



ELSEVIER

Contents lists available at ScienceDirect

## Data in Brief

journal homepage: [www.elsevier.com/locate/dib](http://www.elsevier.com/locate/dib)

Data article

# Comparative analyses of chloroplast genome data representing nine green algae in Sphaeropleales (Chlorophyceae, Chlorophyta)

Karolina Fučíková\*, Louise A. Lewis, Paul O. Lewis

Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, USA

## ARTICLE INFO

## Article history:

Received 16 February 2016

Received in revised form

25 February 2016

Accepted 1 March 2016

Available online 9 March 2016

## ABSTRACT

The chloroplast genomes of green algae are highly variable in their architecture. In this article we summarize gene content across newly obtained and published chloroplast genomes in Chlorophyceae, including new data from nine of species in Sphaeropleales (Chlorophyceae, Chlorophyta). We present genome architecture information, including genome synteny analysis across two groups of species. Also, we provide a phylogenetic tree obtained from analysis of gene order data for species in Chlorophyceae with fully sequenced chloroplast genomes. Further analyses and interpretation of the data can be found in “Chloroplast phylogenomic data from the green algal order Sphaeropleales (Chlorophyceae, Chlorophyta) reveal complex patterns of sequence evolution” (Fučíková et al., In review) [1].

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

## Specifications Table

Subject area	Biology
More specific subject area	Phylogenomics
Type of data	Table, Figures, text file, tree file

DOI of original article: <http://dx.doi.org/10.1016/j.ympbev.2016.01.022>

\* Corresponding author.

E-mail address: [Karolina.fucikova@gmail.com](mailto:Karolina.fucikova@gmail.com) (K. Fučíková).

<http://dx.doi.org/10.1016/j.dib.2016.03.014>

2352-3409/© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

How data were acquired	<i>Phylogenetic analysis of gene order in Badger, Mauve analysis, Geneious genome map</i>
Data format	<i>Analyzed</i>
Experimental factors	<i>Genomic data were collected using Illumina HiSeq and annotated in Geneious</i>
Experimental features	<i>Gene order data analyzed in Badger, Synteny maps in Mauve plugin in Geneious</i>
Data source location	<i>Storrs, CT, U.S.A.</i>
Data accessibility	<i>Data is within this article and at NCBI: <a href="http://www.ncbi.nlm.nih.gov">http://www.ncbi.nlm.nih.gov</a> accessions GenBank: KT199248, GenBank: KT199249, GenBank: KT199250, GenBank: KT199251, GenBank: KT199252, GenBank: KT199253, GenBank: KT199254, GenBank: KT199255, GenBank: KT199256</i>

---

### Value of the data

- Chloroplast genomes of green algae in the order Sphaeropleales are currently sparsely studied, thus new data from nine additional species expands knowledge of the structural variation within this order of algae.
  - Table summarizes the features present in chloroplast genomes of green algae in Sphaeropleales, useful for comparison to other species in future analyses.
  - Figures of nine assembled chloroplast genomes of green algae illustrate the features and their arrangements in these species.
  - New gene order data can be used in future phylogenetic analyses that include information from additional species.
- 

### 1. Data

In this article we present chloroplast genome structural data for nine species of green algae (*GenBank: KT199248, GenBank: KT199249, GenBank: KT199250, GenBank: KT199251, GenBank: KT199252, GenBank: KT199253, GenBank: KT199254, GenBank: KT199255, GenBank: KT199256; NCBI: <http://www.ncbi.nlm.nih.gov>*), including comparison of gene and intron content (Fig. 1). Chloroplast genome maps with all annotated features are presented in Figs. 2–10. Synteny comparisons were performed for two sets of species (Figs. 11 and 12). Gene order data were scored for 15 taxa in the class Chlorophyceae (GeneOrder.txt). A phylogenetic analysis of gene order using Badger is presented in (Fig. 13, GeneOrder.tre).

### 2. Experimental design, materials and methods

Full chloroplast genome sequences were obtained for 9 species of green algae in Sphaeropleales [1]. From these, genome maps were prepared in Geneious (version 6) (Figs. 2–10). Synteny maps were produced using the Mauve plugin in Geneious, which was also used to estimate the numbers of genomic rearrangements among taxa [2] (Figs. 11–12). Gene order data were analyzed phylogenetically using Badger [3] for a total of 15 species in Chlorophyceae with fully sequenced cp genomes (Fig. 13). Only single-copy genes that were present in all included genomes (83 genes) were used for the analysis. Genes present in inverted repeats were counted once and inverted repeats were oriented to have the rRNA genes on the positive strand. For the trans-spliced *psaA* gene, only the first exon was considered. Badger was run for 10,000,000 generations, sampling every 100, with other settings set to default. The first 1000 samples were discarded as burnin.









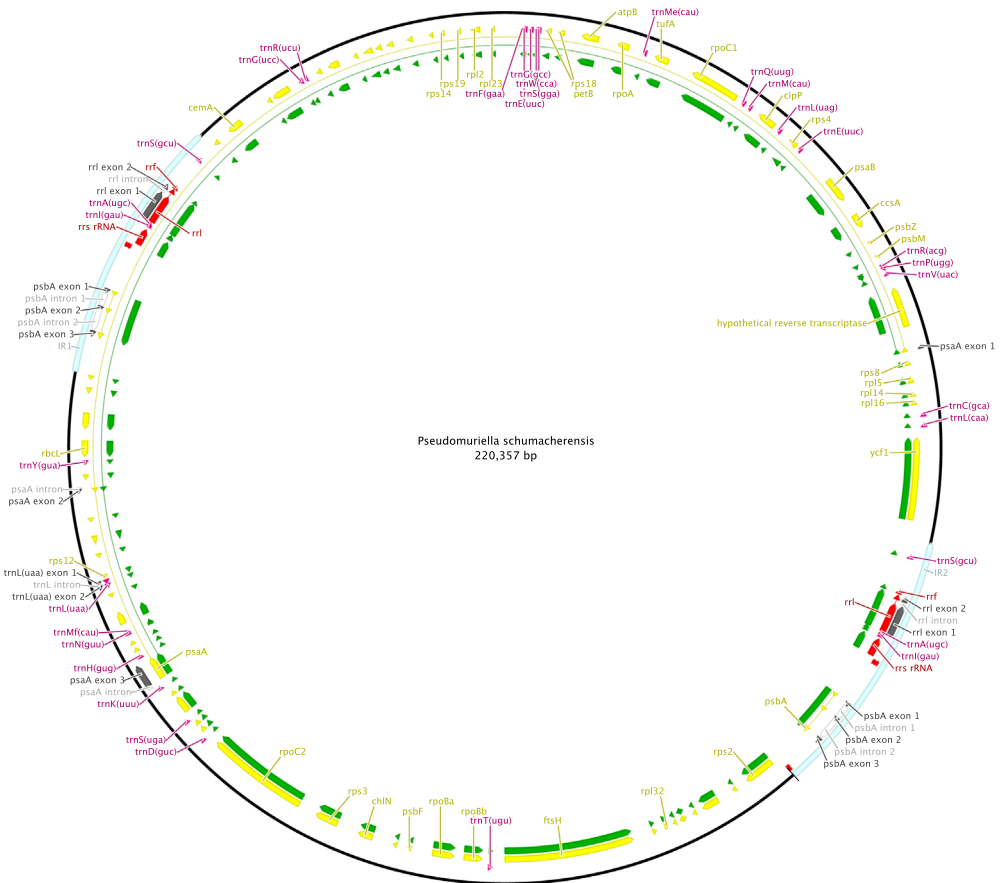




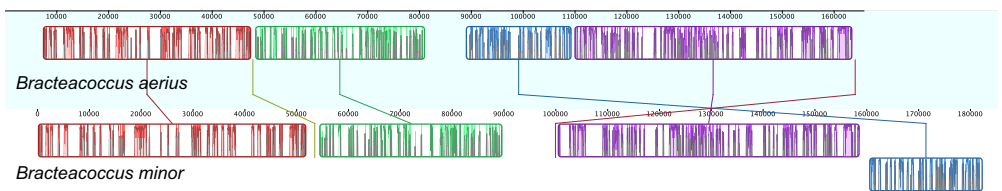




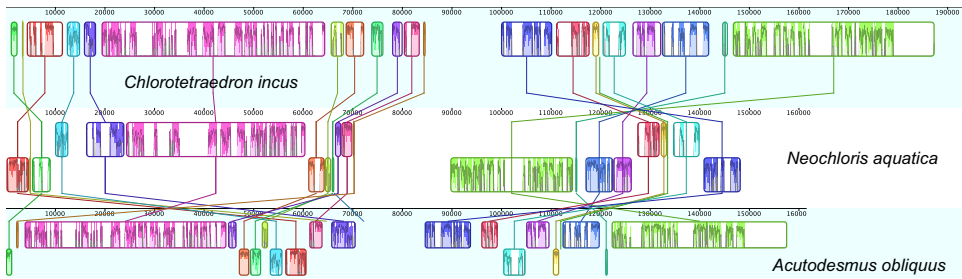




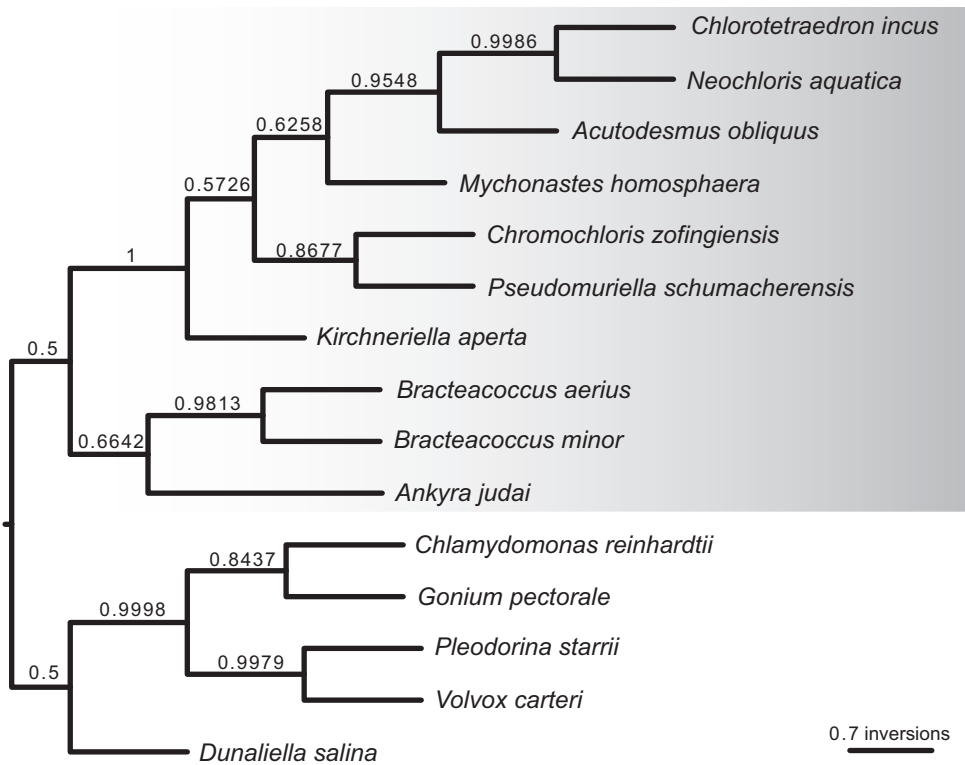
**Fig. 10.** Schematic map of the chloroplast genome of *Pseudomuriella schumacherensis* (KT199256, SAG 2137).



**Fig. 11.** Analysis of synteny between the chloroplast genomes of two species of *Bracteacoccus*, *B. aerius* and *B. minor*, using MAUVE alignments. Colored and outlined blocks surround regions of the genome sequence of one genome that aligned to a corresponding part of the second genome, and lines connect blocks of putative homology. Within the blocks the colored bars indicate the level of sequence similarities.



**Fig. 12.** Synteny analysis among the chloroplast genomes of three species, *Acutodesmus obliquus*, *Neochloris aquatica*, and *Chlorotetraedron incus*. Colored and outlined blocks surround regions of the genome sequence that aligned to a corresponding part of the second genome, and lines connect blocks of putative homology. Within the blocks the colored bars indicate the level of sequence similarities.



**Fig. 13.** Badger phylogenetic tree inferred from gene order data present in completely sequenced chloroplast genomes of Sphaeropleales. Scale bar represents the number of inversions inferred to have occurred along a particular branch.

## Acknowledgments

Analyses were supported with funding from NSF grants DEB-1036448 and DEB-1354146 awarded to LAL and POL at University of Connecticut. Analyses were carried out at the Bioinformatics Facility of the University of Connecticut Biotechnology and Bioservices Center (BBC).

## Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.dib.2016.03.014>.

## References

- [1] K. Fučíková, P.O. Lewis, L.A. Lewis, Chloroplast phylogenomic data from the green algal order Sphaeropleales (Chlorophyceae, Chlorophyta) reveal complex patterns of sequence evolution, *Mol. Phylogenet. Evol.* (2016) (In press).
- [2] A.E. Darling, B. Mau, N.T. Perna, progressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement, *PLoS One* 5 (6) (2010) e11147.
- [3] D. Simon, B. Larget, Bayesian analysis to describe genomic evolution by rearrangement (BADGER), version 1.02 beta, Department of Mathematics and Computer Science, Duquesne University, 2004.