Review

Organization and expression of organellar genomes

Adrian C. Barbrook, Christopher J. Howe*, Davy P. Kurniawan and Sarah J. Tarr

Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, UK

Protist mitochondrial genomes show a very wide range of gene content, ranging from three genes for respiratory chain components in Apicomplexa and dinoflagellates to nearly 100 genes in *Reclinomonas americana*. In many organisms the rRNA genes are fragmented, although still functional. Some protist mitochondria encode a full set of tRNAs, while others rely on imported molecules. There is similarly a wide variation in mitochondrial genome organization, even among closely related groups. Mitochondrial gene expression and control are generally poorly characterized. Transcription probably relies on a ‘viral-type’ RNA polymerase, although a ‘bacterial-type’ enzyme may be involved in some cases. Transcripts are heavily edited in many lineages. The chloroplast genome generally shows less variation in gene content and organization, although greatly reduced genomes are found in dinoflagellate algae and non-photosynthetic organisms. Genes in the former are located on small plasmids in contrast to the larger molecules found elsewhere. Control of gene expression in chloroplasts involves transcriptional and post-transcriptional regulation. Redox poise and the ATP/ADP ratio are likely to be important determinants. Some protists have an additional extranuclear genome, the nucleomorph, which is a remnant nucleus. Nucleomorphs of two separate lineages have a number of features in common.

Keywords: kinetoplast; maxi-circle; mini-circle; plastid; apicoplast; nucleomorph

1. INTRODUCTION

The mitochondrion originated from an α-proteobacterial endosymbiont acquired by a host that was unable to carry out aerobic respiration, but whose nature is otherwise controversial (Embley & Martin 2006). The majority of the symbiont’s genes were either transferred to the nucleus or lost completely. A number of eukaryotic lineages have secondarily adopted an anaerobic lifestyle, with consequent modification of the mitochondrion to form hydrogenosomes or mitosomes, as discussed elsewhere in this issue. This modification has been accompanied by partial or complete loss of the mitochondrial genome. The chloroplast originated from the acquisition of an oxygenic photosynthetic eubacterial endosymbiont by a non-photosynthetic host, with subsequent reduction of the symbiont genome. Although it is frequently supposed that a single primary endosymbiosis gave rise to all chloroplasts, the evidence for this has been questioned (Larkum et al. 2007), and there is increasing evidence that the photosynthetic chromatophore of the amoeba *Paulinella* may represent a clearly independent chloroplast origin (Nakayama & Ishida 2009). It is recognized that multiple ‘secondary’ endosymbiotic acquisitions of photosynthetic eukaryotes have occurred. In some secondary endosymbiotic lineages, the nucleus of the intermediate photosynthetic eukaryote persists as a ‘nucleomorph’ between two of the four membranes surrounding the chloroplast.

Some lineages have lost photosynthetic function, with consequent reduction of the chloroplast genome, analogous to the reduction of the mitochondrial genome in anaerobic eukaryotes. It is not clear whether any formerly photosynthetic eukaryotes have completely lost a chloroplast genome, or indeed a chloroplast compartment (Barbrook et al. 2006). There is tremendous variation among mitochondrial genomes, less among chloroplast genomes, which we illustrate in the following discussion. For the purposes of this discussion, we include macroalgal relatives of the microalgae. For simplicity, we use the term chloroplast to cover all photosynthetic organelles, whatever their pigment type, as well as non-photosynthetic organelles derived from them during evolution or development.

2. GENE CONTENT AND ORGANIZATION OF MITOCHONDRIAL GENOMES

Eukaryotic microbes exhibit highly divergent mitochondrial genome structure and organization. Variable features include genome size (ranging from nothing in the case of mitosomes, as discussed elsewhere in this volume, to thousands of kilobase pairs in the kinetoplastids), gene complement, the presence of a single mitochondrial DNA (mtDNA) unit or
multiple chromosomes, the morphology of the mtDNA (whether existing as discrete units or concatenated in multiple copies) and linear or circular organization (table 1). Mitochondrial genomes of fungi such as Saccharomyces cerevisiae have been well characterized over many years, and will not be discussed here.

(a) Protein-coding genes

Proteins encoded in mitochondrial genomes are typically involved in translation and oxidative phosphorylation (http://www.ncbi.nlm.nih.gov/Genomes/). A core set of mitochondrial genes, encoding subunits of the ribosome, NADH dehydrogenase, succinate dehydrogenase, cytochrome oxidase and ATP synthase, as well as cytochrome b is present to differing extents in different species (Gray et al. 2004). In some cases, proteins involved in protein translation and maturation, RNA processing and transcription are also mitochondrially encoded. Gene content and organization have been used to help establish evolutionary relationships among organisms (e.g. Sankoff et al. 1992). The mitochondrial genome of the jakobid protozoan Reclinomonas americana carries 97 genes encoding proteins and stable RNAs. This is the most gene-rich and least derived mtDNA studied to date (Lang et al. 1997). Eighteen mitochondrial genes in R. americana have not been reported from any other mtDNA. Some are involved in mitochondrial electron transport and respiration but there are genes for proteins involved in other functions, such as tuP (a translation factor) and secY (involved in protein translocation). The observation that the R. americana mtDNA carries rpoA–D encoding the α, β, β’ and σ subunits of prokaryotic RNA polymerases is particularly notable, since other mitochondria are believed to use a viral T3/T7-type single subunit RNA polymerase. Consequently, it has been suggested that the R. americana mitochondrial genome is more reminiscent than others of ancestral mitochondria (Lang et al. 1997).

In terms of gene content, the Apicomplexa (including the malaria parasite Plasmodium and the cattle pathogen Theileria) lie at the opposite end of the scale to R. americana. At around 6 kbp, their mtDNAs represent the smallest discovered, carry the fewest genes and lack introns, representing perhaps a minimal gene complement for mitochondrial genomes. They carry three protein coding genes, for subunits I and III of cytochrome c oxidase (cox1 and cox3, respectively) and cytochrome b (cob) (Wilson & Williamson 1997).

The dinoflagellate algae have a particularly unusual mitochondrial genome. Like their sister group, the Apicomplexa, the mitochondrial genome contains cox1, cox3 and cob, but also contains a large number of (presumably) non-functional fragments of them in a wide range of configurations separated by extensive regions of repetitive non-coding DNA (Nash et al. 2008; Kamikawa et al. 2009). In the early-branched dinoflagellate Oxyrrhis marina, cob and cox3 are encoded and expressed as a fusion protein (Slamovits et al. 2007).

(b) RNA genes

Large and small subunit rRNAs (LSU and SSU respectively) are encoded in essentially all mtDNAs studied to date, although in many species the rRNA genes are fragmented. In a few cases (such as SSU rRNA in dinoflagellate mitochondria), it has not been possible to identify rRNA genes conclusively. It may be that they are present, but highly divergent and fragmented (Nash et al. 2008). In some organisms, a gene rrn5 for 5S rRNA is also found. The sporadic distribution of rrn5 suggests it was lost multiple times from different lineages (Gray et al. 2004).

Fragmentation of rRNA genes has occurred in many groups, most markedly in Apicomplexa, dinoflagellates and many green algal (e.g. Chlamydomonas) lineages. In these instances, rRNA segments are encoded separately and the transcripts associate to form functional ribosomes. In some cases, such as Plasmodium falciparum and Chlamydomonas eugametos, the order of rDNA fragments in the genome is different from the order of their homologues in the Escherichia coli 16S and 23S rRNA. In these organisms, LSU and SSU coding fragments are interspersed with each other as well as other coding sequences (Feagin et al. 1992; Denovan-Wright & Lee 1994; Denovan-Wright et al. 1998). The order of rDNA fragments varies between Chlamydomonas species. In Chlamydomonas reinhardtii, LSU fragments are encoded in the same order as in E. coli 16S rRNA, whereas the LSU fragments are scrambled and in a different order from C. eugametos (Denovan-Wright & Lee 1994). Fragmentation of rRNA genes has occurred to a lesser degree in the ciliates, where LSU and SSU rDNA each split into two separate coding regions. However, in Tetrahymena pyriformis, the LSU rRNA gene is represented twice with sequence differences between the two forms (Burger et al. 2000).

The number of tRNAs encoded in protist mitochondrial genomes varies widely. The mitochondrial genome of the ichthyosporean (an independent group located between animals and fungi in the opisthokonts) Ameobidium parasiticum is organized into many 0.3–8.3 kbp linear chromosomes (Burger et al. 2003) and encodes at least 25 tRNAs in multiple copies. Conversely, tRNA genes are absent from kinetoplastid mitochondrial genomes, although they can be thousands of kilobases in size. Studies in trypanosomes and the apicomplexan Toxoplasma gondii demonstrate that tRNAs are imported from the cytosol (Crausaz Esseiva et al. 2004). At least two pathways exist for tRNA uptake, one of which has elements in common with protein import (Alfonzo & Söll 2009). Obtaining a formylmethionyl-tRNA (fMet-tRNA) for translation initiation is potentially problematic. Trypanosome mitochondria are believed to take up and formylate cytosolic elongator Met-tRNA for this purpose (Tan et al. 2002), whereas it has been suggested that in Apicomplexa the initiator fMet-tRNA may be supplied by the apicoplast (the relic secondary chloroplast maintained within the parasite) (Howe & Purton 2007). In Chlamydomonas, the mitochondrial genome also does not encode an initiator fMet-tRNA. A recent study failed to detect chloroplast initiator fMet-tRNAs in the Chlamydomonas Phil. Trans. R. Soc. B (2010)
<table>
<thead>
<tr>
<th>Organism</th>
<th>Size (bp)</th>
<th>% Coding</th>
<th>Total Protein Coding Genes</th>
<th>Ox. Phos. Genes</th>
<th>Gene Expression Genes</th>
<th>Other Protein Coding Genes</th>
<th>RNA Genes (rRNA/tRNA)</th>
<th>Topology</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Spizellomyces punctatus</em></td>
<td>Chr 1: 58,830</td>
<td>Chr 1: 39</td>
<td>Chr 1: 31</td>
<td>Chr 1: 13</td>
<td>none</td>
<td>Chr 1: 18 ORFs</td>
<td>Chr 1: 10</td>
<td>Circular</td>
</tr>
<tr>
<td>(Chytridiomycota)</td>
<td>Chr 2: 1,381</td>
<td>Chr 2: 16</td>
<td>Chr 2: 1</td>
<td>Chr 2: 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chr 3: 1,136</td>
<td>Chr 3: 0</td>
<td>Chr 3: 0</td>
<td>Chr 3: 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Acanthamoeba castellanii</em></td>
<td>41,591</td>
<td>77</td>
<td>40</td>
<td>16</td>
<td>16</td>
<td>8 ORFs</td>
<td>18</td>
<td>Circular</td>
</tr>
<tr>
<td>(Amoebozoa)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Dictyostelium discoideum</em></td>
<td>55,564</td>
<td>73</td>
<td>42</td>
<td>17</td>
<td>15</td>
<td>9 ORFs plus endonuclease</td>
<td>21</td>
<td>Circular</td>
</tr>
<tr>
<td>(Amoebozoa)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Chlamydomonas reinhardtii</em></td>
<td>15,758</td>
<td>58</td>
<td>8</td>
<td>7</td>
<td>none</td>
<td>rrl</td>
<td>17</td>
<td>Linear</td>
</tr>
<tr>
<td>(Chlorophyta)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Chondrus crispus</em> (Rhodophyta)</td>
<td>25,836</td>
<td>71</td>
<td>29</td>
<td>17</td>
<td>5</td>
<td>6 ORFs plus ymf16</td>
<td>28</td>
<td>Circular</td>
</tr>
<tr>
<td><em>Porphyra purpurea</em> (Rhodophyta)</td>
<td>36,753</td>
<td>70</td>
<td>31</td>
<td>17</td>
<td>4</td>
<td>7 ORFs plus ymf16, dpo, rrl</td>
<td>26</td>
<td>Circular</td>
</tr>
<tr>
<td><em>Catenaria roenbergensis</em></td>
<td>43,159</td>
<td>82</td>
<td>34</td>
<td>18</td>
<td>12</td>
<td>4 ORFs</td>
<td>24</td>
<td>Circular</td>
</tr>
<tr>
<td>(Chromista)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Fucus vesiculosus</em> (Chromista)</td>
<td>36,392</td>
<td>77</td>
<td>38a</td>
<td>17</td>
<td>17</td>
<td>3 ORFs plus tatC</td>
<td>28</td>
<td>Circular</td>
</tr>
<tr>
<td><em>Rhodomonas salina</em> (Crytophyta)</td>
<td>48,063</td>
<td>69</td>
<td>44</td>
<td>22</td>
<td>14</td>
<td>7 ORFs plus ymf16</td>
<td>29</td>
<td>Circular</td>
</tr>
<tr>
<td><em>Plasmodium gallinaceum</em></td>
<td>6003</td>
<td>55</td>
<td>3</td>
<td>3</td>
<td>none</td>
<td>fragmented SSU and LSU rRNA</td>
<td>line and circular</td>
<td></td>
</tr>
<tr>
<td>(Alveolata)</td>
<td>n.d.</td>
<td>n.d.</td>
<td>3 identifiedb</td>
<td>n.d.</td>
<td>n.d.</td>
<td>fragmented LSU rRNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Tetrahymena pyriformis</em></td>
<td>47,296</td>
<td>80</td>
<td>44</td>
<td>15</td>
<td>9</td>
<td>19 ORFs plus yejR</td>
<td>14</td>
<td>Circular</td>
</tr>
<tr>
<td>(Alveolata)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Reclinomonas americana</em></td>
<td>69,034</td>
<td>80</td>
<td>67</td>
<td>25</td>
<td>32</td>
<td>10</td>
<td>30</td>
<td>Circular</td>
</tr>
<tr>
<td>(Jakobida)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Malawimonas pyriformis</em></td>
<td>47,328</td>
<td>67</td>
<td>49</td>
<td>17</td>
<td>22</td>
<td>6 ORFs plus yejR, yejU, yejY</td>
<td>40</td>
<td>Circular</td>
</tr>
<tr>
<td>(Malawimonada)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Naegleria gruberi</em> (Heterolobosea)</td>
<td>49,843</td>
<td>81</td>
<td>46</td>
<td>21</td>
<td>17</td>
<td>4 ORFs, yejR, yejU, ymf16, cox11</td>
<td>23</td>
<td>Circular</td>
</tr>
</tbody>
</table>

aAlso encodes one pseudogene.

bGenes exist in various genetic contexts (fragmented/complete, repeated/alone etc.).
Mitochondrion and suggested that the initiator fMet-tRNA is derived from cytosolic elongator Met-tRNA, as with trypanosomes (Vinogradova et al. 2009).

(c) Genome organization

The organization, morphology and topology of the mtDNA is highly diverse among protists. Although the majority of mtDNAs are topologically circular, linear monomers and linear tandem arrays are also found. Genome organization can differ widely among related protist species. For example (with the exception of *Physarum polycephalum*), mtDNA gene content is largely conserved between members of the Amoebozoa (which include the ‘solitary’ amoeba *Acanthamoeba castellanii* as well as the ‘social’ slime molds *Polysphondylium pallidum* and *Dictyostelium* sp.). They share very similar mtDNA sizes, AT richness and gene content, and the genomes are compact (for example, 93.2% of the *A. castellanii* mtDNA comprises coding sequences) (Burger et al. 1995). They also share a chimeric *cox1* and *cox2* gene and all genes are transcribed from the same strand (Burger et al. 1995; Heidel & Glöckner 2008). However, there are large variations in gene order. While only a single segmental rearrangement occurs between the sister groups that make up the social amoeboae (*Dictyostelium discoideum* and *Dictyostelium citrinum*, and *P. pallidum* and *Dictyostelium fasciculatum*), conservation of gene order in *A. castellanii* is limited to pairs of genes (Heidel & Glöckner 2008).

There is a wide variation in mitochondrial genome size and structure among closely related green algal species. For example, *C. eugametos* and *C. reinhardtii* have similar gene contents, but the order of genes is entirely different. The *C. eugametos* mtDNA contains introns and displays a more dispersed distribution of rRNA fragments (Denovan-Wright et al. 1998). Furthermore, whereas the mtDNA of *C. eugametos* is circular mapping with all genes located on the same strand, in *C. reinhardtii* it exists as linear monomers with genes on both strands. The related green alga *Polytomella parva* also exhibits similar gene content and rDNA organization to *Chlamydomonas*, but its genome exists as two linear DNA molecules—one of 13.5 kbp, which contains the majority of coding DNA, and one of 3.5 kbp, which contains only *nad6* (Fan & Lee 2002). By contrast, mitochondrial genome size and structure are largely conserved among the red algae (Gray et al. 2004). Within individual phyla of the Chromista, mitochondrial genomes are generally reasonably well conserved. For example, the mitochondrial genomes of Phaeophyceae (brown algae) are similar in gene order and content. Differences are primarily due to variation in the presence and length of intergenic regions and introns (Oudot-Le Secq et al. 2006). However, significant diversity exists between phyla, for example in the presence of repeated sequences and introns (Puerta et al. 2004; Kim et al. 2008).

Mitochondrial DNA structure also varies widely within other protist groups. In the Apicomplexa, for example, *Plasmodium* mtDNAs are arranged in linear tandem repeats (Wilson & Williamson 1997), whereas the mtDNA in *Theileria* exists as linear monomers (Kairo et al. 1994). Mitochondrial genome size is largely independent of gene content and organization, and a particularly striking example of this is found in one of the most unusual mitochondrial genomes, in kinetoplastids. All members of the Euglenozoa (i.e. kinetoplastids, diplomemids and euglenids) contain large amounts of mtDNA. In many kinetoplastid flagellates, mitochondrial DNA is present as a compact disc in a region of the mitochondrion known as the kinetoplast (and is often referred to as kinetoplast DNA, or kDNA). Electron microscopic imaging has shown that it exists as a network of large and small circular DNAs known as maxi-circles and mini-circles (Lukeš et al. 2002). Maxi-circles encode proteins for mitochondrial function and exist at around 12 copies per mitochondrion. They are usually 20–40 kbp in size, although coding DNA is always localized within a 17 kbp region known as the ‘conserved region’. The remaining ‘variable’ region contains repeated sequences, some of which are involved in maxi-circle replication (Shapiro & Englund 1995).

Mini-circles are present in thousands of copies per mitochondrion. They encode guide RNAs (gRNAs) that direct the insertion/deletion of Us in RNAs to yield complete coding mRNAs. Some gRNAs are also encoded on maxi-circles. Mini-circles are smaller (0.5–10 kbp) than maxi-circles and contain conserved regions involved in the initiation of their replication and compaction into the kDNA network (Shapiro & Englund 1995). The kDNA is organized differently in different species. In trypanosomatids (which include *Trypanosoma* and *Leishmania*), it exists as a catenated network of maxi- and mini-circles found in a distinct periflagellar region of the mitochondrion, while the kDNA of *Bodo caudatus* and *Cryptobia helicus* are uncatenated and dispersed through the mitochondrion (Hajduk et al. 1986; Lukeš et al. 1998). *Dimastigella mimosa*, *Dimastigella trypaniformis* and *Cruzella marina* display multiple bundles of typically monomeric mini-circles throughout their mitochondria, while in *Trypanoplasma borreli* the mini-circles are linked in tandem to form large circular molecules termed mega-circles (for a detailed review, see Lukeš et al. 2002).

The diplomemids have very different mtDNA structure, although they are sister group to the kinetoplastids. *Diplonema papillatum* appears to have two types of circular mitochondrial chromosomes, one of 6 kbp and one of 7 kbp, both apparently existing in relaxed circular conformation. Fluorescence microscopy studies have shown that the mtDNA is not localized to a distinct region of the mitochondrion as in trypanosomatids, but is distributed throughout the mitochondrion. Furthermore, analysis of *cox1* coding sequence showed that the gene appears to exist in approximately 250 bp fragments distributed between the two types of chromosomes (Marande et al. 2005).

3. MITOCHONDRIAL GENE EXPRESSION AND CONTROL

Relatively little is known about mitochondrial gene expression in eukaryotic micro-organisms (other than in some well-studied species such as *S. cerevisiae*)
and even less about its control. In some lineages, such as the Apicomplexa, studies have elucidated the size and processing of mitochondrial transcripts. There has also been considerable focus on editing in a number of lineages, notably the kinetoplastids. However, advances in our knowledge of mitochondrial gene expression in protists have been rather few in recent years.

Expression of mitochondrial genes appears to be generally mediated by a single subunit T3/T7 phage-like RNA polymerase, which was recruited relatively early in the evolution of the organelle (Shutt & Gray 2006a). Sequences encoding such an enzyme have been specifically amplified from many protists (Cermakian et al. 1996; Li et al. 2001) and can typically be found in sequenced protist nuclear genomes. The corresponding activity has been experimentally verified in some lineages (Grams et al. 2002; Miller et al. 2006). However, as noted above, the mitochondrial genomes of R. americana and other jakobids encode a bacterial-type RNA polymerase (Gray et al. 2004). This is presumed to be involved in mitochondrial gene expression, although the presence of a viral-type enzyme in addition cannot be excluded. In the stramenopile Pylaeidella littoralis a T7-like polymerase is encoded on the mitochondrial genome, but sequences characteristic of promoters of a bacterial-type polymerase are also present (Oudot-Le Secq et al. 2001). It is therefore possible that in certain protists multiple functional polymerases exist within mitochondria. This would be analogous to the presence of multiple polymerases in the chloroplasts of higher plants (Smith & Purton 2002).

In animals and fungi, mitochondrial factors accessory to the core polymerase have been shown to be important for expression of mitochondrial genes (Asin-Cayuela & Gustafsson 2007). Outside the opisthokonts, putative homologues for such transcription factors have been found in amoebozoans and trypanosomatids (Shutt & Gray 2006b). However, it remains to be established whether the phage-type polymerases acting in the mitochondria of other protists require transcription factors for effective mitochondrial gene expression, perhaps unrelated to those found in animals and yeasts.

The identity of promoter sequences has been suggested for a handful of protists (Wolff & Kück 1996; Richard et al. 1998; Oudot-Le Secq et al. 2001). The viral-type promoters are typically proposed to be nonanucleotide sequences close to the transcription initiation site, with little or no sequence similarity to promoter sequences in yeasts, humans and higher plants. Definitive identification of promoters in protists requires the development of mitochondrial transformation systems. Encouragingly, such a system has been reported for C. reinhardtii (Remacle et al. 2006).

Transcript maps have been generated for various protist mitochondria, including members of the Amoebozoa, Chlorophyta (green algae), Rhodophyta (red algae), Chromista, Apicomplexa and Ciliophora (Gray & Boer 1988; Jones et al. 1990; Wolff & Kück 1996; Richard et al. 1998; Edqvist et al. 2000; Rehkof et al. 2000). Taken with gene organization data, these maps are consistent with a model of transcription in which a few transcription initiation sites in each genome give rise to large precursor RNAs that are processed to much smaller mono- or polycistronic mRNAs. Such endonucleolytic processing of larger precursors would appear to be highly specific, since mRNAs are frequently generated with very little or no 5′ untranslated region. In some cases, and similarly to humans, the excision of tRNA sequences appears to generate the mature mRNAs (Wolff & Kück 1996). In Chondrus crispus, there is evidence for specific processing within tRNA sequences to generate the mature flanking cistron ends, as well as alternative processing to generate functional tRNAs (Richard et al. 1999).

The model of transcription described above is by no means universal. The kinetoplastids have a similar system of transcription for maxi-circle non-gRNA genes. However, while the gRNAs are transcribed by the same polymerase (Hashimi et al. 2009), there are numerous distinct transcriptional units present on the mini-circles and maxi-circles of the kDNA for their expression (Lukeš et al. 2005). Diplonemid circular chromosomes carrying different gene fragments also require at least one promoter element each, implying the existence of more than 100 transcription units across the set (Marande & Burger 2007). Given the large spacers between genes, and the indication from expressed sequence tags (ESTs) of predominantly monocistronic transcripts, dinoflagellate mitochondria must also employ a large number of separate promoters (Nash et al. 2008). Similar systems are also likely to be used in the less compact and more disperse mitochondrial genomes such as the alga Pseudendoconium akinetum (Pombert et al. 2004), as with land plants, whose mitochondrial genomes also contain a large number of separate transcriptional units.

A variety of different forms of editing occur in different protist lineages (table 2). Editing has been most closely studied in kinetoplastids (Lukeš et al. 2005). In this system, mRNA editing occurs via the insertion or deletion of U nucleotides. Determination of the positions of editing is carried out via gRNAs that are encoded principally on the kDNA mini-circles. Editing of mitochondrial transcripts also appears to be prevalent within the myxomycetes of the Amoebozoa. The most complex form of mRNA, tRNA and tRNA editing has been elucidated in Physarum polycephalum. In P. polycephalum, a variety of single and dinucleotide insertions take place as well as C to U substitutions (Horton & Landweber 2002). Editing takes place simultaneously with transcription in P. polycephalum and may involve the RNA polymerase itself, whereas kinetoplast editing is post-transcriptional (Byrne & Gutt 2004; Miller & Miller 2008). Editing of tRNAs has also been documented for organisms from the Amoebozoa, Chytridiomycota and Jakobida (Gray et al. 2004). Unusually, in trypanosomatids, the anticodon of cytosolic tryptophanyl-tRNA is edited following import into the mitochondrion, to allow recognition of mitochondrial UGA tryptophan codons (Alfonzo et al. 1999). An apparently unrelated mRNA and rRNA substitutional editing system occurs in dinoflagellates (Nash et al. 2008). However, the mechanism is unknown.
Table 2. Distribution of mitochondrial RNA editing in protist microbes. Arrows indicate the direction of substitutional editing events. Editing has also been characterized in further members of the Dinophyceae and Euglenozoa. However, the editing is of the same RNAs and is essentially the same type in these cases. In many organisms, the full extent of editing is not determined and may extend to further classes of RNA.

<table>
<thead>
<tr>
<th>organism</th>
<th>phylum</th>
<th>kingdom</th>
<th>class of RNA edited</th>
<th>types of editing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physarum polycephalum</td>
<td>Mycetozoa</td>
<td>Amoebozoa</td>
<td>mRNA, rRNA, tRNA</td>
<td>C/U/dinucleotide insertion, C→U</td>
</tr>
<tr>
<td>Didymium nigripes</td>
<td>Mycetozoa</td>
<td>Amoebozoa</td>
<td>mRNA, tRNA</td>
<td>C/U/dinucleotide insertion, C→U</td>
</tr>
<tr>
<td>Stemonitis flavogena</td>
<td>Mycetozoa</td>
<td>Amoebozoa</td>
<td>mRNA</td>
<td>C/U/dinucleotide insertion</td>
</tr>
<tr>
<td>Arcyria cinerea</td>
<td>Mycetozoa</td>
<td>Amoebozoa</td>
<td>mRNA</td>
<td>U insertion, C→U</td>
</tr>
<tr>
<td>Clastoderma debaryanum</td>
<td>Mycetozoa</td>
<td>Amoebozoa</td>
<td>mRNA</td>
<td>U insertion</td>
</tr>
<tr>
<td>Dictyostelium discoideum</td>
<td>Mycetozoa</td>
<td>Amoebozoa</td>
<td>rRNA</td>
<td>C→U</td>
</tr>
<tr>
<td>Acanthamoeba castellani</td>
<td>Centramoebida</td>
<td>Amoebozoa</td>
<td>tRNA</td>
<td>U→A, U→G, A→G</td>
</tr>
<tr>
<td>Spezellomyces punctatus</td>
<td>Chytridiomycota</td>
<td>Opisthokonta</td>
<td>tRNA</td>
<td>A→G, U→G, U→A, C→A</td>
</tr>
<tr>
<td>Alexandrium catenella</td>
<td>Dinophyceae</td>
<td>Chromalveolata</td>
<td>mRNA, rRNA</td>
<td>A→G, U→C, C→U, G→C, G→A</td>
</tr>
<tr>
<td>Cryptophyceum cohnii</td>
<td>Dinophyceae</td>
<td>Chromalveolata</td>
<td>mRNA</td>
<td>A→G, U→C, C→U, G→C</td>
</tr>
<tr>
<td>Procteocentrum minimum</td>
<td>Dinophyceae</td>
<td>Chromalveolata</td>
<td>mRNA</td>
<td>A→G, U→C, C→U, G→C, A→U</td>
</tr>
<tr>
<td>Pfiesteria piscicida</td>
<td>Dinophyceae</td>
<td>Chromalveolata</td>
<td>mRNA</td>
<td>A→G, U→C, C→U, G→C, A→U</td>
</tr>
<tr>
<td>Trypanosoma brucei</td>
<td>Euglenozoa</td>
<td>Excavata</td>
<td>mRNA, tRNA</td>
<td>U insertion/deletion, C→U</td>
</tr>
<tr>
<td>Bodo caudatus</td>
<td>Euglenozoa</td>
<td>Excavata</td>
<td>mRNA</td>
<td>U insertion/deletion</td>
</tr>
</tbody>
</table>

*Plasmoidal slime mould.
*Cellular slime mould.
*Kinetoplast.

Other types of post-transcriptional modification also occur. Mitochondrial transcripts have been found to be polyadenylated in Apicomplexa, dinoflagellates, kinetoplastids (maxi-circle mRNAs) and diplomids (Gray et al. 2004; Luček et al. 2005; Marande & Burger 2007; Nash et al. 2008). In Apicomplexa and dinoflagellates, the poly(A) tails are relatively short (less than 22 nucleotides) and of uncertain function. In kinetoplastids, mitochondrial transcripts have either short (approx. 20–25 nucleotides) or long (approx. 120–250 nucleotides) poly(A) tails (Etheridge et al. 2008). The short tails are added first and are required for maintenance of transcripts undergoing editing. The longer tails are added after the completion of editing and contain a significant proportion of uridyl nucleotides. Polyuridylation of transcripts has also been observed for certain transcripts within the myxomycetes, as well as for gRNAs and rRNAs in trypanosomes (Adler et al. 1991; Horton & Landweber 2000). Furthermore, the presence of a short 5′ poly(U) cap has been demonstrated for transcripts of O. marina (Slamovits et al. 2007). Trans-splicing to generate mature mRNAs has been discovered in diplomids and dinoflagellates (Marande & Burger 2007; Nash et al. 2008).

Control of gene expression in protists is probably post-transcriptional in most cases, given the existence of relatively few promoters and the synthesis of large precursor polycistronic transcripts. Results in *Trypanosoma brucei* support this model (Michelotti et al. 1992; Luček et al. 2005).

4. GENE CONTENT AND ORGANIZATION OF CHLOROPLAST GENOMES

Although there is diversity in the content and organization of chloroplast genomes, there is less variation than among mitochondrial genomes. A ‘core’ of genes is present in the chloroplast genome of essentially all photosynthetic organisms, encoding subunits of the two photosystems, the cytochrome $b_6/f$ complex linking them and the ATP synthase, together with the large subunit of ribulose bis-phosphate carboxylase, components of both ribosomal subunits, a bacterial-type RNA polymerase, rRNAs and tRNAs (table 3).

The genome of green algal chloroplasts typically ranges between 100 and 200 kbp, but with numerous exceptions. Some *Acetabularia* species are reported to have chloroplast genomes up to 1.5 Mbp in size (Simpson & Stern 2002) while the smallest conventional chloroplast genome of a photosynthetic organism belongs to *Ostreococcus tauri*, with only 86 genes (excluding duplications) closely packed into 72 kbp (Robbens et al. 2007). Diversity is also observed within genera. For example, *Chlamydomonas moewusii* (chloroplast genome size 292 kbp) and *Chlamydomonas pitchmanii* (187 kbp) owe their difference to two intergenic insertions. *Chlamydomonas*
reinhardtii (203 kbp) and Chlamydomonas gelatinoa (285 kbp) differ in intergenic spacers and gene order. The difference in the latter is attributable to several inversions and an expansion/contraction of the inverted repeat (Boudreau & Turmel 1996). Sequenced red algal chloroplast genomes range between 150 and 192 kbp and generally contain more genes, up to 251, than the 103–135 chloroplast genes of green algae (http://www.ncbi.nlm.nih.gov/Genomes; Hagopian et al. 2004). The glaucophyte Cyanophora paradoxa has a chloroplast of the same pigment type as red algae, but surrounded by a remnant peptidoglycan wall. However, it has a smaller chloroplast genome than red algae, of 136 kbp (Hagopian et al. 2004). Sequenced stramenopile and other chloroplast genomes are typically smaller than red algal ones, at 105–160 kbp, and with a similar gene content to green chloroplasts (http://www.ncbi.nlm.nih.gov/Genomes/).

Chloroplast genomes are typically represented as circular, with an inverted repeat region (containing rRNA and some other genes) dividing the chloroplast genome into large and small single copy regions. Whether the genome is really topologically circular in vivo has been reviewed elsewhere (Bendich 2004). There are some exceptions to this organization. For example, although the red alga Porphyra yezoensis was reported to have rRNA genes organized in inverted repeats (Shivji 1991), sequence analysis indicates that they are in direct configuration (NCBI accession NC_007932). This could represent a strain difference, but the rRNA genes are also in direct configuration in Porphyra purpurea, although they are not repeated in the red algae Gracilaria tenusistipitata and Cyanidium caldarium (Hagopian et al. 2004). In some Euglena gracilis strains, there are three consecutive rRNA repeats and one additional partial one (Hallick et al. 1993).

Some algal chloroplasts have lost photosynthetic capability, with a reduction and compaction of their genome, as demonstrated in the parasitic, non-photosynthetic green alga Helicosporidium. This has a 37.5 kbp chloroplast genome lacking all genes encoding the photosynthetic machinery, with only 5.1 per cent non-coding DNA, small intergenic spaces and encoding only the minimal complement of tRNAs required (de Koning & Keeling 2006). The non-photosynthetic Apicomplexa generally possess a circular genome of approximately 35 kbp that has inverted repeats and is clearly of chloroplast origin (Gardner et al. 1991; Howe 1992) but no longer encodes any of the proteins directly involved in photosynthesis (Wilson & Williamson 1997). A photosynthetic apicomplexan has also been described, but its chloroplast genome has not yet been fully characterized (Moore et al. 2008). The sister group to the Apicomplexa, the dinoflagellates, have a unique chloroplast genome organization where the single large chloroplast genome is replaced by multiple, small plasmid-like DNA molecules termed ‘minicircles’ (Zhang et al. 1999; Barbrook & Howe 2000). They range between 2 and 4 kbp in size and generally carry one gene per minicircle, although minicircles carrying more than one gene have been reported. The minicircles contain a conserved core region, probably responsible for replication and transcription (Howe et al. 2008; Nisbet et al. 2008). They have a much reduced gene content, with most of the conventional core set of chloroplast genes relocated to the nucleus (Howe et al. 2008).

### Table 3. Core genes (except tRNAs) present in essentially all chloroplast genomes from photosynthetic organisms. Those underlined are present on the mini-circular chloroplast genome of dinoflagellates (adapted from Howe et al. 2008).

<table>
<thead>
<tr>
<th>photosystem I</th>
<th>photosystem II</th>
<th>cytochrome b/f complex</th>
<th>ATP synthase</th>
<th>ribosomal protein large subunit</th>
<th>ribosomal protein small subunit</th>
<th>RNA polymerase</th>
<th>hypothetical protein</th>
<th>Rubisco rRNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>psaA</td>
<td>psbA</td>
<td>psaA</td>
<td>atpA</td>
<td>rpl2</td>
<td>rps2</td>
<td>rpoA</td>
<td>ycf4</td>
<td>rbcL</td>
</tr>
<tr>
<td>psaB</td>
<td>psbB</td>
<td>psaB</td>
<td>atpB</td>
<td>rpl14</td>
<td>rps3</td>
<td>rpoB</td>
<td></td>
<td>16S</td>
</tr>
<tr>
<td>psaC</td>
<td>psbC</td>
<td>psaC</td>
<td>atpE</td>
<td>rpl16</td>
<td>rps4</td>
<td>rpoC1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>psaF</td>
<td>psbD</td>
<td>psaD</td>
<td>atpF</td>
<td>rpl20</td>
<td>rps7</td>
<td>rpoC2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>psbI</td>
<td></td>
<td></td>
<td>rpl36</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>psbH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>psbI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>psbK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>psbL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>psbN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>psbT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>psbZ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. CHLOROPLAST GENE EXPRESSION AND CONTROL

Most of our understanding of gene expression and its control in chloroplasts comes from work on land plants and the unicellular green alga Chlamydomonas. The latter is particularly valuable for such studies as it is (i) readily amenable to genetic analysis and (ii) able to grow heterotrophically on medium containing fixed carbon, so non-photosynthetic mutants remain viable.
Land plant chloroplasts contain a nuclear-encoded, viral-type RNA polymerase (NEP, or nuclear-encoded polymerase), which is used to transcribe genetic system genes in the chloroplasts, such as those for the plastid-encoded RNA polymerase (PEP), which in turn transcribes genes for proteins involved in photosynthesis. However, algae appear only to have a chloroplast-encoded bacterial-type RNA polymerase (with the possible exception of dinoflagellates), so the possibility of control by modulation of the NEP does not exist (Smith & Purton 2002). The PEP is thought to resemble that of *E. coli*, where the RNA polymerase is encoded by the *rpoA, B* and *C* genes. In algae, the *rpoC* gene is normally divided into two separate genes *rpoC1* and *rpoC2*. An exception to this is in *Chlamydomonas* where significant gene rearrangement has occurred and where the location and identity of the *rpoA* and *rpoC1* genes are still unclear (Maul et al. 2002). The PEP presumably requires sigma factors, and there is variation in the number of these. Land plants have several nuclear genes for them, as does the red alga *Cyanidioschyzon merolae* (Minoda et al. 2005). Many other algae, such as *Chlamydomonas reinhardtii*, appear to have only a single sigma factor (Carter et al. 2004), and in *Guillardia theta*, a sigma factor is encoded in the nucleomorph (Smith & Purton 2002).

Early studies on the control of gene expression concentrated on changes in levels of transcripts on illumination, and identified a number of genes whose transcript levels increased. Because chloroplasts were recognized as being of prokaryotic origin, it was assumed that this could be interpreted as an increase in transcription initiation. Subsequent studies determining half-lives of chloroplast transcripts in plants showed that transcript stability and translation efficiency are important determinants of the overall rate of synthesis of many proteins (e.g. Mul et al. 1987), although there are nevertheless examples where transcriptional control is important, such as redox regulation of genes for photosystem components (see below). In *Chlamydomonas reinhardtii*, run-on experiments showed that transcriptional and post-transcriptional processes are both important in determining transcript levels (Lilly et al. 2002). For the red alga *Cyanidioschyzon merolae* transcript levels of a number of chloroplast genes also increase on illumination, and run-on experiments indicate that transcription is a major determinant of these changes (Minoda et al. 2005).

Pfannschmidt et al. (1999) showed that transcription of genes for photosystem II components increased when plants were transferred to light favouring photosystem I and *vice versa*. This was shown to depend on the redox state of the plastocyanine pool. Puthiyaveetil et al. (2008) identified a nuclear-encoded prokaryotic-type sensor kinase in chloroplasts that is required for this regulation. Putative plastidic sensor kinases also exist in green and other algal lineages, so a similar regulation may operate there (Dupleisis et al. 2007; Puthiyaveetil et al. 2008; Puthiyaveetil & Allen 2009).

Sigma factors may also be a means of modulating transcription. Thus the nuclear *SIG5* gene of *Arabidopsis thaliana* encodes a sigma factor that is expressed in blue light and activates the transcription of the chloroplast *psbD* gene (Tsunoyama et al. 2004). Regulation of one of the sigma factors of *C. merolae* has been implicated in light-dependent changes in chloroplast transcription (Minoda et al. 2005).

A number of chloroplast transcripts require splicing. For example, the *Chlamydomonas* chloroplast genome sequence reveals group I and II introns located in the *psbA, psaA* and 23S rRNA genes. Splicing of the group I introns is through a guanosine-dependent mechanism (Herrin et al. 1990) likely to be similar to that reported for the *Tetrahymena* rRNA intron. At physiological temperatures, splicing *in vitro* is much slower than splicing *in vivo*. Moreover, the splicing of these introns was shown to be highly up-regulated by light for photoautotrophically growing *Chlamydomonas* (Herrin & Nickelsen 2004, and references therein). Editing of chloroplast transcripts is largely restricted to land plants, with the possible exception of some dinoflagellates (Zauner et al. 2004). Chloroplast transcripts also undergo complex cleavage and degradation reactions (for which they may be marked by transient polyadenylation), mediated by nuclear-encoded proteins (Zimmer et al. 2008). Translation is by bacterial-type ribosomes, consistent with the origin of chloroplasts.

Nuclear-encoded factors bind to and determine the stability and translation efficiency of many mRNAs in *Chlamydomonas* chloroplasts. A well-characterized example is the MCA1 protein, which stabilizes the mRNA for *petA* (cytochrome *f*) of the electron transfer chain. In transgenic lines expressing different levels of MCA1, the level of *petA* mRNA varies accordingly and the level of PetA correlates closely. In wild-type cells, levels of MCA1 vary with parameters such as culture age and environmental conditions, and levels of the *petA* mRNA vary in the same way. Translation of the *petA* mRNA depends on a factor, TCA1, levels of which are also modulated according to environmental conditions (Raynaud et al. 2007).

Redox poise and the ATP/ADP balance are also likely to be important determinants of translation, a well-studied example being *psbA* mRNA in *Chlamydomonas* (Marin-Navarro et al. 2007) (figure 1). A complex of four proteins, RB60, RB55, RB47 and RB38, binds the 5′ UTR. RB60 regulates the redox state of RB47 (an RNA-binding protein). High light levels lead to reduction of RB60 and increased PsbA synthesis (satisfying the need for increased replacement of PsbA after photodamage), probably through reduction of RB47, which stimulates its binding to the *psbA* mRNA and translation. In addition, an ADP-dependent kinase inhibits binding of the complex by phosphorylation of RB60. In the light the ATP/ADP ratio is high, so there will be less phosphorylation of RB60 and increased PsbA synthesis.

A final determinant of the level of chloroplast-encoded proteins is the fate of protein that has not been assembled into a functional complex (reviewed by Marin-Navarro et al. 2007). For PetA (cytochrome *f*) in *Chlamydomonas*, protein that has not been assembled into the cytochrome *b*/*f* complex represses its own
6. NUCLEOMORPH
An interesting feature of some complex chloroplasts is the retention of a relic eukaryotic nucleus—the nucleomorph—located between the inner two and outer chloroplast membranes. They have been found only within the chloroplasts of chlorarachniophytes and cryptophytes, which were derived through the secondary endosymbiosis of green and red algae, respectively. Despite their different origins, nucleomorphs of green and red algal ancestry show remarkable similarities. They have both retained a nuclear envelope, which and red algal ancestry show remarkable similarities. Despite their different origins, nucleomorphs of green cryptophytes, which were derived through the second-

Figure 1. Light-dependent regulation of pshA mRNA translation in Chlamydomonas. Increased illumination leads to disulphide reduction on RB47 and decreased phosphorylation of RB60, resulting in increased mRNA-binding activity of the complex and stimulation of translation.

synthesis. Deletion of a C-terminal region of PetA leads to an increase in synthesis, consistent with this part of the protein being involved in an autoregulatory circuit. In many cases, the inability to assemble a functional complex because of loss of one or more components leads to degradation of the unassembled components. Thus, the final level of control may simply be turnover of protein that is surplus to requirements.

Redox poise therefore seems to be an important determinant of gene expression in plants and green algae. The requirement for gene expression to respond to redox poise provides an elegant justification for the retention of a chloroplast genome in photosynthetic organisms (Allen 2003), although why non-photosynthetic ones should also retain a genome is less clear (Barbrook et al. 2006). It seems likely that the regulation of chloroplast gene expression in response to redox poise and the ATP/ADP ratio will be a general feature. It is striking that chloroplast gene expression in Chlamydomonas requires many different nuclear-encoded proteins that bind specific transcripts to determine their stability or their translation. It will be interesting to see to what extent this applies to other unicellular photosynthetic organisms.

All nucleomorph genomes characterized to date consist of just three chromosomes, which display characteristics of eukaryotic chromosomes such as telomeric repeats. Subtelomeric regions contain rDNA operons encoding 18S/5.8S/28S rRNAs. In G. theta and B. natans, these occur at all six chromosome ends. However, they occur at only three in the cryptophyte Hemiselmis andersenii, at both ends of chromosome 1 and at one end of chromosome 3 (Lane et al. 2007). The 5S rDNA is absent from B. natans, but is present within G. theta’s telomeric repeats, and is found separately near the telomere ends in H. andersenii. Centromeres have proved hard to identify, although candidate sequences exist within non-coding regions near the midpoint of chromosomes, and centromere-binding proteins are encoded on the nucleomorph genomes (Gilson et al. 2006).

Consistent with their reduction in size, the gene content of nucleomorph genomes is greatly reduced compared with nuclear genomes. Many nucleomorph-targeted proteins are encoded in the cell’s nucleus. Although sequenced nucleomorph genomes have quite dissimilar gene contents, the proteins encoded are primarily involved in the maintenance of the organelle and housekeeping processes such as transcription and translation. The nucleomorph also encodes some chloroplast-targeted proteins. However, of the 17 such proteins encoded on the B. natans nucleomorph genome and the 30 on that of G. theta, only Hsp60, RpoD/Sig2 and isoforms of a Clp protease are common to both (Gilson et al. 2006).

Cryptophyte and chlorarachniophyte nucleomorph genomes differ markedly in the number and sizes of introns present. Cryptophyte nucleomorph genomes appear to have lost the majority of their ancestral introns. The G. theta nucleomorph genome contains only 29 introns, which range in size between 42 and 52 bp. Seventeen of these are within protein coding genes, including 11 in ribosomal protein genes. The remainder are found in tRNA genes (Douglas et al. 2001). The 572 kbp nucleomorph genome of the cryptophyte H. andersenii is slightly larger than that of G. theta (551 kbp) but contains no introns at all, despite having homologues of all but two of G. theta’s intron-containing genes (Lane et al. 2007). Unsurprisingly, the H. andersenii nucleomorph genome has lost numerous genes encoding RNA splicing machinery. Only four genes are present that have very weak homology to spliceosome components and rRNA processing machinery (Lane et al. 2007). In contrast to cryptophytes, chlorarachniophyte nucleomorph genomes have retained many introns, although these introns are very small. The B. natans nucleomorph genome contains 852 introns (Gilson et al. 2006), which range in size from 18 to 21 bp.
They are AT-rich and have typical eukaryotic GT/AG borders. The position of the majority of these introns is conserved when compared with homologues in *Chlamydomonas* or *Arabidopsis*. Concomitantly, a selection of proteins involved in RNA splicing and maturation is also encoded in the nucleomorph. A recent EST study of the nucleomorph genome of the chlorarachniophyte *Gymnochloro cultura* identified 153 introns (Slamovits & Keeling 2009). They are also reduced in size, ranging between 18 and 27 bp. As for *B. natans*, analyses revealed GT/AG intron borders and high AT content within the introns, although no branch-point recognition motifs or polypyrimidine tracts were identified (Slamovits & Keeling 2009). This study also revealed a degree of aberrant splicing of *G. stellata* nucleomorph transcripts—a phenomenon that was not observed in *B. natans* (Gilson et al. 2006). The efficiency of intron splicing may be linked to intron size. A ‘clipper’ model for splicing has been suggested whereby the splicing machinery recognizes introns by virtue of their strict 18–21 bp size (Gilson et al. 2006). However, the presence of 24 and 27 bp introns in *G. stellata* contradicts this proposal, although it appears that smaller introns are more efficiently spliced (Slamovits & Keeling 2009).

Little is known about the regulation of nucleomorph gene expression. The presence of genes for histone acetylases and deacetylases on nucleomorph genomes may indicate the involvement of histone modification. Given the small number of non-housekeeping genes, it is possible that sensitive control of gene expression is not necessary within this compartment. However, any control of nucleomorph gene expression may have to respond to both nuclear and chloroplast signals. It will be interesting to see how far eukaryotic mechanisms of gene regulation have been retained in the evolution of this organelle.

### 7. CONCLUSIONS

There is a huge range of mitochondrial genome organization and content among protists. The chloroplast genome is generally more conserved, with a few notable exceptions in isolated groups. Understanding the mechanism of gene expression and control is generally limited to a small number of model organisms. Given the importance of the group as whole, more information will be valuable. However, this will rely on the availability of better genetic tools.

Work in the authors’ laboratory is funded by the Leverhulme Trust, the BBSRC and the University of Cambridge Isaac Newton Trust.

**REFERENCES**


de Koning, A. P. & Keeling, P. J. 2006 The complete plastid genome of the parasitic green alga *Helicostomum sp.* is highly reduced and structured. *BMC Biol.* 4, 12. (doi:10.1186/1741-7007-4-12)


*Phil. Trans. R. Soc. B* (2010)


1016/j.tplants.2007.03.011)

Li, J., Maga, J. A., Cermekian, N., Cedergren, R. & Feagin, 
J. E. 2001 Identification and characterization of a Plasmo-
dium falciparum RNA polymerase gene with similarity 
to mitochondrial DNA polymerases. Mol. Biochem. Para-
stol. 113, 261–269. (doi:10.1016/S0166-6851(01)00223-7)

Lilly, J. W., Maul, J. E. & Stern, D. B. 2002 The Clyma-
dyomomas reinhardtii organellar genomes respond 
transcriptionally and post-transcriptionally to abiotic 
005595)

Lukesˇ, J., Jirkuˆ, M., Avliyakulov, N. & Benada, O. 1998 
Pankinetoplast DNA structure in a primitive bodonid 
flagellate, Cryptobia helics. EMBO J. 17, 838–846. 
(doi:10.1093/emboj/17.3.838)

Lukesˇ, J., Lys Guilbride, D., Votypka, J., Zikova, A., Benne, 
R. E. & Englund, P. T. 2002 Kinetoplast DNA network: 
evolution of an improbable structure. Eukaryot. Cell 1, 

Lukesˇ, J., Hashimi, H. & Ziková, A. 2005 Unexplained 
complexity of the mitochondrial genome and transcriptome 
(doi:10.1007/s00294-005-0027-0)

Marande, W. & Burger, G. 2007 Mitochondrial DNA as 
a genomic jigsaw puzzle. Science 318, 415. (doi:10.1126 
science.1148033)

Michelotti, E. F., Harris, M. E., Adler, B., Torri, A. F. & 
Hajduk, S. L. 1992 Trypanosoma brucei mitochondrial 
ribosomal RNA synthesis, processing and development-
tally regulated expression. Mol. Biochem. Parasitol. 54, 
31–41. (doi:10.1016/0166-6851(92)90092-X)

Miller, M. L. & Miller, D. L. 2008 Non-DNA-templated 
addition of nucleotides to the 3’ end of RNAs by 
the mitochondrial RNA polymerase of Physarum poly-
1128/MCB.00356-08)

Identification of a putative mitochondrial RNA polymer-
ase from Physarum polycephalum: characterization, 

Minoda, A., Nagasawa, N., Cedergren, R. & Feagin, 
J. E. 2001 Another acquisition of a primary 
photosynthetic organelle is underway in Paulinella 
1016/j.cub.2009.02.043)

Nash, E. A., Nisbet, R. E. R., Barbrook, A. C. & Howe, C. J. 
2008 Dinoflagellates: a mitochondrial genome all at sea. 
001)

Barbrook, A. C. & Howe, C. J. 2008 Transcript analysis of 
mitochondrial plastid gene minicircles. Protist 159, 

Oudot-Le Secq, M. P., Fontaine, J. M., Rousvoal, S., 
Kloareg, B. & Loiseaux-De Goër, S. 2001 The complete 
sequence of a brown algal mitochondrial genome, the 
53, 80–88.

Complete mitochondrial genomes of the three brown 
algae (Heterokonta: Phaeophyceae) Dictyota dichotoma, 
49, 47–58. (doi:10.1007/s00294-005-0031-4)

Pfannschmidt, T., Nilsson, A. & Allen, J. F. 1999 Photosyn-
thetic control of chloroplast gene expression. Nature 
397, 625–628. (doi:10.1038/17624)

The complete mitochondrial DNA sequence of the green 
alga Pseudendoclonium akinetum (Ulvophyceae) 
highlights distinctive evolutionary trends in the chloro-
phyta and suggests a sister-group relationship between 
21, 922–935. (doi:10.1093/molbev/ms099)

The complete mitochondrial genome sequence of the 
haplophyte Emiliania huxleyi and its relation to hetero-
lonks. DNA Res. 11, 1–10. (doi:10.1093/dnares/11.1.1)

Puthiyaveetil, S. & Allen, J. F. 2009 Chloroplast two-
component systems: evolution of the link between 
photosynthesis and gene expression. Proc. R. Soc. B 

Puthiyaveetil, S. et al. 2008 The ancestral symbiont sensor 
kine CSK links photosynthesis with gene expression in 
066. (doi:10.1073/pnas.0803928105)

Raynau, C., Loiselay, C., Wostrikoff, K., Kuras, R., Girard-
for regulatory function of nucleus-encoded 
(doi:10.1128/mbc.07-03-0317)

2002 Identification of the PSB1 gene from 
the red alga Chondrus crispus requires a unique RNA-
processing mechanism, internal cleavage of upstream tRNAs at pyrimidine 

Richard, O., Kloareg, B. & Boyen, C. 1999 mRNA 
expression in mitochondria of the red alga Chondrus 
crispus requires a unique RNA-processing mechanism, 
internal cleavage of upstream tRNAs at pyrimidine 48. 
2725)

Robbens, S., Derelle, E., Ferraz, C., Wuys, J., Moreau, H. & 
Van de Peer, Y. 2007 The complete chloroplast and 
mitochondrial genome of the brown alga Pseudendoclonium akinetum 
(Ulvophyceae) highlights distinctive evolutionary trends in the chloro-
phyta and suggests a sister-group relationship between 
21, 922–935. (doi:10.1093/molbev/ms099)

Remacle, C., Cardol, P., Coosmans, N., Gaisne, M. & Bon-
nefoy, N. 2006 High-efficiency biolistic transformation of 
Chlamydomonas mitochonidia can be used to insert mutations in complex I genes. Proc. Natl Acad. Sci. USA 103, 4771–4776. (doi:10.1073/pnas.05095110103)

Richard, O., Bonnard, G., Grienengerber, J. M., Kloareg, B. & 
Boyen, C. 1998 Transcription initiation and RNA processing 

Richard, O., Kloareg, B. & Boyen, C. 1999 mRNA 
expression in mitochondria of the red alga Chondrus 
crispus requires a unique RNA-processing mechanism, 

Phil. Trans. R. Soc. B (2010)


