




Complete mitogenome of the streptophyte green alga *Coleochaete scutata* (Coleochaetophyceae)

Monique Turmel , Christian Otis  and Claude Lemieux 

Département de Biochimie, de Microbiologie et de Bio-Informatique, Institut de Biologie Intégrative et des Systèmes, Université Laval, Québec, Canada

ABSTRACT

We have sequenced the mitogenome of *Coleochaete scutata* strain SAG 110.80M. This mitogenome is the largest among the streptophyte green algae examined to date. At 242,024 bp, it is 4.3-fold larger than the mitogenome of *Chaetosphaeridium globosum*, the only other mitogenome available for the Coleochaetophyceae. This size difference is mainly explained by differences in the abundance of introns and in the length of intergenic regions containing vestiges of coding sequences thought to be of foreign origin. With 31 group I and 26 group II introns, the *C. scutata* mitogenome is the most intron-rich organelle genome known among streptophyte algae.

ARTICLE HISTORY

Received 31 October 2019
Accepted 9 November 2019

KEYWORDS

Coleochaetophyceae; horizontal transfers; mitochondrial genome; introns; phylogenomics; Streptophyta

Previous studies on the mitogenomes of streptophyte green algae provided insights into the evolutionary dynamics of this genome within five of the six lineages recognized as classes (Turmel et al. 2007, 2013). Such information could not be obtained for the Coleochaetophyceae, which comprises one order and two families (Coleochaetaceae and Chaetosphaeridiaceae), because a single taxon (*Chaetosphaeridium globosum*) had been examined (Turmel et al. 2002). Here, we report the mitogenome sequence of *Coleochaete scutata* and compare it with its *Chaetosphaeridium* homolog.

The axenic strain SAG 110.80M of *C. scutata* was obtained from the Sammlung von Algenkulturen Göttingen and total DNA was isolated using the EZNA HP Plant Mini Kit of Omega Bio-Tek (Norcross, GA, USA). Paired-end reads were generated on the Illumina MiSeq sequencing platform by the 'Plateforme d'Analyses Génomiques' of Laval University. Reads (3.4 million) were assembled using Ray version 2.1.0 (Boisvert et al. 2010) and contigs were visualized, linked, and edited using CONSED version 22 (Gordon et al. 1998). Contigs of mitochondrial origin were identified by BLAST searches against a local database of organelle genomes. Gene annotations were performed as described by Turmel et al. (2017).

At 242,024 bp, the *C. scutata* mitogenome (GenBank: MN613583) is 4.3-fold larger than its *Chaetosphaeridium* homolog, but its gene repertoire is essentially the same. It possesses 67 distinct genes, three of which (*rnn5*, *trnMf*(cau) and *trnP*(ugg)) are duplicated or triplicated. The *rpl6* gene is missing from *C. scutata*, which represents the main difference

in gene content between the two coleochaetophycean mitogenomes. Gene order is highly scrambled between these genomes. The larger size of the *C. scutata* mitogenome is mostly accounted for by a higher abundance of introns (57 versus 11 in *Chaetosphaeridium*) and by the presence of longer intergenic regions near *rns* (56.1 kb versus 8.3 kb). In both algae, the latter regions harbor ORFs showing similarities (*E*-value threshold of 1e-15) with recognized domains of phage/plasmid DNA primase and phage integrase. Considering that genes with these functions are not usually found in green algal mitochondria, we suggest that they were acquired by lateral transfers before the emergence of the Coleochaetophyceae. The *C. scutata* mitogenome contains 31 group I and 26 group II introns, a significant proportion of which (11 group I and 20 group II introns) encode LAGLIDADG homing endonucleases and reverse transcriptases/intron maturases, respectively. The *cox1* and *rnl* genes alone exhibit a total of 19 and 14 introns, respectively. Twenty-four of the *C. scutata* introns are inserted at sites where no introns have been observed in any of the currently available mitogenomes of streptophyte algae, supporting the view that mitochondrial introns proliferated in the Coleochaetaceae.

A maximum-likelihood phylogeny was inferred from 40 concatenated mitogenome-encoded proteins of 22 streptophyte taxa using RAxML version 8.2.3 (Stamatakis 2014) essentially as described in Turmel et al. (2013). The best-scoring tree is congruent with the streptophyte mitochondrial tree reported in the latter study (Figure 1). The two

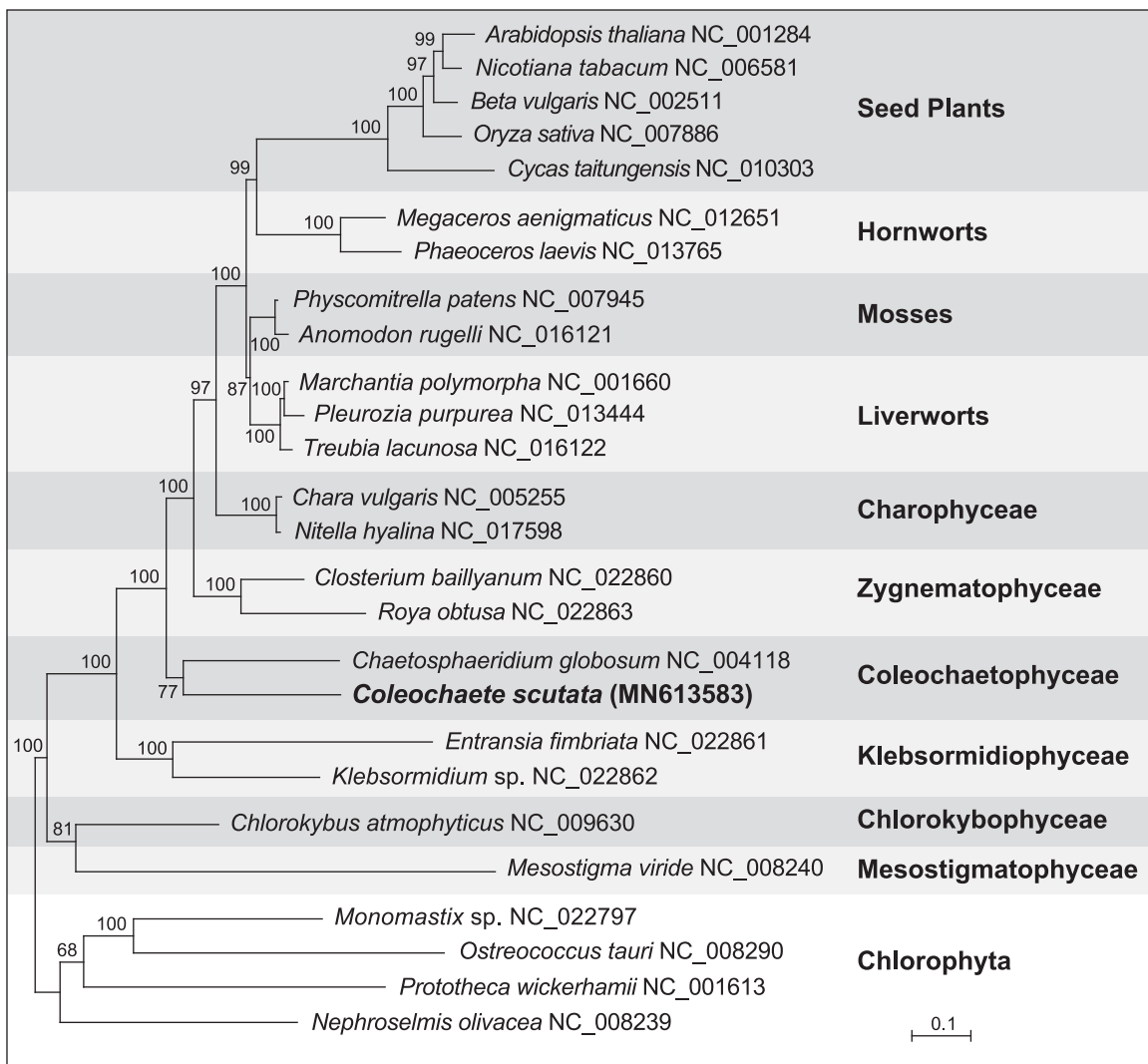


Figure 1. RAxML analysis of 40 concatenated mitogenome-encoded proteins from 22 streptophytes, including 12 land plants. The sequences of four chlorophytes were used to root the tree. The data set was partitioned by gene and each of the 40 partitions was applied to the GTR + I Γ 4 model. The figure shows the best-scoring tree, with bootstrap support values reported on the nodes. GenBank accession numbers for the mitogenomes of all taxa are provided. The scale bar denotes the estimated number of amino acid substitutions per site.

representatives of the Coleochaetophyceae are separated from one another by long branches but as expected, they fall in the same clade.

Disclosure statement

The authors report no conflict of interest. The authors alone are responsible for the content and writing of this article.

Funding

This work was supported by the Natural Sciences and Engineering Research Council of Canada under [Grant RGPIN-2017-04506].

ORCID

Monique Turmel  <http://orcid.org/0000-0001-7060-035X>
 Christian Otis  <http://orcid.org/0000-0001-9680-5863>
 Claude Lemieux  <http://orcid.org/0000-0001-9580-8042>

References

- Boisvert S, Laviolette F, Corbeil J. 2010. Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *J Comput Biol.* 17(11):1519–1533.
- Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. *Genome Res.* 8(3):195–202.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 30(9):1312–1313.
- Turmel M, Otis C, Lemieux C. 2002. The chloroplast and mitochondrial genome sequences of the charophyte *Chaetosphaeridium globosum*: insights into the timing of the events that restructured organelle DNAs within the green algal lineage that led to land plants. *Proc Natl Acad Sci USA.* 99(17):11275–11280.
- Turmel M, Otis C, Lemieux C. 2007. An unexpectedly large and loosely packed mitochondrial genome in the charophycean green alga *Chlorokybus atmophyticus*. *BMC Genomics.* 8(1):137.
- Turmel M, Otis C, Lemieux C. 2013. Tracing the evolution of streptophyte algae and their mitochondrial genome. *Genome Biol Evol.* 5(10):1817–1835.
- Turmel M, Otis C, Lemieux C. 2017. Divergent copies of the large inverted repeat in the chloroplast genomes of ulvophycean green algae. *Sci Rep.* 7(1):994.