The Treasure Trove of Algal Chloroplast Genomes. Surprises in Architecture and Gene Content, and Their Functional Implications[w]

Clare L. Simpson and David B. Stern*

Boyce Thompson Institute for Plant Research, Cornell University, Tower Road, Ithaca, New York 14853

OVERVIEW

The completion of the chloroplast genome sequence of the chlorophyte alga Chlamydomonas reinhardtii in our laboratory has been announced recently (J. Maul, J. Lilly, and D.B. Stern, unpublished data; accession no. AF396929). Because C. reinhardtii is the most genetically and biochemically tractable eukaryotic model system for photosynthesis and chloroplast gene expression (for review, see Harris, 2001), it is appropriate to use this opportunity to reflect briefly upon the history of chloroplast genomics—and more importantly, to take a broad and futuristic view as the stage is set for structure/function studies at a level of detail only recently unimaginable.

The first phase of chloroplast genomics culminated with the completion of the tobacco (Nicotiana tabacum) and liverwort (Marchantia polymorpha) chloroplast genome sequences in 1986. The present chapter has witnessed the discovery of new plastid-encoded traits, the use of plastids for foreign gene expression, and an appreciation of their diversity, particularly outside the vascular plants. Two major foci have emerged: functional studies, ranging from details of photosynthesis to gene expression and cell biology; and genomics, whose major goal is to obtain evolutionary and comparative information through sequence analysis. At present, complete genome sequences have been obtained from virtually all the major algal lineages, and the C. reinhardtii sequence and the Synechocystis sp. PCC 6803 genome, representing its presumed ancestor, are complete. We now envision a new chapter of chloroplast molecular genetics, where the evolutionary forces and intracellular mechanisms that shape genome architecture, gene expression, and ecological adaptation, are revealed.

In this Update, we promote algal plastid genomes as an underutilized resource, particularly the completely sequenced ones, through a discussion of structural and coding diversity. Because the antecedents of land plants are thought to lie within the green algal lineage (Turmel et al., 1999a, 2002; Karol et al., 2001), algal plastid genomics also offers useful experimental guides for Arabidopsis, maize (Zea mays), and other model systems. Importantly, some algal cpDNAs have retained novel or key genes that are absent in land plant cpDNAs, which provides an opportunity to use them to determine gene function, instead of dealing with complex nuclear gene families and technical land mines. This leads us to advocate not only C. reinhardtii, but also other algae as useful and perhaps essential complements to plant-based studies of chloroplast biogenesis and leaf development.

ALGAL LINEAGES HAVE DIVERSE ENDO SYMBIOTIC ORIGINS

The algae are a diverse group of oxygen-evolving organisms, whose evolutionary relationships and morphological variation are summarized in Figure 1. In this Update, we focus on the chloroplasts of eukaryotic algae, and compare them with the embryophytes (or multicellular land plants). Phylogenetic relationships of algae have been reviewed recently (Bhattacharya and Medlin, 1998; Douglas, 1998; Gualtieri, 2001). Additional information is provided in Table I, and classifications used here correspond to those of the National Center for Biotechnology Information taxonomy browser (http://www.ncbi.nlm.nih.gov/Taxonomy), except where recent evidence supports an alternative.

A monophyletic origin of all plastids is no longer a matter of extreme controversy (Palmer, 2000). Molecular evidence (e.g. Turner et al., 1999) clearly indicates that chloroplasts arose from the primary endosymbiotic capture of a cyanobacterium. Three lineages bear plastids derived from primary endosymbiosis: embryophytes and green algae (the green lineage), red algae, and glaucocystophytes. Remaining algal divisions carry secondarily derived chloroplasts, resulting from the engulfment of a primary endosymbiont by a eukaryotic host (see Table I and Fig. 1; Cavalier-Smith, 2000). The euglenoids and the chlorarachniophytes bear plastids secondarily derived from a green alga. Lineages derived from secondary endosymbiosis of a red alga are the cryptophytes, the brown algae or chromophytes (also known as heterokonts), some dinoflagellates, and the...
haptophytes. Dinoflagellates are polyphyletic: In this work, we focus on those within the peridinin-containing (red tide-causing) lineage. Haptophyte chloroplast sequence information is currently limited (except Daugbjerg and Andersen, 1997); hence, ecologically significant organisms such as the prymnesiophytes (some algal blooms) are not discussed here. The land plants, along with green, red, and brown algae, and a few oddballs, will be the subjects of this Update.

PLASTID GENOMES HAVE ASSORTED SIZE AND ARCHITECTURE

Most Genomes Are 100 to 200 kb and Circular

Algal genomes are less homogeneous than those of embryophytes (Fig. 1), which average 140 kb and 110 genes (see supplemental data at www.plantphysiol.org). Algal genomes are consistent within lineages, except the green algae. For example, red algal cpDNAs range from 150 to 191 kb. In the greens, however, despite many classical 100- to 200-kb genomes, extremes are found in the group called the Ulvophyceae. Acetabularia mediterranea, which possesses millions of chloroplasts in a 10-cm, uninucleate cell, has an estimated 1,500-kb chloroplast chromosome. To give a sense of scale, this is only 2.4-fold smaller than the Synechocystis sp. PCC 6803 genome. Because much of the A. mediterranea DNA is repeated, it is doubtful that it contains substantially more genes than other cpDNAs. At the other end of the Ulvophyceae spectrum is Codium fragile, with the smallest known cpDNA at 89 kb. Other greens are challenging C. fragile; for example, the picoplankton Nanochloropsis eukaryotum (90 kb), the charophyte Coleochaete orbicularis (approximately 100 kb), and the prasinophyte Pedinomonas minor (98 kb).

The high-profile genus Chlamydomonas is also non-uniform because intergenic sequences grow and shrink. For example, Chlamydomonas moewusii (292 kb) has two large intergenic insertions relative to its close relative Chlamydomonas pitchmanii (187 kb), which also has smaller intergenic spacers. C. reinhardtii (203 kb) is closely related to Chlamydomonas gelatinosa (285 kb), and shuffled gene order exists in addition to intergenic spacer variability. Large num-
bers of dispersed repeat sequences have been found in both species—more than 1,000 in *C. reinhardtii*—and this may have promoted the rearrangements (Boudreau and Turmel, 1996; see also Fig. 2) and, ultimately, an extensive loss of synteny, including most ancestral operons, in step with increasing numbers of promoters and perhaps new regulatory mechanisms.

Algal cpDNAs, like their plant counterparts, have circular restriction maps. Monomeric circles are also

### Table 1. Major divisions of chloroplast-bearing lineages

<table>
<thead>
<tr>
<th>Division (Common Name)</th>
<th>Major Groups (In This Work)</th>
<th>Pigments (Chlorophylls, Light-Harvesting Complex (LHC), or Phycobilisome)</th>
<th>Typical Plastid Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary endosymbionts Streptophyta (land plants and certain green algae)</td>
<td>- Embryophyta (multicellular land plants)</td>
<td>Chlorophyll (Chl) <em>a, b</em>  Light-harvesting complex (LHC)</td>
<td>Plastid surrounded by two membranes  Stacked thylakoids</td>
</tr>
<tr>
<td></td>
<td>- Charophyta (charophyte algae)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Mesostigmataceae (ancestral green algae)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chlorophyta (green algae)</td>
<td>Chlorophyll (Chl) <em>a, b</em>  Light-harvesting complex (LHC)</td>
<td>Plastid surrounded by two membranes  Stacked thylakoids</td>
</tr>
<tr>
<td></td>
<td>- Chlorophyceae</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Prasinophyceae</td>
<td>LHC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Trebouxiophyceae</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Ulvophyceae</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Diverse nos. of flagella</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhodophyta (red algae)</td>
<td>Bangiophyceidae (microalgae, e.g. <em>Cyanidium caldarium</em> but also <em>Porphyra purpurea</em>)</td>
<td>Chlorophyll (Chl) <em>a</em>  Phycobilisome  LHC</td>
<td>Plastid surrounded by two membranes  Unstacked thylakoids; some plastids have peripheral thylakoid that lies parallel to envelope  Phycobilisome confers brilliant red color</td>
</tr>
<tr>
<td></td>
<td>Florideophyceidae (seaweeds)</td>
<td>Morphologically diverse, lack flagella</td>
<td></td>
</tr>
<tr>
<td>Glaucocystophyta (cyanelle-containing algae)</td>
<td>Glaucocystophyceae (<em>Cyanophora paradoxa</em>)</td>
<td>Chlorophyll (Chl) <em>a</em>  Phycobilisome</td>
<td>Two to four cyanelles (plastids)/cell  Unstacked thylakoids, lack plastocyanin, surrounded by cyanobacterium-like peptidoglycan wall, within two membranes of chloroplast envelope</td>
</tr>
<tr>
<td>Secondary endosymbionts derived from red algae Cryptophyta (cryptomonads)</td>
<td>Cryptomonadaceae (<em>Guillardia theta, Cryptomonas falcate</em>) Unicellular, biflagellate, marine, and freshwater algae</td>
<td>Chlorophyll (Chl) <em>a, c</em>  Phycobilisome</td>
<td>Plastids surrounded by two pairs of membranes, small nucleus-like organelle “nucleomorph” resides between pairs, outer plastid membrane continuous with ER, covered with ribosomes  Secondary endosymbionts of eukaryotic red alga (Douglas and Penny, 1999; Oliveira and Bhattacharya, 2000; Fast et al., 2001)</td>
</tr>
<tr>
<td></td>
<td>- Cryptomonadaceae (Guillardia theta, Cryptomonas falcate) Unicellular, biflagellate, marine, and freshwater algae</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chromophyta (the chromophytes, or brown algae, are also known as heterokonts, and photosynthetic stramenopiles)</td>
<td>- Coscinodiscophyceae (centric diatoms)  Chl <em>a, c, c1, c2, c3</em>  Fucoxanthin</td>
<td>Lack phycobilisome</td>
<td>Diverse, polyphyletic group, abundance of fucoxanthin in LHC confers golden color; four membranes, outer two continuous with ER and outer membrane of nucleus  Secondary endosymbionts of eukaryotic red alga (Oliveira and Bhattacharya, 2000; Fast et al., 2001)</td>
</tr>
<tr>
<td></td>
<td>- Chrysophyceae (golden-brown algae)  Chl <em>a, c, c1, c2, c3</em>  Fucoxanthin</td>
<td>Lack phycobilisome</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Xanthophyceae (yellow-green algae)  Chl <em>a, c, c1, c2, c3</em>  Fucoxanthin</td>
<td>Lack phycobilisome</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Phaeophyceae (brown algae)  Chl <em>a, c, c1, c2, c3</em>  Fucoxanthin</td>
<td>Lack phycobilisome</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Raphidophyceae (e.g. <em>Heterosigma akashiwo</em>)  Chl <em>a, c, c1, c2, c3</em>  Fucoxanthin</td>
<td>Lack phycobilisome</td>
<td></td>
</tr>
<tr>
<td>Alveolata: Dinophyceae (dinoflagellates)</td>
<td>Unicellular, biflagellate, marine, and freshwater algae; notorious for “toxic red tides,” due to production of neurotoxins</td>
<td>Chlorophyll (Chl) <em>a, c, c1, c2</em>  Peridinin (carotenoid)</td>
<td>Highly unusual chloroplast genome structure  Envelope has three membranes  Secondary endosymbionts of eukaryotic red alga (Fast et al., 2001)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>From green algae Euglenophyta (euglenoids)</td>
<td>Unicellular euglenoid flagellate algae (<em>Euglena gracilis</em>)</td>
<td>Chlorophyll (Chl) <em>a, b</em>  Light-harvesting complex (LHC)</td>
<td>Envelope has three membranes  Secondary endosymbionts of green alga (Ishida et al., 1997; Ebel et al., 1999)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
easily found, but multimeric or anomalous forms are so far less prevalent than those revealed by fluorescence in situ hybridization of tobacco and Arabidopsis cpDNAs (Lilly et al., 2001). In contrast, the 35-kb plastid genome of Toxoplasma gondii, an apicomplexan parasite, exists in linear tandem arrays, and dinoflagellate minicircles are discussed below. Why some plastid genomes have switched form, but not others, is an enduring evolutionary mystery.

The rDNA Repeat Exists in Many Versions

A landmark feature of many plastid genomes is an rRNA gene-containing IR, which conveniently divides the genome into so-called large and small single-copy regions. The prototypical rRNA operon in embryophyte chloroplasts is 20 kb and consists of 16S and 23S genes separated by a spacer encoding two tRNAs, with 4.5S and 5S genes immediately downstream. However, algae often stray from this formula. For example, some red and green algae have a single rRNA operon, much as in legumes, where one copy of the IR has been lost. Algae also may have separate small and large subunit operons, transcribed from opposite strands, or divided by massive intergenic spacers. Euglena gracilis has four tandem direct rRNA repeats. In fact, within the red algae, the classic IR-large single copy-small single copy structure is limited to Porphyra yezoensis (Shivji et al., 1992). The presence of non-ribosomal genes in the IR is mostly limited to vascular plants and green algae, which explains the tendency of the IR to be significantly larger in these species. N. olivacea adds extensive non-coding regions, enlarging it still further (Fig. 2). The IR has been proposed as an ancestral chloroplast feature (Turmel et al., 1999b), and a similar version is present in Synechocystis sp. PCC 6301. However, a look at the broader picture suggests that the IR may have been gained and lost on multiple occasions.

The function of the IR could be to increase the relative copy number of rRNA genes, however many plastid genomes thrive without this feature. One proposal that ascribes a parasitic origin to the IR (Glock-
Some Dinoflagellate Genomes Are Composed of Single-Gene Minicircles

Dinoflagellates are best known for causing red tides, but some also have highly unconventional plastid DNA. Instead of single genomes, each gene is contained on its own 2- to 3-kb, separately replicated minicircle (Zhang et al., 1999). This organization is reminiscent of kinetoplast (mitochondrial) DNA in trypanosomes, which have one main genome and numerous catenated minicircles, which encode guide RNAs required for RNA editing (for review, see Simpson et al., 2000). However, dinoflagellates do not appear to have a main genome or RNA editing, and their genome fragmentation has been accompanied by rapid sequence divergence and gene loss. In fact, a close relationship of dinoflagellate plastids to apicomplexan plastids has been evoked (Zhang et al., 2000), with both exhibiting genome reduction and unusual architecture—a feature both also possess in their mitochondria. This leads us to ask whether plastid genome reduction and unusual architecture are widely correlated. If we define normal as a compact circle comprising 120 or more genes, perhaps so. Taking the popular models E. gracilis and C. reinhardtii as examples, we find only 87 and 96 genes, respectively, including tRNAs. At the same time, both have an atypically polar gene distribution (see Fig. 2), and in the case of E. gracilis, a pervasive invasion of introns—as many as 150. In C. reinhardtii, we find many hundreds of small dispersed repeats and diverging gene fragments (Fig. 2). The reader is referred to the primary literature for other tidbits on the fascinating and biologically important dinoflagellates.

Some Algal Plastid Genomes Are Dense and Gene Rich

Outside the green lineage, gene density is uniform and large non-coding sequences, introns, and pseudogenes are exceptional. The 122-kb G. theta genome may be the epitome of compactness: It encodes 180 genes, which account for 90% of its DNA. Cyanophora paradoxa (182) and Odontella sinensis (165) are equally gene rich, among others (see supplemental data at www.plantphysiol.org). Although plant cpDNAs have generally retained subsets of the genes required for photosynthesis and gene expression, nongreen algal cpDNAs frequently encode additional functions, a few of which are highlighted below as tempting targets for in-depth study. In fact, the two fully sequenced red algal cpDNAs each encode 30 to 40 unique genes, which may have novel functions.

What is the significance of increased coding capacity? One could speculate that keeping genes in the plastid versus transferring them to the nucleus is an adaptive advantage in certain environments. Perhaps less romantic is the notion that the transfer of plastid genes to the nucleus is simply slower in the red lineage, which is among the evolutionarily oldest. There are further twists in the reds: Even the largest genomes, for example P. purpurea, lack genes that are ubiquitous in embryophyte cpDNAs, such as infA and the ndh genes. P. purpurea also lacks clpP, which is retained in the highly reduced cpDNA of the holoparasite Epifagus virginiana (Wolfe et al., 1992). Given all these perturbations, it is hard to find rhyme or reason in biochemical explanations. On the other hand, gene content reflects interesting evolutionary forces and, in some cases, gives clues as to functions retained or lost in the organism.

It is also worth noting the contribution of horizontal transfer (the transfer of genes from other species) to plastid genome content. Most plastids can be assigned to either of two groups, based on the presence of green type (form I) or red type (form II) Rubisco, where the former is of cyanobacterial origin and has been inherited vertically (through “traditional” evolution), and the latter is of purple bacterial origin and entered chloroplast genomes of the red lineage (red algae, and most secondary endosymbionts) horizontally, although this is an oversimplification that does not take into account a very complicated evolutionary history (Delwiche and Palmer, 1996). In addition, certain dinoflagellates have form II Rubisco (Palmer, 1996). Mechanisms underlying gene transfer to the nucleus, and hypotheses about the associated evolutionary forces are much debated (for review, see Martin et al., 1998; Race et al., 1999; Blanchard and Lynch, 2000; Selosse et al., 2001).

In the overall chloroplast genome picture, where do green plants stand? Because they were studied first and more intensively, many of us find normalcy in their conserved gene clusters, nearly uniform transfer of genes to the nucleus, and fairly constant coding capacity. However, the exceptional preservation of gene arrays in green plants and the early diverging green alga Mesostigma viride is not a feature of the other green algae, including Chlorella vulgaris. Thus, we can argue that vascular plant chloroplast genomes are actually unusual—giving us impetus to delve more deeply into other groups. The newly sequenced genome of C. reinhardtii, which has a reduced coding capacity and atypical genome organization, has already stimulated much debate on whether nuclear genes that regulate plastid gene expression will truly be orthologous with those of higher plants. So far, the answer is usually “no,” with the one interesting exception being a nuclear gene
ALGAL CHLOROPLAST GENOMES MAY GIVE NEW INSIGHTS INTO CHLOROPLAST PROCESSES

Chloroplast gene expression in C. reinhardtii and vascular plants has been the subject of numerous recent reviews (e.g., Barkan and Stern, 1998; Goldschmidt-Clermont, 1998; Zerges, 2000; Cahoon and Stern, 2001). Chloroplast gene expression in most other algae has not been comprehensively investigated, and here we use them to draw attention to potentially exciting targets for reverse genetics. Given the proper molecular tools, substantial progress will be possible in both a basic and applied sense.

Two-Component Transcriptional Regulators Are Found in Some cpDNAs

Although transcription rates change globally during chloroplast biogenesis, regulation by classic “on/off” switches is unknown. For this reason, we draw attention to several two-component transcriptional regulators encoded by nongreen algal cpDNAs. Two-component regulators are widespread in bacteria, including Synechocystis, and mediate responses to environmental changes such as osmolarity and nutrient deficiency. Although two-component systems are redox-responsive (Oh and Kaplan, 2000), as has been proposed for plant chloroplast transcription (Pfannschmidt et al., 1999), two-component genes had not been found in cpDNA, nor have nucleus-encoded chloroplast-localized versions been identified (see below). Therefore, the nongreen algae may have an unexpected regulatory mechanism, an unrecognized form of which might occur in land plants (for information on other potential transcriptional regulators encoded in plastid DNA, see supplemental data at www.plantphysiol.org).

A canonical Escherichia coli two-component system is the EnvZ-OmpR osmolarity response phosphorelay, where EnvZ is a cytoplasmic membrane sensor kinase and OmpR is its substrate and the regulatory transcription factor. EnvZ homologs are found in red plastid DNA, whereas OmpR homologs are found in all phycobilin-containing algae, i.e. the reds plus cryptomonads and C. paradoxa. Gene disruption experiments have implicated two Synechocystis sp. PCC 6803 relatives of ompR in energy transfer between phycobilisomes, although whether they act as transcriptional regulators is unknown (Ashby and Mullineaux, 1999). If the red algal genes are transcriptional regulators, it will be exciting to find their stimuli and targets; for example, by measuring transcription rates over the whole genome using arrays, under different growth conditions.

ALGAL CHLOROPLAST GENOMES MAY GIVE NEW INSIGHTS INTO CHLOROPLAST PROCESSES

that regulates the only operon conserved between plants and C. reinhardtii (Vaistij et al., 2000; Felder et al., 2001).

Another putative transcriptional regulator is cxQ. What stands out is its apparent clustering in an expression module with its target genes, in the prokaryotic style. In most nongreen cpDNAs, cxQ is part of the rbcLS operon. In turn, cxQ is the counterpart of purple bacterial cbbX genes, which encode ATP-binding polypeptides necessary for photoautotrophic growth, thought to be involved in expression of Rubisco genes (Maier et al., 2000). The implication, then, is that CfxQ may transcriptionally regulate rbcL and/or rbcS in nongreen plastids, perhaps diurnally.

If two-component systems are common in nongreen, what has become of them in the chlorophyte lineage? Certainly, plants contain many such regulators encoded in the nucleus; for example, in hormone responses (for review, see Urao et al., 2000). However, none so far has been found to be organelle targeted. Taken at face value, this implies that two-component control of chloroplast transcription is extinct in green plants and algae, but we hesitate to draw hasty conclusions before significantly more complete chloroplast proteomics has been accomplished.

Algal Plastids Encode a Variety of RNA Maturation and Decay Functions

A Ribozyme Component for tRNA Processing

RNAse P is the tRNA 5′ maturation endonuclease, containing an essential RNA subunit in bacteria, archaea, and some eukaryotes, whereas spinach (Spinacia oleracea) chloroplast RNAse P lacks RNA (Schön, 1999). Therefore, it was surprising to discover rnpB, which encodes the RNA subunit, in diverse cpDNAs such as N. ollvacea, C. paradoxa, and P. purpurea. In C. paradoxa, the cyanelle RNAse P holoenzyme is destroyed by nuclease treatment, indicating an essential role of the RNA component (Baum et al., 1996). It will be interesting to discover whether algae lacking a plastid rnpB gene also lack an RNAse P RNA component. If they do turn out to have an RNAse P RNA, it would have to be imported from the cytoplasm. Although chloroplast RNA import has not been demonstrated, import into mitochondria of both RNAse P RNA and its relative, MRP RNA, has been well-documented in animals (Puramun and Attardi, 2001).

Other Small Non-Coding RNAs Are Found in Plastids

One small RNA, so far unique to C. reinhardtii psaA, is the tscA trans-splicing transcript. In chloroplasts, splicing occurs either in the familiar cis pathway or in trans, where a complete intron forms through hybridization of separately transcribed molecules. The tscA RNA forms part of the first intron, which is therefore tripartite (the other two RNA molecules encode exon 1 and the first domain of intron 1, and the final domains of intron 1 and exon 2; Gold-
In bacteria, division proteins interact with DNA-binding proteins, partly coordinating DNA synthesis and cell division. Two relevant genes found in algal cpDNAs are hlpA and dnaB. hlpA is unique to cryptomonads and the primitive red microalga Cyanidioschyzon merolae, and recombinant cryptomonad HlpA behaves as a functional homolog of the E. coli DNA-binding and -bending HU- and HMGI-like proteins (Grasser et al., 1997). HU helps determine global nucleoid (chromosome) structure, which is required among other things for cell division, via interactions with MinCDE (Jaffe et al., 1997). The colocalization of division (minDE) and DNA packaging (hlpA) functions in G. theta cpDNA is unique outside bacteria, providing appealing targets for reverse genetics.

In E. coli, DnaB unwinds DNA at replication forks, and dnaB is found in most nongreen algal cpDNAs. Chloroplast dnaB genes are curious because although they have close relatives in cyanobacteria, no other chloroplast-localized DnaB proteins are known. For example, the only apparent Arabidopsis dnaB gene is nuclear, and its product appears to be mitochondrially targeted (Leipe et al., 2000). Thus, dnaB may have been lost from most chloroplast genomes without having been stably transferred to the nucleus, perhaps having been replaced by a novel activity.

An Apicoplast Gene Found in Algae

Algal cpDNAs can provide desperately needed clues about potential apicoplast targets in parasites such as Plasmodium falciparum, which causes nearly one-half billion cases of malaria each year. For example, ycf24, a ubiquitous bacterial gene also found in all nongreen cpDNAs, corresponds to P. falciparum open reading frame (ORF) 470. Some intriguing Synecocystis sp. PCC 6803 results suggest that ycf24/ ORF470 might be involved in plastid division, and thus an essential housekeeping gene (Law et al., 2000). Naturally, one would also like to disrupt the gene in the algae themselves, to see whether ycf24 is involved in chloroplast division or has some other function.

Proteolysis Uses Both Protein and RNA Components

FtsH is a metalloprotease, chloroplast-encoded by all nongreen algal cpDNAs, but part of a nuclear gene family in higher plants. Large ORFs in green algal and embryophyte genomes have partial FtsH homology. FtsH proteins play an important yet poorly understood role in chloroplast biogenesis, and because their functions appear to be complex and redundant in higher plants (Chen et al., 2000), we suggest that the simpler algae could be used to study FtsH, although they might have additional nucleus-encoded family members.

Some algal chloroplasts encode tmRNAs, small RNAs that combine the functions of tRNA and...
mRNA, and are widespread in bacteria. Ribosomes at the 3’ end of damaged mRNAs translate the proteolysis-inducing tag encoded on the tmRNA and are then released, allowing the polypeptide to be digested. The tmRNAs in several algal cpDNAs can be found in the tmRNA database, and represent a fascinating mode of proteolysis unknown elsewhere in eukaryotes.

GENOMICS

Algal Chloroplast Genomics Is Biologically Relevant

There are compelling ecological, economic, and medical reasons to pursue algal chloroplast molecular biology. Certain groups of algae have a major role as primary producers (phytoplankton), or as components of the human diet and sources of derived products (e.g. alginates, which are used widely in frozen foods and salad dressings). Global warming is potentially significantly impacted by the carbon fixation of oceanic phytoplankton such as green algae, diatoms, and haptophytes, but these processes in the marine biome remain incompletely understood. A report of C_4 photosynthesis based on phosphoenolpyruvate carboxylase activity and ^14C incorporation into C_4 compounds in a diatom (Reinfelder et al., 2000) has been provocative. It remains to be seen how key elements of a C_4 system such as the cellular partitioning seen in C_3 land plants would be configured in eukaryotes.

Gene Discovery Will Be a Major Activity

Overall genomic resources have a major impact on the experimental utility of any organism. In the case of the algae, the best endowed system is C. reinhardtii, for which >170,000 expressed sequence tags exist, mostly in the context of two major projects (Asamizu et al., 2000; Harris, 2001). Because much of chloroplast biology involves nucleus-encoded proteins, the genes encoding them should not be neglected. Array technologies will surely be applied; for example, to identify important genes whose expression is altered in response to known chloroplast mutations or environmental changes. A pilot array has been developed in our laboratory for C. reinhardtii that includes both organellar genes and nuclear genes encoding organellar proteins (for the most recent updates, see the C. reinhardtii genome project chloroplast page (http://www.biology.duke.edu/chlamy_genome/chloro.html).

Chloroplast Transformation Is a Key Technology

Using new organisms for molecular studies requires certain technologies, and a key to some of the experiments envisioned here will be chloroplast transformation, which has been routine for some time in C. reinhardtii and tobacco. Chloroplast transformation occurs by homologous recombination, permitting a reproducibility and precision of results that have led to profound insights. The use of this technology to dissect the role of ycf s is reviewed by Davies and Grossman (1998). Additional applications, including site-directed mutagenesis of chloroplast genes, are reviewed by Simpson and Stern (2001). Chloroplast transformation mostly uses the gene gun, but polyethylene glycol-mediated DNA delivery has also been used (Kofer et al., 1998). Other considerations include selectable markers, obtaining homoplasmicity, and the need to deal with complex life cycles. Although there are barriers, we note recent reports of chloroplast transformation in E. gracilis (Doetsch et al., 2001), and the first stable chloroplast transformation of a red microalga, Porphyridium sp. (Lapidot et al., 2002). Perhaps the first targets among the organisms discussed here should be unicellular algae, especially those with single chloroplasts such as Chlorella vulgaris and Cyanidium caldarium, followed by more challenging ventures.

PERSPECTIVE

The algae recommend themselves to the biologist for their economic and ecological importance. They are a major component of the Earth’s biomass, responsible for significant carbon fixation, and occupy an extreme diversity of climatic niches. To the scientist interested in genome evolution, chloroplast biogenesis and functional genomics, the algae are a rich source of experimental material that has significantly accrued as sequencing technology has accelerated. Because of space limitations, we have only highlighted a few of the genes and diverse organisms worth pursuing for the biologist: The reader is referred to the primary literature and Web-only supplements for further details and ideas.

Received October 5, 2001; accepted February 17, 2002.

LITERATURE CITED

Simpson and Stern


Turmel M, Otis C, Lemieux C (2002) The complete mitochondrial DNA sequence of *Mesostigma viride* identifies this green alga as the earliest green plant divergence and predicts a highly compact mitochondrial genome in the ancestor of all green plants. Mol Biol Evol 19: 24–38


