Current Biology
Dispatches

Evolution: In Chloroplast Genomes, Anything Goes

David Roy Smith
Department of Biology, University of Western Ontario, London, ON, Canada
Correspondence: dsmit242@uwo.ca
https://doi.org/10.1016/j.cub.2017.10.049

A new study shows that Cladophorales green algae have the most unconventional chloroplast DNAs ever observed, whereby genes are located on small linear single-stranded palindromic elements. This puzzling architecture has parallels with mini-circular chloroplast genomes of dinoflagellates and raises many questions about how it arose and is maintained.

> *"Improvement makes straight roads, but the crooked roads without improvement, are roads of genius."*
>
> —William Blake

When it comes to weird genomes, chloroplast DNAs (ptDNAs) are nearly always outdone by their mitochondrial counterparts [1]. Indeed, the 100–200 kb circular-mapping structure that typifies most ptDNAs is unremarkable, rarely inspiring a second glance. Mitochondrial genomes, on the other hand, regularly break well-established rules in genetics, and come in every shape and size imaginable [1,2]. But in this issue of *Current Biology*, Del Cortona and colleagues [3] present a ptDNA that could give even the most extreme mitochondrial DNA (mtDNA) a run for its money.

Using a next-generation sequencing approach, the authors describe what is perhaps the most unusual chloroplast genome yet observed, one in which all the genes are found on small linear single-stranded chromosomes with elaborate secondary structures. Yes, you read that right — a ptDNA in the form of a shattered jigsaw puzzle.

This strange genomic architecture belongs to the little-known marine multicellular green alga *Boodlea composita* (Figure 1), which is found in an equally poorly understood ulvophycean order, the Cladophorales, whose members are renowned for having giant multinucleated cells containing numerous chloroplasts [4]. Altogether, 34 *bona fide* ptDNA chromosomes were identified in *Boodlea*. They vary in length from about 1–7 kb, were all shown to be transcriptionally active and, with a few exceptions, contain only a single gene apiece.

The arrangement of genes on these single-stranded elements is surely unlike anything you have heretofore seen. Each chromosome contains an inverted repeat and is organized into a perfect or imperfect palindrome, meaning that the coding and noncoding regions fold onto themselves forming a bobby-pin-like secondary structure. The inverted repeat sequence is similar among the different chromosomes and was also found inside dozens of contigs lacking any identifiable chloroplast genes, suggesting that the *Boodlea* chloroplast harbours 'empty' ptDNA molecules in addition to gene-containing ones, and that the true chromosome number might be much greater than 34.

Normally, when a study presents for the first time such a complex and unconventional genome, the sequencing and assembly data need to be backed up by detailed molecular biology work. In this case, the authors are lucky in that much of the painstaking bench work has already been carried out by previous researchers [5–8]. The high-throughput sequencing analyses described by Del Cortona *et al.* are, therefore, all the more convincing when placed alongside these earlier pioneering experiments on Cladophorales, including ones clearly demonstrating the presence of single-stranded plasmid-like DNAs within chloroplasts [6–8].

If the story of the *Boodlea* ptDNA ended here, it wold be a classic. But the list of eccentricities goes on. In addition to a fragmented architecture, it has one of the highest guanine and cytosine contents (57%) ever recorded for a chloroplast genome [9]. It also employs a nonstandard genetic code — a rarity among ptDNAs — whereby the canonical termination codon 'UGA' has a double meaning, acting as a stop codon in some genes and representing an amino acid in others, but precisely what amino acid it encodes remains to be determined. Equally as extraordinary, the 16S rRNA gene is split in two pieces, each located on a different palindromic element.

Despite considerable effort, the authors were not able to uncover a chloroplast 5S or 23S rRNA in the sequencing data, nor could they find a single chloroplast tRNA. If these genes are truly missing from the *Boodlea* ptDNA (and are imported from somewhere else), it would be, to the best of my knowledge, the first example of such wholesale loss of RNA-coding regions from a photosynthetic chloroplast. The *Boodlea* chloroplast appears to have proclivity for shedding protein-coding genes as well — only 21 were detected in the ptDNA contigs, all but one (*rbcL*) representing components of the major thylakoid transmembrane protein complexes, and some of which are duplicated and located on different chromosomes. The deduced amino acid sequences of these 21 genes are extremely divergent compared with their orthologues in other photosynthetic organisms, which is a recurring theme among bizarre organelle genomes [1].

Green algal ptDNAs typically encode more than 50 different proteins [10]. So, what has happened to the missing protein-coding genes from the *Boodlea* chloroplast? It seems they have been relocated to the nuclear genome. Del Cortona *et al.* identified the transcripts of an additional 66 proteins known to be chloroplast-encoded in other algae, but they were all deemed to be nuclear encoded and chloroplast targeted based on various sequence-based features, such as a high mRNA to total-RNA read ratio. Why *Boodlea* has been so successful at moving chloroplast genes to

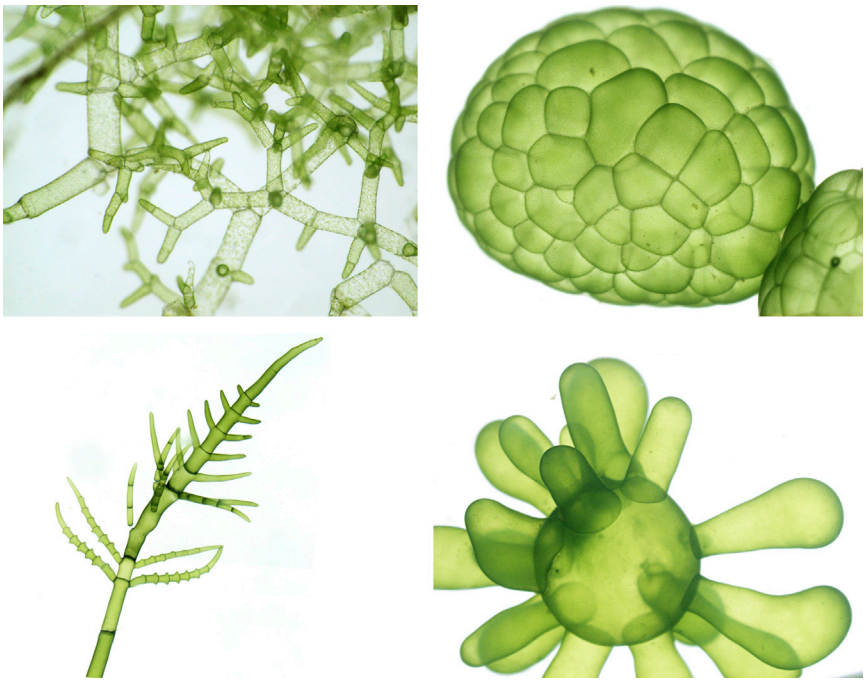

Figure 1. Cladophorales green algae.
Clockwise from top left: *Boodlea composita*, *Dictyosphaeria cavernosa*, *Strucea elegans*, *Valonia utricularis*. Photos by Frederik Leliaert.

the nuclear genome is a mystery but might be related to the fact that, unlike a number of other algae, it has numerous chloroplasts per cell [11].

If the depiction of the *Boodlea* ptDNA leaves you craving more data from strange chloroplast genomes, don't fret. Cursory sequence analyses of nine other species representing main lineages within the Cladophorales (Figure 1) indicate that a fragmented ptDNA architecture is a common feature of the entire order. It is noteworthy that other kinds of green algae can have extravagant chloroplast genomes [1], although nothing that comes close to that seen in *Boodlea*. The ulvophyte *Acetabularia acetabulum*, for example, is estimated to have a ptDNA in excess of 1 Mb, with a high noncoding content (>85%) [12], and possibly small circular plasmid-like molecules that exist alongside the conventional chloroplast genome [13]. The chlorophycean green alga *Koshicola spirodelophila* has a fragmented ptDNA, but the level of fragmentation is minor: three large circular-mapping chromosomes with a cumulative length of 385 kb, and a standard gene content [14].

The only known algae with ptDNAs that are somewhat similar to that of *Boodlea* are peridinin-containing dinoflagellates, whose chloroplasts arose from a secondary-endosymbiosis with a red alga and have highly fragmented genomes comprising miniature (1–12 kb) circular chromosomes with zero, one, or multiple genes, depending on the species [15,16]. Like the *Boodlea* ptDNA, dinoflagellate minicircles encode components for the major photosynthetic complexes but lack protein-coding regions involved in gene expression; but unlike *Boodlea*, they are made up of AT-rich, double-stranded DNA and often undergo RNA editing [15,16].

In the case of *Boodlea*, one must ultimately ask: how could any self-respecting chloroplast genome find itself in such a chaotic and precarious state? Del Cortona *et al.* believe that the nuclear genome might be partly at fault. Some of the same non-coding motifs on the chloroplast chromosomes were present exclusively on nuclear-derived sequencing reads containing long terminal repeat retrotransposons (RT-LTRs). One take on this is that DNA transfer from the nucleus to the chloroplast seeded the ptDNA with RT-LTRs resulting in the expansion of the chloroplast genome and its subsequent fragmentation into hairpin elements (through recombination between repeats and displacement of the palindromic sequences from the lagging strand during replication).

Whatever the roots of this remarkable genome architecture, they reinforce a growing idea in organelle genomics — that when things go wonky, they go really wonky [1]. The reasons for this are unknown but might be related to error-prone and capricious organelle DNA maintenance machinery, most of which are nuclear encoded, and sometimes dual targeted to the chloroplast and mitochondrion [1].

And what of the *Boodlea* mitochondrion — does it, too, have a peculiar genome? The authors located 52 contigs likely representing mtDNA. Together, these sequences hint at a very large mitochondrial genome with many repeats (but no RT-LTRs), a high GC content, and a reduced coding capacity. But, thankfully, the mtDNA does not, for now, appear to be nearly as cool as the ptDNA (score one for the chloroplast).

A popular saying in the field of organelle biology is "mitochondrial genomes: anything goes" [17]. With the work of Del Cortona *et al.*, it is now fair to say the same of chloroplast genomes — anything goes.

REFERENCES

1. Smith, D.R., and Keeling, P.J. (2015). Mitochondrial and plastid genome architecture: reoccurring themes, but significant differences at the extremes. Proc. Natl. Acad. Sci. USA *112*, 10177–10184.

2. Burger, G., Gray, M.W., and Lang, B.F. (2003). Mitochondrial genomes: anything goes. Trends Genet. *19*, 709–716.

3. Del Cortona, A., Leliaert, F., Bogaert, K.A., Turmel, M., Boedeker, C., Janouskovec, J., Lopez-Bautista, J.M., Verbruggen, H., Vandepoele, K., and De Clerck, O. (2017). The plastid genome in Cladophorales green algae is encoded by hairpin chromosomes. Curr. Biol. *27*, 3771–3782.

4. Leliaert, F., and Boedeker, C. (2007). Cladophorales. In Green Seaweeds of Britain and Ireland, J. Brodie, and C.A. Maggs, eds. (The British Phycological Society), pp. 131–183.

5. La Claire, J.W., Zuccarello, G.C., and Tong, S. (1997). Abundant plasmid-like DNA in various members of the orders Siphonocladales and Cladophorales (Chlorophyta). J. Phycol. *33*, 830–837.

6. La Claire, J.W., and Wang, J. (2000). Localization of plasmidlike DNA in giant-celled marine green algae. Protoplasma *213*, 157–164.

7. La Claire, J.W., Loudenslager, C.M., and Zuccarello, G.C. (1998). Characterization of novel extrachromosomal DNA from giant celled marine green algae. Curr. Genet. *34*, 204–211.

8. La Claire, J.W., and Wang, J.S. (2004). Structural characterization of the terminal domains of linear plasmid-like DNA from the green alga *Ernodesmis* (Chlorophyta). J. Phycol. *40*, 1089–1097.

9. Smith, D.R. (2012). Updating our view of organelle genome nucleotide landscape. Front. Genet. *3*, 175.

10. Lang, B.F., and Nedelcu, A.M. (2012). Plastid genomes of algae. In *Genomics of Chloroplasts and Mitochondria*, Volume 35, R. Bock, and V. Knoop, eds. (Dordrecht: Springer Netherlands), pp. 59–87.

11. Smith, D.R., Crosby, K., and Lee, R.W. (2011). Correlation between nuclear plastid DNA abundance and plastid number supports the limited transfer window hypothesis. Genome Biol. Evol. *3*, 365–371.

12. de Vries, J., Habicht, J., Woehle, C., Huang, C., Christa, G., Wägele, H., Nickelsen, J., Martin, W.F., and Gould, S. (2013). Is *ftsH* the key to plastid longevity in sacoglossan slugs? Genome Biol. Evol. *5*, 2540–2548.

13. Green, B.R. (1976). Covalently closed minicircular DNA associated with *Acetabularia* chloroplasts. Biochim. Biophys. Acta *447*, 156–166.

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

15. Howe, C.J., Nisbet, R.E.R., and Barbrook, A.C. (2008). The remarkable chloroplast genome of dinoflagellates. J. Exp. Bot. *59*, 1035–1045.

16. Wisecaver, J.H., and Hackett, J.D. (2011). Dinoflagellate genome evolution. Annu. Rev. Microbiol. *65*, 369–387.

17. Burger, G., Gray, M.W., and Lang, B.F. (2003). Mitochondrial genomes: anything goes. Trends Genet. *19*, 709–716.

Microbial Biodiversity: Straight from the Dolphin's Mouth

Adrienne Narrowe[1,*], Christopher S. Miller[1,*], and Catherine Lozupone[2,*]
[1]Department of Integrative Biology, University of Colorado Denver, Denver, CO 80217, USA
[2]Department of Medicine, University of Colorado Denver, Denver, CO 80045, USA
*Correspondence: adrienne.narrowe@ucdenver.edu (A.N.), chris.miller@ucdenver.edu (C.S.M.), Catherine.Lozupone@ucdenver.edu (C.L.)
https://doi.org/10.1016/j.cub.2017.10.068

Advances in metagenomic sequencing and bioinformatics have vastly expanded our knowledge of microbial phylogenetic and functional diversity. In this issue, Dudek *et al.* show that shotgun metagenomic sequencing of a less-well-studied environment — dolphin gums — uncovers surprising novelty in the bacterial tree of life, underscoring the promise of future discovery.

When thinking of biodiversity, certain locales quickly come to mind: the Amazon rain forest, the deep ocean, or the plains of the Serengeti. At the microbial scale, an Amazon's worth of diversity is regularly found in a myriad of sometimes surprising habitats—for example, in near-boiling hot springs [1], your showerhead [2], kilometers underground [3], or even in the human belly button [4]. However, some habitats have been especially revealing, both in terms of the total amount of biodiversity present and in the degree of novel diversity that characterization of these environments has revealed. For instance, probing microbial diversity in the hypersaline microbial mats of Guerrero Negro just over a decade ago led to the discovery of 15 novel candidate phyla [5]. Dudek *et al.* [6] now add to this list of unusual habitats of high diversity, with a genome-enabled survey of a place few would have thought to look—the gums of a dolphin. In doing so, they produced a plethora of novel genomes, including many from bacterial candidate phyla (deep branching lineages that lack cultured isolates). This work significantly expands our knowledge of the phylogenetic and metabolic diversity of a large part of the bacterial tree of life and fleshes out our understanding of these poorly understood lineages in a host-associated community.

With the introduction of methods to amplify and sequence ribosomal small subunit (16S) RNA genes to explore microbial communities *in situ*, studies of microbial ecology and biodiversity underwent a molecular revolution, during which known microbial phylogenetic diversity exploded [7]. However, novel lineages known only from 16S rRNA-targeted sequencing lack genomic representatives, such that their potential functional significance cannot be predicted. Additionally, over time it has become apparent that 16S rRNA gene-based studies are hindered by the lack of truly 'universal' PCR primers [8]. Because primers are designed based on known database sequences, there is an inherent risk that we are missing unknown biological diversity with such an approach [9,10]. In recent years, advances in DNA-sequencing technology and bioinformatics have enabled a second molecular revolution in microbial ecology, this time leveraging shotgun metagenomics and genome assembly. Genome-centric metagenomics methods that circumvent 16S rRNA-sequencing bias have led to the discovery of vast swaths of previously hidden microbial phylogenetic and, notably, functional diversity. For example, repeated deep metagenomic sequencing at a single subsurface aquifer has revealed dozens of new phyla, including those in the bacterial Candidate Phyla Radiation