SPECIAL ISSUE REVIEW PAPER

The remarkable chloroplast genome of dinoflagellates

Christopher J. Howe*, R. Ellen Nisbet and Adrian C. Barbrook

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

Received 1 June 2007; Revised 29 October 2007; Accepted 31 October 2007

Abstract

Dinoflagellates are an economically and ecologically important eukaryotic algal group. The organization of their chloroplast genome appears to be radically different from that in plants and other algae. The gene content has been dramatically reduced in dinoflagellates, with the large-scale transfer of genes to the nucleus. Most of the remaining genes encode subunits of Photosystems I and II, the cytochrome b6f complex, and ATP synthase, as well as rRNAs and a few tRNAs. Whereas conventional chloroplast genomes have all their genes physically linked on one molecule, dinoflagellate chloroplast genes are located on small plasmids, termed ‘minicircles’. Each minicircle has at most a few genes, and a distinguishable ‘core’ region. Genes are always in the same orientation with respect to the core region. There are also non-coding minicircles, including aberrant forms of minicircles apparently derived from other minicircles by rearrangement. The evidence that the minicircles are located in the chloroplast and that there is no conventional chloroplast genome in addition to the minicircles is discussed. Transcription of minicircles is probably initiated close to the core, generating transcripts corresponding to an almost entire minicircle. The transcripts are then cleaved to molecules corresponding to individual genes. Post-transcriptional modifications include editing and addition of a polyU tail. It is discussed why these particular genes have been retained in the dinoflagellate chloroplast, together with the possibility that the chloroplast supplies fMet-tRNA to the mitochondrion.

Key words: Chloroplast, dinoflagellate, minicircle, photosynthesis.

Dinoflagellates

Dinoflagellates are a widespread group of single-celled eukaryotic algae, occurring in marine and freshwater environments. Aspects of their biology are summarized here, and more detailed descriptions are available in excellent books (Taylor, 1987; van den Hoek et al., 1995; Lee, 1999). They are motile, with longitudinal and transverse flagella, and usually photosynthetic. They are of major economic and environmental importance, not least because North Sea oil deposits are likely to owe their origin to dinoflagellates (Downie, 1956; Gallois, 1976). Present day proliferation of cells results in blooms, or ‘red tides’, and in many instances the dinoflagellate strains responsible are toxic. These blooms can lead to the accumulation of toxins in shellfish, causing potentially fatal poisoning to humans eating them. Dinoflagellates of the genus Symbiodinium form symbionts (‘zooxanthellae’) in corals and other benthic marine animals, and the expulsion of these symbionts under stress conditions is widely recognized as coral ‘bleaching’ (Douglas, 2003; Coffroth and Santos, 2005). Such bleaching may in fact be a beneficial mechanism for adaptation to environmental change (Baker et al., 2004).

Dinoflagellates generally have polysaccharide-based thecal plates in vesicles below the plasma membrane. In addition, some species have external scales. Other structural features include trichocysts, projectiles that can be fired from the cell, and eyespots, composed of lipid globules. The nucleus is remarkable in that the chromosomes remain condensed throughout the cell cycle and lack conventional organization into histone-containing nucleosomes. In many species, there is a high level of modified bases such as 5-hydroxymethyluracil in the nuclear genome. The chloroplasts are surrounded by a total of three membranes, and are believed to be of secondary origin from a red algal source (Fast et al., 2001) (reviewed in Larkum et al., 2007), and generally contain chlorophylls a and c. Many species also contain the carotenoid
peridinin for light harvesting. This is present both in an intrinsic protein of the thylakoid membrane and as a soluble trimeric complex, the peridinin-chlorophyll protein (Hiller et al., 1993; Hofmann et al., 1996). Several other photosynthetic species contain the carotenoid hexa-fucoxanthin instead of peridinin. Establishing the detailed phylogenetic history of these groups, as well as non-photosynthetic strains, is controversial (Bachvaroff et al., 2005) and will not be discussed in detail here. However, rRNA trees combined with the presence of genes apparently encoding plastid proteins in early-diverging heterotrophic (i.e. non-photosynthetic) strains such as Cryptothecodinium cohnii and Perkinsus marinus indicate that heterotrophic species are likely to have been ancestrally photosynthetic (Saldarriaga et al., 2001; Sanchez-Puerta et al., 2007; Stelter et al., 2007). A number of species have plastids of tertiary origin, replacing their secondary ones (Saldarriaga et al., 2001; Ishida and Green, 2002), and some of these contain chlorophylls a and b (Watanabe et al., 1987).

**Chloroplast genomes**

It is over 20 years since the first complete sequence of a chloroplast genome was reported, from the liverwort Marchantia polymorpha (Ohyma et al., 1986). There are over 100 chloroplast genome sequences now available, from land plants as well as green, red and ‘brown’ (i.e. chlorophyll c-containing) algae. A core of about 40 protein-coding genes is present in essentially all chloroplast genomes from photosynthetic organisms (Martin et al., 2002; Hagopian et al., 2004), including the alga Ostreococcus tauri, which has one of the smallest chloroplast genomes known (Robbens et al., 2007). These core genes encode subunits of the main complexes of the light reactions of photosynthesis (Photosystems I and II, the cytochrome b_{6}f complex, and the ATP synthase complex) as well as polypeptides of the small and large subunits of the ribosome and a eubacterial-type RNA polymerase (Martin et al., 2002). Chloroplast genomes also encode tRNAs and rRNAs. Other genes are present in particular lineages, giving a typical genome size of 120–200 kbp. The molecule is widely regarded as topologically circular, typically containing two single-copy regions separated by inverted repeated regions (containing the rRNA genes and others), the extent of which accounts for much of the variation in chloroplast genome size, at least among land plants (Palmer, 1985). One arm of the inverted repeat has been lost on a number of independent occasions, for example, within legumes, in conifers, and in the chlorophycean algal lineage that includes Stigeoclonium helveticum (Palmer and Thompson, 1981; Lidholm and Gustafsson, 1991; Belanger et al., 2006). In Euglena gracilis, the rRNA genes are tandemly repeated with three complete copies and a partial one (Hallick et al., 1993). However, the view of the chloroplast genome as a topologically circular molecule containing a single copy of each gene (with the exception of those in the repeated regions) and a small fraction of dimeric molecules has been challenged, and the molecule may be more accurately envisaged as occurring primarily in linear and branched multimeric forms (Bendich, 2004). Nevertheless, this does not alter the general principle that the chloroplast genome contains many tens of physically linked genes.

**The dinoflagellate chloroplast genome: minicircular DNA**

The dinoflagellates were the last major algal group to have their chloroplast genome analysed in detail. Initial studies on the dinoflagellates Protogonyaulax catenella, P. tamarensis, and Glenodinium sp. indicated the presence of a chloroplast genome of a conventional unit size of the order of 120 kbp (Boczar et al., 1991). A subsequent analysis of the gene for ribulose bisphosphate carboxylase indicated that, although in all other photosynthetic organisms a gene for the large subunit of this enzyme is located in the chloroplast, there is a different organization in the dinoflagellate Gonyaulax polyedra (now Lingulodinium polyedrum), as the gene for the enzyme is located in the nucleus (Morse et al., 1995). This inference was made on the grounds that the coding sequence was represented in a polyA+ cDNA library, and had a high GC content. Furthermore, the G. polyedra ribulose bisphosphate carboxylase was found to be of Form II, which comprises a single type of subunit, in contrast to the Form I found in plants, algae, and cyanobacteria. These observations thus hinted that the dinoflagellate chloroplast genome might show other significant differences from conventional ones, although a probe for atpB (located in the chloroplast in other organisms) hybridized to DNA obtained from a chloroplast-containing fraction.

A very different organization of the dinoflagellate chloroplast genome was demonstrated in 1999 by Zhang et al. (1999). They reported sequences of psaA, psaB, psbA, psbB, psbC, atpA, petB, 23S rRNA, and 16S rRNA genes (located in the chloroplast in plants and other algae) from a satellite fraction obtained by CsCl density gradient centrifugation of DNA from the peridinin-containing dinoflagellate Heterocapsa triquetra. The genes were shown to be located on individual circular molecules of 2.2–3.1 kbp. The non-coding regions contained a tripartite conserved region, that could be folded in silico into elaborate secondary structures reminiscent of replication origins. The coding regions of all the genes identified were in the same orientation in relation to the core sequence. Probing of Southern blots of uncut DNA from the dinoflagellates Amphidinium carterae, H. pygmaea
(the same strain designated as Glenodinium sp. by Boczar et al., 1991) and H. rotunda indicated that the psbA gene was on molecules of 3 kbp or less in these species, and the 23S rRNA gene was on molecules of 6 kbp or less. For Gonyaulax grindleyi, there was some evidence of larger molecules. The presence of chloroplast genes on small circles was confirmed independently by Barbrook and Howe with Amphidinium operculatum (Barbrook and Howe, 2000). [This strain was subsequently redesignated Amphidinium carterae, and will be referred to as such throughout the rest of this review. Note that this is a different strain of Amphidinium carterae from that used by Hiller (2001).] Barbrook and Howe (2000) demonstrated by Southern blotting and screening of libraries that psaA, psbA, psbB, petD, and atpB genes were present on minicircles with conserved core regions. Further experiments by Hiller (2001) and Barbrook et al. (2001) with two A. carterae strains showed the presence on minicircles of other genes (psbC, psbD, psbE, petB, and atpA) usually present in the chloroplast in other organisms. PCR using primers based on the core regions revealed the presence of ‘empty’ minicircles that contained recognizable core regions but either no recognizable coding regions or only small fragments of coding regions. Hiller (2001) and Barbrook et al. (2001) also found the first examples of minicircles containing more than one gene, namely atpA/petB and psbD/psbE. In both minicircles, the genes were arranged in tandem. Northern analysis of the atpA/petB pair indicated that the predominant transcripts were of sizes corresponding to the single genes, suggesting that the individual genes on a minicircle were either transcribed independently, or were part of a larger transcript that was rapidly cleaved to generate monocistronic transcripts. These pairs of genes had not been previously reported as adjacent in other chloroplast DNAs. The presence of chloroplast genes on minicircles has now been reported from many other dinoflagellate genera, including Protoceratium (Zhang et al., 2002), Symbiodinium (Moore et al., 2003; Takishita et al., 2003), and Ceratium (Laatsch et al., 2004). The organization of an example of a minicircle containing multiple genes is shown in Fig. 1.

**Coding content of minicircles**

The protein-coding genes identified on minicircles in the studies referred to above and subsequent reports are listed in Table 1. These represent a subset of the core of genes present in the chloroplast genome of all other photosynthetic organisms (Table 2). The ycf16 and ycf24 genes reported from Ceratium (Laatsch et al., 2004) are not part of the core retained in all organisms, as they are generally absent from green chloroplast lineages, although they are retained in other lineages. Unidentified open reading frames that may be unique to dinoflagellates have also been reported (Barbrook et al., 2006b). It is possible that they may be involved in functions specific to minicircles. The coding sequences seen in minicircles are generally highly divergent compared with other chloroplast DNA sequences. The psaA gene of A. carterae shows a number of small deletions compared with homologues from other chloroplasts (Barbrook and Howe, 2000). The location of these deletions on structural models of the protein indicates that they correspond to short stretches in hydrophilic loops of the protein, rather than in membrane spanning regions, and suggests that they do not interfere with the function of the protein (Barbrook and Howe, 2000). Given the generally large size of the PsaA protein (typically 70 kDa) and thus the gene, it is possible that the deletions may be necessary to allow the gene to be accommodated within a typically-sized minicircle from A. carterae. In Heterocapsa by contrast, where the minicircles are typically slightly larger than A. carterae, the psaA gene shows fewer deletions. The highly divergent nature of dinoflagellate chloroplast coding regions complicates their use in phylogenetic analysis.

The identity of the rRNA genes remained uncertain for some time, as the similarity to other rRNA sequences as revealed by BLAST searches was very low (Hiller, 2001; Koumandou et al., 2004). More detailed comparison of putative rRNA genes from two strains of A. carterae to typical eubacterial sequences revealed ‘minimal’ rRNA gene sequences in the dinoflagellate lacking a number of regions found in eubacterial rRNAs (Barbrook et al., 2006b). In many cases, secondary structural elements were conserved, even though the primary sequence was very divergent. In spite of the unusual nature of the rRNA
Table 1. Reported minicircular sequences from dinoflagellates

Genes located on the same minicircles are indicated by superscript numbering. *Ceratium horridum* minicircles have a reported nuclear localization.

<table>
<thead>
<tr>
<th>Organism</th>
<th>atpA</th>
<th>atpB</th>
<th>petB</th>
<th>petD</th>
<th>psaA</th>
<th>psaB</th>
<th>psbA</th>
<th>psbB</th>
<th>psbC</th>
<th>psbD</th>
<th>psbE</th>
<th>psbI</th>
<th>LSU-rRNA</th>
<th>SSU-rRNA</th>
<th>tRNA</th>
<th>Unidentified/ycf</th>
<th>Non-coding References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adenoides eludens</strong></td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Amphidinium carterae</strong></td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Heterocapsa triqueta</strong></td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Ceratium horridum</strong></td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Heterocapsa niei</strong></td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Heterocapsa pygmaea</strong></td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Protoceratium reticulatum</strong></td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Symbiodinium sp., clade A</strong></td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Symbiodinium sp., clade B</strong></td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Symbiodinium sp., clade C</strong></td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The core region

There is considerable variation in the core region among certain dinoflagellates. In general, the core regions for minicircles with the same genes within the same strain are identical, but core regions are not identical within a given gene in different strains. Likewise, minicircles with some genes within the same strain are identical, but core regions are not identical between different growth stages. However, there are large changes in the copy number of minicircles, as discussed below (Koumandou and Howe, 2007).

Although there is probably some variation in the copy number of minicircles, this does not seem to be more than 4-fold. The largest example so far is from A. carterae, where psbD, psbE, and psbI genes occur on one circle, along with an unidentified open reading frame (Koumandou and Howe, 2007). As described above, a number of instances have been reported of multiple genes, both for RNAs and proteins, occurring on a single minicircle. One example is from C. horridum (Nelson et al., 2007), or in other chlalogal species, such as *Caenorhabditis* and *Heterocapsa* (Nelson et al., 2007). There is considerable variation in the core region among different strains, as discussed below (Koumandou and Howe, 2007).
of Symbiodinium isolates, as well as Adenoides eludens, the core regions show a particularly complex pattern of direct and inverted repeats (Moore et al., 2003; Nelson and Green, 2005). It is interesting that different isolates of the C phylotype of Symbiodinium all have a much more complex organization of the core region than the other phylotypes of Symbiodinium, and core regions from the C phylotype show promise as genetic markers for studies of differential susceptibility to bleaching among strains (Barbrook et al., 2006c).

The detailed function of the core region remains obscure. Given that the coding regions of all minicircles within a strain are in the same orientation with respect to the core, it seems probable that the core contains transcription initiation signals. Other probable functions include an involvement in replication and possibly in membrane attachment and partitioning of minicircles in chloroplast division.

Aberrant minicircles

In addition to ‘conventional’ minicircles carrying a single core and a recognizable coding region, a number of ‘aberrant’ minicircles have been reported. These can be divided into a number of categories. Many of them appear to be generated by recombination events, reminiscent of the ‘sublimons’ found at low abundances in plant mitochondrial genomes that also appear to be generated by recombination (Small et al., 1987).

Small deletions

Zhang et al. (1999) reported the existence of multiple forms of minicircles carrying putative 23S rRNA sequences, with deletions of 25 and 35 nucleotides in the core region. Similarly, deletions of up to 300 nucleotides were reported in different forms of the 23S rRNA minicircle from Symbiodinium sp. by Santos et al. (2002, 2003). Length heterogeneity has also been found in other minicircles from Symbiodinium (AC Barbrook, unpublished results).

Empty minicircles

Hiller (2001) and Barbrook et al. (2001) used primers based on the conserved regions of the core of two strains of A. carterae to amplify a population of minicircles without prior assumptions as to the coding regions present. They found a number of minicircles that had recognizable core regions, but had no recognizable coding regions apart from short stretches showing some similarity to parts of previously identified sequences, or short open reading frames of at most 300 bp or so. They designated these as ‘empty’ minicircles. Between five and 10 empty minicircles have been identified from the two strains of A. carterae, and one has also been reported from H. triqueta (Green, 2004; Barbrook et al., 2006b). The function, if any, of these is not clear, and it is possible they may represent ‘parasitic’ DNA elements.

Jumbled minicircles

Zhang et al. (2001) identified a family of minicircles from H. triqueta that contained a core region and fragments of two or more identifiable coding regions. They termed...
these ‘jumbled’ minicircles and proposed that they were generated by recombination between different minicircles to give heterodimers with two core regions (and presumably two replication origins). In their model, the heterodimers then underwent deletions that resulted in loss of one of the core regions. Presumably this allowed the molecules to be more stably propagated, avoiding problems caused by the existence of multiple replication origins in the same minicircle. These may also represent parasitic DNA molecules (Zhang et al., 2001).

**Microcircles**

Nisbet et al. (2004) reported a series of minicircles in *A. carterae* of 400–600 bp, which they called ‘microcircles’. The microcircles appeared to be derived from other minicircles by recombination across stretches of homology, without requiring an intermediate heterodimeric minicircle (although an intermediate heterodimer could also be postulated). They also reported that the microcircles could be recovered from the same culture over a period of 18 months, suggesting that they were stably retained in the organisms (or arose continually by recombination). Again, the function of the microcircles is unknown.

**Are the minicircles really the chloroplast genome?**

At present, there are several lines of evidence suggesting that the minicircles represent a genuine chloroplast genome, but they are indirect. They are discussed below.

1. The genes identified on minicircles are in the chloroplast genome in most or all other organisms. (Although the *ycf16* and *ycf24* sequences from *Ceratium* are not found in green chloroplast genomes, they are in the chloroplast in other lineages.) Obvious exceptions to this are the unidentified open reading frames that are unique to dinoflagellates.

2. There is no evidence of transit sequences. If the minicircles were located outside the chloroplast, and there were no copies of these sequences inside the chloroplast, the proteins encoded should have N-terminal transit sequences, to direct their import into the organelle.

3. Sequences of many of the ‘core’ protein genes that are located in the chloroplast in other organisms, but have not yet been identified on minicircles, are found in cDNA libraries that are made from poly(A) RNA. The chloroplast proteins encoded by these cDNAs generally include N-terminal extensions (Bachvaroff et al., 2004; Hackett et al., 2004). The polyA tail on the RNA and the N-terminal extension of the protein indicate that the coding sequences have been transferred from the chloroplast to the nucleus.

4. The minicircle coding sequences have different codon preferences from nuclear genes (such as *rbcL, psbO* and many others) and the cDNA sequences referred to above. This suggests that they are located in a different compartment.

5. Probes for transcripts to *psbA* hybridize *in situ* to chloroplasts of *Symbiodinium* (Takishita et al., 2003). Unless the mRNA is imported, this indicates that the gene is also present in the organelle.

6. Synthesis of PsbA protein in *Amphidinium* and *L. polyedrum* (G. polyedra) is inhibited by chloramphenicol, but not cycloheximide, and is not regulated in a circadian fashion (Wang et al., 2005). These are typical features of proteins encoded and synthesized in the chloroplast.

7. There is no evidence for a ‘conventional’ chloroplast genome. Although Boczar et al. (1991) reported an apparently conventional chloroplast genome organization in some species, this finding has not generally been reproduced. For example, with *A. carterae*, both Southern analysis and PCR experiments using dinoflagellate DNA with primers for genes that are adjacent in other chloroplast genomes fail to give any indication of a chloroplast genome of conventional organization (AC Barbrook, unpublished results).

Two studies conflict with the view that minicircles represent a genuine chloroplast genome. Laatsch et al. (2004) studied minicircles and other DNA molecules from the peridinin-containing species *C. horridum*. They reported that isolated chloroplasts contained a high molecular weight (HMW) DNA species. In Southern analysis, a 66 bp probe from the chloroplast HMW DNA hybridized only to that fraction, whereas a probe for *psbB* derived from a minicircle failed to hybridize to the HMW chloroplast DNA fraction but did hybridize to a fraction including nuclear material. *In situ* hybridization was also interpreted as showing that the *psbB* probe hybridized to the cell nucleus. Laatsch et al. (2004) therefore concluded that the minicircles in *C. horridum* are located in the nucleus, and that a distinct HMW DNA species exists in the chloroplast. However, the sequence of the 66 bp probe derived from the HMW chloroplast DNA has an unusual nucleotide composition (58 of the 66 residues being G or T), and does not encode any peptide recognizable as homologous to chloroplast proteins. Together with the failure of the *psbB* probe to hybridize to the HMW chloroplast DNA, this makes it unlikely that the HMW DNA has a conventional chloroplast genome coding function. Thus the minicircle genes are likely to represent functional copies. If so, and they are really located outside the chloroplast (although this is arguably not demonstrated unambiguously by the hybridization data), it is not clear how the proteins they encode enter the chloroplast without N-terminal import sequences.
Wang and Morse (2006b) analysed the \( psbA \) gene of \( L.\ polyedrum \) (\( G.\ polyedra \)). They were unable to demonstrate its location on a conventionally sized minicircle by PCR (as reactions using primers directed outward from the \( psbA \) gene failed to generate a product). Southern hybridization and pulse field gel electrophoresis indicated that the gene was present on a molecule of 50–150 kbp, in an AT-rich fraction. This fraction did not give discrete bands on restriction digestion, however, indicating that even though \( psbA \) may be encoded on a molecule different from minicircles it is not present on a conventional chloroplast genome. In addition, EST databases of \( L.\ polyedrum \) using polyA\(^+\) (and therefore presumably nuclear-encoded) RNA include sequences such as \( psaC \), which is located in the chloroplast genome in all other species (Tanikawa et al., 2004). This is a further indication there is not a conventional chloroplast genome in \( L.\ polyedrum \).

How do we reconcile all these observations? It is possible that there is significant variation in genome organization among dinoflagellates, especially those with chloroplasts of tertiary origin. However, \( L.\ polyedrum \) and \( C.\ horridum \) are both peridinin-containing species, so there is a need to reconcile the observations described above with the existence of minicircles in other peridinin-containing species. A general principle for all these species is that the chloroplast genome is dramatically reduced in coding content, with much of the ‘core’ chloroplast gene complement relocated to the nucleus. In many species, the genes left in the chloroplast are located on minicircles, as already described. It is possible that in some species the remaining chloroplast genes are present in high molecular weight DNA, which is nevertheless different from the conventional chloroplast genome found in other algae and plants and contains a significant amount of non-coding sequence. This DNA could represent the DNA observed by Laatsch et al. (2004) and the 50–150 kbp AT-rich fraction observed for \( L.\ polyedrum \). It will be important to carry out in situ hybridization in a wider range of species, to look for the intracellular distribution of minicircles reported for \( C.\ horridum \). It will also be important to study genes other than \( psbA \) in species such as \( L.\ polyedrum \) to see if their arrangement is similar to that of \( psbA \).

It is possible that some species may contain both minicircles and high molecular weight DNA, with the balance shifting according to environmental or other conditions. Koumandou and Howe (2007) found that the copy number of minicircles in \( A.\ carterae \) varied dramatically with growth phase. Quantitative hybridization analyses indicated that, during exponential growth of a culture, minicircles were present in only a few copies per cell. However, as cell growth slowed, with increasing age of a culture, the copy number rose so that each minicircle was present in tens to hundreds of copies per cell. Could the minicircles be generated from a high molecular weight species and accumulate with increasing age of a culture (perhaps through coupling of transcription and replication as described below)? Although this is an attractive idea, there was no evidence in the experiments of Koumandou and Howe (2007) for a high molecular weight form of the minicircle genes, even at low copy number. Nevertheless the possibility that one exists and is particularly labile, and therefore difficult to detect, cannot be excluded.

### Gene expression

The details of dinoflagellate chloroplast gene expression have not been well studied, not least because of the lack of a chloroplast transformation system. It has been suggested that the core region contains a promoter (explaining why coding regions are always in the same orientation relative to it), but this remains to be demonstrated. The predominant transcripts seen from minicircles correspond reasonably in size to individual genes, even in cases where there are multiple genes on a minicircle (Barbrook et al., 2001) but these may well derive from larger transcripts, corresponding to entire (or nearly so) minicircles, that are subsequently processed (Nelson et al., 2007; Nisbet et al., 2008) and Koumandou and Howe (2007) suggested that replication and transcription are linked, with a transcript being able to serve as a primer for DNA synthesis as with mammalian mitochondria (Lee and Clayton, 1998). This might account for the increase in copy number of minicircles when a culture is growing only slowly. Continued transcription would be needed to replace protein that was turned over. If replication and transcription are coupled, continued transcription would result in continued replication, even though there would be little overall cell growth or division. Little is known about the RNA polymerase responsible for transcription. In algae and land plants there is a multi-subunit eubacterial-type RNA polymerase that is chloroplast encoded, and which is supplemented in land plants by one or more nuclear-encoded phage-type RNA polymerases (Smith and Purton, 2002). No dinoflagellate genes or ESTs have yet been identified for subunits of eubacterial-type RNA polymerases or sigma factors. Given this, and the lack of identifiable –35 and –10 sequences upstream from transcript start sites on minicircles, it is likely that transcription does not involve a typical eubacterial-type RNA polymerase. Whether a phage-type polymerase is used instead remains to be seen. Although an EST for a phage type polymerase in an \( Alexandrium \) EST database has been identified (AC Barbrook, unpublished results; Hackett et al., 2004), it is not clear where the protein would be targeted. In any case, other algal chloroplasts do not appear to use a nuclear
encoded phage-type polymerase (Smith and Purton, 2002).

Zauner et al. (2004) reported extensive substitutional editing of transcripts in *C. horridum*. The most frequent events were A to G transitions, followed by U to C transitions. C to U and G to A transitions were less common, with transversions such as G to C and A to C at even lower frequencies. A similar pattern of editing was reported for *L. polyedrum* (Wang and Morse, 2006a). On the other hand, Barbrook et al. (2001) did not find editing for *A. carterae*. Thus the extent of editing may vary from species to species. Perhaps the most remarkable modification seen for dinoflagellate transcripts is the addition in a number of species of a 3′ polyU tail of 25–40 residues (Wang and Morse, 2006a; Nelson et al., 2007). Wang and Morse suggested that the motifs AGAAA and AAUUA might be used as signals for polyuridylation. The function of the polyU tail is unknown. It has not been reported in any other chloroplasts, although RNAs of trypanosome and myxomycete mitochondria are also polyuridylylated (Adler et al., 1991). In trypanosomes, the polyU tail of guide RNAs may be involved in editing by insertion of U residues (Cech, 1991), but editing is by substitution rather than insertion in dinoflagellate chloroplasts. Poly(U) polymerases have been identified in a range of eukaryotes, including Arabidopsis thaliana, Schizosaccharomyces pombe, Caenorhabditis elegans, and humans (Kwak and Wickens, 2007). Their substrate RNAs are not clear but may include snRNAs and microRNAs.

Little is known about translation in dinoflagellate chloroplasts. As with other chloroplast systems, it appears to be sensitive to chloramphenicol and resistant to cycloheximide, and the synthesis of PsbA has been shown to be light regulated (Wang et al., 2005). Remarkably few tRNAs have been identified on dinoflagellate chloroplast minicircles, as described earlier. Presumably the remainder are imported from the cytosol. Import of tRNAs by chloroplasts has been inferred for other organisms, such as the non-photosynthetic land plant Epifagus virginiana (Wolfe et al., 1992). It is possible that some of the ORFs seen on minicircles may encode divergent ribosomal proteins, including Rpl15 and Rps3 (Barbrook et al., 2006b).

**Evolutionary implications**

Both dinoflagellates and their sister group, the Apicomplexa, have a dramatically reduced mitochondrial genome, with only three protein coding genes (for cytochrome *b* and subunits I and III of cytochrome oxidase) retained, perhaps along with fragmented rRNA genes (Kamikawa et al., 2007; Nash et al., 2007). Chloroplast genome reduction has been taken much further in the dinoflagellate lineage than in the Apicomplexa (Wilson, 2005). Many potential selective advantages have been put forward for the reduction of organelle genomes, such as placing genes in a nuclear gene pool, and the removal of genes from DNA-damaging processes associated with photosynthesis or respiration (Allen and Raven, 1996). Alternatively, the reduction may simply be a non-selective consequence of DNA being able to transfer from the chloroplast to the nucleus but not the other way. Why so many more genes should have been transferred to the nucleus in the dinoflagellates is not clear. It is presumably linked in some way to the dramatic reorganization of the genome, but whether the reorganization facilitated the transfer or whether both phenomena reflect some other event remains to be seen.

The large-scale transfer of genes to the nucleus raises the question of why some genes have been retained in the chloroplast (Barbrook et al., 2006a). The fact that the retained genes generally encode polypeptides of the light reactions of photosynthesis is consistent with the proposal that organelle genomes are retained to allow regulation of expression of a few key genes in response to the redox state of the photosynthetic machinery (Allen, 2003). The genes encode a mix of membrane intrinsic and soluble proteins, suggesting that retention of particular genes is not dictated by the hydrophobicity of the protein they encode (von Heijne, 1986). It has been suggested that some organisms whose mitochondria lack a f-Met tRNA gene obtain the tRNA for mitochondrial protein synthesis from the plastid, and that this places a requirement for the chloroplast to retain the f-Met tRNA gene (Barbrook et al., 2006a). There is indirect experimental evidence that this is so in the Apicomplexa (Howe and Purton, 2007), which are a sister group to the dinoflagellates, and so it is interesting that one of the few tRNA genes detected in dinoflagellate chloroplast minicircles is the f-Met species.

**Outstanding questions**

As discussed above, one of the most important questions is to settle the location of the minicircles, and to determine whether there is indeed an additional DNA species in the dinoflagellate chloroplast. It seems unlikely that such a DNA species will turn out to be a conventional algal-type chloroplast genome with the standard coding complement. The answers to these questions may differ among species. It is also necessary to understand how the replication of the minicircles is controlled (if it is!) and how a full set of molecules is partitioned during cell division. It will be fascinating to determine the function, if any, of the empty and other aberrant minicircles. Many aspects of gene expression deserve attention, such as the role of the polyU tail. It will also be interesting to understand the significance of this large-scale genome
reorganization for the evolution of the group. Perhaps the large-scale transfer of so many genes to the nucleus explains the apparent ease with which dinoflagellates can lose plastids and transiently or permanently acquire new ones (Saldarriaga et al., 2001; Green, 2004; Patron et al., 2006; Larkum et al., 2007). If the cell is unable to guarantee the correct partitioning of minicircles during cell division, the result may be cells that lack a complete set of minicircle genes. If so, the only way for them to survive would be to acquire a new chloroplast from outside, and this would increase the frequency with which dinoflagellates lose plastids in exchange for new ones. Finally, the possibility that dinoflagellate minicircles may be useful as taxonomic markers for studies of environmentally important phenomena such as coral bleaching (Barbrook et al., 2006c) shows that this topic has potential practical significance and is far from an evolutionary curiosity.

Acknowledgements

We thank the Leverhulme Trust, the Biotechnology and Biological Sciences Research Council, and the Newton Trust for financial support.

References


