Nadia Martins Lemes da Silva

Filogenômica e caracterização genômica de algas vermelhas continentais da ordemBatrachospermales (Rhodophyta)

Relatório de Pós-doutorado, realizado na Universidade Estadual Paulista (UNESP), Instituto de Biociências, Letras e Ciências Exatas de São José do Rio Preto.

Supervisor(a): Prof. Dr. Orlando Necchi Júnior

São José do Rio Preto
2026

Relatório Final Referente ao Estágio de Pós Doutorado– SisPROPe

Título: Filogenômica e caracterização genômica de algas vermelhas continentais da ordem Batrachospermales (Rhodophyta)

Supervisor: Orlando Necchi Júnior

Pós-doutoranda: Nadia Martins Lemes da Silva

Colaboradores: Morgan L. Vis, Ohio University, EUA – estará envolvida nas etapas de intercâmbio de amostras, trabalho de laboratório (sequenciamento de alto rendimento) e análise de dados (características dos genomas e filogenômica) de táxons complementares aos gerados neste projeto.

Instituição sede: Instituto de Biociências, Letras e Ciências Exatas, UNESP, Câmpus de São José do Rio Preto, Departamento de Ciências Biológicas, Laboratório de Biologia, Ecologia e Taxonomia de Algas.

Período: 01 de março de 2023 a 31 de outubro de 2025

Resumo do projeto

Atualmente, as informações disponíveis sobre a organização genômica de rodófitas de água doce das ordens Batrachospermales são esparsas e a relação filogenética entre os representantes da ordem não está bem resolvida. O projeto propõe análise filogenômica e caracterização de genomas plastidiais e mitocondriais para ambas as ordens de Nemaliophycidae. Os resultados gerados permitirão testar as seguintes hipóteses: 1) Os genomas de Batrachospermales serão altamente conservados, em concordância com outros membros de Florideophyceae em geral e mais particularmente da subclasse Nemaliophycidae; outras características são esperadas, como genomas de tamanho entre 170 e 200 kpb e com cerca de 200 genes codificantes, ambos dentro da faixa relatada para Florideophyceae e mais semelhantes aos grupos mais próximos, como Nemaliales; 2) os dados atualmente disponíveis para membros da ordem Batrachospermales não permitem definir clados distintos dentro da ordem, razão pela qual apenas uma família é reconhecida no esquema atual; as análises filogenômicas revelarão dois ou mais grupos bem sustentados, que associados às características morfológicas vegetativas e reprodutivas, permitirão um novo e melhor rearranjo de gêneros e famílias; 3) a inclusão de maior número de táxons de Batrachospermales nas análises permitirá um melhor entendimento das relações

filogenéticas entre as ordens exclusivamente de água doce (Batrachospermales, Thoreales e Balbianiales) dentro de Nemaliophycidae. O estudo proposto aborda aspectos de grande relevância: a) é uma ampla investigação de filogenômica e caracterização genômica dos dois grupos mais representativos de rodófitas de água doce; 2) fornecerá informações valiosas para análises filogenômicas comparativas da ordem Batrachospermales com outros grupos de Rhodophyta, especialmente dentro da subclasse Nemaliophycidae; viabilizando o refinamento das relações filogenéticas do grupo; 3) fornecerá um robusto e amplo conjunto de dados para uma avaliação mais geral sobre a transição de ambientes marinhos para de água doce para membros das duas ordens, bem como as possíveis relações com as outras ordens exclusivamente de água doce com genoma já sequenciado (Balbianiales e Thoreales).

Sobre o projeto:
Como resultado do desenvolvimento do projeto, foram produzidos dois artigos científicos, que estão em anexo e apresentam de maneira detalhada os resultados alcançados: Investigation of mitochondrial genomes of members of the order Batrachospermales (Rhodophyta) - já publicado **(Anexo A)**; Plastid genome structure and phylogenomics of the freshwater red algal order Batrachospermales (Rhodophyta) – em fase de revisão pela revista **(Anexo B)**.

Outras atividades acadêmicas desenvolvidas pela pós-doutoranda no período.

1. Minicurso ministrado. A pós-doutoranda ministrou o mini-curso "O uso de genomas de cloroplastos em estudos evolutivos e ecológicos", 8 h/a, no VI Congresso Paulista de Ciências Biológicas do Instituto de Biociências, Letras e Ciências Exatas, IBILCE – UNESP realizado entre os dias 28 de agosto e 01 de setembro de 2023.

2. **Minicurso ministrado.** A pós-doutoranda ministrou o mini-curso "Bioinformática aplicada a microbiologia", 4 h/a, no V Simpósio de Microbiologia, no Instituto de Biociências, Letras e Ciências Exatas, IBILCE – UNESP realizado entre os dias 29 e 30 de outubro de 2024.

3. Co-orientação de alunos. Mestrado: aluna Yasmin Maria da Silva Lopes (Programa de pós-graduação em Biodiversidade, IBILCE/UNESP) - "Filogenômica e caracterização genômica de espécies dos gêneros de algas vermelhas continentais

Kumanoa, Paludicola e *Sheathia* (Batrachospermales, Rhodophyta). Defesa marcada para 14/11/2025.

Iniciação científica: aluna Letícia Camillo Costa (curso de Ciências Biológicas, IBILCE/UNESP) - “Metodologia para pesquisa em sistemática molecular de macroalgas continentais”.

4. Aulas ministradas. Aula na disciplina “Sistemática Biológica”, curso de Ciências Biológicas, IBILCE/UNESP – “Genômica e Filogenômica”. Aula na disciplina “Morfologia, Fisiologia e Genética de microrganismos eucariotos e Virologia”, Programa de pós-graduação em Microbiologia – “Genômica – enfoque em algas”. Aula ministrada na disciplina “Tópicos avançados em Botânica”, Programa de pós-graduação em Biodiversidade - “Genômica em algas, fungos e plantas”.

5. Bancas. Membro das bancas avaliadoras de trabalhos de conclusão de curso (Bacharelado em Ciências Biológicas IBILCE/UNESP):

Letícia Fernandes de Santana – “Estudo taxonômico de algas verdes da ordem Cladophorales (Chlorophyta) para a região noroeste do estado de São Paulo”.

Pedro Matos Farinha - “Caracterização taxonômica de populações de cianobactérias de crostas biológicas da caatinga”.

6. Avaliação de trabalhos. Participação na comissão avaliadora de resumos e painéis do VII Congresso Paulista de Ciências Biológicas.

Phycologia

Phycologia

Anexo A

ISSN: 0031-8884 (Print) 2330-2968 (Online) Journal homepage: www.tandfonline.com/journals/uphy20

Investigation of mitochondrial genomes of members of the order Batrachospermales (Rhodophyta)

Nadia M. Lemes Da Silva, Roseanna M. Crowell, Morgan I. Vis, Wayne B. Chiasson, Monica o. Paiano, Yasmin M. Silva Lopes & Orlando JR. Necchi

To cite this article: Nadia M. Lemes Da Silva, Roseanna M. Crowell, Morgan I. Vis, Wayne B. Chiasson, Monica o. Paiano, Yasmin M. Silva Lopes & Orlando JR. Necchi (2025) Investigation of mitochondrial genomes of members of the order Batrachospermales (Rhodophyta), Phycologia, 64:2, 118-126, DOI: 10.1080/00318884.2025.2473285

To link to this article: https://doi.org/10.1080/00318884.2025.2473285

PHYCOLOGIA
2025, VOL. 64, NO. 2, 118–126
https://doi.org/10.1080/00318884.2025.2473285

Investigation of mitochondrial genomes of members of the order Batrachospermales (Rhodophyta)

Nadia M. Lemes Da Silva [1], Roseanna M. Crowell [2], Morgan I. Vis [2], Wayne B. Chiasson [2], Monica o. Paiano [1], Yasmin M. Silva Lopes [1] and Orlando JR. Necchi [1]

[1]Biological Sciences Department, São Paulo State University, Rua Cristóvão Colombo, 2265, São José do Rio Preto 15054-000, Brazil
[2]Environmental and Plant Biology Department, Ohio University, Porter Hall, Athens, Ohio 45701, USA

ABSTRACT

Data on genome characteristics and phylogenomics are becoming more readily available for the Rhodophyta, especially from organelles. However, for the freshwater red algal order Batrachospermales, information is still scarce with only nine organellar genomes available from six genera. A total of 22 new mitochondrial genomes representing 13 genera were generated and combined with the previously published data. Based on this large mitogenome data set, this study aimed to: (i) describe the structural characteristics of the mitogenomes across the Batrachospermales and compare with other red algal mitogenomes; (ii) infer phylogenetic relationships among the Batrachospermales using a broad dataset of taxa and genes. We determined that structure of the mitogenomes showed little variation, particularly genome size (22,863–29,785 bp), total number of genes (39–49), protein-coding genes (22–27), and GC content (25.3–32.4%). The largest genomes were observed in three species of *Paralemanea*, which also had larger intergenic regions. Synteny among members of Batrachospermales was highly conserved with the exceptions of three species of *Paralemanea* having a group II intron in the *cox*1 gene and the five *Kumanoa* species having a small divergent region. Only three gene losses were detected in the mitogenomes of members of Batrachospermales and the gene content was otherwise similar to other orders of Nemaliophycidae. The phylogenomic analysis based on 17 concatenated genes showed two large groups within the Batrachospermales, which were also recovered in previous studies of single or multigene phylogenetic analyses. The existence of two groups can have taxonomic implications to be considered in future studies.

ARTICLE HISTORY
Received 15-11-2024
Accepted 25-02-2025
Version of Record 21-03-2025

KEYWORDS
Freshwater red algae; genome structure; mitogenome; organellar genome; phylogenomics

INTRODUCTION

With the increased ease and decreased cost of high-throughput sequencing, the number of organellar genomes, both plastomes and mitogenomes, available in the Rhodophyta has been steadily growing. These organellar genomes have been employed to infer phylogenetic relationships in Rhodophyta, both at the higher taxonomic levels – phyla and classes (Janouškovec *et al.*, 2013; Lee *et al.*, 2015, 2016; Muñoz-Gómez *et al.*, 2017; Yang *et al.*, 2016), as well as for orders and families (Costa *et al.*, 2016; Diaz-Tapia *et al.*, 2017; Iha *et al.*, 2018; Lee *et al.*, 2018). In addition, the generation of new organellar genomes has also facilitated investigations of other aspects of red algal genomes, such as synteny, lateral gene transfers, and transposable elements (Iha *et al.*, 2018).

Due to their ubiquity, distribution, and conserved functions, complete mitochondrial genome (mtDNA) data have been used in diverse studies of population genetics, phylogeny, and evolution of these genomes (Yang *et al.*, 2015). In red algae, similarly to plants, mtDNA data have been employed for systematics and population genetics studies as molecular markers (Saunders, 2005). Although a reasonable number of red algal mitochondrial genomes have been sequenced, reconstructing the broader evolutionary history of these genomes has been challenging due to insufficient taxon sampling and the high level of structural variation observed in available mtDNA (Van Beveren *et al.*, 2022; Yang *et al.*, 2015).

The mitochondrial genomes of red algae exhibit unique characteristics that distinguish them from the mitochondrial genomes of other groups of algae: (1) large genome size – mitochondrial genomes of red algae are known for their relatively large size compared to other algal groups, can range from approximately 20,000 to over 100,000 bp in size, and are relatively complex (Janouškovec *et al.*, 2013; Van Beveren *et al.*, 2022; Yang *et al.*, 2015); (2) gene content – mitochondrial genomes of red algae contain a diverse range of genes involved in essential cell functions and the total number of genes ranges from 13 to 79 and the number of protein coding genes from 9 to 36 (Van Beveren *et al.*, 2022; Yang *et al.*, 2015); (3) presence of introns; mitochondrial genomes of red algae may harbour introns adding complexity to their genetic structure (Matsuzaki *et al.*, 2004); the number of introns in mtDNA in red algae is generally low (1–3), but can reach a few dozen (10–58), especially in some Protorhodophytina species (Van Beveren *et al.*, 2022). In general, the mitochondrial genomes of red algae present

CONTACT Orlando Jr. Necchi o.necchi@unesp.br
Supplemental data for this article can be accessed online at https://doi.org/10.1080/00318884.2025.2473285

a unique combination of features that reflect their evolutionary history and complexity, making them a topic of great interest in algal genomic research (Van Beveren *et al.*, 2022).

The subclass Nemaliophycidae in addition to marine orders, houses the strictly freshwater orders Batrachospermales, Thoreales and Balbianiales (Lam *et al.*, 2016). The order Batrachospermales (Rhodophyta) is currently treated as having a single family Batrachospermaceae (Entwisle *et al.*, 2009) with 21 genera recognized in the most recent revision (Vis & Necchi, 2021). Batrachospermales is the most diverse group in terms of number of taxa, morphology, and reproductive characters among the exclusively freshwater orders of red algae (Entwisle *et al.*, 2009; Kumano, 2002; Lam *et al.*, 2016; Vis & Necchi, 2021). Although the order is well supported in all phylogenetic analyses (Entwisle *et al.*, 2009; Lam *et al.*, 2016; Vis & Necchi, 2021), the infra-ordinal classification is not yet well supported. The classic three-family system (Batrachospermaceae, Lemaneaceae and Psilosiphonaceae, Pueschel & Cole, 1982; Sheath *et al.*, 1996) was reduced to a single family (Batrachospermaceae) in a molecular phylogeny analysis based on two molecular markers (Entwisle *et al.*, 2009).

Data on mitogenomes for members of Batrachospermales are scarce. The only specific investigation for the group was by Paiano *et al.* (2018) that sequenced genomes of six members of the order Batrachospermales. The other studies reporting mitogenomes for the group only reported one member as part of organellar genome studies (Nan *et al.*, 2017; Wolf *et al.*, 2017). To date, Ten mitogenomes of members of the order are available in NCBI GenBank (Sayers *et al.*, 2020).

This study is part of a larger project investigating organellar genomes of Batrachospermales. The research goals were: to describe the structural characteristics of the mitogenomes to clarify their evolution from a comparative analysis with other red algal groups; and to infer the phylogenetic relationships based on a representative taxon sampling including a broad range of vegetative and reproductive morphology and taxonomic position within the group.

MATERIAL AND METHODS

Taxon sampling

To add to the ten previously published mitogenomes from six genera, new mitogenomes of 22 species from 13 genera were sequenced, including representative genera of the wide variety of vegetative and reproductive characters of the Batrachospermales (Table S1). Each live specimen was divided into two portions: (i) for DNA sequencing by immediately being placed in silica gel; and (ii) for vouchering by either liquid preservation or pressed on herbarium paper and deposited in recognized herbaria (Table S1). Specimens were identified based on current diagnostic characters using recent taxonomic literature (Vis & Necchi, 2021). In addition, taxa were checked for sequence identity using COI-5P barcode region against sequences in the GenBank database. From the 10 mitogenomes available in GenBank, MG787094 genome sequence for *Montagnia macrospora* was excluded after a BLAST search of the COI-5P sequence showed that the sequence did not belong to the Batrachospermales and does not represent that species.

Sequencing and assembling of mitochondrial DNA

Specimens processing differed between the two laboratories: one set of 8 samples was handled in Necchi's lab, whereas another set of 14 samples was processed in Vis' lab (Table S1). In Necchi's lab genomic DNA (gDNA) was extracted using the Qiagen Magattract HMW (High Molecular Weight) Kit (Qiagen, Hilden, Germany), following the manufacturer protocol. gDNA was quantified using a Qubit 2.0 fluorometer (Thermo Fisher, Waltham, USA) and Quant-it dsDNA High-Sensitivity Assay kit. Quality checking of gDNA was conducted in a TapeStation 4200 system (Agilent, Santa Clara, USA). Sequencing of the gDNA was performed on an Illumina NextSeq 2000 platform (Illumina, San Diego, USA) with paired-end 2 × 150 pb and coverage of 10 million reads. Libraries were prepared from gDNA with the Illumina DNA prep protocol.

In Vis' lab samples were ground in liquid nitrogen and the Nucleospin Plat II kit (Macherey–Nagel, Düren, Germany) employed following the manufacturer's protocol, except for an extended lysis step. For *Paralemanea blumii* M.L. Vis, *Tuomeya americana* (Kützing) Papenfuss and *Volatus carrionii* I.S. Chapuis and M.L. Vis the sample was extracted using a modified CTAB extraction protocol with polyvinylpyrrolidone-10 (Schenk *et al.*, 2023). The quantity and quality of extracted DNA was assessed using one of three methods: a Nanodrop ND-1000 (ThermoFisher Scientific, Waltham, Massachusetts, USA), Qubit 3 fluorometer (ThermoFisher Scientific, Waltham, Massachusetts, USA) or Agilent 4150 TapeStation (Agilent, Santa Clara, California). Sequencing of the genomic DNA was performed on one of the following Illumina platforms: NovaSeq, MiSeq or NovaSeq X Plus. For sequencing on the NovaSeq platform, the Ohio University Genomics Facility completed libraries of 350 nucleotide fragments using a Nextera XT DNA Sample Prep kit. For sequencing on the Illumina MiSeq platform, libraries of 200–450 nucleotide fragments were completed using the New England BioLabs NEBNext Ultra II FS DNA Library Kit following the protocol for inputs ≤ 100ng. Sequencing on the Illumina NovaSeq Plus was completed by SeqCenter (Pittsburgh, Pennsylvania) using libraries of 280 nucleotide fragments with an Illumina DNA Prep protocol.

Genome assembling and annotation

For samples assembled in Necchi's lab the quality of trimmed Illumina reads was checked with FastQC (v0.12.1; Andrews, 2016). Illumina raw reads were trimmed with Trimmomatic version 0.39 (Bolger *et al.*, 2014), to remove adapters and low-quality reads, with the following parameters: ILLUMINACLIP: adapters.fa: 2:30:10; SLIDINGWINDOW: 4:20; MINLEN: 36. After trimming, the clean reads were used for the *de novo* genome assembly. In order to deal with the specificities of the genomes of each organism, three different assemblers were used: NOVOPlasty version 4.3.1 (Dierckxsens *et al.*, 2017), with the *Kumanoa ambigua* (NC_039405) mitogenome sequence in the assembly as the seed sequence and k-mer 29 and insert size 300, GetOrganelle version 1.5.1c (Jin *et al.*, 2020), with k-mers 21, 45, 65, 85, 105,

-R 100, -F other _pt and seeds *K. ambigua* (NC_039405), *K. malahcensis* (MK611801 MG787096) and SPAdes version 3.15.5 (Nurk *et al*., 2017; Prjibelski *et al*., 2020). The resulting scaffolds obtained with SPAdes were sorted by tBLASTn (e-value: 1e-10) using a database of mitogenomes from the same genus or, when not available, from closely related genera of Batrachospermales, to separate the scaffolds corresponding to mitogenomes. Genome assemblies obtained with GetOrganelle and SPAdes were visualized and manually inspected using Bandage version 0.8.1 (Wick *et al*., 2015).

For samples assembled in Vis' lab gene assembling and annotation was conducted as follows: paired-end reads were trimmed and the first 15 million were used for a *de novo* assembly using low-sensitivity parameters in Geneious Prime 2024.0.7 (Biomatters, geneious.com). The published mitochondrial genome of *Lympha mucosa* (MF488959) was used as a reference to identify contigs containing mitochondrial sequence for *Batrachospermum gelatinosum*. Subsequently, the *B. gelatinosum* reference was used to identify contigs for other species. These mitochondrial contigs were trimmed and *de novo* assembled. To close gaps in the assembled sequence, contigs were extended by iteratively mapping reads with low sensitivity and a minimum mapping quality of 30 until complete mitochondrial genome sequences could be circularized.

Genome annotation was performed by MFannot (Beck & Lang, 2010), the tRNAs were identified with tRNAscan-SE version 2.0.7 annotator in GeSeq with default parameters (Chan & Lowe, 2019) and the maps were generated with OGDRAW version 1.3.1 (Greiner *et al*., 2019). All annotations were carefully reviewed and manually corrected in Geneious Prime 2023.2 (http://www.geneious.com).

The genomic features (Table 1) were obtained from the.-gff files of the annotated genomes using R version 4.2.2 (R Core Team, 2021). Analysis and plots used the packages READXL, TIDYVERSE, GGPLOT2, DPLYR (Wickham, 2016; Wickham *et al*., 2019; Wickham & Bryan, 2023; Wickham *et al*., 2023).

Synteny among genomes was investigated via whole-genome alignment with the progressiveMauve algorithm version 2.4.0 (Darling *et al*., 2010) using automatically calculated seed weights, and automated calculation of locally collinear block (LCB) scores.

In order to test correlation among the genomic features, the Pearson coefficient (Zar, 2010) was calculated. Tests were performed using R version 4.4.

Phylogenetic analyses

A total of 31 Batrachospermales and one outgroup (*Thorea hispida* KY083066) taxa were used in the phylogenetic analyses. For nine taxa, data were retrieved from GenBank and 22 comprised the newly sequenced genomes. Seventeen coding genes present in all the 32 taxa (*atp*4, *atp*8, *atp*9, cob, *cox*1, *cox*2, *cox*3, *nad*1, *nad*2, *nad*3, *nad*4, *nad*4L, *nad*5, *nad*6, *sdh*2, *sdh*4 and *rpl*16) were used in the analyses, which were manually concatenated in Geneious and aligned at the nucleotide level using MUSCLE (Edgar, 2004). The final alignment was 14,084 nt positions in length. Maximum likelihood (ML) phylogenetic tree inferred from the nucleotides was obtained with RAxML version 8 (Stamatakis, 2014), using the following parameters: 1,000 fast bootstraps. All analyses were performed using the computational resources from the Center for Scientific Computing (NCC/GridUNESP) of the São Paulo State University (UNESP). The scripts and pipelines used for command-line of all analyses are stored on the Notion platform (www.notion.so).

RESULTS

Genome structure

The 22 mitogenomes were fully assembled, annotated, and all mapped as circular molecules (Figs. S1-S2). The structure of the mitogenomes was similar for gene content and gene order, with little variation observed in most characteristics, particularly genome size, number of genes, and GC content (Table 1, Figs S3-S4). Genome sizes ranged from 22,863 bp in *Montagnia australis* (Collins) Necchi & Vis to 29,785 bp in *Paralemanea sp*. (25,701 ± 1,494, overall mean ± standard deviation for all Batrachospermales taxa). The largest genomes were observed in the three species of *Paralemanea*, which also had larger intergenic regions (Table 1, Figs. 2, 3, S3-S4). GC content ranged from 25.3% in *Paludicola diamantinensis* Necchi & M.L. Vis to 32.4% in *Paralemanea sp*. (29.1 ± 1.7%).

The number of genes varied from 39 in *M. australis* to 49 in *Paralemanea blumii* M.L. Vis (45.3 ± 2.4), while the number of protein-coding genes ranged from 22 in four species (*Kumanoa ambigua* (Montagne) Entwisle, M.L. Vis, W.B. Chiasson, Necchi and A.R. Sherwood, *L. borealis* Atkinson, *M. australis* and *Psilosiphon scoparius* Entwisle) to 27 in *Batrachospermum qujingense* (23.6 ± 1.5). Gene lengths among mitogenomes ranged from 21,027 bp in *M. australis* to 27,078 bp in *P. blumii* (23,032 ± 1,373), with average lengths varying from 469.4 bp in *Kumanoa procarpa* (Skuja) *Kumanoa ambigua* (Montagne) Entwisle to 554.7 bp in *Paralemanea* sp. (505.6 ± 26.4). The proportion of intergenic regions in the genomes was highly variable, ranging from 5.0% in *S. delicatuliformis* Necchi, Rossignolo & Paiano to 19.4% in *Nocturama novamundensis* Necchi & Entwisle (10.2 ± 3.3%).

In all species, most coding genes were between 500–1,000 bp in length, with the *nad*4, *nad*5, and *cox*1 (genes encoding respiratory chain complexes) reaching up to approximately 2,000 bp (Fig. S4). In species of *Paralemanea*, a maturase gene (an intron-encoded protein) was detected, with a length of 1,998 bp (Fig. 1) in a group II introns within the *cox*1 gene (with the full intron length of 2,271 bp).

The number of tRNA genes ranged from 14 in *Virescentia viride-americana* Necchi, D.C. Agostinho & M.L. Vis to 20 in several other species (18.8 ± 1.9). The tRNA content was similar among species (Fig. 4), with *trn*R-UCU being the least distributed, while *trn*M-AUG was duplicated in all species. In most species (12), mitogenomes contained three rRNA genes: *rrn*S (small subunit), *rrn*L (large subunit), and 5S rRNA. In the remaining ten species, only two rRNA genes were observed, with no 5S rRNA gene present. Repetitive elements were absent in 10 species and present in 12 others, with the number of repeats ranging from 1 to 5 and total

Table 1. Characteristics of the mitogenomes of members of the Batrachospermales generated (in this study in boldface) and previously published.

Species	Genome Size (bp)	GC Content (%)	Total Gene Number	Protein Coding Genes	TRNA; rRNA	Total Gene Length (bp)	% of Genome (Genes)	Intergenic Region Length (bp)	Repetitive Elements Number; Total Length (bp); % of Genome	Coverage (mean; min; max)
Batrachospermum gelatinosum	25,340	29.7	45	24	19; 2	23,272	91.85	2,067	–	710; 23; 2297
B. qujingense MN882544	27,268	29.3	48	27	19; 2	23,572	86.52	3,672	1; 14; 0.05	–
B. naiadis	25,291	30.4	46	24	20; 2	23,218	91.81	2,072	–	1134; 171; 3168
Kumanoa ambigua MG787095	25,862	26.9	41	22	16; 3	21,984	85.08	3,855	5; 54; 0.21	–
K. capensis	24,997	29.2	45	23	19; 3	21,433	85.82	3,541	1; 16; 0.06	1754; 16; 4056
K. equisetoidea	25,111	28.6	45	23	19; 3	21,650	86.29	3,438	2; 71; 0.27	346; 1; 1946
K. mahlacensis NC039406	25,069	28.4	45	23	19; 3	23,034	91.97	2,012	–	51; 20; 81
K. mahlacensis MK611801	24,992	28.4	47	25	19; 3	23,641	94.68	1,328	–	–
K. procarpa	25,017	29.1	45	23	19; 3	21,594	86.32	3,422	–	464; 26; 686
Lemanea borealis	25,946	31.2	45	22	20; 2	22,568	86.98	3,377	2; 26; 0.10	780; 44; 2189
L. occidentalis	26,726	31.1	48	24	20; 2	23,291	87.15	3,434	3; 42; 0.15	4478; 280; 10843
Lympha mucosa MF488959	25,191	27.5	46	24	20; 2	23,351	92.70	1,839	1; 14; 0.05	–
Montagnia australis	22,863	26.8	39	22	14; 3	21,027	93.22	1,528	2; 22; 0.96	333; 49; 576
Nocturama novamundensis	26,682	28.1	44	24	16; 3	21,491	80.61	5,168	1; 18; 0.07	640; 40; 5324
Paludicola aquanigra	24,815	26.8	45	23	19; 3	22,483	90.69	2,307	1; 16; 0.06	88; 4; 263
P. communis	26,017	26.0	45	23	18; 3	22,425	86.28	3,567	3; 30; 0.11	966; 43; 1156
P. diamantinensis	24,828	25.3	44	23	19; 3	22,370	90.19	2,433	2; 9; 0.10	72; 1; 138
Paralemanea sp. MG787097	29,785	32.4	47	25	20; 3	26,624	89.46	3,136	–	125; 16; 535
P. blumii	29,673	30.0	49	26	20; 3	27,078	91.33	2,569	2; 23; 0.09	2045; 37; 3746
P. grandis	29,321	30.8	47	24	20; 3	25,714	87.77	3,582	2; 23; 0.09	151; 23; 598
Psilosiphon scoparius	24,751	30.7	40	22	15; 3	22,018	89.04	2,710	1; 17; 0.07	71; 4; 113
Sheathia dispersa MG787098	25,036	28.4	47	24	20; 3	23,648	94.54	1,365	2; 25; 0.99	419; 133; 537
S. dispersa	25,071	28.5	48	25	20; 3	23,085	92.16	1,963	2; 25; 0.99	321; 3; 597
S. involuta	25,122	29.1	47	24	20; 3	22,396	89.19	2,714	2; 27; 0.11	850; 72; 2503
S. longipedicellata NC_035349	25,086	29.7	46	25	19; 2	23,736	94.62	1,349	1; 10; 0.03	–
Sirodotia delicatuliformis NC039407	25,150	30.7	46	25	18; 3	23,874	95.01	1,252	1; 16; 0.06	419; 133; 537
S. suecica	25,140	31.6	45	23	19; 2	23,009	91.53	2,130	–	4094; 438; 13897
Tuomeya americana	25,207	27.7	46	24	20; 2	23,042		2,139	–	1987; 988; 3069
Virescentia viride-americana	25,044	29.0	40	23	14; 2	22,137	88.49	2,880	–	940; 34; 1333
Volatus carrionii	25,136	29.2	47	24	20; 3	23,339	92.93	1,774	–	450; 5; 974
V. personatus	25,342	29.0	46	24	20; 2	23,040	90.91	2,301	1; 10; 0.03	2100; 187; 2794

lengths between 9 and 71 bp, accounting for 0.03%–0.70% of the genome (Table 1). Pearson's correlation test applied to all genome characteristics (values of each genome parameter for all taxa as described in Table 1) revealed no significant correlations among them.

Genome arrangement

Synteny among members of Batrachospermales was shown to be highly conserved (Fig. S5). Some exceptions were the five *Kumanoa* species that had a small, divergent region and more notably the three *Paralemanea* species which had a group II intron in the *cox*1 gene (Figs. 2, 3). Whole-genome alignments revealed few rearrangements among the mitogenomes, as indicated by DCJ values. Three collinear blocks were predicted with the ProgressiveMAUVE algorithm for whole-genome alignments (Fig. S5).

In terms of gene losses, only three genes that occur in other orders of Nemaliophycidae were absent in the Batrachospermales: trnS-AGC tRNA (present in Palmariales and Thoreales); trnU tRNA and secY (present in Palmariales; Fig. 4).

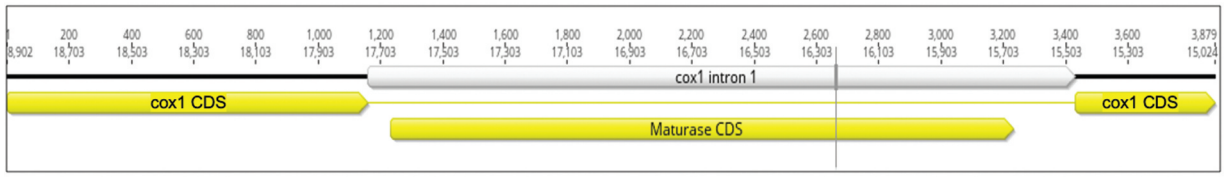
Phylogeny

The ML phylogenetic tree for the mitogenome dataset (Fig. 5) resulted in full support for the order Batrachospermales. Within the Batrachospermales, two groups with high support (99–100%) were recovered: Group 1 – the genera *Kumanoa, Montagnia, Paludicola, Psilosiphon* and *Virescentia*; Group 2 – the genera *Batrachospermum, Lemanea, Lympha, Nocturama, Paralemanea, Sheathia, Sirodotia, Tuomeya* and *Volatus*. All genera with more than one mitogenome sequenced consisted of fully supported clades.

DISCUSSION

This research has added 22 new mitogenomes and has extended the number of genera with these data to 13 members of the Batrachospermales. The variations of the newly sequenced and previously published mitogenomes of members of the Batrachospermales were slightly expanded for most characteristics, particularly genome size, number of genes, and GC content, but within the range reported for the Nemaliophycidade of the class Florideophyceae (Kim *et al.*, 2022; Van Beveren *et al.*, 2022).

1



2

3

Fig. 1-3. Major genome rearrangements found in the Batrachospermales. **1**. Sequence of *cox*1 in the three species of *Paralemanea* showing the insertion of the intron group II. **2**. Partial data of ProgressiveMauve with the species of the genus *Kumanoa* showing the small, divergent region (arrows). **3**. Partial data of ProgressiveMauve with the species of the genus *Paralemanea* showing their larger genomes (highlighted in the red line rectangle).

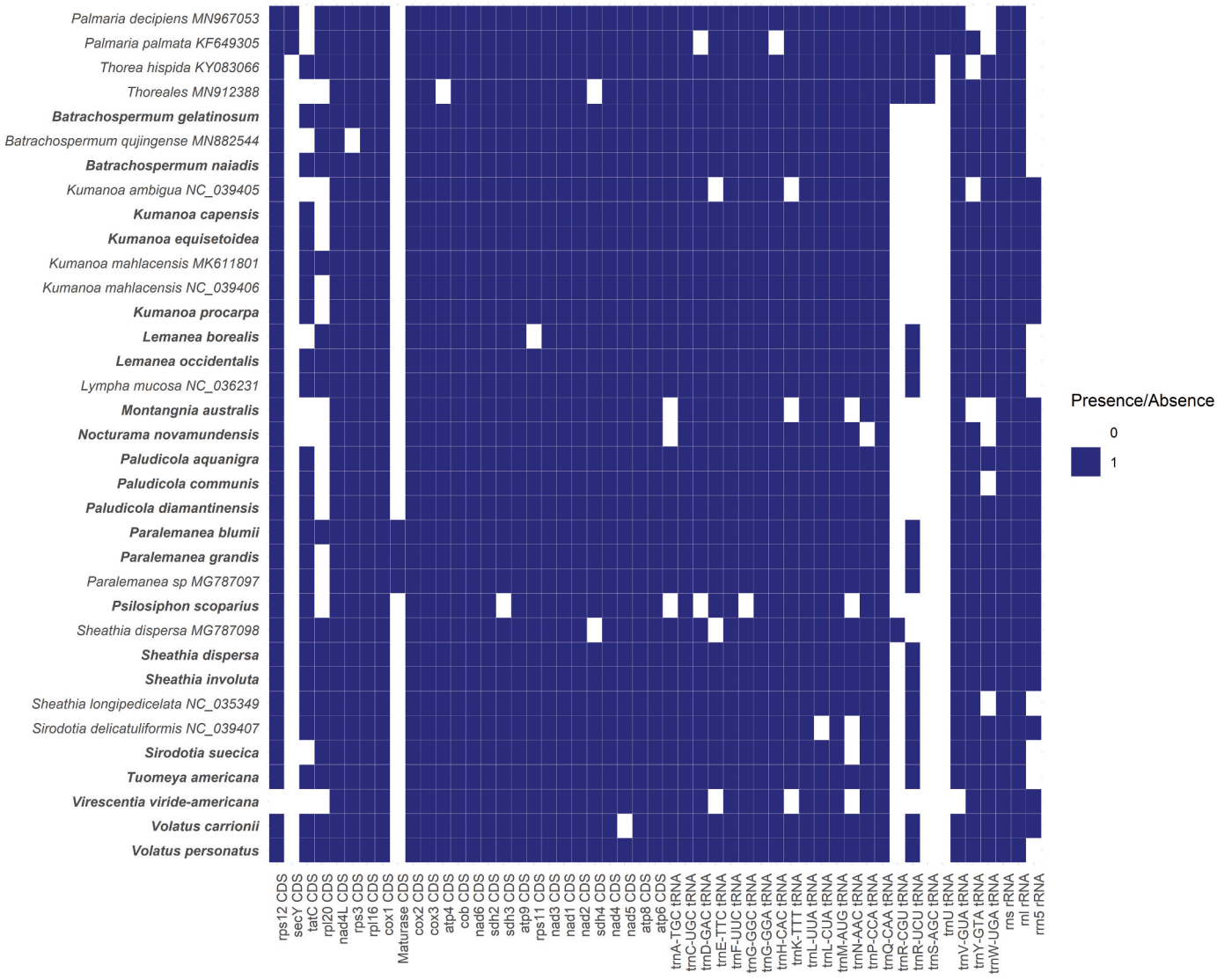


Fig. 4. Heatmap of coding genes occurrence of the mitogenomes in species of Batrachospermales and of the closest related orders (Palmariales and Thoreales).

A particularly interesting characteristic was observed in the three species of *Paralemanea*, where we found a group II intron carrying an ORF coding for the corresponding intron maturase that interrupted the *cox*1 gene. Kim *et al.* (2022) demonstrated that the main factors responsible for mitogenome expansion are introns and intergenic regions. The current data on *Paralemanea* species corroborated this pattern since they have the largest genomes within the Batrachospermales species sequenced to date, the largest intergenic regions, and are the only ones that contain introns. No other member of the Batrachospermales had this intron, not even the sister genus *Lemanea*. However, Florideophyceae taxa that are much more distantly related to the Batrachospermales possess the *cox*1 gene with an intron containing an ORF (*Ahnfeltia* – Ahnfeltiales and *Hildenbrandia* – Hildenbrandiales) and without the ORF (*Plocamium* – Plocamiales, Yang *et al.*, 2015). Furthermore, while the *cox*1 gene in these florideophytes is interrupted by a single intron, other species have more than one group II introns: two in *Pyropia* spp. (Bangiophyceae) and four or five in *Porphyridium* spp. (Porphyridiophyceae; Kim *et al.*, 2022). Additionally, other groups I and II introns in different regions of the mitogenome have been reported in other members of red algae (Depriest *et al.*, 2014; Hancock *et al.*, 2010; Harden *et al.*, 2016; Kim *et al.*, 2022; Seo *et al.*, 2023; Yang *et al.*, 2016). The analysis of group II introns in *Porphyridium* mitogenomes by Kim *et al.* (2022) demonstrated the dynamic nature of group II intron evolution, supporting the occurrence of lateral movement of group II introns among diverse eukaryotes, and revealed their ability to proliferate, once integrated in mitochondrial DNA. Although the origin and evolution of group II introns remains unknown, more genome data are necessary to provide a more accurate picture of the dynamic history of these mobile genetic elements (Kim *et al.*, 2022).

Our findings revealed only three minor gene losses in the mitogenomes of members of the order Batrachospermales, whose gene content is otherwise similar to other orders of

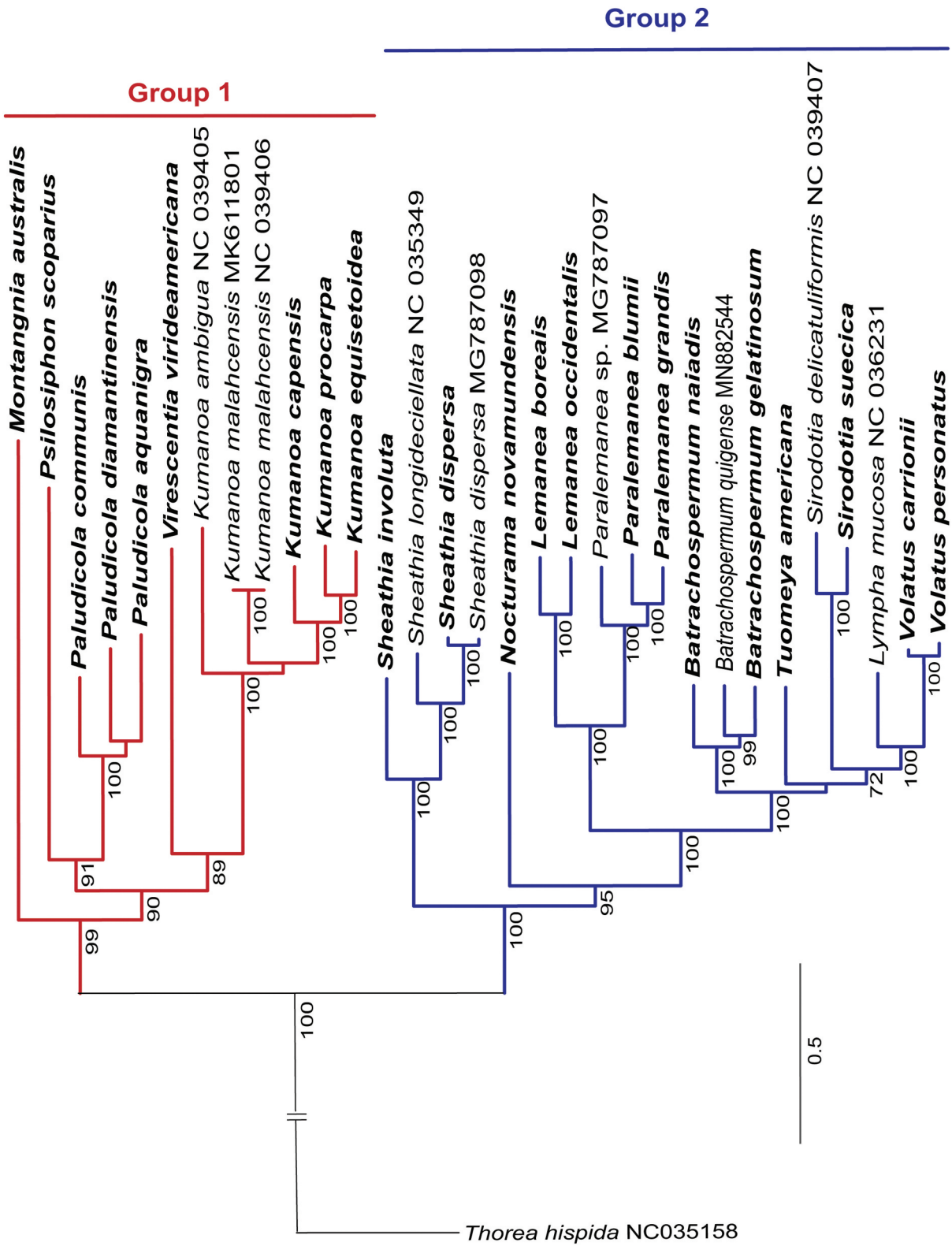


Fig. 5. Maximum Likelihood tree of the taxa of Batrachospermales with Thoreales as the outgroup inferred from the concatenated alignment of 17 mitochondrial genes. Newly sequenced taxa are in boldface. Support values are bootstrap % and the branches with no number had < 70% support. Scale represents substitutions per site.

the Nemaliophycidae (Palmariales and Thoreales) and the class Florideophyceae (Kim *et al.*, 2022; Van Beveren *et al.*, 2022): 22–27 vs 18–34 protein-coding genes, respectively. Synteny among members of Batrachospermales was shown to be highly conserved, which corroborates previously published data for the order (Paiano *et al.*, 2018) and for the Florideophyceae (Van Beveren *et al.*, 2022). One exception within the order was the finding of a small, divergent region

in the five *Kumanoa* species sequenced to date. This new finding might have some potential taxonomic significance in future comparative analyses of mitogenomes with other genera of the order.

The phylogenetic tree revealed two large groups within the Batrachospermales. It is noteworthy that in previous phylogenetic analyses, these two groups were also recovered containing the same set of genera in either in single gene analysis (Entwisle *et al.*, 2009) or in multigene (Lam *et al.*, 2016) and genomic analyses (Paiano *et al.*, 2017, 2018). The existence of two groups may have taxonomic significance in the infraordinal rank, potentially allowing the recognition of more than one family. However, a proposal for rearrangement of the genera within the order is premature at the current. In addition, no clear-cut morphological features are evident to distinguish each major group. All genera with more than one mitogenome sequenced were confirmed as monophyletic groups reinforcing their status of good taxonomic entities as indicated in previous phylogenetic analyses (Entwisle *et al.*, 2009; Vis & Necchi, 2021).

ACKNOWLEDGEMENTS

We sincerely thank: FAPESP for the grant-in-aid to the project and Technical Training Scholarship to NMLS (Proc 2023/00544-2), Postdoc Scholarship to MOP (Proc 2023/10342-8), MS scholarship to YMSL (Proc 2022/12124-5); CNPq for Senior Researcher Scholarship to ON (Proc 303077/2021-2); the staff of the Life Sciences Core Facility (LaCTAD) from State University of Campinas (UNICAMP), for the genomic analysis; Center for Scientific Computing (NCC/GridUNESP) of the São Paulo State University (UNESP) for allowing the use of high performance computing resources; G.W. Saunders and S.J. Shainker-Connelly for help collecting taxa; L. Ellie Becklund for her assistance with CTAB extractions, and library preparations.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

FUNDING

This project was supported by FAPESP grant to ON (Proc [2020/11064-3]; Northeast Algal Society Graduate Research Grant to RMC, an International Phycological Society Paul C. Silva Student Grant to RMC, the Ohio University Graduate Student Senate Original Work Grant to RMC and Ohio University research incentive funds to MLV.

ORCID

Nadia M. Lemes Da Silva http://orcid.org/0000-0001-7769-8601
Roseanna M. Crowell http://orcid.org/0000-0003-3117-2094
Morgan I. Vis http://orcid.org/0000-0003-3087-1563
Wayne B. Chiasson http://orcid.org/0009-0005-4980-6536
Monica o. Paiano http://orcid.org/0000-0001-9200-3433
Orlando JR. Necchi http://orcid.org/0000-0001-7354-2124

REFERENCES

Andrews, S. (2016). *FastQC: A quality control tool for high throughput sequence data*. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Beck, N., & Lang, B. F. (2010). *MFannot, organelle genome annotation websever*. https://github.com/BFL-lab/Mfannot

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Chan, P. P., & Lowe, T. M. (2019). tRNAscan-SE: Searching for tRNA Genes in genomic sequences. In M. Kollmar, (Ed.), *Gene prediction: methods and protocols* (Vol. 1962, pp 1–14). Humana. https://doi.org/10.1007/978-1-4939-9173-0_1

Costa, J. F., Lin, S., Macaya, E. C., Fernández-García, C., & Verbruggen, H. (2016). Chloroplast genomes as a tool to resolve red algal phylogenies: A case study in the Nemaliales. *BMC Ecology and Evolution*, *16*, 1–13. https://doi.org/10.1186/s12862-016-0772-3

Darling, A. E., Mau, B., Perna, N. T., & Stajich, J. E. (2010). ProgressiveMauve: Multiple genome alignment with gene gain, loss, and rearrangement. *PLOS ONE*, *5*, Article e11147. https://doi.org/10.1371/journal.pone.0011147

Depriest, M. S., Bhattacharya, D., & Lopez-Bautista, J. (2014). The mitochondrial genome of *Grateloupia taiwanensis* (Halymeniaceae, Rhodophyta) and comparative mitochondrial genomics of red algae. *Biological Bullettin*, *227*, 191–200. https://doi.org/10.1086/BBLv227n2p191

Diaz-Tapia, P., Maggs, C. A., West, J. A., & Verbruggen, H. (2017). Analysis of chloroplast genomes and a supermatrix inform reclassification of the Rhodomelaceae (Rhodophyta). *Journal of Phycology*, *53*, 920–937. https://doi.org/10.1111/jpy.12553

Dierckxsens, N., Mardulyn, P., & Smits, G. (2017). NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research*, *45*(4), e18. https://doi.org/10.1093/nar/gkw955

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*, 1792–1797. https://doi.org/10.1093/nar/gkh340

Entwisle, T. J., Vis, M. L., Chiasson, W. B., Necchi, O., Jr, & Sherwood, A. R. (2009). Systematics of the Batrachospermales (Rhodophyta) – A synthesis. *Journal of Phycology*, *45*, 704–715. https://doi.org/10.1111/j.1529-8817.2009.00686.x

Greiner, S., Lehwark, P., & Bock, R. (2019). Organellar Genome DRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research*, *47*(W1), W59–W64. https://doi.org/10.1093/nar/gkz238

Hancock, L., Goff, L., & Lane, C. (2010). Red algae lose key mitochondrial genes in response to becoming parasitic. *Genome Biology and Evolution*, *2*, 897–910. https://doi.org/10.1093/gbe/evq075

Harden, L. K., Morales, K. M., & Hughey, J. R. (2016). Identification of a new marine algal species *Pyropia nitida* sp. nov. (Bangiales: Rhodophyta) from Monterey, California. *Mitochondrial DNA, Part A*, *27*, 3058–3062. https://doi.org/10.3109/19401736.2015.1063137

Iha, C., Grassa, C. J., Lyra, G. M., Davis, C. C., Verbruggen, H., Oliveira, M. C., & Müller, K. (2018). Organellar genomics: A useful tool to study evolutionary relationships and molecular evolution in Gracilariaceae (Rhodophyta). *Journal of Phycology*, *54*, 775–787. https://doi.org/10.1111/jpy.12765

Janouškovec, J., Shao-Lun, L., Martone, P. T., Carré, W., Leblanc, C., Collén, J., Keeling, P. J., & Bhattacharya, D. (2013). Evolution of red algal plastid genomes: Ancient architectures, introns, horizontal gene transfer, and taxonomic utility of plastid markers. *PLOS ONE*, *8*, Article e59001. https://doi.org/10.1371/journal.pone.0059001

Jin, J.-J., Yu, W. B., Yang, J. B., Song, Y., Pamphilis, C. W., Yi, T. S., & Li, D. Z. (2020). GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology*, *21*, 241. https://doi.org/10.1186/s13059-020-02154-5

Kim, D., Lee, J., Cho, C. H., Kim, E. J., Bhattacharya, D., & Yoon, H. S. (2022). Group II intron and repeat-rich red algal mitochondrial genomes demonstrate the dynamic recent history of autocatalytic RNAs. *BMC Biology*, *20*, 2. https://doi.org/10.1186/s12915-021-01200-3

Kumano, S. (2002). *Freshwater red algae of the world*. Biopress.

Lam, D. W., Verbruggen, H., Saunders, G. W., & Vis, M. L. (2016). Multigene phylogeny of the red algal subclass Nemaliophycidae. *Molecular Phylogenetics and. Evolution*, *94*, 730–736. https://doi.org/10.1016/j.ympev.2015.10.015

Lee, J. M., Boo, S. M., Mansilla, A., & Yoon, H. S. (2015). Unique repeat and plasmid sequences in the mitochondrial genome of *Gracilaria chilensis* (Gracilariales, Rhodophyta). *Phycologia*, *54*(54), 20–23. https://doi.org/10.2216/PH14-97.1

Lee, J. M., Cho, C. H., Park, S. I., Choi, J. W., song, H. S., West, J. A., Bhattacharya, D., & Yoon, H. S. (2016). Parallel evolution of highly conserved plastid genome architecture in red seaweeds and seed plants. *BMC Biology*, *14*, 75. https://doi.org/10.1186/s12915-016-0299-5

Lee, J. M., Song, H. J., Park, S. I., Lee, Y. M., Jeong, S. Y., Cho, T. O., Yoon, H. S., Choi, H.-G., Choi, C. G., Nelson, W. A., Fredericq, S., Bhattacharya, D., & Yoon, H. S. (2018). Mitochondrial and plastid genomes from coralline red algae provide insights into the incongruent evolutionary histories of organelles. *Genome Biology and Evolution*, *10*, 2961–2972. https://doi.org/10.1093/gbe/evy222

Matsuzaki, M., Misumi, O., Shin-I, T., Maruyama, S., Takahara, M., Miyagishima, S., Kuroiwa, T., Nishida, K., Yagisawa, F., Nishida, K., Yoshida, Y., Nishimura, Y., Nakao, S., Kobayashi, T., Momoyama, Y., Higashiyama, T., Minoda, A., Sano, M., Nomoto, H., … Kuroiwa, T. (2004). Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature*, *428*, 653–657. https://doi.org/10.1038/nature02398

Muñoz-Gómez, S., Mejía-Franco, F. G., Durnin, K., Colps, M., Grisdale, C. J., Archibald, J. M., & Slamovits, C. H. (2017). The new red algal subphylum Proteorhodophytina comprises the largest and most divergent plastid genomes known. Current Biology, *27*, 1677–1684. https://doi.org/10.1016/j.cub.2017.04.054

Nan, F., Feng, J., Lv, J., Liu, Q., Fang, K., Gong, C., & Xie, S. (2017). Origin and evolutionary history of freshwater Rhodophyta: Further insights based on phylogenomic evidence. *Scientific Reports*, *7*, Article 2934. https://doi.org/10.1038/s41598-017-03235-5

Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: A new versatile metagenomic assembler. *Genome Research*, *27*, 824–834. https://doi.org/10.1101/gr.213959.116

Paiano, M. O., Del Cortona, A., Costa, J. F., Liu, S.-L., Verbruggen, H., De Clerck, O., Necchi, O., Jr, & Amsler, C. (2017). Organization of plastid genomes in the freshwater red algal order Batrachospermales (Rhodophyta). *Journal of Phycology*, *54*, 25–33. https://doi.org/10.1111/jpy.12602

Paiano, M. O., Del Cortona, A., Costa, J. F., Liu, S.-L., Verbrugen, H., De Clerck, O., & Necchi, O., Jr. (2018). Complete mitochondrial genomes of six species of the freshwater red algal order Batrachospermales (Rhodophyta). *Mitochondrial DNA, Part B*, 2,:607–610. https://doi.org/10.1080/23802359.2018.1473734

Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A., & Korobeynikov, A. (2020). Using SPAdes de novo assembler. *Current Protocols in Bioinformatics*, *70*, Article e102. https://doi.org/10.1002/cpbi.102

Pueschel, C. M., & Cole, K. M. (1982). Rhodophycean pit plugs: An ultrastructural survey with taxonomic implications. *American Journal of Botany*, *69*, 703–720. https://doi.org/10.1002/j.1537-2197.1982.tb13310.x

R Core Team. (2021). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Saunders, G. W. (2005). Applying DNA barcoding to red macroalgae: A preliminary appraisal holds promise for future applications. *Philosophical Transactions of the Royal Society B*, *360*, 1879–1888. http://doi.org/10.1098/rstb.2005.1719

Sayers, E. W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K. D., & Karsch-Mizrachi, I. (2020). GenBank. *Nucleic Acids Research*, *48*, D84–D86. https://doi.org/10.1093/nar/gkz956

Schenk, J. J., Becklund, L. E., Carey, S. J., & Fabre, P. P. (2023). What is the "modified" CTAB protocol? Characterizing modifications to the CTAB DNA extraction protocol. *Applications in Plant Sciences*, *11*(3), Article e11517. https://doi.org/10.1002/aps3.11517

Seo, M. H., Kang, S. C., Kim, K. M., Kwak, M. S., Jo, J., Choi, H. G., Boo, G. H., & Yoon, H. S. (2023). Novel rearrangements in the mitochondrial genomes of the Ceramiales (Rhodophyta) and evolutionary implications. *Algae*, *38*(4), 253–264.

Sheath, R. G., Müller, K. M., Vis, M. L., & Entwisle, T. J. (1996). A re-examination of the morphology, ultrastructure and classification of genera in the Lemaneaceae (Batrachospermales, Rhodophyta). *Phycological Research*, *44*, 233–246. https://doi.org/10.1111/j.1440-1835.1996.tb00053.x

Stamatakis, A. (2014). rAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

van Beveren, F., Eme, L., López-García, P., Ciobanu, M., Moreira, D., & Sloan, D. (2022). Independent size expansions and intron proliferation in red algal plastid and mitochondrial genomes. *Genome Biology and Evolution*, *14*, Article evac037. https://doi.org/10.1093/gbe/evac037

Vis, M. L., & Necchi, O., Jr. 2021. Subphylum eurhodophytina, Class Florideophyceae, Subclass Nemaliophycidae, order Batrachospermales. In M. L. Vis & O. Necchi Jr. (Eds.), *Freshwater red algae: Phylogeny, taxonomy and biogeography* (pp 129–332). Springer Nature https://doi.org/10.1007/978-3-030-83970-3_5

Wick, R. R., Schultz, M. B., Zobel, J., & Holt, K. E. (2015). Bandage: Interactive visualization of de novo genome assemblies. *Bioinformatics*, *31*, 3350–3352.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag. https://ggplot2.tidyverse.org

Wickham, H., Averick, M., Brya, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*, Article 1686. https://doi.org/10.21105/joss.01686

Wickham, H., François, R., Henry, L., & Müller, K. (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.3. https://CRAN.R-project.org/package=dplyr

Wickham, H., & Bryan, J. (2023). *readxl: Read Excel Files*. R package version 1.4.3. https://CRAN.R-project.org/package=readxl

Wolf, D. I., Evans, J. R., & Vis, M. L. (2017). Complete mitochondrial genome of the freshwater red alga Lympha mucosa (Rhodophyta). *Mitochondrial DNA Part B*, *2*(2), 707–708.

Yang, E. C., Kim, K. M., Kim, S. Y., Lee, J. M., Boo, G. H., Lee, J. H., Yoon, H. S., Yi, G., Schmidt, W. E., Fredericq, S., Boo, S. M., Bhattacharya, D., & Yoon, H. S. (2015). Highly conserved mitochondrial genomes among multicellular red algae of the Florideophyceae. *Genome Biology and Evolution*, *7*, 2394–2406. https://doi.org/10.1093/gbe/evv147

Yang, E. C., Boo, S. M., Bhattacharya, D., Saunders, G. W., Knoll, A. H., Fredericq, S., Graf, L., & Yoon, H. S. (2016). Divergence time estimates and the evolution of major lineages in the florideophyte red algae. *Scientific Reports*, *6*, Article 21361. https://doi.org/10.1038/srep21361

Zar, J. H. (2010). *Biostatistical Analysis* (5th ed). Prentice Hall, Upper Saddle River.

Anexo B

PLASTID GENOME STRUCTURE AND PHYLOGENOMICS OF THE FRESHWATER RED ALGAL ORDER BATRACHOSPERMALES (RHODOPHYTA)

Journal:	*Journal of Phycology*
Manuscript ID	Draft
Manuscript Type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Crowell, Roseanna; Ohio University, Environmental and Plant Biology Lemes da Silva, Nadia; São Paulo State University, Biological Sciences Department Paiano, Monica; São Paulo State University, Biological Sciences Department Vis, Morgan; Ohio University, Environmental and Plant Biology Necchi Jr., Orlando; São Paulo State University, Biological Sciences Department
Keywords:	morphology, phylogeny, Rhodophyta
Alternate Keywords:	Batrachospermum, introns, plastid genome, phylogenomics

PLASTID GENOME STRUCTURE AND PHYLOGENOMICS OF THE FRESHWATER RED ALGAL ORDER BATRACHOSPERMALES (RHODOPHYTA)

Roseanna M. Crowell[1,3], Nadia M. Lemes Da Silva[2], Monica O. Paiano[2], Morgan L. Vis[1*], Orlando Necchi Jr.[2]

[1] Department of Environmental and Plant Biology, Ohio University, Athens, OH, 45701, USA

[2] Biological Sciences Department, São Paulo State University, Rua Cristóvão Colombo, 2265, São José do Rio Preto 15054-000, Brazil

[3]Current Address: Department of Natural Sciences, University of Houston–Downtown, Houston, TX, 77002, USA

[*]Author to whom correspondence should be addressed: Morgan L. Vis (vis-chia@ohio.edu)

RUNNING TITLE: PLASTID GENOMES OF THE BATRACHOSPERMALES

KEY WORDS: Batrachospermum, introns, plastid genome, morphology, phylogenomics

ABSTRACT

For the freshwater red algal order Batrachospermales, the number of plastid genomes available is relatively few compared to the number of genera. Fully assembled plastid genomes can provide insights into plastid evolution and crucial data for phylogenetic reconstruction. In the

present study, 18 plastid genomes were generated for a total of 40 plastid genomes from 38 species representing 18 of the 23 genera. The greatly expanded dataset allowed for comparison of the plastid genome structural characteristics with the other orders in the Nemaliophycidae and inference of the phylogenetic relationships of the genera within the order. Results showed the genomes had either one or two RNA operons and this variation could be intrageneric. All genomes had the *chl*B gene with an intron like all Nemaliophycidae but lacked the *apc*F gene present in all Nemaliophycidae. The loss of the *pbs*A gene was variable in the Batrachospermales and the Nemaliophycidae. Phylogenetic analysis using a 126-gene concatenated dataset produced a fully-supported Batrachospermales. In addition, generally high support for the relationships among the genera resulted in the most robust phylogeny to date. Nevertheless, it also highlighted that potentially more data will be needed to resolve the relationship among sections of *Nothocladus* and other related genera. Overall, the batrachospermalean genera were split into two well-supported lineages which has been noted in other studies using plastid and mitochondrial genomes. However, we are lacking a combination of characters to distinguish these two lineages as the morphological characters to describe taxa are shared between them.

INTRODUCTION

The number of organellar genomes, both plastomes and mitogenomes, available in the Rhodophyta has been steadily growing primarily due to the increased ease and decreased cost of high-throughput sequencing. The increased data available from organellar genomes have been valuable for phylogenetic inference at the higher taxonomic levels (phyla and classes) in red algae (Janouškovec et al., 2013; Lee et al., 2015, 2016; Muñoz-Gómez et al., 2017; Yang et al.,

2016). Additionally, studies at lower taxonomic levels (orders and families) of red algae have also advanced due to increased data provided by organellar genome sequencing (Costa et al., 2016; Iha et al., 2018; Lee et al., 2018; Díaz-Tapia et al., 2019; Lemes da Silva et al., 2025).

Complete plastid genome data have been used in an array of studies focused on phylogeny, structural rearrangements and evolution of red algae (Costa et al., 2016; Lee et al., 2016; van Beveren et al., 2022; Borg et al., 2023; San et al., 2025). The conserved structure of plastid genomes in red algae, along with their predominantly uniparental inheritance, have made plastids the main targets for understanding the evolutionary relationships of Rhodophyta (Janouškovec et al., 2013; Costa et al., 2016). In addition, red algal plastid genomes are particularly interesting, because they contain the largest gene complement among all known plastids, are compactly organized, and slow evolving compared with other plastid containing lineages (Janouškovec et al., 2013).

Red algae have plastid genomes that vary greatly in size especially in the early divergent subphyla. The genome sizes typically range from 150,000 nucleotides (nt) in *Cyanidioschyzon merolae* (Cyanidiophyceae) to 259,000 nt in *Porphyridium sordidum* (Porphyridiophyceae) (van Beveren et al., 2022). Notably, some members of Proteorhodophytina have the largest genomes sequenced to date from 205,000 to 1,127,000 nt (Muñoz-Gómez et al., 2017). However, the highly morphologically diverse and species rich class Florideophyceae of the Eurhodophytina have plastid genomes reported with a much narrower range (167,000–195,000 nt) (Lee et al., 2016; van Beveren et al., 2022).

The number of protein coding genes in red algae is also the highest among algal groups (184 to 235, Lee et al., 2016; van Beveren et al., 2022). Plastid genomes of red algae not only have the highest gene content among eukaryotes with primary plastids (Archaeplastida), but their

genomes are also the most conserved (Janouškovec et al., 2013; Lee et al., 2016). Plastid genomes are highly suitable for phylogenomic studies as they are present in multiple copies per cell making sequences easily obtained from genomic DNA extractions. Additionally, they appear to contain sufficient phylogenetic signal to resolve relationships across various taxonomic levels (Costa et al., 2016; Iha et al., 2018; Muñoz-Gómez et al., 2018). Studies using multiple genes or phylogenomic data have shown that addressing individual orders of Nemaliophycidae can elucidate relationships within and between closely related groups (Costa et al., 2016; Lam et al., 2016; Paiano et al., 2017; Díaz-Tapia et al., 2019).

The strictly freshwater orders Batrachospermales, Thoreales and Balbianiales are classified within the subclass Nemaliophycidae (Lam et al., 2016; Vis & Necchi, 2021). The order Batrachospermales (Rhodophyta) previously was divided into three families (Batrachospermaceae, Lemaneaceae and Psilosiphonaceae, Pueschel & Cole, 1982; Sheath et al., 1996) but is currently treated as having a single family Batrachospermaceae (Entwisle et al., 2009) with 23 genera currently recognized (Vis & Necchi, 2021; Jayalakshmi & John, 2023; Necchi et al., 2024). The Batrachospermales is the most diverse group in terms of number of taxa, morphology, and reproductive characters among the exclusively freshwater orders of red algae (Entwisle et al., 2009; Kumano, 2002; Lam et al., 2016; Vis & Necchi, 2021). Although the order is well supported in all phylogenetic analyses (Entwisle et al., 2009; Lam et al., 2016; Paiano et al., 2017; Vis & Necchi, 2021), the infra-ordinal classification is not yet well supported. Phylogenetic analyses have revealed two lineages containing the same set of genera either in single gene (Entwisle et al., 2009), multigene (Lam et al., 2016) and most recently in genomic analyses as well (Paiano et al., 2017, 2018; Lemes da Silva et al., 2025). These two lineages may have taxonomic significance in the infraordinal rank, potentially allowing the

recognition of more than one family. However, a robust phylogeny and an analysis of character evolution are needed to assess morphological features that can be potentially applied to recognize each clade as a family.

Despite recent advances in the study of organellar genomes, plastid genome data for Batrachospermales remain limited. Paiano et al. (2017) generated seven plastid genomes of taxa within the order but each distantly related from the other. Two other studies produced a single plastid genome each and there have been a few others without publications bringing the total to 12 available in GenBank (Nan et al., 2017; Cho et al., 2018; Evans & Vis, 2020; Sayers et al., 2020). Crowell et al. (in review) recently expanded this dataset by focusing on seven closely related genera, demonstrating that whole plastid genome data can represent phylogenetic relationships among these taxa with high statistical support. In addition, genera with similar morphological characters were shown to be most closely related (Crowell et al., in review).

Given the well supported phylogeny among a small number of genera of previous studies, a more expansive approach including the majority of the genera in the Batrachospermales was undertaken. The research goals were as follows: to document the structural characteristics of the plastid genomes in the Batrachospermales for comparison with the other red algal orders in the Nemaliophycidae; and to infer the phylogenetic relationships of the genera using taxon sampling that represents a broad range of vegetative and reproductive morphology within the Batrachospermales.

MATERIALS AND METHODS

Taxon sampling

A total of 18 plastid genomes were generated from 11 genera including seven genera without previously published data (Table S1). When added to previously published data for the order, the data set includes 40 plastid genomes for 38 species from 18 genera, including genera representative of the wide variety of vegetative and reproductive characters of the Batrachospermales (Table 1). Each live specimen was divided into two portions: (i) for DNA sequencing by immediately being placed in silica gel; and (ii) for vouchering by either liquid preservation or pressed on herbarium paper and deposited in recognized herbaria (Thiers, 2025) (Table S1). Specimens were identified based on current diagnostic characters using recent taxonomic literature (Vis & Necchi, 2021). In addition, the morphological identifications were confirmed by comparing *rbc*L gene sequence with those in GenBank (Sayers et al., 2020).

DNA extraction and sequencing

Three methods for extraction and sequencing were used. For 11 specimens, genomic DNA (gDNA) was extracted using the Qiagen Magattract High Molecular Weight Kit (Qiagen, Hilden, Germany), following the manufacturer's protocol. The gDNA was quantified using a Qubit 2.0 fluorometer (ThermoFisher Scientific, Waltham, Massachusetts, USA) and Quant-it dsDNA High- Sensitivity Assay kit. Quality checking of gDNA was conducted in a TapeStation 4200 system (Agilent, Santa Clara, USA). Libraries were prepared with the Illumina DNA prep protocol and sequencing was performed on an Illumina NextSeq 2000 platform (Illumina, San Diego, California, USA) with paired-end 2 × 150 bp. For *Montagnia australis*, *Paludicola communis* and *Virescentia viride-americana*, the tissue was ground in liquid nitrogen and gDNA was extracted using the Nucleospin Plant II kit (Macherey–Nagel, Düren, Germany) following the manufacturer's protocol. The gDNA quantity and quality was assessed using a Qubit 3.0

fluorometer (ThermoFisher Scientific, Waltham, Massachusetts, USA). Libraries of 200–450 nucleotide fragments were completed using the New England BioLabs NEBNext Ultra II FS DNA Library Kit following the protocol for inputs ≤ 100ng and sequenced using the Illumina MiSeq platform (Illumina, San Diego, California, USA) with paired-end 2 x 150 bp reads. For *Nothocladus kraftii*, *N. ranulifer*, *N. theaquus* and *Notohesperus serendipidus*, the material was homogenized in a bead-beater and a modified CTAB extraction protocol with Polyvinylpyrrolidone-10 was utilized (Schenk et al., 2023). The gDNA quantity and quality was assessed using a NanoDrop ND-1000 spectrophotometer (ThermoFisher Scientific, Waltham, Massachusetts, USA). Sequencing of these samples was completed by SeqCenter (Pittsburgh, Pennsylvania) on the Illumina NovaSeq X Plus using libraries of 280 nucleotide fragments generating paired end 2 x 150 bp reads with an Illumina DNA Prep protocol.

Plastid genome assembling and annotation

For 14 samples, the quality of trimmed Illumina reads was checked with FastQC (v0.12.1; Andrews, 2016). Illumina raw reads were trimmed with Trimmomatic version 0.39 (Bolger *et al*., 2014), to remove adapters and low-quality reads, with the following parameters: ILLUMINACLIP: adapters.fa: 2:30:10; SLIDINGWINDOW: 4:20; MINLEN: 36. After trimming, the clean reads were used for the *de novo* genome assembly. Three assemblers were used depending on the species. NOVOPlasty version 4.3.1 (Dierckxsens et al., 2017) was used with the *rbc*L sequence from closely related genera as the seed sequence for assembly, with a k-mer size of 29 and an insert size of 300. GetOrganelle version 1.5.1c (Jin et al., 2020) was run with k-mers 21, 45, 65, 85, and 105, along with the parameters -R 100 and -F other_pt, using closely related plastomes as seed references. SPAdes version 3.15.5 (Nurk et al., 2017; Prjibelski

et al., 2020) was also employed. The resulting scaffolds from SPAdes were sorted using tBLASTn (e-value: 1e-10) against a database of plastomes from the same genus, or, when unavailable, from closely related genera within Batrachospermales, to identify plastome-related scaffolds. Genome assemblies obtained with GetOrganelle and SPAdes were visualized and manually inspected using Bandage version 0.8.1 (Wick et al., 2015).

For *Nothocladus kraftii*, *N. ranulifer*, *N. theaquus* and *Notohesperus serendipidus*, paired-end reads were trimmed using the trim ends function in Geneious Prime 2024.0.7 with default parameters and the first 15 million were used for a *de novo* assembly using low-sensitivity parameters in Geneious Prime 2024.0.7 (Biomatters, geneious.com). The plastid genome of *Batrachospermum gelatinosum* was used as a reference to identify contigs containing plastid sequence. These plastid contigs were trimmed and *de novo* assembled. To close gaps in the assembled sequence, contigs were extended by iteratively mapping reads with low sensitivity and a minimum mapping quality of 30 until complete plastid genome sequences could be circularized.

For all genomes, annotation was performed using MFannot (Beck & Lang, 2010) or Geneious Prime 2024.0.7. Annotations generated in MFannot were carefully reviewed and manually refined in Geneious Prime 2024.0.7 using the 'Find ORFs' function. For genomes annotated directly in Geneious Prime, all rRNA and protein-coding genes were identified by extracting sequences from the published plastid genome of *Batrachospermum gelatinosum* (PX060844, Crowell et al. in review). A second RNA operon, when present, was identified using the 'find repeats' function in Geneious Prime with a minimum repeat length of 100 nt. If a second RNA operon was present, the large single copy (LSC) and small single copy (SSC) region were annotated. The tRNAs were identified with tRNAscan-SE version 2.0.7 annotator in

GeSeq with default parameters (Chan & Lowe, 2019). Genome maps were generated in Geneious Prime 2024.0.7 (Biomatters geneious.com).

To assess the plastid genome structure and gene content within the Batrachospermales, the genome annotations for the 20 previously published plastid were reviewed and updated as needed. In GenBank there are plastid data available for members of the Acrochaetiales, Balbianiales, Nemaliales, Palmariales and Thoreales that could be used as a comparison with the Batrachospermales. Representative genomes from these orders were also reviewed and updated for comparison (Tables 2, S2). Synteny among Batrachospermales plastid genomes was investigated via whole-genome alignment with the progressiveMauve algorithm version 2.4.0 (Darling *et al*., 2010) within Geneious Prime using automatically calculated seed weights, and automated calculation of locally collinear block (LCB) scores.

Phylogenomic analyses

A total of 40 Batrachospermales sequences were used in the phylogenetic analyses. These genomes included two each for *Kumanoa mahlacensis* and *Sheathia dispersa* and one each for 38 species representing 18 genera. For the outgroup, the following plastid genomes were used: *Acrochaetium subsecundum* (MH026107), *Balbiania investiens* (MH026108), *Nemalion* sp. (LT622871), *Palmaria palmata* (KX284726) and *Thorea hispida* (KX284714) (Table 2). A conservative approach was taken and only genes that were present in all taxa or only missing in one taxon were considered for phylogenetic analyses. Alignment of the individual genes was conducted in Geneious Prime using the MUSCLE v5.1 algorithm (Edgar, 2022).

Phylogenetic analyses were conducted using the genes that readily aligned including 118 genes present in all taxa and eight genes that were present in 44 of the 45 taxa. These 126 genes

were concatenated using AMAS (Alignment Manipulation and Summary; Borowiec, 2016). The final alignment was 89,919 nt positions in length. The best-fit model, GTR+F+R6, was determined for the concatenated alignment using ModelFinder in IQ-Tree version 2.3.5 (Kalyaanamoorthy et al., 2017; Minh et al., 2020). A maximum likelihood tree with 1,000 ultrafast bootstraps (Hoang et al., 2018) was generated with the concatenated alignment partitioned by gene and codon in IQ-Tree 2.3.5.

Morphological Character Assessment

Morphological characters for the 18 genera under investigation in the Batrachospermales were compiled from the literature (Vis & Necchi, 2021 and references therein). Within the Batrachospermales, the gametophyte thallus structure, the carpogonia and carpogonial branch morphology and size and carposporophyte structure, size, and placement were considered as taxonomically important. Typically, a suite of these characters distinguishes the genera from one another.

RESULTS

DNA Extraction and Sequencing

Three extraction and sequencing methods were used to produce the plastid genomes in this study. For species sequenced on Illumina NextSeq 2000 platform, the number of paired-end reads were as follows: *Acarposporophycos brasiliensis* (8,450,584)*, Kumanoa capensis* (54,966,038)*, K. equisetoidea* (13,746,414)*, K. procarpa* (15,863,342)*, Nocturama novamundensis* (22,122,012), *Paludicola aquanigra* (4,602,468), *P. diamantinensis* (7,988,020), *Psilosiphon scoparius* (17,193,174), *Sheathia dispersa* (17,792,160), *S. involuta* (19,356,674)

and *Torularia puiggariana* (52,463,408). For species sequenced on Illumina MiSeq platform, the number of paired reads is as follows: *Montagnia australis* (14,616,660), *Paludicola communis* (15,265,102) and *Virescentia viride-americana* (11,111,410). For samples sequenced on an Illumina NovaSeq X Plus, the number of reads were as follows: *Nothocladus kraftii* (105,879,318), *N. ranulifer* (125,797,194), *N. theaquus* (160,756,500) and *Notohesperus serendipidus* (196,152,700).

Plastid Structure and genome arrangement

The 18 newly sequenced plastid genomes were fully assembled, annotated and mapped as circular molecules (Figs. S1-18) bringing the total number of Batrachospermales plastid genomes to 40. The plastid structure for all Batrachospermales genomes were similar in size, number of genes and GC content (Table 2). The Batrachospermales differ by < 20 kb in size ranging from 171,682 nt in *Virescentia viride-brasilensis* to >189,825 nt in *Lympha mucosa* (mean = 184,715 nt ±3,519). Number of protein coding genes ranged from 189 in *Virescentia viride-brasilensis* to 202 in *Kumanoa equisetoidea* (mean = 199 ±3). Similarly, the GC content ranged from 28.1 in *Montagnia macrospora* to 31.0 in *Paralemanea* sp. (mean = 29.3 ±0.8). The three *Paralemanea* species had the lowest number of tRNA (28) and *Virescentia viride-brasilensis* the highest with 34 (mean = 31 ±1.3). Among the Batrachospermales, the presence of one or two RNA operons was variable across taxa (Table 2). Approximately half of the taxa contained two RNA operons, while the remainder lacked this feature. In taxa with two RNA operons, the genome has a quadripartite structure. Both the previously published and new genome of *Sheathia dispersa* had two RNA operons, likewise both genomes of *Kumanoa mahlacensis* had only one RNA operon.

Within the Batrachospermales, taxa contained either one or two introns as well as notable patterns of gene presence or absence among the genera and species (Table 2, Table S2). For all taxa, an intron in the *chl*B gene was present (Fig. S19). An intron was also present in a *trn*M gene when that gene was present in the genome (Fig. S20). This gene and intron were present in seven species from four genera (Fig. 1, Table S2). The *dfr* gene was absent from the three species of *Paludicola*, whereas the *mla*E gene was missing from all sampled species of *Montagnia* (2)*, Paludicola* (3) and *Virescentia* (2) (Fig. 1, Table S2). The *pbs*A gene had the broadest pattern of loss, missing from the monospecific genera *Acarposporophycos*, *Notohesperus* and *Psilosiphon*, and in *Nothocladus ranulifer* and *N. theaquus* as well as all sampled species of *Kumanoa* (6), *Montagnia* (2) and *Paludicola* (3) (Fig. 1, Table S2). The *his*S gene was missing from *Torularia puiggariana, Nothocladus theaquus* and *Psilosiphon scoparius* (Fig. 1, Table S2). There local collinear blocks were predicted with the MAUVE algorithm for the 40 plastid genomes (Figure S21).

The batrachospermalean plastid genomes and the eight representatives of five orders (Acrochaetiales, Balbianiales, Nemaliales, Palmariales and Thoreales) in the subclass Nemaliophycidae were generally similar in size and number of protein coding genes (Table 2). In terms of gene content, *apc*F was present in the other Nemaliophycidae plastid genomes but absent from the Batrachospermales (Fig. 1, Table S2). The *gr*x gene was only present in *Acrochaetium secundatum* (Acrochaetiales), *Balbiania investiens* (Balbianiales) and *Palmaria palmata* (Palmariales) (Table S2). The genes *ycf*34, *ycf*35 and *ycf*46 were absent from the Batrachospermales and *Balbiania investiens* (Balbianiales) (Table S2). Of the three genes (*dfr*, *pbsA*, *mlaE*) that showed a pattern among taxa in the Batrachospermales, *drf* and *mla*E were present in the other Nemaliophycidae taxa (Fig. 1, Table S2). The *pbs*A gene was missing in

Thorea hispida (Thoreales) and *Nemalion* sp. but was present in other members of Nemaliales and Nemaliophycidae (Fig. 1, Table S2). Either one or two introns were detected in all the other Nemaliophycidae genomes except *Balbiania investiens* (Balbianiales) with none. The seven genomes had the same intron in *chl*B as found in all Batrachospermales. The *trn*M gene was present in *Acrochaetium secundatum* (Acrochaetiales)*, Palmaria palmata* (Palmariales) and *Thorea hispida* (Thoreales) and the intron was detected in those taxa like in the Batrachospermales taxa (Fig. 1, Table S2).

Phylogenomics

The maximum likelihood (ML) tree produced from the concatenated alignment recovered a fully–supported Batrachospermales [ultrafast bootstrap (bs) =100] (Fig. 1). In addition, there was generally high support for the relationship among the genera. There were two well-supported (bs = 100) lineages: the first containing the genera *Batrachospermum*, *Lemanea*, *Lympha*, *Nocturama*, *Nothocladus*, *Paralemanea*, *Sheathia*, *Sirodotia*, *Torularia*, *Tuomeya* and *Volatus*; the second comprising of the genera *Acarposporophycos*, *Kumanoa*, *Montagnia*, *Notohesperus*, *Paludicola*, *Psilosiphon* and *Virescentia* (Fig. 1). There were six branches that were not well supported (bs < 95). The placement of *Tuomeya americana* as sister to *Batrachospermum* spp. had weak support (bs = 79). The genera *Nothocladus*, *Torularia*, *Nocturama*, and *Sheathia* formed a clade, but it was only moderately supported (bs = 90). Likewise, two other relationships within that clade were unsupported: *Nocturama* sister to *Sheathia* spp. (bs = 89) and *Torularia* sister to *Nothocladus ranulifer* (bs = 55) (Fig. 1). Among the genera in the second lineage, the sister relationship of *Kumanoa* and *Virescentia* was

297 unsupported (bs = 60), as well as *Notohesperus serendipidus* sister to *Paludicola,*
298 *Acarposporophycos* and *Psilosiphon* (bs = 57).
299
300 *Morphological Characters*
301 Within the Batrachospermales, characteristics related to the gametophyte thallus,
302 carpogonial branch, carpogonium and carposporophytes may differentiate the genera but also can
303 be shared among genera (Table 1). Pseudoparenchymatous gametophyte thallus was present in
304 *Tuomeya*, *Lemanea* and *Paralemanea* from lineage 1 and *Psilosiphon* in lineage 2 of the
305 phylogeny (Fig. 1). Three genera, *Tuomeya*, *Volatus* (lineage 1) and *Kumanoa* (lineage 2) had
306 curled carpogonial branches. In lineage 1, pedunculate carposporophytes was common being
307 present in *Batrachospermum*, *Nocturama*, *Sheathia* as well as *Nothocladus* (Sections *Kraftii* and
308 *Australicus*), but only in *Montagnia* from lineage 2. The character state of large
309 carposporophytes was prevalent in both lineages with *Tuomeya*, *Lympha*, *Volatus*, *Nothocladus*
310 (Sections *Australicus* and *Theaquus*), and *Torularia* in lineage 1 and *Kumanoa*, *Virescentia*,
311 *Notohesperus* and *Paludicola* in lineage 2 (Fig. 1).
312
313 *DISCUSSION*
314 *Genome structure*
315 Our analysis of 40 plastid genomes (including 18 newly sequenced) has expanded variations of
316 members of the Batrachospermales for most characteristics, particularly genome size, number of
317 coding genes, and GC content, but within the range reported for other orders of Nemaliophycidae
318 (Table 2; Kim et al., 2022; Van Beveren et al., 2022). It is noteworthy that the values for these
319 three genome characteristics, were mostly close to the lowest ranges for the other orders:

171.682–189.825 vs 181.215– 192.960 bp for genome size, 189-202 vs 199-206 coding genes and 28.1-31.0 vs 29.6-35.9 % GC. These data support the pattern observed by Paiano et al. (2017) that Batrachospermales plastid genomes tend to be among the smallest and least gene-dense within the Florideophyceae.

The greatly expanded number of plastid genomes in the Batrachospermales has provided corroboration of previous observations about gene losses in the order and new insights. Paiano et al. (2017) documented the lack of *apc*F in eight members of the order, and this finding is confirmed in the current study with 40 plastid genomes. These researchers suggested that this may be interpreted as a gene loss, a gene transfer to the nucleus or potentially the function was replaced by another type of allophycocyanin (Glazier et al., 1997; Chang et al., 2015). The new data indicate that *apc*F is likely lacking in the order as a whole, but which mechanism might be responsible is not clear. The gene *apc*F encodes the allophycocyanin beta subunit, a phycobilin component of the light-harvesting proteins characteristic of red algae (Apt & Grossman, 1993). The hypothesis that a new type of allophycocyanin appeared in the Batrachospermales seems most likely, since a new type of phycocyanobilin-containing phycoerythrin was described for several bluish freshwater red algae species (Glazer et al., 1997), including the “Chantransia” stages of members of Batrachospermales.

Within the Batrachospermales, either one or two RNA operons are present in the plastid genome. Most taxa in lineage 1 having two operons and all taxa in lineage 2 having one. Of the three *Nothocladus* species sequenced, *N. kraftii* had one operon and *N. ranulifer* and *N. theaquus* had two, but the phylogenetic relationships were not well supported. However, the genus *Sheathia* was monophyletic, and the number of operons differed among species confirming that this trait can be variable within a genus. In species for which two genomes were available

(*Sheathia dispersa* and *Kumanoa mahlacensis*) both genomes had the same number which suggests the presence of one or two RNA operons is consistent within species even if it varies among taxa.

Duplicated rRNA operons are widespread in red algae, having been reported in four of the seven classes (Lee et al., 2016), with Compsopogonophyceae, Porphyridiophyceae, and Rhodellophyceae) being the exceptions. Within the Florideophyceae, duplicated rRNA operons occur in all subclasses except the Ahnfeltiophycidae, whereas in the Nemaliophycidae, they are absent only in the Thoreales and the genus *Nemalion* (Nemaliales) (Table 1; Lee et al., 2016). The presence of single or partially inactivated rRNA operons in some florideophycean species suggests independent losses of one operon copy in these taxa (Lee et al., 2016).

Numerous gene features varied within the Batrachospermales and among the Batrachospermales and the five other Nemaliophycidae orders available for comparison. Of interest was the loss of the *pbs*A gene which encodes a heme oxygenase enzyme that is activated under iron limitation (Richaud & Zabulon, 1997). Costa et al. (2016) noted its absence from some members of the Nemaliales and Cho et al. (2018) reported its absence in *Thorea hispida* (Thoreales) and *Kumanoa americana* (Batrachospermales). The present study confirmed the loss of this gene in all members of *Kumanoa* sequenced to date (6 species, 7 genomes), the genus *Montagnia*, *Nothocladus ranulifer*, *N. theaquus* and a clade containing *Paludicola*, *Acarposporophycos*, *Notohesperus* and *Psilosiphon*. Cho et al. (2018) hypothesized that absence of this gene may be linked to freshwater habitats in which iron is not limited and therefore the gene may be redundant or unnecessary. With many more plastid genomes now available, it appears to be more difficult to link to freshwater habitats as its presence or absence is in about half of Batrachospermales sequenced to date as well as being present in the freshwater *Balbiania*

366 *investiens* (Balbianiales). A *trn*M with an intron is present in the batrachospermalean *Nocturama*
367 *novamundensis* and some species of *Kumanoa* and *Paludicola* and also in the representatives of
368 the Thoreales, Acrochaetiales and Palmariales. The functional role and the importance of its
369 presence or absence is not clear at this stage.
370
371 *Phylogeny*
372 The current study produced the most robust phylogeny of the Batrachospermales to date.
373 Represented on the tree were 18 of the 23 genera currently recognized in the order (Vis &
374 Necchi, 2021). In previous studies, genera were supported as monophyletic but the relationship
375 among the genera often had low support and could vary depending on genes and taxon sampling
376 (Entwisle et al., 2009; Vis & Necchi, 2021).
377 Although most branches across the phylogeny were highly supported, *Tuomeya* sister to
378 *Batrachospermum* received low support. The relationship of this genus to others has been
379 difficult to determine. Recently, Lemes da Silva et al. (2025) using mitochondrial genomes
380 showed it sister to the clade of *Sirodotia*, *Lympha* and *Volatus* but with no support. Likewise,
381 Crowell et al. (in review) found the same placement with higher support. The relationship of the
382 monospecific *Tuomeya* may continue to be difficult to resolve as there are no other species to
383 add to the data set.
384 The Australasian genus *Nothocladus* is currently divided into morphologically diverse
385 four sections (Entwisle et al., 2016; Vis & Necchi, 2021). The genus is often paraphyletic in
386 phylogenetic analyses using one to many genes, especially in relation to *Torularia* (Entwisle et
387 al., 2016; Vis & Necchi, 2021). In the current study, three sections (*Kraftii*, *Australicus* and
388 *Theaquus*) are in a clade with *Torularia* rendering the genus paraphyletic. In addition, the larger

389 clade containing *Nocturama* and *Sheathia* as well is not highly supported. Future research should
390 focus on sequence data from *Nothocladus* section *Nothocladus* as well as more species of the
391 other sections to clarify the relationships among the sections and if the sections should be
392 elevated to genera.

393 Within the Batrachospermales clade, there were two well supported lineages that
394 potentially might have taxonomic significance. These two lineages have been recovered in other
395 phylogenetic studies using one to multiple genes and organellar genomes (Entwisle et al., 2016;
396 Lam et al., 2016; Paino et al., 2016, 2017; Lemes da Silva et al., 2025). In the present study,
397 lineage 1 contains 11 genera and lineage 2 has seven genera. The morphological characters used
398 to describe taxa are shared between the two lineages. In summary, there is no single
399 morphological character or combination of characters that could be applied to define each of
400 these two lineages.

401 We have been able to make comparisons in these gene features among the
402 Batrachospermales that now have been well sampled with five other orders in the
403 Nemaliophycidae that have representative plastid genomes. However, there are seven other
404 orders in the subclass (Lam et al., 2016; Nan et al., 2022) that have no genome sequenced so far.
405 Therefore, more genomes will be needed in order to describe general gene features as well as
406 phylogenetic inferences in comparison to other orders in the Nemaliophycidae.

407

408 *ACKNOWLEDGEMENTS*

409 We would like to acknowledge colleagues who collected specimens and are listed in Table S1.
410 We would like to thank Ellie Becklund for assistance with CTAB extractions and library
411 preparations, the staff of the Life Sciences Core Facility (LaCTAD) from State University of

Campinas (UNICAMP), for the genomic analysis; Center for Scientific Computing (NCC/GridUNESP) of the São Paulo State University (UNESP) for allowing the use of high-performance computing resources. Funding sources that made this research possible was as follows: Ohio University research incentive funds to MLV, FAPESP for the grant-in-aid to the project and Technical Training Scholarship to NMLS (Proc 2023/00544-2), Postdoc Scholarship to MOP (Proc 2023/10342-8), CNPq for Senior Researcher Scholarship to ON (Proc 303077/2021-2).

REFERENCES

Andrews, S. (2016). FastQC: A quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Apt, K. E. & Grossman, A. R. 1(993). The phycobilisome b subunit gene of allophycocyanin is located on the plastid genome in *Aglaothamnion neglectum* (Rhodophyta) and co-transcribed with an unidentifiied open reading frame. *Journal of Phycology*, 29: 716–718. https://doi.org/10.1111/j.0022-3646.1993.00716.x

Beck, N. & Lang, B. F. (2010). MFannot, organelle genome annotation websever. https://github.com/BFL-lab/Mfannot

Borg, M., Krueger-Hadfield, S.A., Destombe, C., Collén, J., Lipinska, A. & Coelho, S.M. (2023). Red macroalgae in the genomic era. *New Phytologist*, 240: 471-488. https://doi.org/10.1111/nph.19211

Borowiec, M. L. (2016). AMAS: A fast tool for alignment manipulation and computing of summary statistics. *PeerJ*, *4*, e1660. https://doi.org/10.7717/peerj.1660

Chan, P. P. & Lowe, T. M. (2019). tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. *Methods in Molecular Biology 1962*, 1–14. DOI: 10.1007/978-1-4939-9173-0_1

Cho, C. H., Choi, J. W., Lam, D. W., Kim, K. M., & Yoon, H. S. (2018). Plastid genome analysis of three Nemaliophycidae red algal species suggests environmental adaptation for iron limited habitats. *PLoS ONE*, *13*(5), e0196995. https://doi.org/10.1371/journal.pone.0196995

Costa, J. F., Lin, S.-M., Macaya, E. C., Fernández-García, C., & Verbruggen, H. (2016). Chloroplast genomes as a tool to resolve red algal phylogenies: A case study in the Nemaliales. *BMC Evolutionary Biology*, *16*(1), 205. https://doi.org/10.1186/s12862-016-0772-3

Crowell, R. C., Karol, K. G., & Vis, M. L. Phylogenetic relationships among seven freshwater red algal genera in the Batrachospermaceae (Batrachospermales, Rhodophyta) using complete chloroplast genome data. *Journal of Phycology* (in reveiw).

Darling, A. E., Mau, B., & Perna, N. T. (2010). progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLOS ONE*, *5*(6), e11147. https://doi.org/10.1371/journal.pone.0011147

Díaz-Tapia, P., Pasella, M. M., Verbruggen, H., & Maggs, C. A. (2019). Morphological evolution and classification of the red algal order Ceramiales inferred using plastid phylogenomics. *Molecular Phylogenetics and Evolution*, *137*, 76–85. https://doi.org/10.1016/j.ympev.2019.04.022

455 Dierckxsens, N., Mardulyn, P. & Smits, G. (2017). NOVOPlasty: De novo assembly of organelle
456 genomes from whole genome data. *Nucleic Acids Research*, 45(4), e18.
457 https://doi.org/10.1093/nar/gkw955
458 Edgar, R. C. (2022). Muscle5: High-accuracy alignment ensembles enable unbiased assessments
459 of sequence homology and phylogeny. *Nature Communications*, *13*(1), 6968.
460 https://doi.org/10.1038/s41467-022-34630-w
461 Entwisle, T. J., Vis, M. L., Chiasson, W. B., Necchi, O., Jr & Sherwood, A. R. (2009).
462 Systematics of the Batrachospermales (Rhodophyta) – A synthesis. *Journal of Phycology*,
463 45, 704–715. https://doi.org/10.1111/j.1529-8817.2009.00686.x
464 Evans, J. R., St. Amour, N., Verbruggen, H., Salomaki, E. D., & Vis, M. L. (2019). Chloroplast
465 and mitochondrial genomes of *Balbiania investiens* (Balbianiales, Nemaliophycidae).
466 *Phycologia*, *58*(3), 310–318. https://doi.org/10.1080/00318884.2019.1573349
467 Evans, J. R., & Vis, M. L. (2020). Relative expression analysis of light-harvesting genes in the
468 freshwater alga *Lympha mucosa* (Batrachospermales, Rhodophyta). *Journal of*
469 *Phycology*, *56*(2), 540–548. https://doi.org/10.1111/jpy.12967
470 Fang, K., Nan, F., Feng, J., LV, J.-P., Liu, Q., Liu, X., & Xie, S. (2020). *Batrachospermum*
471 *qujingense* (Batrachospermales, Rhodophyta), a new freshwater red algal species from
472 Southwest China. *Phytotaxa*, *461*, 1–11. https://doi.org/10.11646/phytotaxa.461.1.1
473 Glazer, A. N., Chan, C. F. & West, J. A. 1997. An unusual phycocyanobilin-containing
474 phycoerythrin of several bluish colored acrochaetioid, freshwater red algal species.
475 *Journal of Phycology*, 3:617–624. https://doi.org/10.1111/j.0022-3646.1997.00617.x

476 Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2:
477 Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*,
478 *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281
479 Iha, C., Grassa, C. J., Lyra, G. M., Davis, C. C., Verbruggen, H., Oliveira, M. C., & Müller, K.
480 (2018). Organellar genomics: A useful tool to study evolutionary relationships and
481 molecular evolution in Gracilariaceae (Rhodophyta). *Journal of Phycology*, 54, 775–787.
482 https://doi.org/10.1111/jpy.12765
483 Janouškovec, J., Liu, S.-L., Martone, P. T., Carré, W., Leblanc, C., Collén, J., & Keeling, P. J.
484 (2013). Evolution of red algal plastid genomes: ancient architectures, introns, horizontal
485 gene transfer, and taxonomic utility of plastid markers. *PLOS ONE*, *8*(3), e59001.
486 https://doi.org/10.1371/journal.pone.0059001
487 Jayalakshmi, P. S., & John, J. (2023). *Macrosporophycos sahyadricus* (Batrachospermales,
488 Rhodophyta), a new genus and species of freshwater red algae from the Western Ghats of
489 India. *Phycologia*, *62*(2), 109–116. https://doi.org/10.1080/00318884.2023.2177062
490 Jin, J.-J., Yu, W. B., Yang, J. B., Song, Y., Pamphilis, C. W., Yi, T. S. & Li, D. Z. (2020).
491 GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle
492 genomes. Genome Biology, 21, 241. https://doi.org/10.1186/s13059-020-02154-5
493 Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017).
494 ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*,
495 *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285
496 Kumano, S. (2002). Freshwater red algae of the world. Biopress.

Lam, D. W., Verbruggen, H., Saunders, G. W., & Vis, M. L. (2016). Multigene phylogeny of the red algal subclass Nemaliophycidae. *Molecular Phylogenetics and Evolution*, *94*(Pt B), 730–736. https://doi.org/10.1016/j.ympev.2015.10.015

Lee, J. M., Boo, S. M., Mansilla, A. & Yoon, H. S. (2015). Unique repeat and plasmid sequences in the mitochondrial genome of *Gracilaria chilensis* (Gracilariales, Rhodophyta). Phycologia, 54(54), 20–23. https://doi.org/10.2216/PH14-97.1

Lee, J. M., Cho, C. H., Park, S. I., Choi, J. W., song, H. S., West, J. A., Bhattacharya, D., & Yoon, H. S. (2016). Parallel evolution of highly conserved plastid genome architecture in red seaweeds and seed plants. BMC Biology, 14, 75. https://doi.org/10.1186/s12915-016-0299-5

Lee, J. M., Song, H. J., Park, S. I., Lee, Y. M., Jeong, S. Y., Cho, T. O., Yoon, H. S., Choi, H.-G., Choi, C. G., Nelson, W. A., Fredericq, S., Bhattacharya, D. & Yoon, H. S. (2018). Mitochondrial and plastid genomes from coralline red algae provide insights into the incongruent evolutionary histories of organelles. Genome Biology and Evolution, 10, 2961–2972. https://doi.org/10.1093/gbe/evy222

Lemes da Silva, N. M., Crowell, R. M., Vis, M. L., Chiasson, W. B., Paiano, M. O., Lopes, Y. M. S., & Necchi, O. Jr. (2025). Investigation of mitochondrial genomes of members of the order Batrachospermales (Rhodophyta). *Phycologia*, *64*(2), 118–126. https://doi.org/10.1080/00318884.2025.2473285

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, *37*(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

Muñoz-Gómez, S., Mejía-Franco, F. G., Durnin, K., Colps, M., Grisdale, C. J., Archibald, J. M. & Slamovits, C. H. (2017). The new red algal subphylum Proteorhodophytina comprises the largest and most divergent plastid genomes known. *Current Biology*, 27, 1677–1684. https://doi.org/10.1016/j.cub.2017.04.054

Nan F., Feng, J., Liu, Q., Fang, K., Gong, C., & Xie, S. (2017). Origin and evolutionary history of freshwater Rhodophyta: further insights based on phylogenomic evidence. *Scientific Reports*, *7,* 2394. https://doi.org/10.1038/s41598-017-03235-5

Necchi Jr, O., Entwisle, T. J., Lopes, Y. M. S., & Vis, M. L. (2024). Reappraisal of the genus *Visia* (Batrachospermales, Rhodophyta) and proposal for a new genus and species. *Phycologia*, *63*(3), 271–277. https://doi.org/10.1080/00318884.2024.2330108

Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: A new versatile metagenomic assembler. *Genome Research*, 27, 824–834. https://doi.org/10.1101/gr.213959.11

Paiano, M. O., Del Cortona, A., Costa, J. F., Liu, S.-L., Verbruggen, H., De Clerck, O., & Necchi Jr, O. (2017). Organization of plastid genomes in the freshwater red algal order Batrachospermales (Rhodophyta). *Journal of Phycology*, *54*(1), 25–33. https://doi.org/10.1111/jpy.12602

Paiano, M. O., Del Cortona, A., Costa, J. F., Liu, S.-L., Verbrugen, H., De Clerck, O. & Necchi, O., Jr. (2018). Complete mitochondrial genomes of six species of the freshwater red algal order Batrachospermales (Rhodophyta). *Mitochondrial DNA*, Part B, 2, 607–610. https://doi.org/10.1080/23802359.2018.1473734

Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A., & Korobeynikov, A. (2020). Using SPAdes de novo assembler. *Current Protocols in Bioinformatics*, 70, Article e102. https://doi.org/10.1002/cpbi.102

Pueschel, C. M. & Cole, K. M. (1982). Rhodophycean pit plugs: an ultrastructural survey with taxonomic implications. *American Journal of Botany*, 69, 703-720. https://doi.org/10.1002/j.1537-2197.1982.tb13310.x

Richaud, C. & Zabulon, G. 1997. The heme oxygenase gene (*pbs*A) in the red alga *Rhodella violacea* is discontinuous and transcriptionally activated during iron limitation. *Proceedings of the National Academy of Science USA*, 94, 11736-41. https://www.pnas.org/doi/10.1073/pnas.94.21.11736

San, M. H., Kawamura, Y., Kimura, K., Witharana, E. P., Shimogiri, T. Fujiyoshi, E. & Nagano, Y. 2025. Gene tree conflict in chloroplast-based phylogenies in *Pyropia* species. *Phycological Research* https://doi.org/10.1111/pre.12576

Sayers, E. W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K. D. & Karsch-Mizrachi, I. (2020). GenBank. Nucleic Acids Research, 48, D84–D86. https://doi.org/10.1093/nar/gkz956

Schenk, J. J., Becklund, L. E., Carey, S. J., & Fabre, P. P. (2023). What is the "modified" CTAB protocol? Characterizing modifications to the CTAB DNA extraction protocol. *Applications in Plant Sciences*, *11*(3), e11517. https://doi.org/10.1002/aps3.11517

Sheath, R. G. (1984). The Biology of Freshwater Red Algae. In Round, F. E. & Chapman, D. J. (Eds.), *Progress in Phycological Research* (Vol. 3, pp. 89–156). Biopress Ltd.

Sheath, R. G., Müller, K. M., Vis, M. L. and Entwisle, T. J. (1996). A re-examination of the morphology, ultrastructure and classification of genera in the Lemaneaceae

(Batrachospermales, Rhodophyta). *Phycological Research*, 44, 233-246. https://doi.org/10.1111/j.1440-1835.1996.tb00053.x

Thiers, B. M. (updated continuously). *Index Herbariorum.* https://sweetgum.nybg.org/science/ih/

van Beveren, F., Eme, L., López-García, P., Ciobanu, M., Moreira, D. & Sloan, D. (2022). Independent size expansions and intron proliferation in red algal plastid and mitochondrial genomes. *Genome Biology and Evolution*, 14, Article evac037. https://doi.org/10.1093/gbe/evac037

Verbruggen, H., Maggs, C. A., Saunders, G. W., Le Gall, L., Yoon, H. S., & De Clerck, O. (2010). Data mining approach identifies research priorities and data requirements for resolving the red algal tree of life. *BMC Evolutionary Biology*, *10*(1), 16. https://doi.org/10.1186/1471-2148-10-16

Vis, M. L., Tiwari, S., Evans, J. R., Stancheva, R., Sheath, R. G., Kennedy, B., Lee, J. & Eloranta, P. (2020). Revealing hidden diversity in the Sheathia arcuata morphospecies (Batrachospermales, Rhodophyta) including four new species. *Algae*, *35*(3), 213–224. https://doi.org/10.4490/algae.2020.35.8.31

Vis, M. L., & Necchi Jr, O. (2021). *Freshwater Red Algae: Phylogeny, Taxonomy and Biogeography*. Springer International Publishing. https://doi.org/10.1007/978-3-030-83970-3

Wick, R. R., Schultz, M. B., Zobel, J., & Holt, K. E. (2015). Bandage: Interactive visualization of de novo genome assemblies. *Bioinformatics*, 31, 3350–3352.

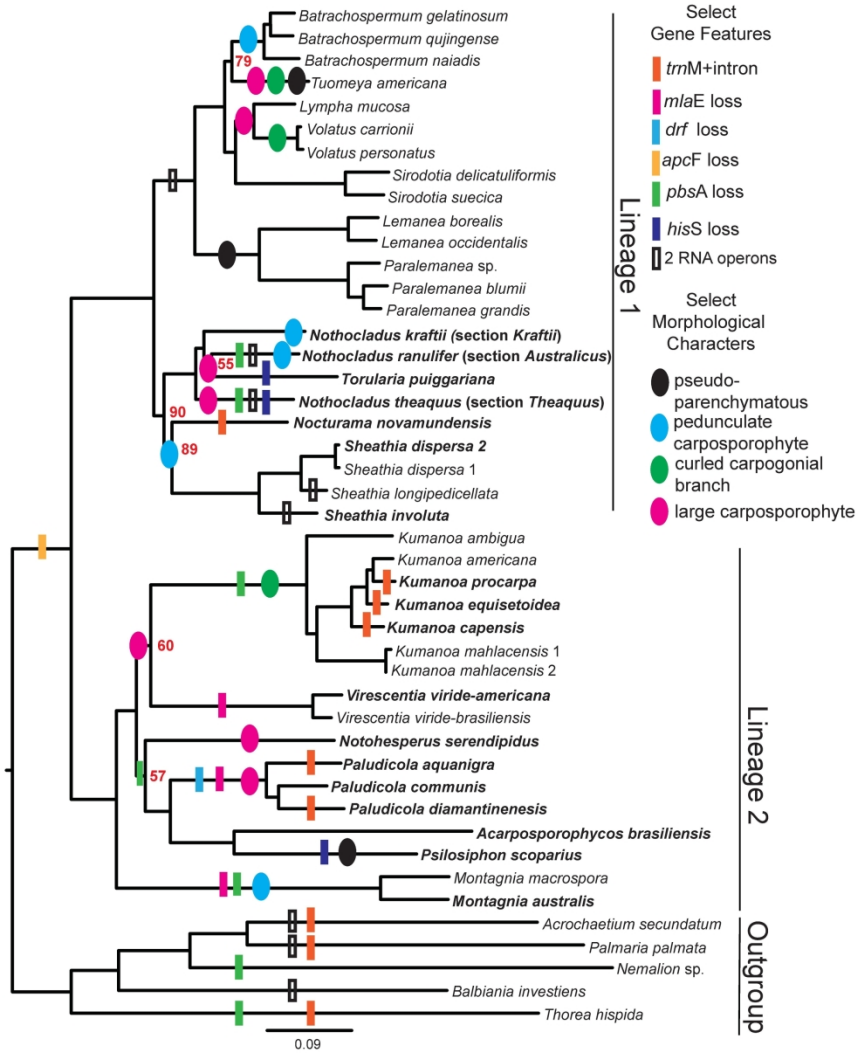
Yang, E. C., Boo, S. M., Bhattacharya, D., Saunders, G. W., Knoll, A. H., Fredericq, S., Graf, L. & Yoon, H. S. (2016). Divergence time estimates and the evolution of major lineages in

585 the florideophyte red algae. *Scientific Reports*, 6, Article 21361.

586 https://doi.org/10.1038/srep21361

Figure Legend

Figure 1. Maximum likelihood phylogeny based on the concatenation of 126 plastid genes depicting relationships among genera in the Batrachospermales and five outgroup taxa from the Nemaliophycidae. Newly sequenced taxa in bold. Support values are ultrafast bootstraps (bs); most branches had 100 bs (not shown) and only those <100 bs are provided adjacent to the branch (in red). Most gene features that differed among taxa were added to the phylogeny; not shown were genes only present in the outgroup taxa (*grx*, ycf34, ycf35, ycf45) and the *chl*B gene because it was present in all taxa except *Balbiania investiens*. Select morphological characters were added to the phylogeny to illustrate characters present in both lineage 1 and 2.



215x279mm (300 x 300 DPI)

Table 1. Character states of genera of Batrachospermales within the two lineages shown in the phylogenetic analyses. Characteristics not present in a genus are represented by "N/A" (not applicable) and if unknown represented by "-".

Genus	Thallus construction	Carpogonial branch		Carpogonium		Carposporophyte			
		length[1]	orientation	relative length[2]	symmetry	arrangement	origin	position	size[3]
Lineage 1									
Batrachospermum	Bead-like, no outer cortex	Long	Straight	Short	Symmetric	Compact	External	Pedunculate	Small
Lemanea	Pseudoparenchymatous, with outer cortex	Short	Straight	Short	Putatively symmetric	Compact	Inside thallus	Inside thallus	Small
Lympha	Bead-like, no outer cortex	Short	Straight	Short	Symmetric	Compact	External	Axial	Large
Nocturama	Bead-like, no outer cortex	Predominantly long	Straight	Short	Symmetric	Compact	External	Pedunculate	Small
Nothocladus section *Australasicus*	Bead-like, no outer cortex	Short or long	Straight	Short or long	Symmetric	Compact or diffuse	External	Pedunculate	Small or Large
Nothocladus section *Kraftii*	Bead-like, no outer cortex	Long	Straight	Short	Symmetric	Compact	External	Pedunculate	Small
Nothocladus section *Theaquus*	Bead-like, no outer cortex	Short	Straight	Short or long	Symmetric	Compact	External	Axial	Large
Paralemanea	Pseudoparenchymatous, with outer cortex	Short	Straight	Short	Putatively symmetric	Compact	Inside thallus	Inside thallus	Small

Sheathia	Bead-like, no outer cortex	Long	Straight	Short	Symmetric	Compact	External	Pedunculate	Small
Sirodotia	Bead-like, no outer cortex	Short	Straight	Short or long	Asymmetric	Diffuse	External	N/A	N/A
Torularia	Bead-like, no outer cortex	Short	Straight	Short	Symmetric	Compact	External	Axial	Large
Tuomeya	Pseudoparenchymatous, with outer cortex	Short	Contorted or curved	Short	Asymmetric	Compact	Inside thallus	Axial	Large
Volatus	Bead-like, no outer cortex	Short	Contorted or curved	Short or long	Asymmetric	Compact	External	Axial	Large
Lineage 2									
Acarposporophycos	Bead-like, no outer cortex	Short	Straight	Short	Symmetric	N/A	N/A	N/A	N/A
Kumanoa	Bead-like, no outer cortex	Short	Contorted or curved	Short or long	Asymmetric	Compact	External	Axial	Large
Montagnia	Bead-like, no outer cortex	Short	Straight	Short	Symmetric	Compact	External	Pedunculate	Small
Notohesperus	Bead-like, no outer cortex	Short	Straight	Short	Symmetric	Compact	External	Axial	Large
Paludicola	Bead-like, no outer cortex	Short	Straight	Long	Symmetric	Compact	External	Axial	Large
Psilosiphon	Pseudoparenchymatous, with outer cortex	-	-	-	-	-	-	-	-
Virescentia	Bead-like, no outer cortex	Short	Straight	Long	Symmetric	Compact	External	Axial	Large

[1]Carpogonial branch: short < one-third of the whorl diameter; long ≥ one-third of the whorl diameter; short carpogonial branches are typically < 40 μm in length, whereas long carpogonial branches are typically ≥ 40 μm in length.
[2]Carpogonium relative length: short < 1.5 times shorter than the carpogonial branch; long ≥ two times longer than the carpogonial branch. Short carpogonia are typically ≤ 35 μm in length, whereas long carpogonia are typically > 35 μm in length
[3]Carposporophyte size: small < half of the whorl diameter; large > half of the whorl diameter.

Table 2. Overview of plastid genome characteristics generated in this study (in bold) and previously published for members of the Batrachospermales and other Nemaliophycidae orders.

Species	Genome Size (nt)	% GC content	Protein coding genes	YCF[1]	ORF[2]	tRNA	RNA Operon	Intron	Coverage (mean)	GenBank Accession number	Reference
Batrachospermales											
Acarposporophycos brasiliensis	183,356	30.6	196	18	5	31	1	1	50	TBA	This study
Batrachospermum gelatinosum	188,711	29.5	201	20	6	32	2	1	855	PX060844	Crowell et al., 2025
B. naiadis	188,526	29.9	201	20	6	32	2	1	2390	PX060845	Crowell et al., 2025
B. qujingense[3,5]	184,902	29	201	20	6	29	2[8]	1	–	MN905507	Unpublished
Kumanoa ambigua[3]	182,932	28.2	201	21	6	31	1	1	345	MG252485	Paiano et al., 2018
K. americana[3]	184,025	29.3	201	21	6	31	1	1	215	KX284725	Cho et al., 2018
K. capensis	184,267	29.2	201	20	7	32	1	2	103	TBA	This study
K. equisetoidea	184,008	29.3	202	21	7	31	1	2	352	TBA	This study
K. mahlacensis[3]	181,368	29.8	199	21	6	31	1	1	176	MG252486	Paiano et al., 2018
K. mahlacensis	183,893	29.6	199	19	7	31	1	1	–	MK641509	Unpublished
K. procarpa	184,368	29.4	199	19	6	31	1	2	448	TBA	This study
Lemanea borealis	188,030	30.8	200	19	6	32	2	1	614	PX060846	Crowell et al., 2025

L. occidentalis	188,242	30.8	200	19	6	32	2	1	2275	PX060847	Crowell et al., 2025
Lympha mucosa[3,4]	≥189,825	29.5	198	19	5	32	2	2	146	MN509464	Evans & Vis, 2020
Montagnia australis	179,420	28.4	193	17	5	30	1	1	65	TBA	This study
M. macrospora	176,626	28.1	192	18	6	33	1	1	478	MG252483	Paiano et al., 2018
Nocturama novamundensis	185,426	28.2	200	21	5	31	1	2	40	TBA	This study
Nothocladus kraftii	184,530	28.8	200	19	6	30	1	1	999	TBA	This study
N. ranulifer	183,398	29.3	199	20	5	30	1	1	3875	TBA	This study
N. theaquus	187,490	29.1	199	20	6	31	2	1	12810	TBA	This study
Notohesperus serendipidus	183,582	28.3	193	19	4	30	1	1	19480	TBA	This study
Paludicola aquanigra	179,217	28.5	194	18	4	31	1	2	310	TBA	This study
P. communis	183,091	28.6	197	20	5	30	1	1	966	TBA	This study
P. diamantinensis	181,653	28.7	194	18	5	31	1	2	137	TBA	This study
Paralemanea sp.[3]	187,837	31	201	19	6	28	2	1	364	MG252487	Paiano et al., 2018
P. blumii	187,783	30.8	201	20	6	28	2	1	1320	PX060848	Crowell et al., 2025
P. grandis[4]	≥183,183	30.3	201	20	6	28	2	1	210	PX060849	Crowell et al., 2025
Psilosiphon scoparius	182,178	30.9	197	20	5	31	1	1	70	TBA	This study
Sheathia dispersa[7]	179,671	29.1	196	18	5	31	1	1	1708	MG252488	Paiano et al., 2018
S. dispersa	183,683	29.2	198	18	5	30	1	1	361	TBA	This study
S. involuta	188,048	29.7	198	18	5	32	2	1	874	TBA	This study

Sheathia longipedicellata[3,6]	187,354	29.9	200	19	5	31	2	1	–	KY033529	Nan et al., 2017
Sirodotia delicatulaformis[3]	185,489	29.2	200	20	5	32	2^8	1	1984	MG252489	Paiano et al., 2018
S. suecica	186,915	29.6	200	19	6	32	2	1	6536	PX060850	Crowell et al., 2025
Tolularia puiggariana	186,712	29.1	199	20	5	32	2	1	284	TBA	This study
Tuomeya americana	189,523	29.3	200	20	5	32	2	1	2600	PX060851	Crowell et al., 2025
Virescentia viride-americana	184,608	28.6	199	20	4	30	1	1	970	TBA	This study
Virescentia viride-brasiliensis[3]	171,682	28.5	189	17	3	34	1	1	961	MG252484	Paiano et al., 2018
Volatus carrionii	188,856	29.1	200	19	6	32	2	1	502	PX060852	Crowell et al., 2025
Volatus personatus[4]	≥183,580	28.6	200	19	6	29	2^8	1	1858	PX060853	Crowell et al., 2025
Acrochatiales											
Acrochaetium secundatum	192,311	32.0	206	24	6	33	2	2	–	MH026107	Evans et al., 2019
Balbianiales											
Balbiania investiens	186,778	34	199	21	5	33	2	0	562	MH026108	Evans et al., 2019
Nemaliales											
Galaxaura rugosa	181,215	29.6	204	24	5	31	1	1	327	LT622865	Costa et al., 2016
Liagora harveyana	182,933	33.9	204	23	5	31	1	1	3239	LT622869	Costa et al., 2016
Nemalion sp.	182,930	35.5	202	24	4	31	1	1	3297	LT622871	Costa et al., 2016
Scinaia undulata	183,795	35.9	206	24	4	31	1	1	1479	LT622873	Costa et al., 2016
Palmariales											

Palmaria palmata	192,960	33.9	204	23	5	33	2	2	343	KX284726	Cho et al., 2018
Thoreales											
Thorea hispida	175,193	28.3	193	19	3	30	1	2	1003	KX284714	Cho et al., 2018

[1]YCF = Hypothetical chloroplast open reading frame
[2]ORF = Open reading frame
[3]Statistics reflect updated annotations made in this study
[4]Incomplete chloroplast genome
[5]Identification updated from *Batrachospermum* sp. to *B. qujingense* (Fang et al., 2020)
[6]Identification updated from *Sheathia arcuata* to *S. longipedicellata* (Vis et al., 2020)
[7]Identification updated from *Sheathia arcuata* to *S. dispersa* (Vis et al., 2020)
[8] Assumed to be two operons due to at least one RNA gene duplicate.

Table S1. Herbarium voucher, collection information, and GenBank accession numbers for chloroplast genomes generated in the present study.

Species	Herbarium voucher[1]	Locality	Latitude, Longitude	Collector(s)	Collection date	GenBank accession number
Acarposporophycos brasiliensis	SJRP-Algae 31476	Pousada Alto da Serra, Negro River, Paraibuna, São Paulo State, Brazil	23.564761 S, 45.460100 W	O. Necchi Jr.	23.ix.2006	XXXXXX
Kumanoa capensis	SJRP-Algae 32607	'Chapada dos Veadeiros' National Park, Alto Paraíso de Goiás, Goiás State, Brazil (PNCV 10)	14.051833 S, 47.512722 W	L.H.Z. Branco	iv.2010	
Kumanoa equisetoidea	SJRP-Algae 32608	Igarapé Acará, 'Adolfo Ducke' Forest Reserve, Manaus, Amazonas State, Brazil (AM 8)	2.570600 S, 59.570100 W	O. Necchi Jr., M.L. Vis., J.C. Hodge	19.viii.2004	
Kumanoa procarpa	SJRP-Algae 32600	Upstream of 'Lagoa das Ninféias', São Paulo Botanical Garden, São Paulo, São Paulo State, Brazil	23.639361 S, 46.619647 W	O. Necchi Jr., J. Prado	25.iv.2023	
Montagnia australis	BHO A-1565	Small stream crossing Lobelia Rd. North Carolina, USA	35.189694 N, 79.135833 W	M.L. Vis, W.B. Chiasson	6.xii.2003	
Nocturama novamudensis	SJRP-Algae 32610	Iguaçu National Park, Céu Azul, Paraná State, Brazil	---	O. Necchi Jr., C.K. Peres	24.v.2022	

Nothocladus kraftii	NSW799510	Acheron River, Victoria, Australia	37.750000 S, 145.700000 E	T.J. Entwisle (coll. 3391)	05.i.2010
Nothocladus ranulifer	NSW706505	Westerway Creek, c. 22 km NNW of Queenstown, Tasmania, Australia	41.928861 S, 145.432000 E	T.J. Entwisle (coll. 3345)	27.x.2005
Nothocladus theaquus	NSW742917	Melaleuca Creek, Southwest National Park, Tasmania, Australia	43.421111 S, 146.150556 E	T.J. Entwisle (coll. 3351)	25.ii.2007
Notohesperus serendipidus	NSW747863	Trib. of Red Tape Ck, Southwest National Park, Tasmania, Australia	43.018033 S, 146.371467 E	T.J. Entwisle (coll. 3364)	05.vii.2007
Paludicola aquanigra	SJRP-Algae 32596	Trail to the Upper part of 'Cachoeira da Fumaça', 'Chapada Diamantina' National Park, Palmeiras, Bahia State, Brazil	12.603889 S, 41.466389 W	O. Necchi Jr., C.C. Necchi	02.vii.2018
Paludicola communis	BHO A-0368	Sterling Lake, New York, USA	41.214694 N, 74.254389 W	J.D. Hall	13.xi.2010
Paludicola diamantinensis	SJRP-Algae 32597	Riachinho Stream, Vale do Capão, Palmeiras, Bahia State, Brazil	12.572222 S, 41.514722 W	O. Necchi Jr., C.C. Necchi	02.vii.2018
Psilosiphon scoparius	MEL2036082A	Waikotahu Stream, North Island, New Zealand	35.626667 S, 173.560000 E	N.G. Walsh (coll. 4686)	11.i.1997

Sheathia dispersa	SJRP-Algae 32578	Ward No. 5, Khayer Khola, Province No. 1, Koshi Zone, Morang District, Sundarharaicha Municipality, Nepal	26.690222 N, 87.321694 E	O. Necchi Jr., J.A. West, E.K. Ganesan, S.K. Rai	21.ii.2018
Sheathia involuta	BHO A-1966	Unnamed outflow of Myosotis Lake upstream of Rensselaerville Falls, Huyck Preserve, New York, USA	42.515083 N, 74.144083 W	M.L. Vis, R.M. Crowell, G.A. Lindsey, M. French	17.v.2023
Torularia puiggariana	SJRP 31909	State Park, near to the visitor center, Campos do Jordão, São Paulo State, Brazil	22.690278 S, 45.479722 W	O. Necchi Jr.	22.v.2008
Virescentia viride-americana	BHO A-1103	Whiskey Creek near Marengo, Indiana, USA	38.371056 N, 86.330944 W	M.L. Vis, W.B. Chiasson	12.v.2013

[1]Herbarium abbreviations as follows: BHO = Bartley Herbarium, Ohio University, MEL = National Herbarium of Victoria, NSW = National Herbarium of New South Wales, SJRP = São Paulo State University, Campus São José Rio Preto.

	Acrocarposporophycos brasiliensis	Batrachospermum gelantinosum
accA		
accB		
accD		
acpP		
acsF		
apcA		
apcB		
apcD		
apcE		
apcF		
argB		
atpA		
atpB		
atpD		
atpE		
atpF		
atpG		
atpH		
atpI		
bas1		
carA		
cbbX/cfxQ		
cbiQ/ycf92		
ccs1		
ccsA		
cemA		
chlB		
chlI		
chlL		
chlN		
clpC		
cpcA		
cpcB/rpcB		
cpcG		
cpcS/orf156		
cpeA		
cpeB		
dfr/ycf26		
dnaB		
dnaK		
dsbD/ccdA		
fabH		
ftrB		
ftsH		
gltB		
groEL		
grx/orf26		
hisS/syh		
ilvB		

ilvH		
infB		
infC		
mlaE/ycf63		
moeB		
nblA		
ntcA		
ompR/ycf27		
orf52		
orf55		
orf59		
orf61		
orf63		
orf88		
orf109		
orf136		
orf149		
orf166		
orf167		
orf169		
orf170		
orf172		
orf176		
orf177		
orf266		
orf273		
orf287		
orf307		
orf334		
orf350		
orf361		
orf366		
orf371		
orf391		
orf392		
orf398		
orf399		
orf407		
orf418		
orf466		
orf664		
orf735		
orf748		
orf753		
orf761		
orf768		
orf772		
orf773		
pafI/ycf3		
pafII/ycf4		

pbsA		
pdhA/odpA		
pdhB/odpB		
petA		
petB		
petD		
petF		
petG		
petJ		
petL		
petM		
petN		
petP/orf9		
pgmA		
preA		
psaA		
psaB		
psaC		
psaD		
psaE		
psaF		
psaI		
psaJ		
psaK		
psaL		
psaM		
psb30		
psbA		
psbB		
psbC		
psbD		
psbE		
psbF		
psbH		
psbI		
psbJ		
psbK		
psbL		
psbN/pbfI		
psbT		
psbV		
psbW		
psbX		
psbY		
psbZ		
rbcL		
rbcR		
rbcS		
rne		
rnz/ycf56		

rpl1		
rpl2		
rpl3		
rpl4		
rpl5		
rpl6		
rpl9		
rpl11		
rpl12		
rpl13		
rpl14		
rpl16		
rpl18		
rpl19		
rpl20		
rpl21		
rpl22		
rpl23		
rpl24		
rpl27		
rpl28		
rpl29		
rpl31		
rpl32		
rpl33		
rpl34		
rpl35		
rpl36		
rpoA		
rpoB		
rpoC1		
rpoC2		
rpoZ/ycf61		
rps1		
rps2		
rps3		
rps4		
rps5		
rps6		
rps7		
rps8		
rps9		
rps10		
rps11		
rps12		
rps13		
rps14		
rps16		
rps17		
rps18		

rps19		
rps20		
secA		
secG		
secY		
sufB/ycf24		
sufC/ycf16		
syfB		
tatC/ycf43		
thiG		
thiS		
tilS		
trpA		
trpG		
trxA		
tsf		
tufA		
upp		
ycf19		
ycf20		
ycf21/orf90/orf113		
ycf22		
ycf23		
ycf29		
ycf33		
ycf34		
ycf35		
ycf36		
ycf37/orf108/orf145		
ycf38		
ycf39		
ycf41		
ycf45		
ycf46		
ycf52		
ycf53		
ycf54		
ycf55		
ycf60		
ycf65		
ycf80		
ycf91		

Batrachospermum naiadis	Batrachospermum qujingense	Kumanoa ambigua	Kumanoa americana
		***	
		***	
		***	
		***	***

		***	
		***	
		***	

		***	
		***	***
		***	
		***	

		***	
		***	
		***	
		***	

		***	
		***	
		***	
		***	

Kumanoa capensis	Kumanoa equisetoidea	Kumanoa mahlacensis (MK641509)
***		

		***

***		***

Kumanoa mahlacensis	Kumanoa procarpa	Lemanea borealis	Lemanea occidentalis
	***		
***			

***			
	***		

Lympha mucosa	Montagnia australis	Montagnia macrospora	Nocturama novamundensis

**			
**			
		***	
		***	
**			

**			

**			

Nothocladus kraftii	Nothocladus ranulifer	Nothocladus theaquus

Notohesperus serendipidus	Paludicola aquanigra	Paludicola communis

	*	

Paludicola diamantinensis	Paralemanea blumii	Paralemanea grandis	Paralemanea sp.
*			

*			

Psilosiphon scoparius	Sheathia dispersa	Sheathia dispersa (MG252488)
		***

		***

		***
		***

Sheathia involuta	Sheathia longipedicellata	Sirodotia delicatulaformis	Sirodotia suecica
		***	
		***	
		***	
		***	

		***	
		***	

		***	
		***	
		***	
		***	
		***	
		***	

		***	
		***	
		***	

Torularia puiggariana	Tuomeya americana	Volatus carionii	Volatus personatus

Virescentia viride-americana	Virescentia viride-brasiliense	Thorea hispida (KX284714)
	***	
	***	
***		
	*	
	***	
	***	
	***	
	***	
	***	

	***	
***	***	
	*	
	*	
	*	

	*	
	***	
	*	
	*	

	***	
	***	
	***	
	***	
	***	
	***	
	***	
	***	
	***	
	***	

	*	
	***	
	***	
	***	
	***	
	***	
	*	
	*	
	***	
	***	
	***	
	*	

Acrochaetium secundatum (MH026107)	Balbiana investens (MH026108)

Palmaria palmata (KX284726)	Nemalion sp (LT622871)	Galaxaura rugosa (LT622865)

Scinaia undulata (LT622873)	Liagora harveyana (LT622869)	
		*
		**
		***

Key
likey present in the genome, but no ORF present due to poor sequence
likely present in the genome, but there is missing sequence data (gap)
partial gene due to poor sequence data

Supplemental Figures

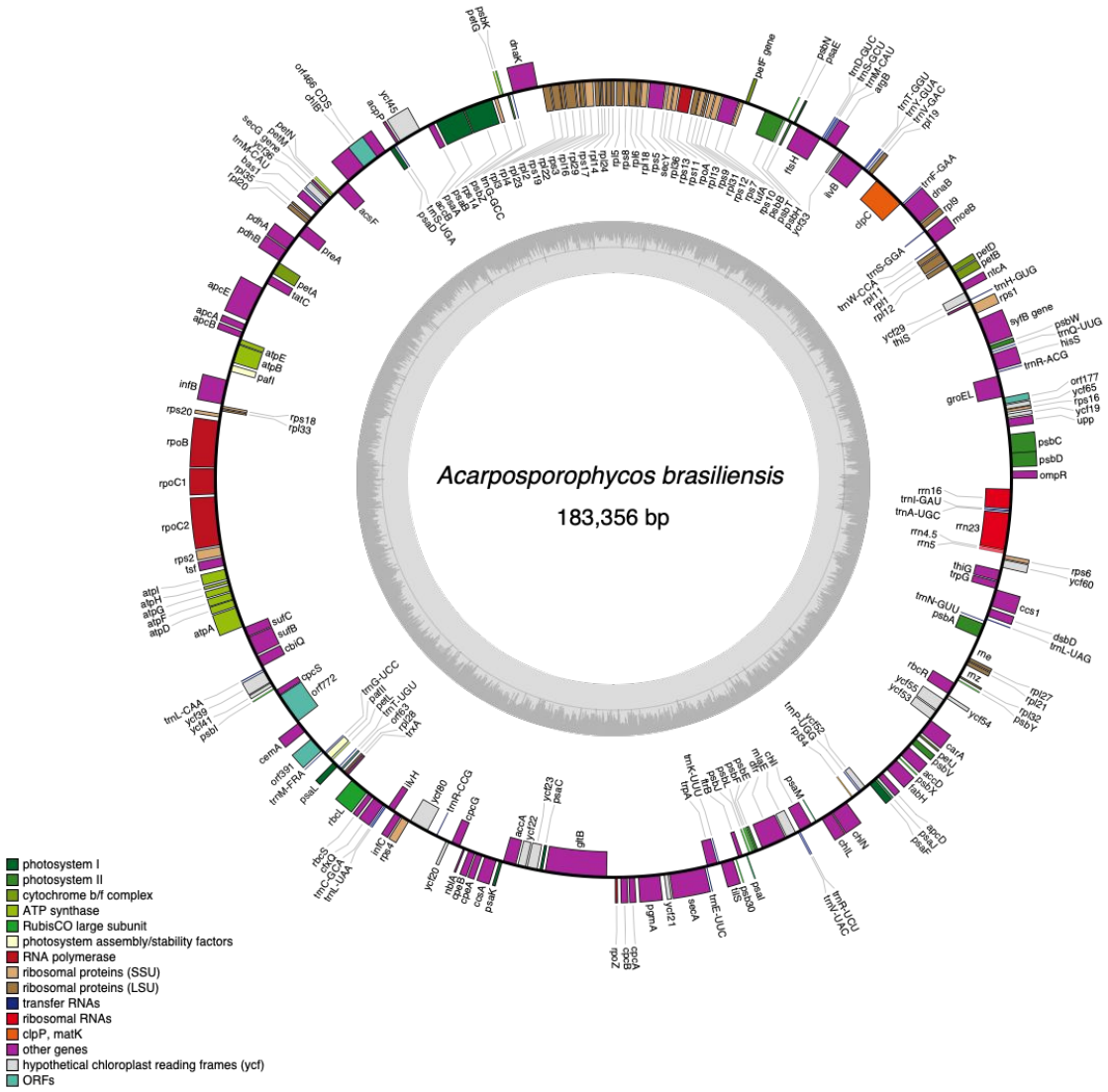


Figure S1. Gene map of *Acarposporophycos brasiliensis* chloroplast genome. Genes labeled on the outside of the outer circle are transcribed clockwise, genes on the inside counterclockwise. A gene with a star (*) indicates an intron within the gene. The inner circle represents the nucleotide content (G/C = dark gray, A/T = light gray). Gene boxes are color-coded by functional groups shown in the key.

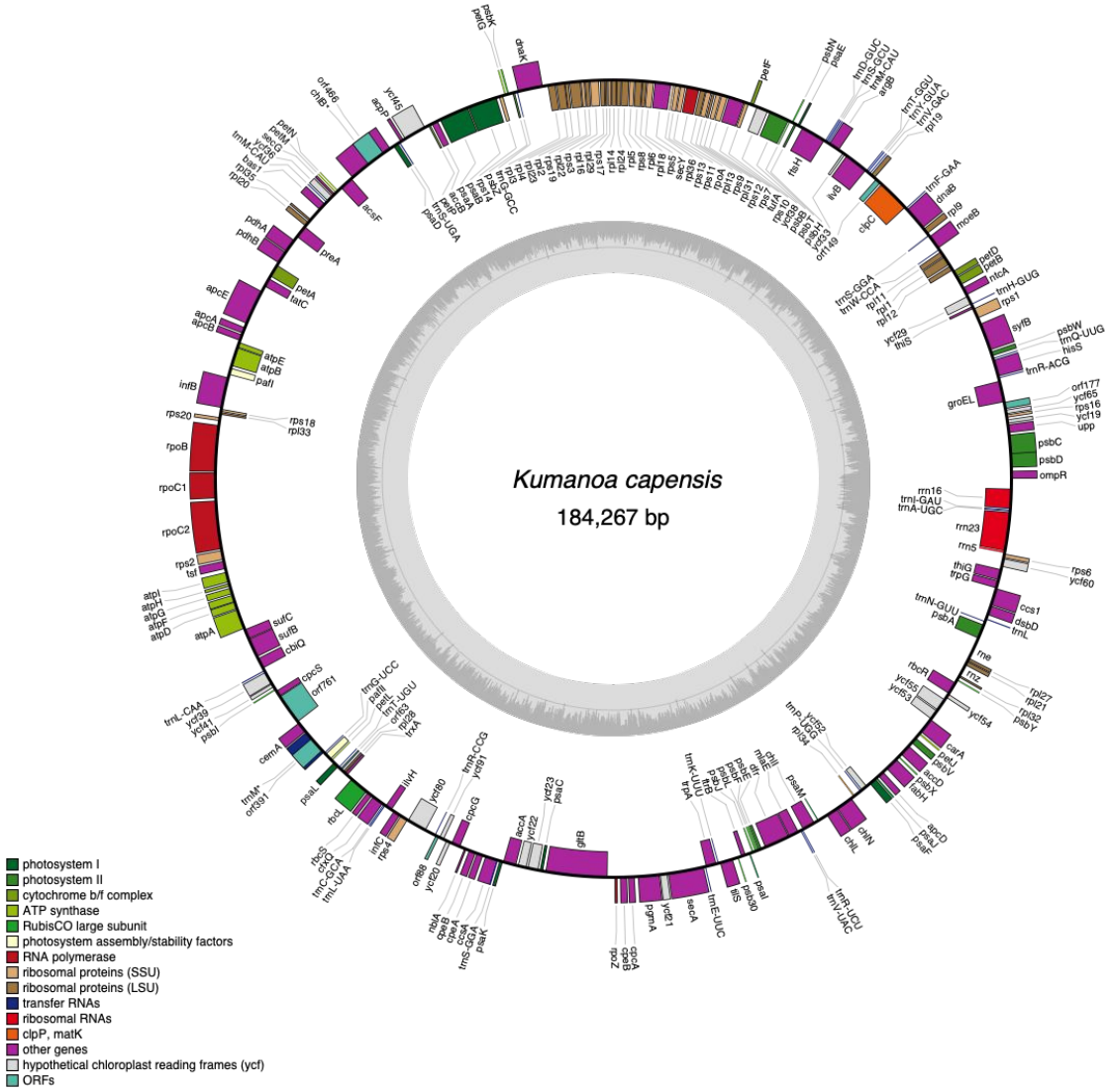


Figure S2. Gene map of *Kumanoa capensis* chloroplast genome. Genes labeled on the outside of the outer circle are transcribed clockwise, genes on the inside counterclockwise. A gene with a star (*) indicates an intron within the gene. The inner circle represents the nucleotide content (G/C = dark gray, A/T = light gray). Gene boxes are color-coded by functional groups shown in the key.

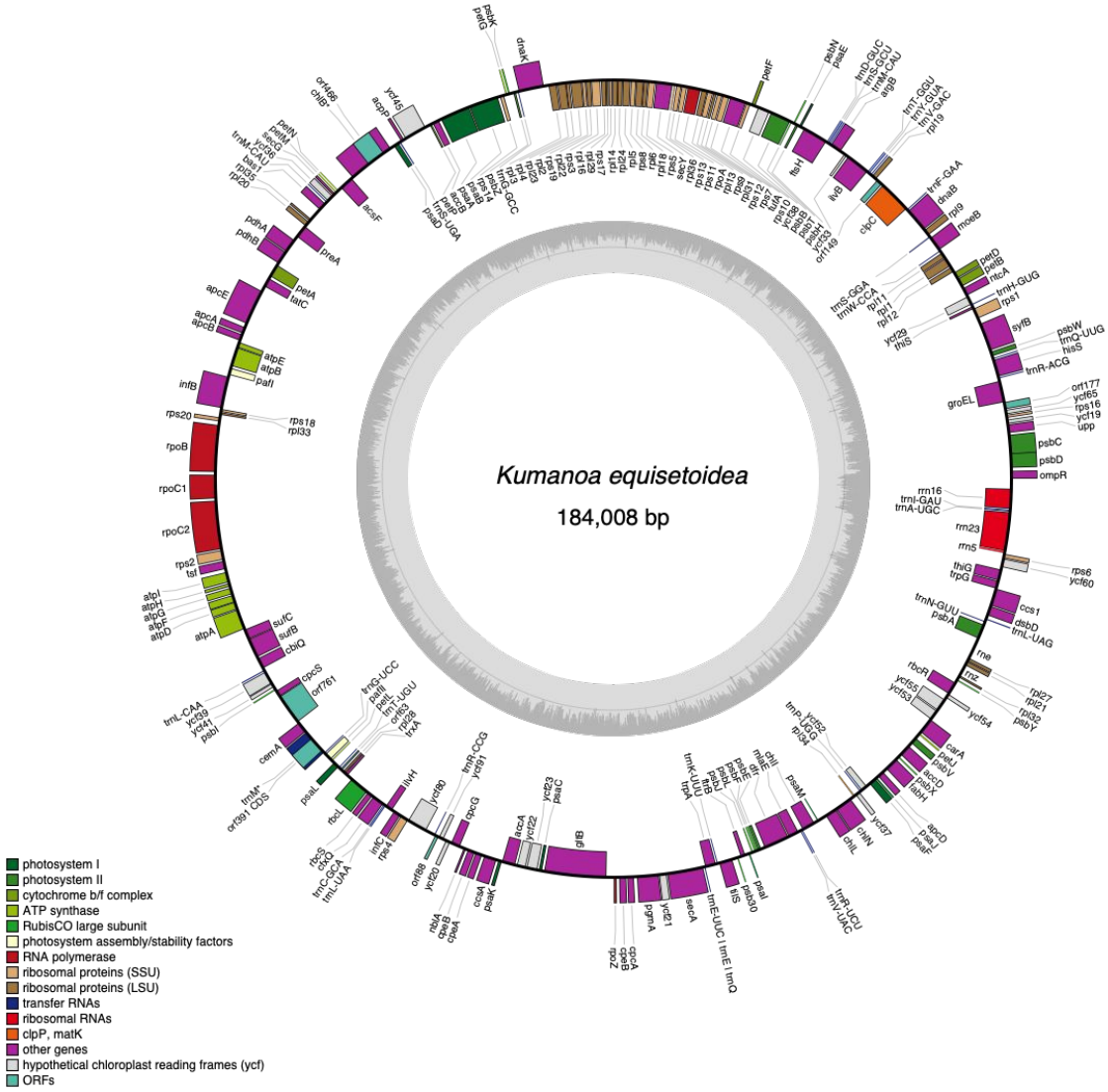


Figure S3. Gene map of *Kumanoa equisetoidea* chloroplast genome. Genes labeled on the outside of the outer circle are transcribed clockwise, genes on the inside counterclockwise. A gene with a star (*) indicates an intron within the gene. The inner circle represents the nucleotide content (G/C = dark gray, A/T = light gray). Gene boxes are color-coded by functional groups shown in the key.

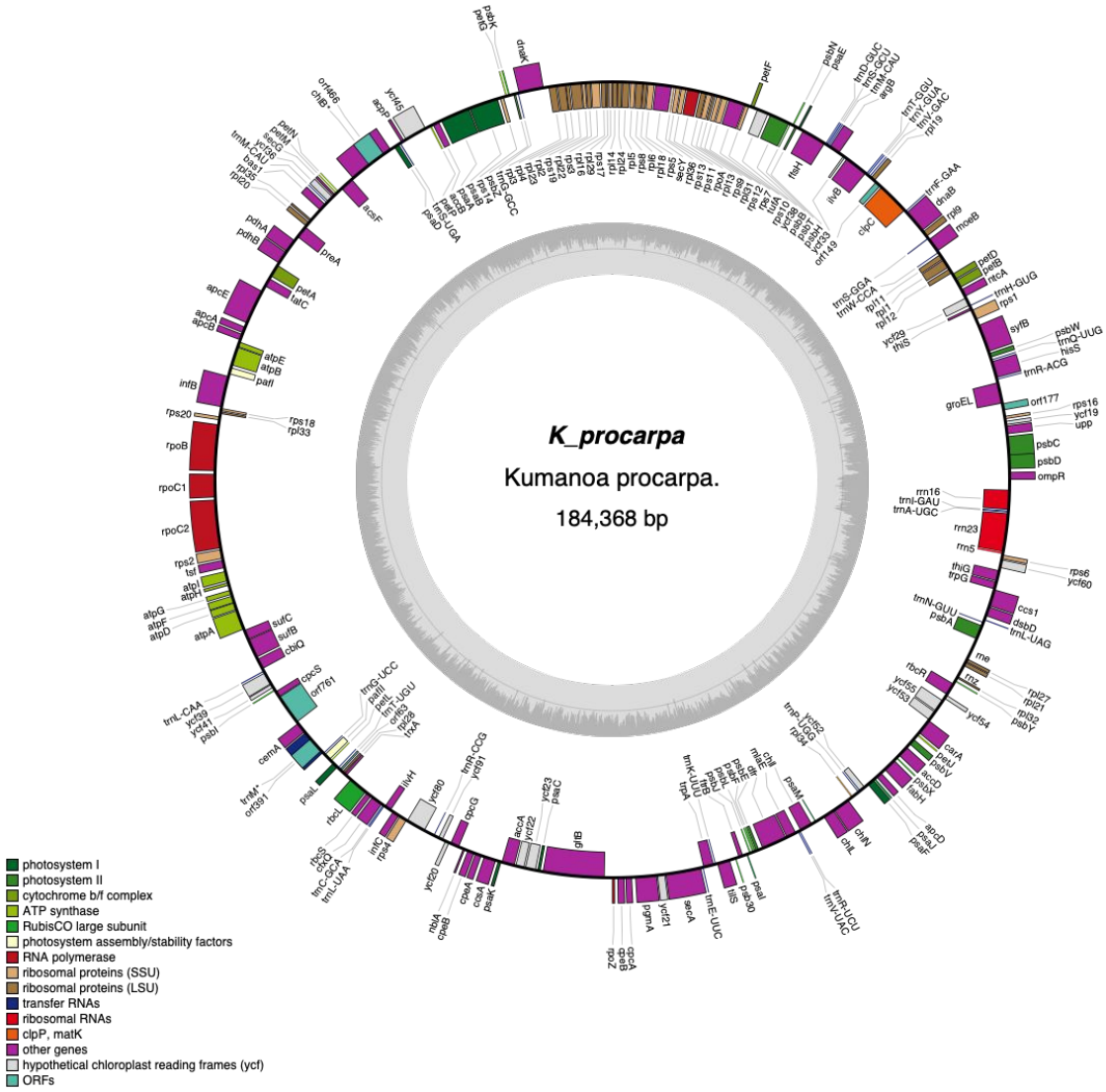


Figure S4. Gene map of *Kumanoa procarpa* chloroplast genome. Genes labeled on the outside of the outer circle are transcribed clockwise, genes on the inside counterclockwise. A gene with a star (*) indicates an intron within the gene. The inner circle represents the nucleotide content (G/C = dark gray, A/T = light gray). Gene boxes are color-coded by functional groups shown in the key.

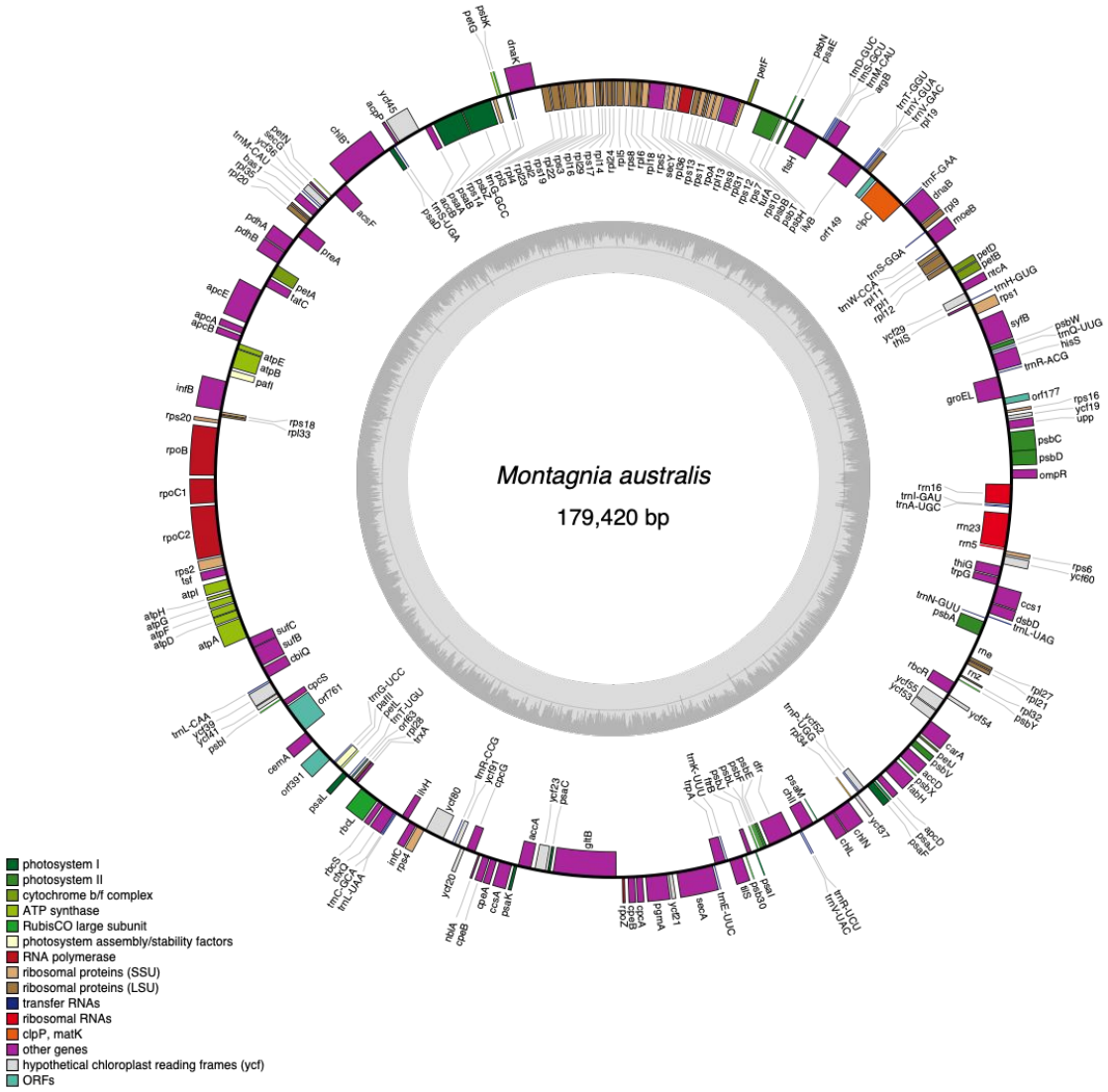


Figure S5. Gene map of *Montagnia australis* chloroplast genome. Genes labeled on the outside of the outer circle are transcribed clockwise, genes on the inside counterclockwise. A gene with a star (*) indicates an intron within the gene. The inner circle represents the nucleotide content (G/C = dark gray, A/T = light gray). Gene boxes are color-coded by functional groups shown in the key.

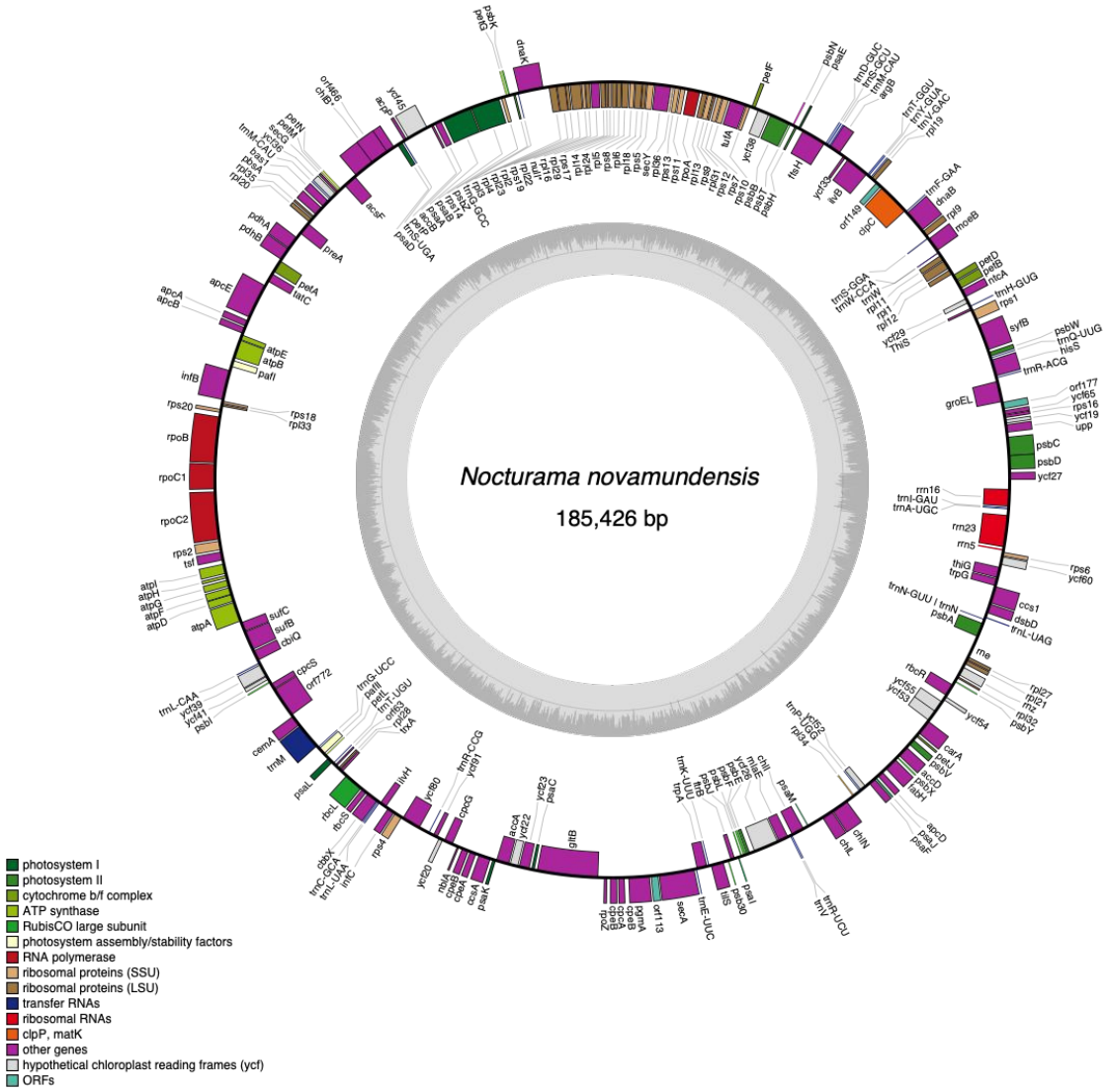


Figure S6. Gene map of *Nocturama novamundensis* chloroplast genome. Genes labeled on the outside of the outer circle are transcribed clockwise, genes on the inside counterclockwise. A gene with a star (*) indicates an intron within the gene. The inner circle represents the nucleotide content (G/C = dark gray, A/T = light gray). Gene boxes are color-coded by functional groups shown in the key.

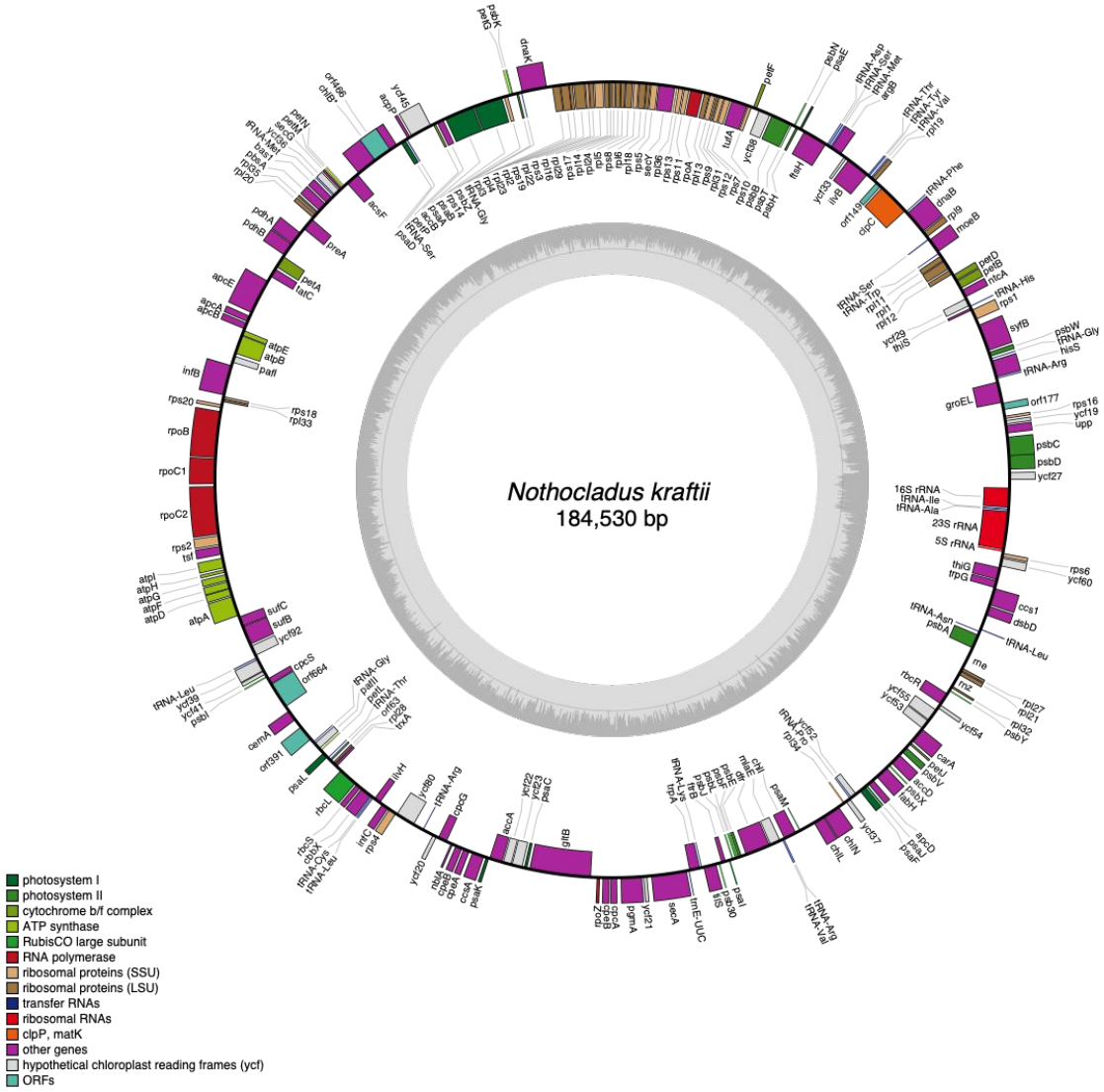


Figure S7. Gene map of *Nothocladus kraftii* chloroplast genome. Genes labeled on the outside of the outer circle are transcribed clockwise, genes on the inside counterclockwise. A gene with a star (*) indicates an intron within the gene. The inner circle represents the nucleotide content (G/C = dark gray, A/T = light gray). Gene boxes are color-coded by functional groups shown in the key.

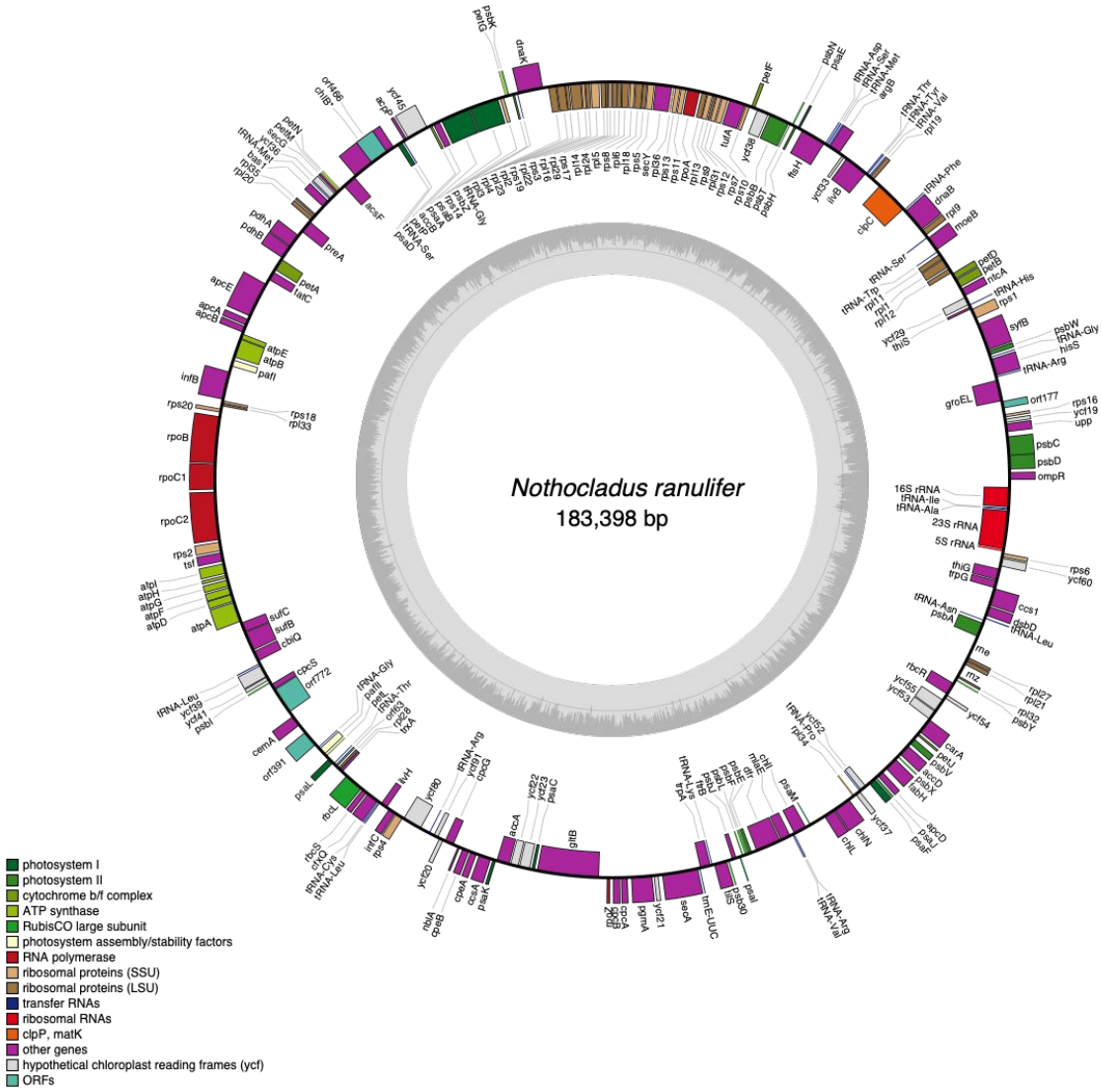


Figure S8. Gene map of *Nothocladus ranulifer* chloroplast genome. Genes labeled on the outside of the outer circle are transcribed clockwise, genes on the inside counterclockwise. A gene with a star (*) indicates an intron within the gene. The inner circle represents the nucleotide content (G/C = dark gray, A/T = light gray). Gene boxes are color-coded by functional groups shown in the key.

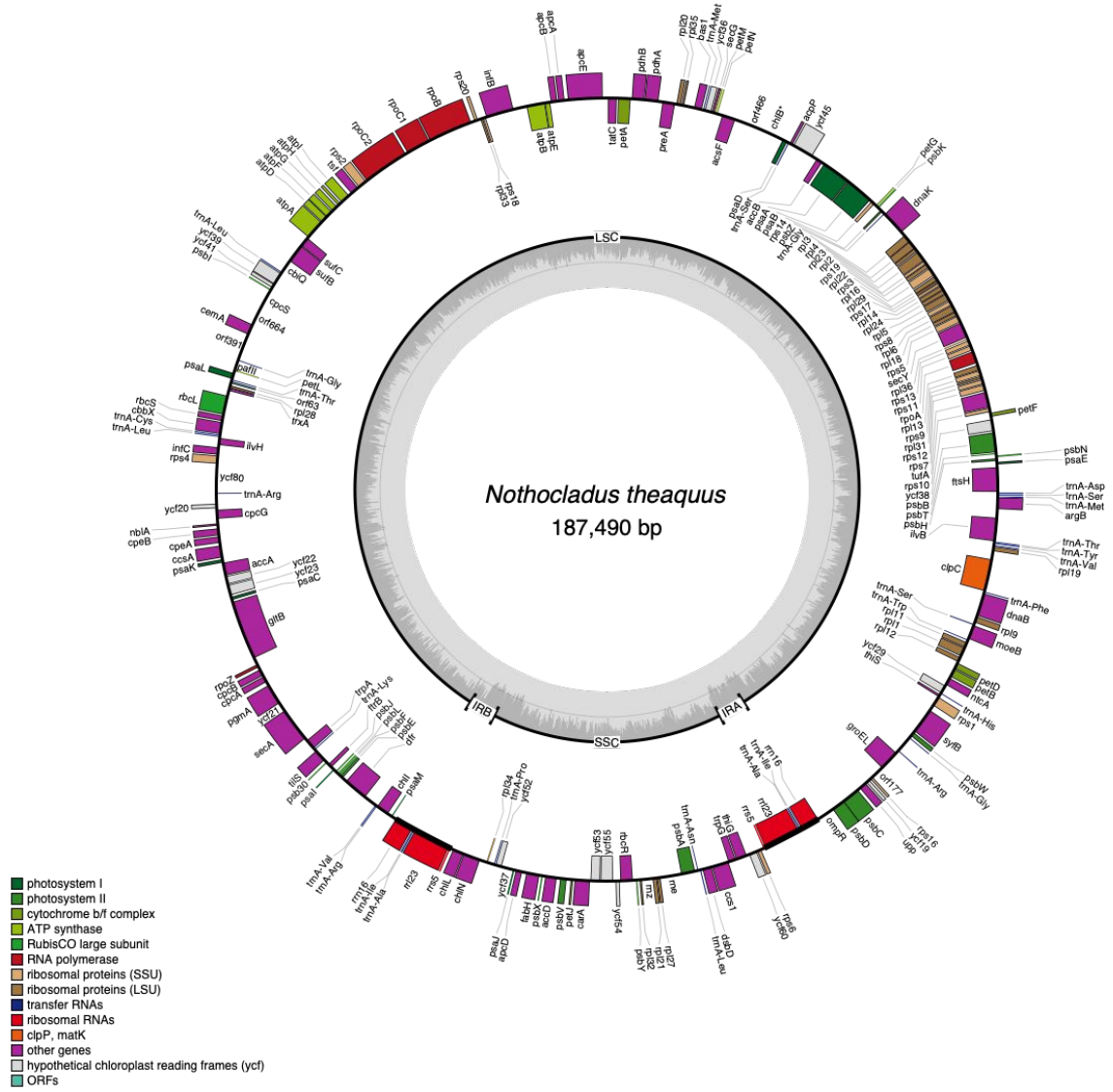


Figure S9. Gene map of *Nothocladus theaquus* chloroplast genome. Genes labeled on the outside of the outer circle are transcribed clockwise, genes on the inside counterclockwise. A gene with a star (*) indicates an intron within the gene. The inner circle represents the nucleotide content (G/C = dark gray, A/T = light gray). Gene boxes are color-coded by functional groups shown in the key.

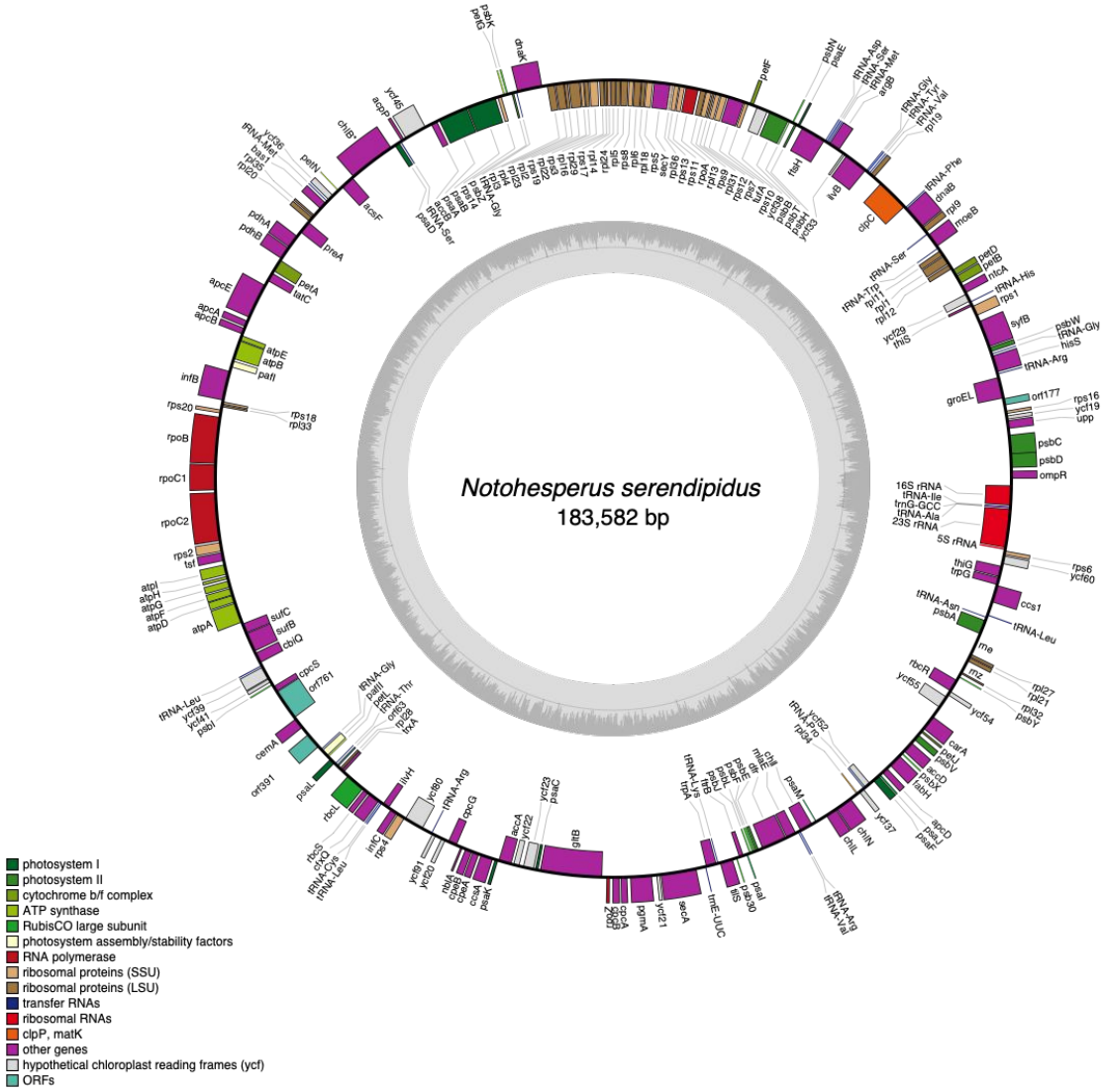


Figure S10. Gene map of *Notohesperus serendipidus* chloroplast genome. Genes labeled on the outside of the outer circle are transcribed clockwise, genes on the inside counterclockwise. A gene with a star (*) indicates an intron within the gene. The inner circle represents the nucleotide content (G/C = dark gray, A/T = light gray). Gene boxes are color-coded by functional groups shown in the key.

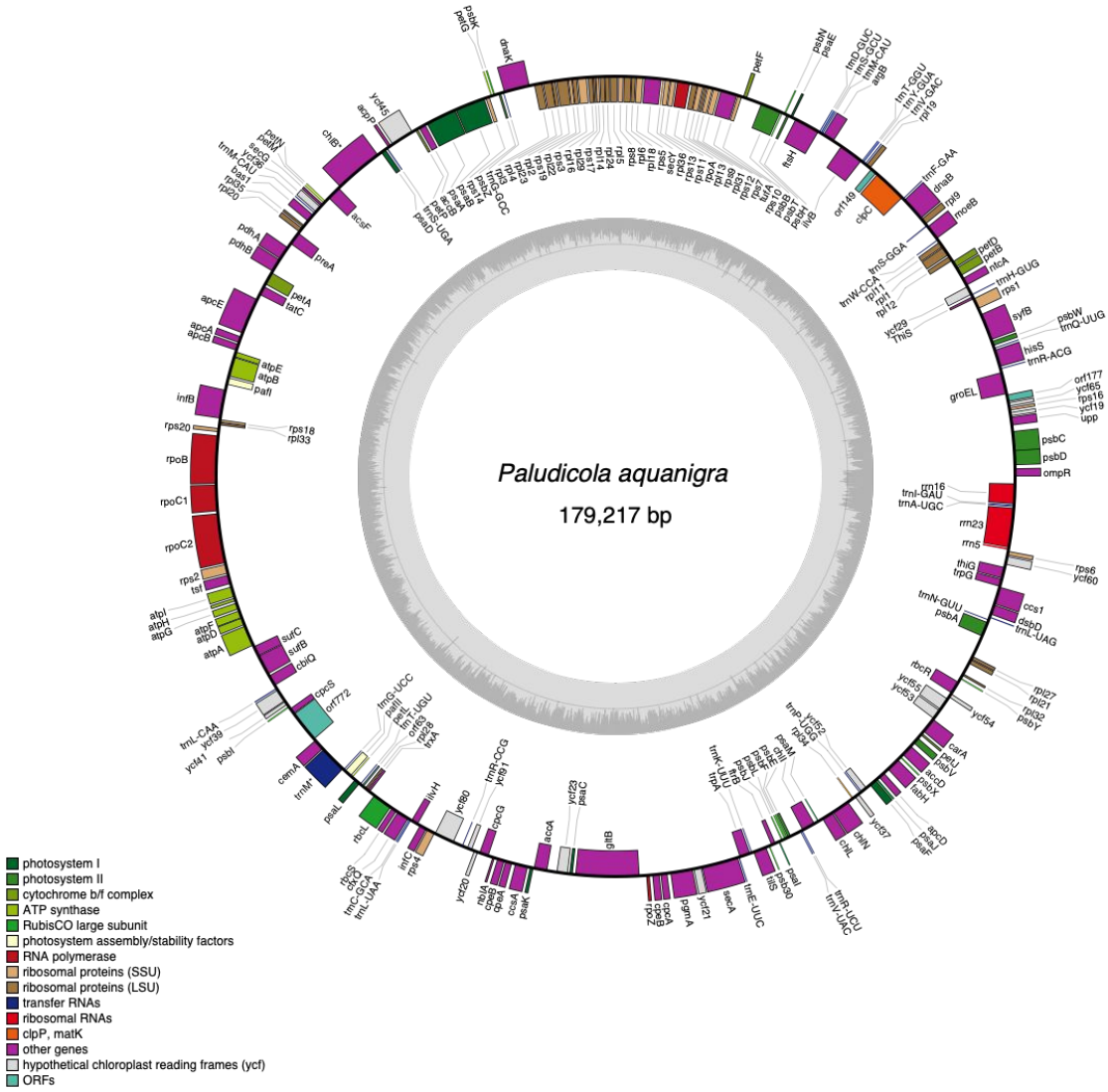


Figure S11. Gene map of *Paludicola aquanigra* chloroplast genome. Genes labeled on the outside of the outer circle are transcribed clockwise, genes on the inside counterclockwise. A gene with a star (*) indicates an intron within the gene. The inner circle represents the nucleotide content (G/C = dark gray, A/T = light gray). Gene boxes are color-coded by functional groups shown in the key.

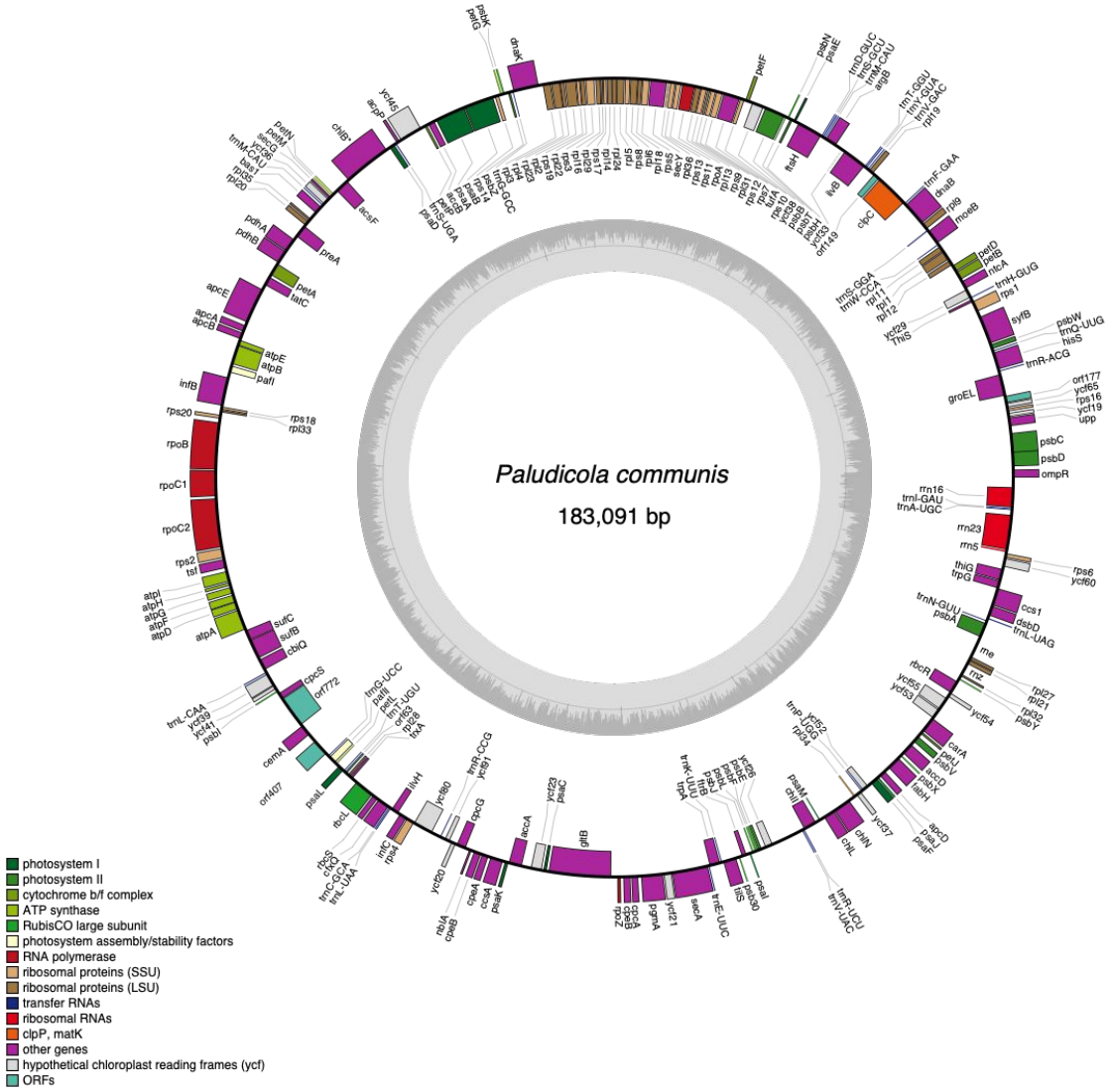


Figure S12. Gene map of *Paludicola communis* chloroplast genome. Genes labeled on the outside of the outer circle are transcribed clockwise, genes on the inside counterclockwise. A gene with a star (*) indicates an intron within the gene. The inner circle represents the nucleotide content (G/C = dark gray, A/T = light gray). Gene boxes are color-coded by functional groups shown in the key.

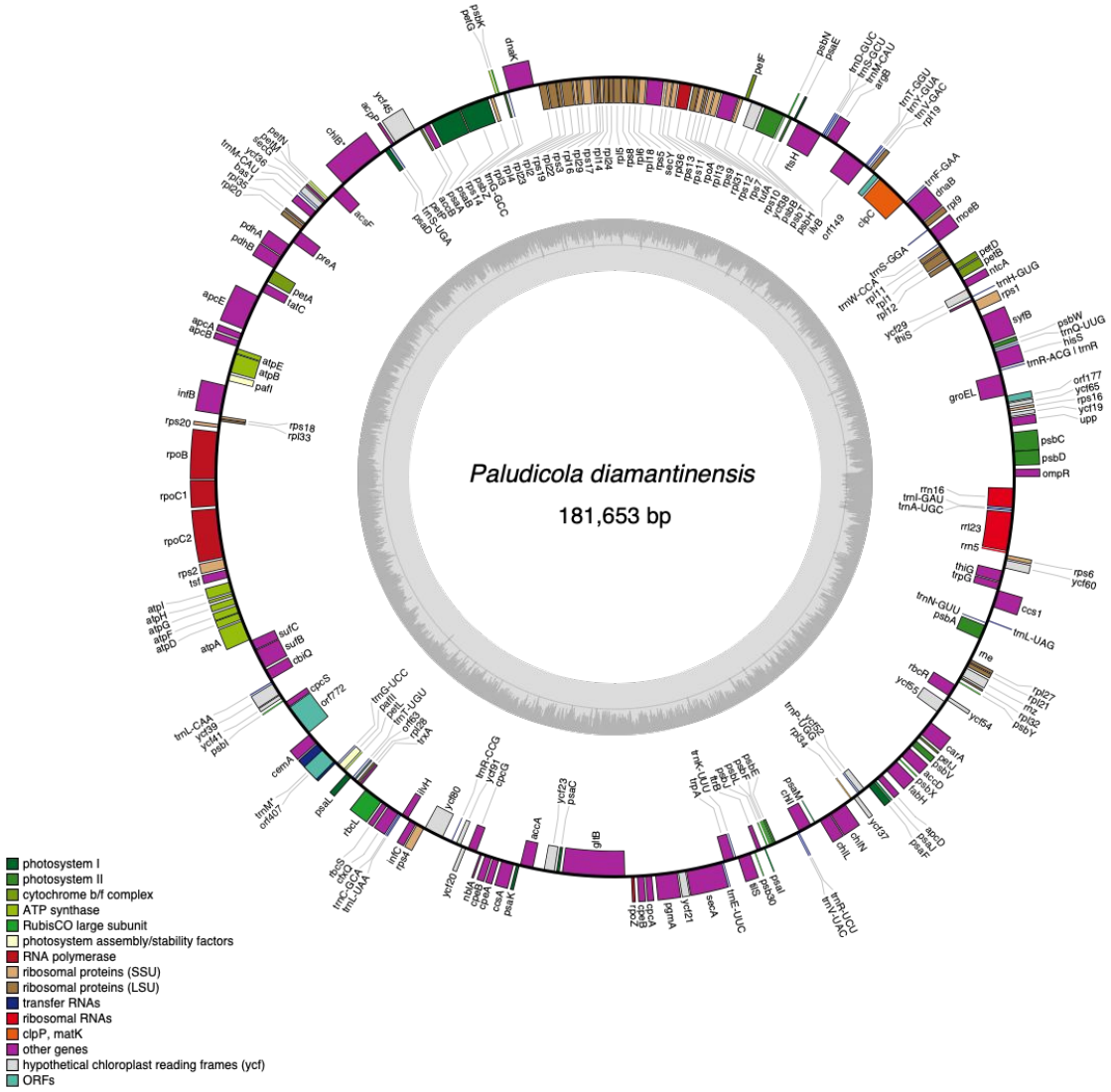


Figure S13. Gene map of *Paludicola diamantinensis* chloroplast genome. Genes labeled on the outside of the outer circle are transcribed clockwise, genes on the inside counterclockwise. A gene with a star (*) indicates an intron within the gene. The inner circle represents the nucleotide content (G/C = dark gray, A/T = light gray). Gene boxes are color-coded by functional groups shown in the key.

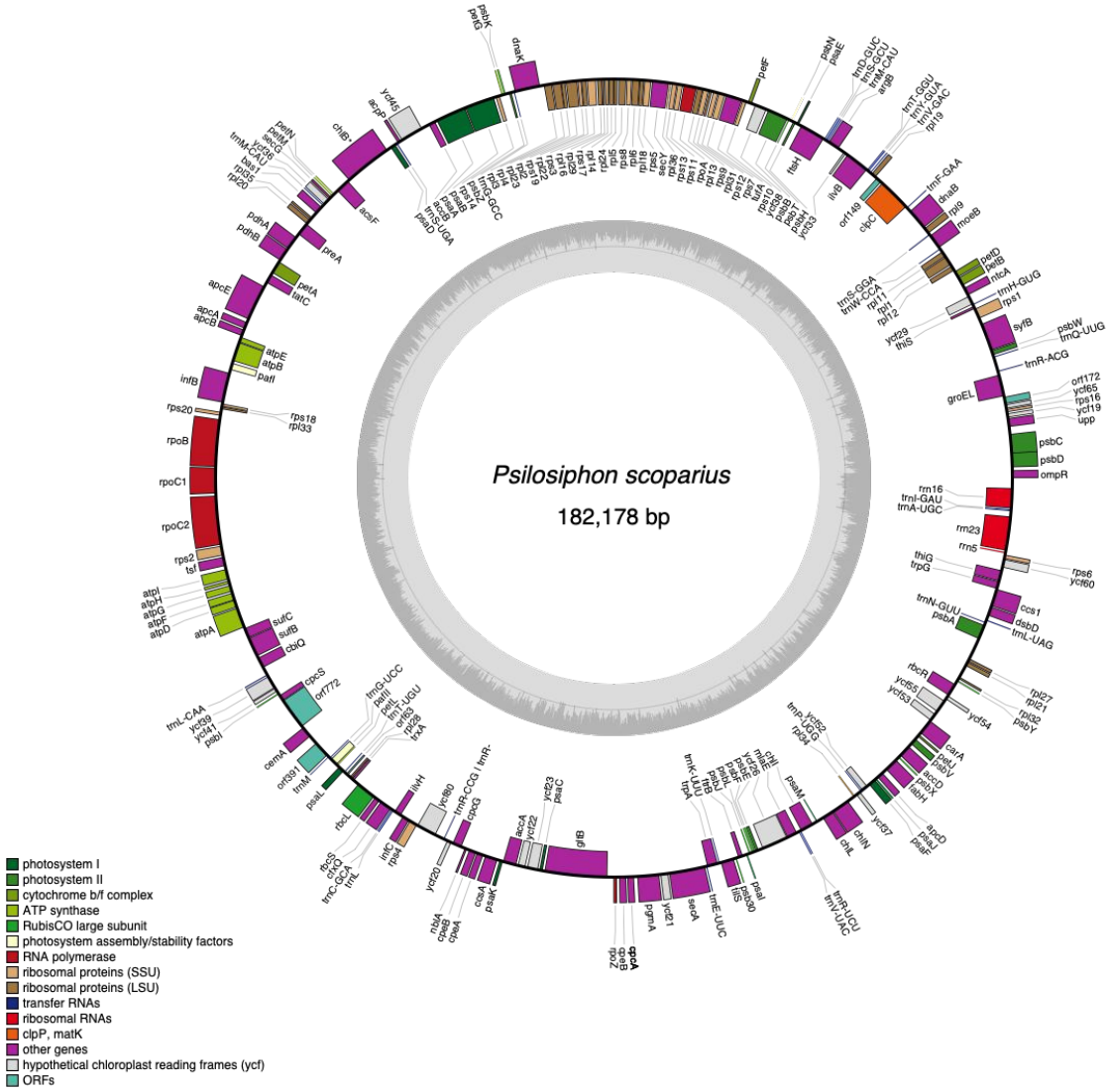


Figure S14. Gene map of *Psilosiphon scoparius* chloroplast genome. Genes labeled on the outside of the outer circle are transcribed clockwise, genes on the inside counterclockwise. A gene with a star (*) indicates an intron within the gene. The inner circle represents the nucleotide content (G/C = dark gray, A/T = light gray). Gene boxes are color-coded by functional groups shown in the key.

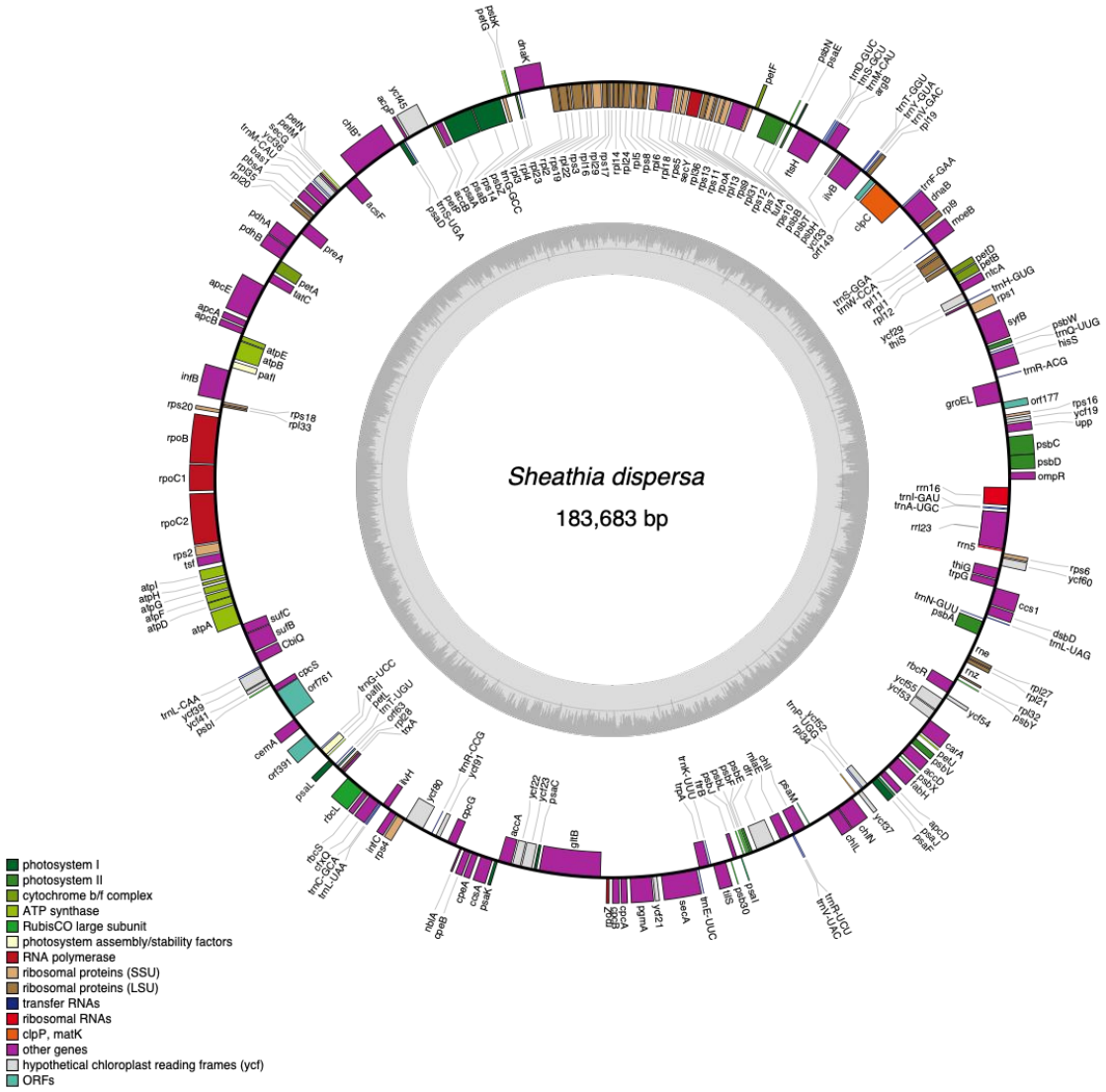


Figure S15. Gene map of *Sheathia dispersa* chloroplast genome. Genes labeled on the outside of the outer circle are transcribed clockwise, genes on the inside counterclockwise. A gene with a star (*) indicates an intron within the gene. The inner circle represents the nucleotide content (G/C = dark gray, A/T = light gray). Gene boxes are color-coded by functional groups shown in the key.

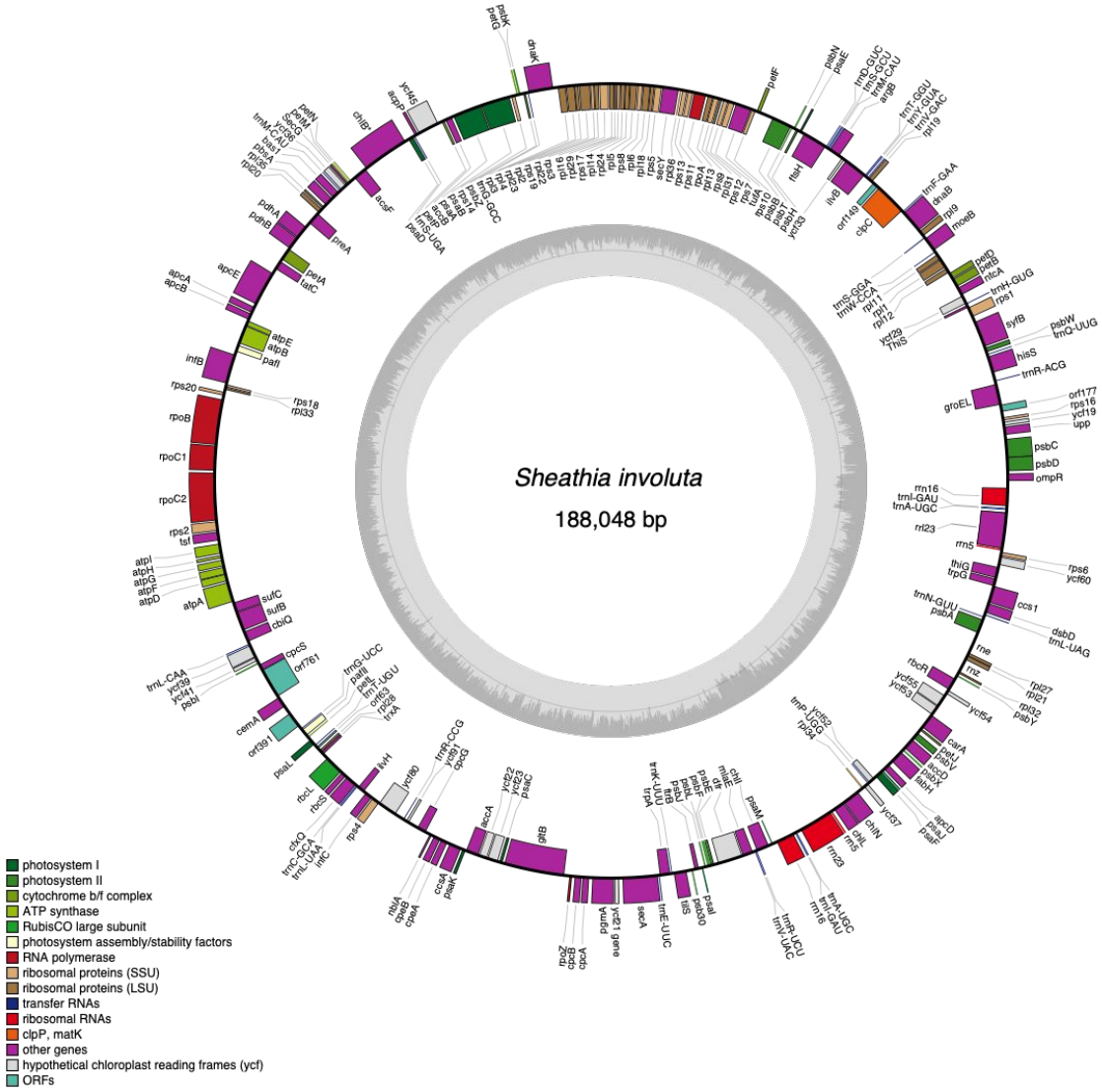


Figure S16. Gene map of *Sheathia involuta* chloroplast genome. Genes labeled on the outside of the outer circle are transcribed clockwise, genes on the inside counterclockwise. A gene with a star (*) indicates an intron within the gene. The inner circle represents the nucleotide content (G/C = dark gray, A/T = light gray). Gene boxes are color-coded by functional groups shown in the key.

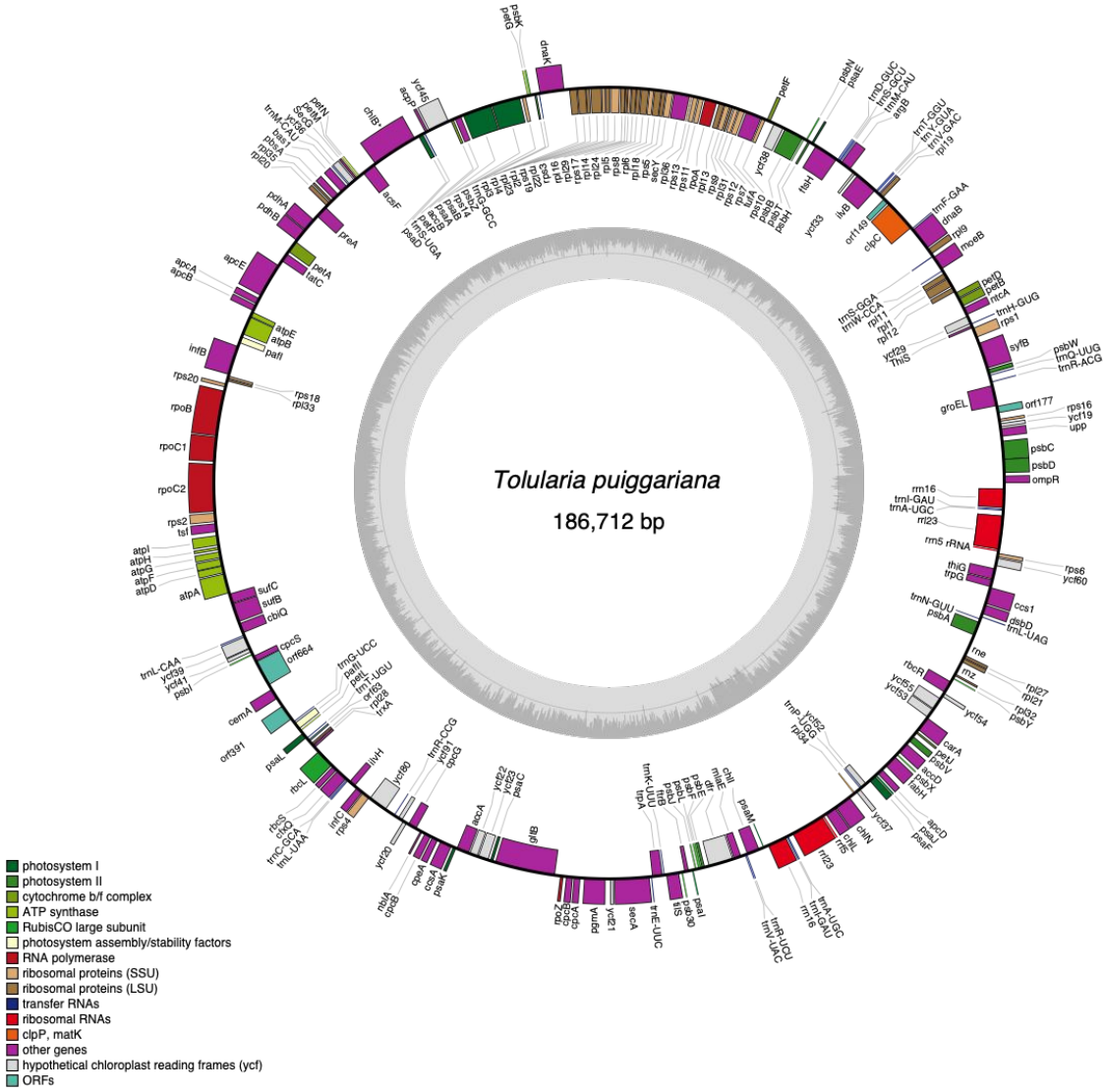


Figure S17. Gene map of *Torularia puiggariana* chloroplast genome. Genes labeled on the outside of the outer circle are transcribed clockwise, genes on the inside counterclockwise. A gene with a star (*) indicates an intron within the gene. The inner circle represents the nucleotide content (G/C = dark gray, A/T = light gray). Gene boxes are color-coded by functional groups shown in the key.

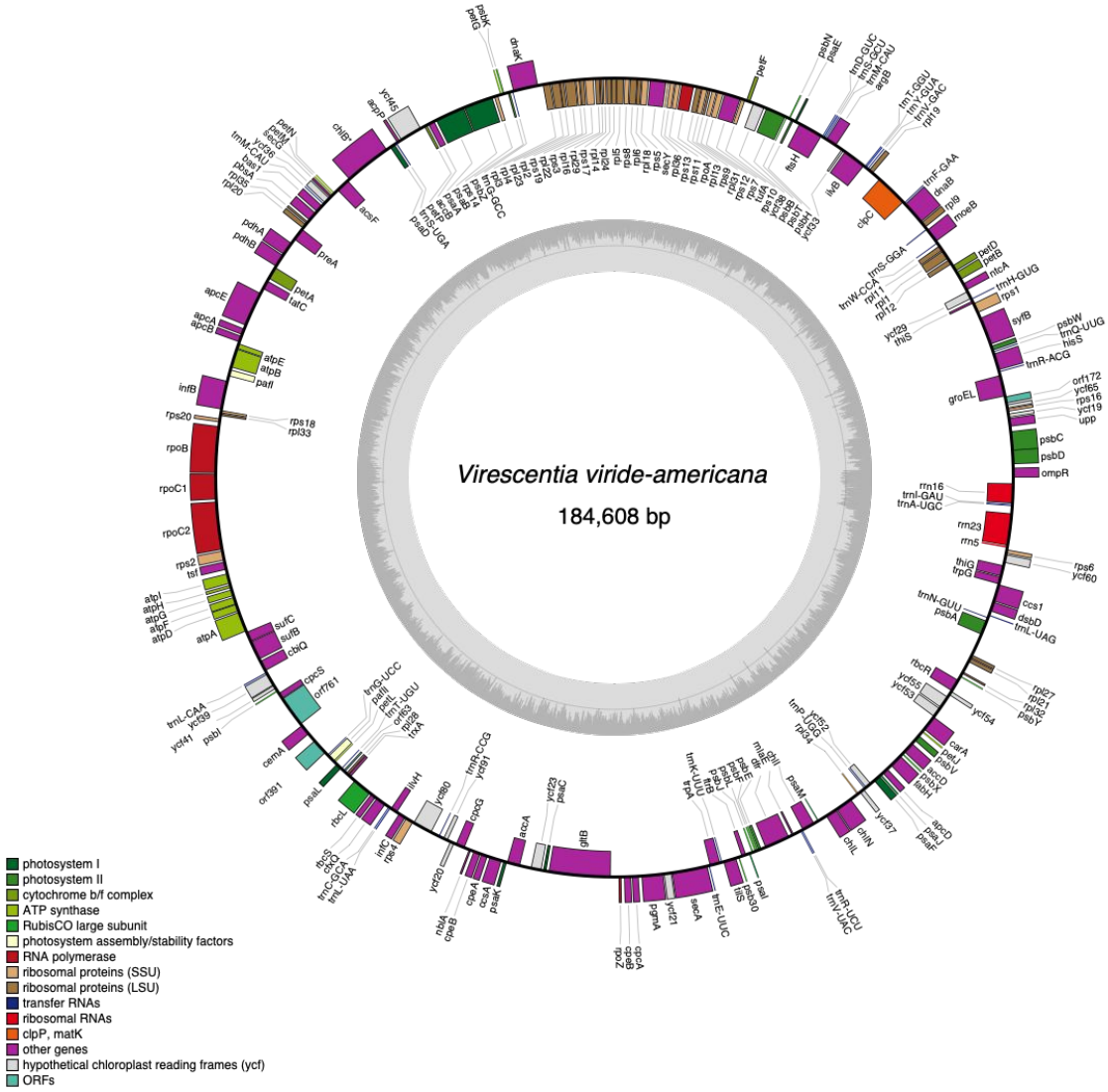


Figure S18. Gene map of *Virescentia viride-americana* chloroplast genome. Genes labeled on the outside of the outer circle are transcribed clockwise, genes on the inside counterclockwise. A gene with a star (*) indicates an intron within the gene. The inner circle represents the nucleotide content (G/C = dark gray, A/T = light gray). Gene boxes are color-coded by functional groups shown in the key.

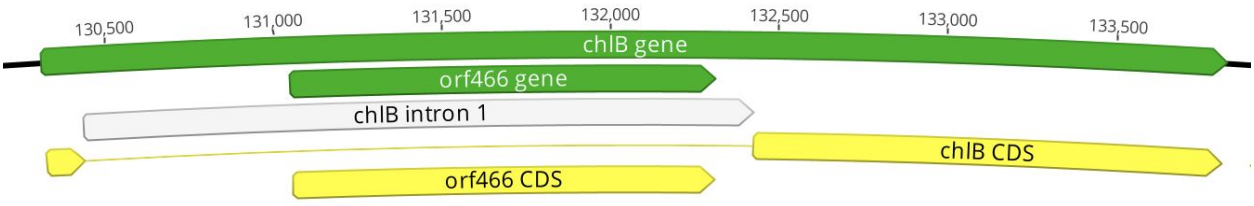


Figure S19. Depiction of the *chl*B gene with an intron that was present in all 40 Batrachospermales plastid genomes. Shown *Acarposporophycos brasiliensis*.

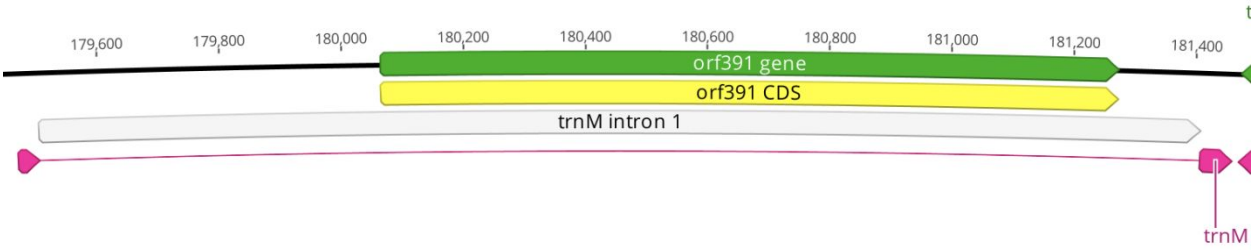


Figure S20. Depiction of the *trn*M gene with an intron. When the gene was present in the plastid genome, the intron was also present. Shown *Kumanoa capensis*.

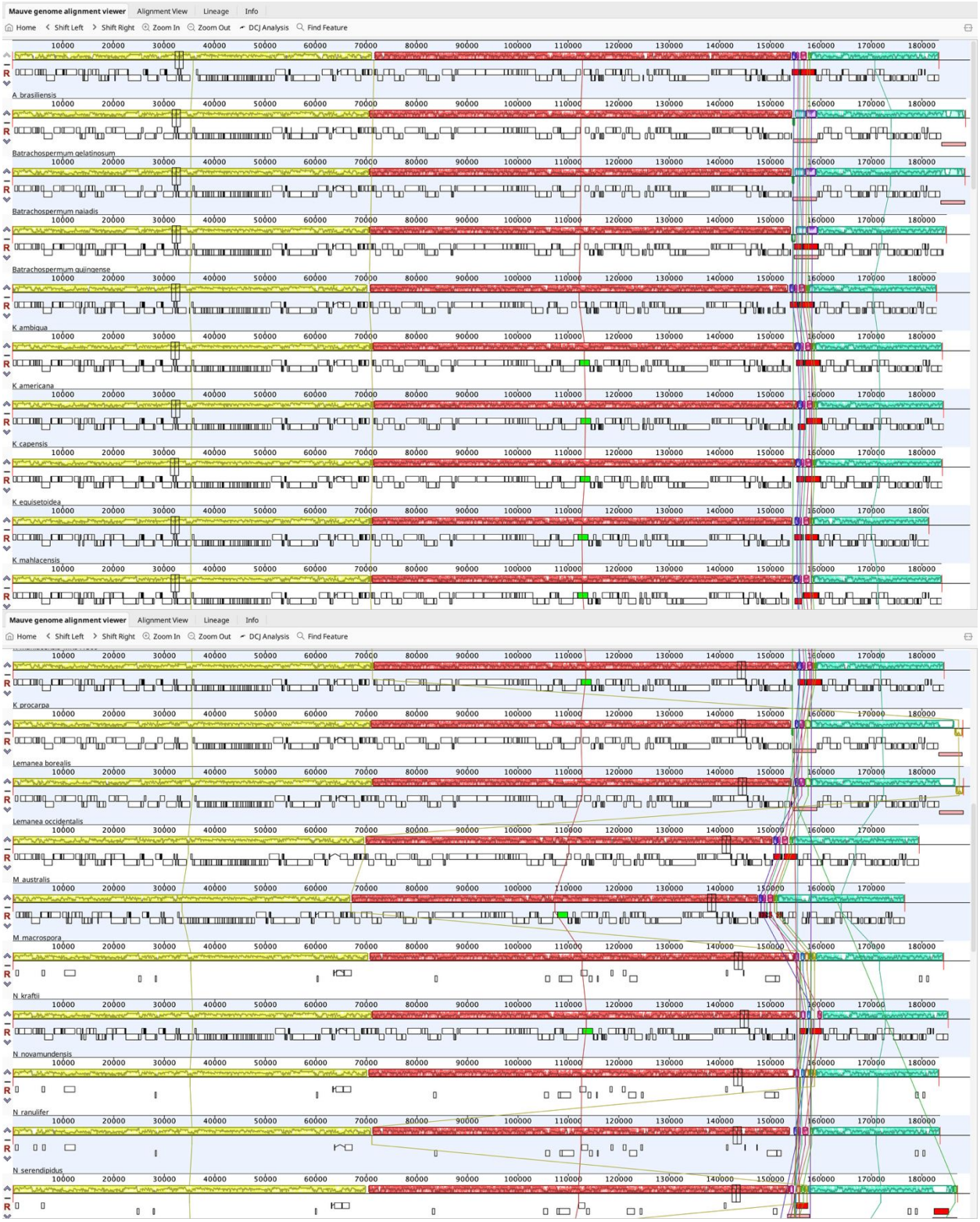

Mauve genome alignment viewer
Alignment View
Lineage
Info
Home
Shift Left
Shift Right
Zoom In
Zoom Out
DCJ Analysis
Find Feature
A. brasiliensis
Batrachospermum gelatinosum
Batrachospermum naiadis
Batrachospermum quiingense
K. ambigua
K. americana
K. capensis
K. equisetoidea
K. mahlacensis
K. procarpa
Lemanea borealis
Lemanea occidentalis
M. australis
M. macrospora
N. kraftii
N. novamundensis
N. ranulifer
N. serendipidus

Figure S21. Alignment of the 37 Batrachospermales plastid genomes. There were three large locally collinear blocks (LCB) predicted, and they are indicated by corresponding boxes. Within each LCB, a sequence similarity profile is reported. Annotations are represented by the boxes below the LCB: protein-coding genes and tRNAs are represented by white boxes, while rRNAs

genes are represented by filled (shaded) boxes. ORFs are represented by light green boxes. The position of the boxes above or below the line refers to the orientation of the gene.