga *Clamydomonas*, where translation of its mRNA depends on the presence of PetB and PetD, which is again an example of CES (Kuras and Wollman, 1994). A prediction of COCOA is that all organelle genes, even if they are regulated in response to redox poise or by epistasy of synthesis, can in principle be relocated to the nucleus, as long as their products are not required for initiating assembly. Tables 1 and 2 indicate which genes are expected to be retained in all chloroplasts and mitochondria (as long as the organelles carry out oxygenic photosynthesis and aerobic respiration respectively).

What caused fragmentation and genome reduction?

A number of hypotheses have been proposed for the initiation of fragmentation. One possibility is the invasion of organelles by transposable elements (which might correspond to the conserved regions of different minicircles). The presence of multiple copies of a repeat within an organelle genome, combined with recombination, might then lead to fragmentation (Lonsdale et al., 1984: Blount et al., 2023). If smaller molecules replicated faster than larger ones, minicircles containing one or a few genes might come to dominate (Shao et al., 2009). Molecules containing multiple origins of replication might also be unstable. A number of studies have commented on the association between fragmentation and rapid sequence evolution, suggesting that the latter reflects a breakdown of normal repair and recombination processes, somehow leading to fragmentation. For example, the chloroplast genomes of Trentepohliales, which are close relatives of the Cladophorales (del Cortona et al., 2017), show an increased level of sequence divergence and genome rearrangements, but are not fragmented (Fang et al., 2021). The booklice genus *Liposcelis* shows less mitochondrial DNA fragmentation than the sucking lice, but nevertheless shows high levels of sequence divergence (Feng et al., 2022).

In chloroplasts, fragmentation was accompanied by large-scale transfer of genes to the nucleus. It is not clear whether fragmentation preceded gene loss or *vice versa*. As noted above, fragmentation may lead to a risk of loss of genes on organelle division. That risk may be particularly great if the copy number of individual subgenomic molecules is low. This is the case for dinoflagellate chloroplast minicircles, for example, where the copy number may be as low as a few per cell, depending on growth conditions (Koumandou and Howe, 2007; He et al., 2024). One hypothesis is that movement of genes to the nucleus preceded fragmentation. This made subsequent fragmentation less risky, with fewer genes to maintain in the organelle. A second hypothesis is that fragmentation occurred first and was followed by streamlining of the genome to reduce the risk of loss. This transfer could in principle have happened in a short time, because transfer of DNA to the nucleus is surprisingly frequent, at least in strains with multiple chloroplasts (Lister et al., 2003). It is also possible that fragmentation and streamlining happened concomitantly. Comparisons between taxa that are closely related but differ in the extent of organelle genome fragmentation may be very informative in understanding fragmentation and its timing with regard to gene transfer.

Future directions

For fragmented genomes, one of the most obvious questions remaining is how organisms ensure (if they do) that daughter cells retain a full set of the subgenomic molecules on division. The molecules may be present in low copy number, but they encode proteins that are essential for organelle function, so loss of one would be expected to be very deleterious. A better understanding of how subgenomic molecules are replicated and expressed will also be important. For minicircles, rolling circle replication is an obvious possibility, supported by observations in some instances of concatemers (e.g. Yu et al., 2022). Identification of *trans*-acting factors required for replication will also be important. Rolling circle transcription may also occur (Dang and Green, 2010; Barbrook et al., 2012).

In organisms where genes have apparently been lost from an organelle, it will be important to confirm if they have been relocated to the nucleus, or at least

functionally replaced by a nuclear gene, rather than being lost altogether. For example, dinoflagellate mitochondrial genomes encode fewer proteins than those of most other aerobic mitochondria. However, Complexes I and V of the dinoflagellate electron transfer chain appear to be very different from the canonical forms, so the reduction in coding content may reflect outright gene loss rather than transfer to the nucleus (Butterfield et al., 2013; Butterfield et al., 2016; Raven and Beardall, 2017).

A number of studies have looked at the consequence of artificially relocating a gene that is usually located in an organelle. For example, in a classic study, insertion of a gene encoding an atrazine-resistant form of PsbA (fused to a chloroplast-targeting sequence) from *Amaranthus* into the tobacco nucleus resulted in plants with some atrazine tolerance, although *psbA* is a chloroplast gene in both organisms (in accordance with COCOA). This indicated successful import and assembly of the PsbA protein, although the atrazine tolerance was incomplete, suggesting that the relocated gene was not as effective as a chloroplast gene would have been. However, there are many possible reasons for this, not least variability of expression in the nucleus (Cheung et al., 1988). Some attempts to relocate mitochondrial genes have also met with only partial success (e.g. Björkholm et al., 2017). Again, interpretation of these experiments is complex. Getting a rigorous understanding of the consequences for assembly of complexes in an organelle of relocating genes to the nucleus will be an important challenge.

There are a number of possibilities for exploiting the fragmentation of organelle genomes. Where they are present at high copy number, and especially when they show a high level of sequence divergence, they may be particularly useful as taxonomic markers (Barbrook et al., 2006) or for diagnostic purposes, as with trypanosomatid minicircles (e.g. Weigle et al., 2002). In addition, by analogy with small self-replicating plasmids in other microorganisms, minicircles may also have value as the basis for shuttle vectors in transformation experiments, as has been shown for dinoflagellate chloroplasts (Nimmo et al., 2019), opening up a wide range of studies in otherwise intractable organisms.

Acknowledgements

We thank other members of our lab for helpful comments, and the Natural Environment Research Council (NE/X010503/1) and Churchill College, Cambridge, for financial support.

The Authors have no conflicting interest to declare in the paper.

References

Allen, J.F., 2003. The function of genomes in bioenergetic organelles. Phil. Trans. Roy. Soc. B 358, 19-37.

Barbrook, A.C., Dorrell, R.G., Burrows, J., Plenderleith, L.J., Nisbet, R.E.R., Howe, C.J., 2012. Polyuridylylation and processing of transcripts from multiple gene minicircles in chloroplasts of the dinoflagellate *Amphidinium carterae*. Plant Molecular Biology 79, 347-357.

Barbrook, A.C., Howe, C.J., 2000. Minicircular plastid DNA in the dinoflagellate *Amphidinium operculatum*. Mol. Gen. Genet. 263, 152-158.

Barbrook, A.C., Howe, C.J., Purton, S., 2006a. Why are plastid genomes retained in non-photosynthetic organisms? Trends Plant Sci. 11, 101-108.

Barbrook, A.C., Nisbet, R.E.R., Howe, C.J., 2019. Breaking up is hard to do: the complexity of the dinoflagellate chloroplast genome. Perspectives Phycol. 6, 31-37.

Barbrook, A.C., Santucci, N., Plenderleith, L.J., Hiller, R.G., Howe, C.J., 2006b. Comparative analysis of dinoflagellate chloroplast genomes reveals rRNA and tRNA genes. BMC Genomics 7, 297.

Barbrook, A.C., Symington, H., Nisbet, R.E.R., Larkum, A., Howe, C.J., 2001. Organisation and expression of the plastid genome of the dinoflagellate *Amphidinium operculatum*. Mol. Genet. Genom. 266, 632-638.

Barbrook, A.C., Visram, S., Douglas, A., Howe, C.J., 2006. Molecular diversity of dinoflagellate symbionts of Cnidaria: the psbA minicircle of Symbiodinium. Protist 157, 159-171.

Björkholm, P., Ernst, A.M., Hagström, E., Andersson, S.G., 2017. Why mitochondria need a genome revisited. FEBS Lett 591, 65-75.

Blount, B.A., Liu, X., Driessen, M.R.M. et al., 2023. Synthetic yeast chromosome XI design provides a testbed for the study of extrachromosomal circular DNA dynamics. Cell Genomics 3, 100418.

Burger, G., Forget, L., Zhu, Y., Gray, M.W., Lang, B.F., 2003. Unique mitochondrial genome architecture in unicellular relatives of animals. Proc. Natl Acad. Sci. USA 100, 892-897.

Burger, G., Gray, M.W., Forget, L., Lang, B.F., 2013. Strikingly bacteria-like and gene-rich mitochondrial genomes throughout jakobid protists. Genome Biol. Evol. 5, 418-438.

Butterfield, E.R., Howe, C.J., Nisbet, R.E.R., 2013. An analysis of dinoflagellate metabolism using EST data, Protist 164, 218-236.

Butterfield, E.R., Howe, C.J., Nisbet, R.E.R., 2016. Identification of sequences encoding *Symbiodinium minutum* mitochondrial proteins. Genome Biol. Evol. 8, 439-445.

Callejas-Hernandez, F., Herreros-Cabello, A., del Moral-Salmoral, J., Fresno, M., Gironès, N., 2021. The complete mitochondrial DNA of *Trypanosoma cruzi:* maxicircles and minicircles. Front. Cell. Inf. Microbiol. 11, 672448.

Cheung, A.Y., Bogorad, L., van Montagu, M., Schell, J., 1988. Relocating a gene for herbicide tolerance: a chloroplast gene is converted into a nuclear gene. Proc. Natl Acad. Sci. USA. 85, 391-395.

Cho, C.H., Park, S.I., Ciniglia, C., Yang, E.C., Graf, L., Bhattacharya, D., Yoon, H.S., 2020. Potential causes and consequences of rapid mitochondrial genome evolution in thermoacidophilic *Galdieria* (Rhodophyta). BMC Evol. Bio. 20, 112.

Cui, H., Ding, Z., Zhu, Q., Wu, Y., Qiu, B., Gao, P., 2021. Comparative analysis of nuclear, chloroplast, and mitochondrial genomes of watermelon and melon provides evidence of gene transfer. Sci. Rep. 11, 1595.

Dang, Y., Green, B.R., 2010. Long transcripts from dinoflagellate chloroplast minicircles suggest “rolling circle” transcription. J. Biol. Chem. 285, 5196-5203.

Del Cortona, A., Leliaert F., Bogaert, K.A., Turmel, M., Boedeker, C., Janouškovec, J., Lopez-Bautista, J.M., Verbruggen, H., Vandepoele, K., De Clerck, O., 2017.

The plastid genome in Cladophorales green algae is encoded by hairpin chromosomes. Curr. Biol. 27, 3771-3782.

Dobákova, E., Flegontov, P., Skalicky, T., Lukeš, J., 2015. Unexpectedly streamlined mitochondrial genome of the euglenozoan *Euglena gracilis*. Genome Biol. Evol. 7, 3358–3367.

Dobákova, M., Tichý, M., Komenda, J., 2007. Role of the PsbI protein in Photosystem II assembly and repair in the cyanobacterium *Synechocystis* sp. PCC 6803. Plant Physiol. 145, 1681-1691.

Dorrell, R.G., Howe, C.J., 2012. Functional remodeling of RNA processing in replacement chloroplasts by pathways retained from their predecessors. Proc. Natl Acad. Sci. USA. 109, 18879-18884.

Dorrell, R.G., Howe, C.J., 2015. Integration of plastids with their hosts: lessons learned from dinoflagellates. Proc. Natl Acad. Sci. USA 112, 10247-10254.

Drapier, D., Rimbault, B., Vallon, O., Wollman, F-A., Choquet, Y., 2007. Intertwined translational regulations set uneven stoichiometry of chloroplast ATP synthase subunits. EMBO J. 26, 3581-3591.

Ellis, R.J., 1977. Protein synthesis by isolated chloroplasts. Biochim. Biophys. Acta 463, 185-215.

Fan, J., Lee, R.W., 2002. Mitochondrial genome of the colorless green alga *Polytomella parva*: two linear DNA molecules with homologous inverted repeat termini. Mol. Biol. Evol. 19, 999-1007.

Fang J., Liu, B., Liu, G., Verbruggen, H., Zhu, H., 2021. Six newly sequenced chloroplast genomes from Trentepohliales: the inflated genomes, alternative genetic code and dynamic evolution. Front. Plant Sci. 12, 780054.

Feng, S., Pozzi, A., Stejskal, V., Opit, G., Yang, Q., Shao, R., Dowling, D.K., Li, Z., 2022. Fragmentation in mitochondrial genomes in relation to elevated sequence divergence and extreme rearrangements. BMC Biology, 20, 7.

Formaggioni, A., Luchetti, A., Plazzi, F., 2021. Mitochondrial genome landscape: a portrait of the mitochondrial genome 40 years after the first complete sequence. Life 11, 663.

Fu, Y-Y., Shao, R., Suleman, Wang, W., Wang, H-M., Liu, G-H., 2023. The fragmented mitochondrial genomes of two *Linognathus* lice reveal active

minichromosomal recombination and recombination hotspots. iScience 26, 107351.

Gómez, F., 2012. A quantitative review of the lifestyle, habitat and trophic diversity of dinoflagellates (Dinoflagellata, Alveolata). Systematics and Biodiversity 10, 267-275.

Gornik, S.G., Ford, K.L., Mulhern, T.D., Bacic, A., McFadden, G.I., Waller, R.F., 2012. Loss of nucleosomal DNA condensation coincides with appearance of a novel nuclear protein in Dinoflagellates. Curr. Biol. 22, 2303-2312.

Hackett, J.D., Yoon, H.S., Soares, M.B., Bonaldo, M.F., Casavant, T., Scheetz, T.E., Nosenko, T., Bhattacharya, D., 2004. Migration of the plastid genome to the nucleus in a peridinin dinoflagellate. Curr. Biol. 14, 213–218.

Hassan, S.S., Yamashita, E., Cramer, W.A., 2013. Transmembrane signaling and assembly of the cytochrome b_6f-lipidic charge transfer complex. Biochim. Biophys. Acta 1827, 1295-1308.

Hassanin, A., 2006. Phylogeny of Arthropoda inferred from mitochondrial sequences: strategies for limiting the misleading effects of multiple changes in pattern and rates of substitution. Mol. Phyl. Evol., 38, 100-116.

He., J., Ford, H.C., Carroll, J., Douglas, C., Gonzales, E., Ding, S., Fearnley, I.M., Walker, J.E., 2018. Assembly of the membrane domain of ATP synthase in human mitochondria. Proc. Natl Acad. Sci. USA 115, 2988-2993.

He., J., Huang, Y., Li, L., Lin, S., Ma, M., Wang, Y., Lin, S., 2024. Novel plastid genome characteristics in *Fugacium kawagutii* and the trend of accelerated evolution of plastid proteins in dinoflagellates. Genome Biol. Evol. 16, evad237.

Hofmann, E., Wrench, P.M., Sharples, F.P., Hiller, R.G., Welte, W., Diederichs, K., 1996. Structural basis of light harvesting by carotenoids: peridinin-chlorophyll-protein from *Amphidinium carterae*. Science 272, 1788-1791.

Janouškovec, J., Horák, A., Oborník, M., Lukeš, J., Keeliing, P.J., 2010. A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. Proc. Natl Acad. Sci. USA 107, 10949-10954.

Kaur, B., Záhonová, K., Valach, M., Faktorová, D., Prokopchuk, G., Burger, G., Lukeš, J., 2020. Gene fragmentation and RNA editing without borders:

eccentric mitochondrial genomes of diplonemids. Nucleic Acids Res. 28, 2694-2708.

Kayal, E., Bentlage, B., Collins, A.G., Kayay, M., Pirro, S., Lavrov, D.V., 2012. Evolution of linear mitochondrial genomes in medusozoan cnidarians. Genome Biol. Evol. 4, 1-12.

Komenda, J., Sobotka, J., Nixon, P.J., 2024. The biogenesis and maintenance of photosystem II: recent advances and current challenges. Plant Cell accepted manuscript doi.org/10.1093/plcell/koae082, accessed 9 April 2024.

Koumandou, V.L., Howe, C.J., 2007. The copy number of chloroplast gene minicircles changes dramatically with growth phase in the dinoflagellate *Amphidinium operculatum*. Protist 158, 89-103.

Kuras, R., Wollman, F.A., 1994. The assembly of cytochrome b6/f complexes: an approach using genetic transformation of the green alga Chlamydomonas reinhardtii. EMBO J 13, 1019-1027.

Lavrov, D.V., Adamski, M., Chevaldonné, P., Adamska, M., 2016. Extensive mitochondrial mRNA editing and unusual mitochondrial genome organization in calcaronean sponges. Curr. Biol. 26, 86-92.

Lavrov, D.V., Pett, W., 2016. Animal mitochondrial DNA as we do not know it: mt-genome organization and evolution in non-bilaterian lineages. Genome Biol. Evol. 8, 2896-2913.

Lavrov, D.V., Pett, W., Voigt, O., Wörheide, G., Forget, L, Lang, B.F., Kayal, E., 2012. Mitochondrial DNA of *Clathrina clathrus* (Calcarea, Calcinea): six linear chromosomes, fragmented rRNAs, tRNA editing, and a novel genetic code. Mol. Biol. Evol. 30, 865-880.

Lee, Y., Cho, C.H., Noh, C., Yang, J.H., Park, S.I., Lee, Y.M., West, J.A., Bhattacharya, D., Jo, K., Yoon, H.S., 2023. Origin of minicircular mitochondrial genomes in red algae. Nat. Comm. 14, 3363.

Li, X., Li, Z., Wang, F. et al., 2024. Structures and organizations of PSI–AcpPCI supercomplexes from red tidal and coral symbiotic photosynthetic dinoflagellates. Proc. Natl Acad. Sci USA 121, e23154761.

Ling, M.K., Yap, N.W.L., Iesa, I.B., Yip, Z.T., Huang, D., Quek, Z.B.R., 2023. Revisiting mitogenome evolution in Medusozoa with eight new mitochondrial genomes. iScience 26, 108252.

Lister, D.L., Bateman, J.M., Purton, S., Howe, C.J., 2003. DNA transfer from chloroplast to nucleus is much rarer in *Chlamydomonas* than in tobacco. Gene 316, 33-38.

Lonsdale, D.M., Hodge, T.P., Fauron, C.M., 1984. The physical map and organisation of the mitochondrial genome from the fertile cytoplasm of maize. Nucleic Acids Res. 12, 9249-9261.

Luo, H., Xu, J., Jiao, S-Q., Zhang, R-G., Mao, J.F., 2020. The complete mitochondrial genome of an endangered tree: *Malania oleifera*, Mitochondrial DNA Part B, 5, 3829-3830.

Mackiewicz, P., Bodył, A., Moszczyński, K., 2013. The case of horizontal gene transfer from bacteria to the peculiar dinoflagellate plastid genome. Mobile Genetic Elements 3, e25845.

Marande, W., Burger, G., 2007. Mitochondrial DNA as a genomic jigsaw puzzle. Science 318, 415.

Marande, W., Lukeš, J., Burger, G., 2005. Unique mitochondrial genome structure in diplonemids, the sister group of kinetoplastids. Euk. Cell. 4, 1137-1146.

Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M., Penny, D., 2002. Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc. Natl Acad. Sci. USA 99, 12246-12251.

Meade, M.J., Proulex, G.C.R., Manoylov, K.M., Cahoon, A.B., 2020. Chloroplast mRNAs are 3' polyuridylylated in the green alga *Pithophora roettleri* (Cladophorales). J. Phycol. 56, 1124-1134.

Morse, D., Salois, P., Markovic, P., Woodland Hastings, J., 1995. A nuclear-encoded form II rubisco in dinoflagellates. Science 268, 1622-1624.

Nash, E.A., Barbrook, A.C., Edwards-Stuart, R.K., Bernhardt, K., Howe, C.J., Nisbet, R.E.R., 2007. Organization of the mitochondrial genome in the dinoflagellate *Amphidinium carterae*. Mol. Biol. Evol. 24, 1528-1536.

Nash, E.A., Nisbet, R.E.R., Barbrook, A.C., Howe, C.J., 2008. Dinoflagellates: a mitochondrial genome all at sea. Trends Genet. 24, 328-335.

Nelson, M.J., Green, B.R., 2005. Double hairpin elements and tandem repeats in the non-coding region of *Adenoides eludens* chloroplast gene minicircles. Gene 358, 102-110.

Nelson, M.J., Dang, Y., Filek, E., Zhang, Z., Yu, V.W.C., Ishida, K-I., Green B.R., 2007. Identification and transcription of tRNA genes in dinoflagellate plastid minicircles. Gene 392, 291-298.

Nimmo, I.C., Barbrook, A.C., Lassadi, I., Chen, J.I., Geisler, K., Smith, A.G., Aranda, M., Purton, S., Waller, R.F., Nisbet, R.E.R., Howe, C.J., 2019. Genetic transformation of the dinoflagellate chloroplast. eLife 8, e45292.

Palmer, J.D., Shields, C.R., 1984. Tripartite structure of the *Brassica campestris* mitochondrial genome. Nature 307, 437-440.

Pfannschmidt, T., Nilsson, A., Allen, J.F., 1999. Photosynthetic control of chloroplast gene expression. Nature 397, 625-628.

Raven, J.A., Beardall, J., 2017. Consequences of the genotypic loss of mitochondrial complex I in dinoflagellates and of phenotypic regulation of complex I content in other photosynthetic organisms. J. Exp. Bot., 68, 2683-2692.

Rosengarten, R.D., Sperling, E.A., Moreno, M.A., Leys, S.P., Dellaporta, S.L., 2008. The mitochondrial genome of the hexactinellid sponge *Aphrocallistes vastus*: evidence for programmed translational frameshifting. BMC Genomics 9, 33.

Sana, S., Hardouin, E.A., Paley, R., Zhang, T., Andreou, D., 2020. The complete mitochondrial genome of a parasite at the animal-fungal boundary. Parasites & Vectors 13, 81.

Schelkunov, M.I., Nuraliev, M.S., Logacheva, M.D., 2019. *Rhopalocnemis phalloides* has one of the most reduced and mutated plastid genomes known. PeerJ 7, e7500.

Schnepf, E., Elbrächter, M., 1999. Dinophyte chloroplasts and phylogeny – a review. Grana 38, 81-97.

Shao, R., Kirkness, E.F., Barker, S.C., 2009. The single mitochondrial chromosome typical of animals has evolved into 18 minichromosomes in the human body louse, *Pediculus humanus*. Genome Res. 19, 904-912.

Shao, R., Zhu, X-Q., Barker, S.C., Herd, K., 2012. Evolution of extensively fragmented mitochondrial genomes in the lice of humans. Genome Biol. 4:1088-1101.

Shoguchi, E., Shinzato, C., Hisata, K., Satoh, N., Mungpakdee, S., 2015. The large mitochondrial genome of *Symbiodinium minutum* reveals conserved noncoding sequences between dinoflagellates and apicomplexans. Genome Biol. Evol. 7, 2237-2244.

Signes, A., Fernandez-Vizarra, E., 2018. Assembly of mammalian oxidative phosphorylation complexes I-V and supercomplexes. Essays in Biochemistry 62, 255-270.

Slamovits, C.H., Saldarriaga, J.F., Larocque, A., Keeling, P.J., 2007. The highly reduced and fragmented mitochondrial genome of the early branching dinoflagellate *Oxyrrhis marina* shares characteristics with both apicomplexan and dinoflagellate mitochondrial genomes. J. Mol. Biol. 372, 356-368.

Sloan, D.B., Alverson, A.J., Chuckalovcak, J.P., Wu, M., McCauley, D.E., Palmer, J.D., Taylor, D.R., 2012. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. PLoS Biol. 10, e1001241.

Smith, D.R., Kayal, E., Yanagihara, A.A., Collins, A.G., Pirro, S., Keeling, P.J., 2011. First complete mitochondrial genome sequence from a box jellyfish reveals a highly fragmented linear architecture and insights into telomere evolution. Genome Biol. Evol. 4, 1-11.

Tan B.G., Gustafsson, C.M., Falkenberg, M., 2024. Mechanisms and regulation of human mitochondrial transcription. Nat. Rev. Mol. Cell. Biol. 25, 119-132.

Valach, M., Moreira, S., Faktorová, D., Lukeš, J., Burger, G., 2016. Post-transcriptional mending of gene sequences: looking under the hood of mitochondrial gene expression in diplonemids. RNA Biology, 13, 1204-1211.

van Kranken, J.G., Na, U., Winige, D.R., Rutter, J., 2015. Protein-mediated assembly of succinate dehydrogenase and its cofactors. Crit. Rev. Biochem. Mol. Biol. 50, 168-180.

Watanabe, K.I., Bessho, Y., Kawasaki, M., Hori, H., 1999. Mitochondrial genes are found on minicircle DNA molecules in the mesozoan animal *Dicyema*. J. Mol. Biol. 286, 645-650.

Watanabe, S., Fučíková, K., Lewis, L.A, Lewis, P.O., 2016. Hiding in plain sight: *Koshicola spirodelophila* gen. et sp. nov. (Chaetopeltidales, Chlorophyceae), a novel green alga associated with the aquatic angiosperm *Spirodela polyrhiza.* Am. J. Bot. 103-865-875.

Wei, D-D., Shao, R., Yuan, M-L., Dou, W., Barker, S.C., Wang, J-J., 2012. The multipartite mitochondrial genome of *Liposcelis bostrychophila*: insights into the evolution of mitochondrial genomes in bilateral animals. PLoS One, 7, e33973.

Weigle, K.A., Labrada, L.A., Lozano, C., Santrich, C., Barker, D.C., 2002. PCR-based diagnosis of acute and chronic cutaneous leishmaniasis caused by *Leishmania* (*Viannia*). J. Clin. Microbiol. 40, 601-606.

Wietrzynski, W., Traverso, E., Wollman, F-A., Wostrikoff, K., 2021.The state of oligomerization of Rubisco controls the rate of synthesis of the Rubisco large subunit in *Chlamydomonas reinhardtii*. Plant Cell, 33, 1706–1727.

Wilson, R.J.M., 2005. Parasite plastids: approaching the endgame. Biol. Rev. 80, 129-153.

Yahalomi, D., Haddas-Sasson, M., Rubinstein, N., Feldstein, T., Diamant, A., Huchon, D., 2017. The multipartite mitochondrial genome of *Enteromyxum leei* (Myxozoa); eight fast-evolving megacircles. Mol. Biol. Evol. 34, 1551-1556.

Yang, H., Liu, J., Wen, X., Lu, C., 2015. Molecular mechanism of photosystem I assembly in oxygenic organisms. Biochim. Biophys. Acta 1847, 838-848.

Yu, R., Sun, C., Zhong, Y., Liu, Y., Sanchez-Puerta, M.V., Mower, J.P., Zhou, R., 2022. The minicircular and extremely heteroplasmic mitogenome of the holoparasitic plant *Rhopalocnemis phalloides*. Curr. Biol. 32, 470–479.

Zhang, Z., Cavalier-Smith, T., Green, B.R., 2001. A family of selfish minicircular chromosomes with jumbled chloroplast gene fragments from a dinoflagellate. Mol. Biol. Evol. 18, 1558-1565.

Zhang, Z., Green, B.R., Cavalier-Smith, T., 1999. Single gene circles in dinoflagellate chloroplast genomes. Nature 400, 155-159.

Table 1 Retention of genes in chloroplasts of peridinin dinoflagellates and *Boodlea composita*. The table lists components of the photosynthetic electron transfer chain that are generally chloroplast encoded in all other photosynthetic lineages (Martin et al., 2002). In addition, 23S and 16S rRNA are chloroplast-encoded in dinoflagellates, but only the 16S gene has been identified so far in *Boodlea* chloroplasts. The dinoflagellate ribulose *bis*-phosphate carboxylase is a different form from other oxygenic photosynthetic organisms and is nuclear encoded (Morse et al., 1995). (Carbon dioxide fixation is a sink for photosynthetically generated electrons, so can also be thought of as part of the electron transfer chain.) Grey boxes = gene present; ? = gene reported present in one dinoflagellate strain to date. COCOA = minimal gene set predicted to be retained under COCOA hypothesis. [PsbI is not included in the predictions, as its role and importance differ between taxa (Dobákova et al., 2007]).

	PHOTOSYSTEM I (Psa)				**PHOTOSYSTEM II** (Psb)													
	A	B	C	J	A	B	C	D	E	F	H	I	J	K	L	N	T	Z
Dinoflagellates																	?	
Boodlea																		
COCOA																		

	CYT b_6f COMPLEX (Pet)				**ATP SYNTHASE** (Atp)						**RUBISCO** (Rbc)		
	A	B	D	G	A	B	E	F	H	I		L	
Dinoflagellates													
Boodlea													
COCOA													

Table 2A Summary of which components of the electron transfer complexes I-IV and ATP synthase (V) are encoded in mitochondrial genomes. Dark grey box = gene present, unshaded box = gene absent. Light grey boxes in left hand column indicate taxa with unfragmented (or partially fragmented) mitogenomes, for comparison with the taxa above (unshaded boxes). *Cycas taitungensis* mitogenome is included as an example of a larger mitogenome. (Jakobid flagellates have even more gene-rich mitogenomes (Burger et al., 2013), but are not included as an example here, for simplicity). COCOA = minimal gene set predicted to be retained under COCOA hypothesis.

	ELECTRON TRANSFER COMPLEXES																				
	I										II			III	IV			V				
	1	2	3	4	4 L	5	6	7	8	9	B	C	D	B	1	2	3	1	4	6	8	9
Cycas taitungensis																						
Lice																						
Arthropod ancestor																						
Rhopalocnemis phalloides																						
Malania oleifera																						
Rhodosorus marinus Chroodactylon ornatum																						
Florideophyceae Bangiophyceae																						
Diplonemids																						
Euglena																						
Enteromyxum leei																						
Alatina moseri																						
Cnidaria																						
Calcaronea (sponges)																						
Calcinea (sponges)																						
Hexactinellids (sponges)																						
Amoebidium parasiticum																						
Sphaerothecum destruens																						
COCOA																						

Table 2B Summary of which components of the cytochrome maturation machinery (CCM), ribosome (RPL and RPS) or protein translocation machinery (T) are encoded in mitochondrial genomes. Other genes that do not differ between fragmented and unfragmented genomes are not included, for simplicity. Shading as for Table 2A.

.

	CCM			RPL					RPS											T
	B	C	F	2	5	10	16	20	1	2	3	4	7	10	11	12	13	14	19	
Cycas taitungensis																				
Lice																				
Arthropod ancestor																				
Rhopalocnemis phalloides																				
Malania oleifera																				
Rhodosorus marinus Chroodactylon ornatum																				
Florideophyceae Bangiophyceae																				
Diplonemids																				
Euglena																				
Enteromyxum leei																				
Alatina moseri																				
Cnidaria																				
Calcaronea (sponges)																				
Calcinea (sponges)																				
Hexactinellids (sponges)																				
Amoebidium parasiticum																				
Sphaerothecum destruens																				
COCOA																				

Table 3 Summary of features of fragmented organelle genomes. For mitochondrial genomes, 'Reduced content' is by comparison to closely related mitochondrial genomes (see Table 2). *Dicyema* is not included, for lack of detailed information. N.R. = not reported

	Linear or circular	Common core region	Diverged sequences	polyU'd transcripts	Reduced content
Dinoflagellate chloroplast	Circular	Yes	Yes	Yes	Major reduction
Cladophorales chloroplast	Linear	No	Yes	Yes	Major reduction
Lice mitochondria	Circular	Yes	Yes	N.R.	Little or no reduction
Rhopalocnemis mitochondria	Circular	Yes	Yes	N.R.	Some reduction
Rhodosorus mitochondria	Circular	Yes	Yes	N.R.	Reduction
Diplonemid mitochondria	Circular	Yes	Yes	Yes	Little or no reduction
Enteromyxum mitochondria	Circular	Yes	Yes	N.R.	Not clear
Box jellyfish mitochondria	Linear	No	Yes	N.R.	No
Sponge mitochondria	Linear	No	Yes	N.R.	No
Amoebidium mitochondria	Linear	No	Yes	N.R.	No