ORGANIZATION OF PLASTID GENOMES IN THE FRESHWATER RED ALGAL ORDER BATRACHOSPERMALES (RHODOPHYTA)[1]

Monica Orlandi Paiano[3]

Zoology and Botany Department, São Paulo State University, São José do Rio Preto 15054-000, Brazil

Andrea Del Cortona[3]

Phycology Research Group, Ghent University, Krijgslaan 281-S8, 9000 Ghent, Belgium

Joana F. Costa

School of Biosciences, University of Melbourne, Melbourne Victoria 3010, Australia

Shao-Lun Liu iD

Department of Life Science, Tunghai University, Taichung 40704, Taiwan

Heroen Verbruggen iD

School of Biosciences, University of Melbourne, Melbourne Victoria 3010, Australia

Olivier De Clerck

Phycology Research Group, Ghent University, Krijgslaan 281-S8, 9000 Ghent, Belgium

and Orlando Necchi Jr iD[2]

Zoology and Botany Department, São Paulo State University, São José do Rio Preto 15054-000, Brazil

Little is known about genome organization in members of the order Batrachospermales, and the infra-ordinal relationship remains unresolved. Plastid (cp) genomes of seven members of the freshwater red algal order Batrachospermales were sequenced, with the following aims: (i) to describe the characteristics of cp genomes and compare these with other red algal groups; (ii) to infer the phylogenetic relationships among these members to better understand the infra-ordinal classification. Cp genomes of Batrachospermales are large, with several cases of gene loss, they are gene-dense (high gene content for the genome size and short intergenic regions) and have highly conserved gene order. Phylogenetic analyses based on concatenated nucleotide genome data roughly supports the current taxonomic system for the order. Comparative analyses confirm data for members of the class Florideophyceae that cp genomes in Batrachospermales is highly conserved, with little variation in gene composition. However, relevant new features were revealed in our study: genome sizes in members of Batrachospermales are close to the lowest values reported for Florideophyceae; differences in cp genome size within the order are large in comparison with other orders (Ceramiales, Gelidiales, Gracilariales, Hildenbrandiales, and Nemaliales); and members of Batrachospermales have the lowest number of protein-coding genes among the Florideophyceae. In terms of gene loss, *apc*F, which encodes the allophycocyanin beta subunit, is absent in all sequenced taxa of Batrachospermales. We reinforce that the interordinal relationships between the freshwater orders Batrachospermales and Thoreales within the Nemaliophycidae is not well resolved due to limited taxon sampling.

Key index words: Batrachospermales; conserved genomes; genome rearrangements; infra-ordinal classification; Nemaliophycidae; Rhodophyta

Abbreviations: *apc*F, allophycocyanin beta subunit; bp, base pair; cp, plastid; LCB, locally collinear block; ML, Maximum Likelihood; mt, mitochondria; mya, million years ago; nt, nucleotide; *omp*, transcriptional regulatory protein; ORF, open reading frame; *rbc*L, ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit

The red algae (Rhodophyta) are a diverse group of around 7,150 species of photosynthetic eukaryotes with a 1.2 billion-year-old fossil record assigned to *Bangiomorpha pubescens* (Butterfield 2000, Guiry and Guiry 2017). Most red algae inhabit marine

environments (98%), but many well-known taxa are exclusively from freshwater habitats or acid hot springs, and some are also reported from tropical rainforests as members of the subaerial community (Gurgel and Lopez-Bautista 2007). The phylum Rhodophyta is currently composed of seven classes (Bangiophyceae, Compsopogonophyceae, Cyanidiophyceae, Florideophyceae, Phorphyridiophyceae, Rhodellophyeceae, and Stylonematophyceae) with the most species-rich class Florideophyceae containing five subclasses: Hildenbranndiophycidae, Nemaliophycidae, Corallinophycidae, Ahnfeltiophycidae, and Rhodymeniophycidae (Saunders and Hommersand 2004, Yoon et al. 2006, Yang et al. 2016).

The conservation among cp genomes in red algae, in addition to the fact that they are predominantly inherited uniparentally, has made chloroplasts prime targets for understanding evolutionary relationships across and within the Rhodophyta (Janouškovec et al. 2013, Costa et al. 2016).

Red algae have the largest plastid (cp) genomes among the groups of algae, ranging from 149,987 nt in length in *Cyanidioschyzon merolae* (Cyanidiophyceae) to 259,000 nt in *Porphyridium sordidum* (Porphyridiophyceae), but with narrow ranges (175,000–194,000 nt) in the Florideophyceae (Lee et al. 2016). However, much larger cp genomes have been recently described (Muñoz-Gómez et al. 2017) in some members of the new subphylum Proteorhodophytina composed of the classes Compsopogonophyceae, Porphyridiophyceae, Rhodellophyceae, and Stylonematophyceae and comprising the largest cp genomes sequenced (ranging from 205,000 to 1,127,000 nt). The number of coding genes in red algae is also the highest among the groups of algae (184–235; Lee et al. 2016). Red algal cp genomes not only have the highest gene content among eukaryotes with primary cp (Archaeplastida), but their genomes are also the most conserved (Janouškovec et al. 2013, Lee et al. 2016). The only known exception is the cp of the alloparasite *Choreocolax polysiphoniae*, which lost all photosynthetic pathways, and its size was reduced to 90,000 nt (Salomaki et al. 2015). Among the advantages of analyzing cp genomes is the fact that they are present in multiple copies in each cell; therefore, cpDNA data are easily obtained from bulk DNA extractions (Costa et al. 2016). In addition, the non-recombinant nature of cp genomes makes them a good tool when inferring ancient phylogenetic relationships (Leliaert et al. 2016).

Among the five subclasses of Florideophyceae, the Nemaliophycidae is the most taxonomically diverse, providing an ideal case study for red algal evolution (Lam et al. 2015). Freshwater red algae occur in early diverging members of the phyllum, for example, *Compsopogon* (Compsopogonophyceae), *Bangia* (Bangiophyceae), *Chroodactylon* and *Chroothece* (Stylonematophyceae), *Kyliniella* (Porphyridiophyceae),

and in some Florideophyceae with mostly marine representatives, for example, Hildenbrandiophycidade (*Hildenbrandia*) and Rhodymeniophycidade (*Bostrychia, Caloglossa,* and *Polysiphonia*). Furthermore, Nemaliophycidae is the only subclass within the class Florideophyceae comprising orders with exclusively freshwater taxa, that is, Balbianiales, Batrachospermales, and Thoreales. These three orders have been shown to be relatively phylogenetically distant (Lam et al. 2015, Yang et al. 2016). To elucidate the phylogeny of the Nemaliophycidae, Lam et al. (2015) performed a phylogenetic analysis based on a nine-gene dataset comprised of nuclear, cp, and mitochondrial markers. They demonstrated that some orders (Nemaliales, Entwisleiales, Colaconematales, Palmariales, and Acrochaetiales) formed a highly supported clade. In contrast, all other relationships among the orders had low support. They concluded that phylogenomic approaches are necessary to provide a well-supported phylogeny for this subclass. They pointed out the need to resolve all the phylogenetic relationships in order to trace the evolution of freshwater species from marine ancestors, as well as explore the evolution of diverse reproductive traits.

Recent multigene or phylogenomic studies of individual orders of Nemaliophycidae (Lam et al. 2015, Costa et al. 2016) have elucidated the relationships within and among closely related groups. Costa et al. (2016) investigated Nemaliales and found that the six currently circumscribed families are clustered into two evolutionary lineages with strong support based on chloroplast phylogenomic analyses.

Batrachospermales is the most diverse in terms of morphology, reproductive characters, and number of taxa among the freshwater red algal orders (Kumano 2002, Entwisle et al. 2009, Lam et al. 2015). Thoreales was originally included in Batrachospermales (Pueschel and Cole 1982), but was separated into its own order (Muller et al. 2002). It is presently characterized by the following combination of characters (Pueschel and Cole 1982, Garbary and Gabrielson 1990, Kumano 2002, Entwisle et al. 2009): (i) thalli heterotrichous, uniaxial, gelatinous, or cartilaginous; (ii) axial cells bearing determinant lateral assimilatory filaments; (iii) pit plugs with two cap layers, with an expanded dome-shaped outer layer; tetraspores lacking, meiosis in diploid vegetative cells giving rise to haploid axes; (iv) multiple discoid chloroplasts lacking pyrenoids; (v) exclusively freshwater distribution. Although the order is well supported in all phylogenetic analyses (Entwisle et al. 2009, Lam et al. 2015), the infra-ordinal classification is not well established yet. The classical system with three families (Batrachospermaceae, Lemaneaceae, and Psilosiphonaceae; Pueschel and Cole 1982, Sheath et al. 1996) was reduced to only one family (Batrachospermaceae) in a molecular phylogeny analysis based on two molecular markers

(Entwisle et al. 2009). In contrast to some other Nemaliophycidae orders, data on cp genomes for members of the Batrachospermales are scarce and have been reported only in studies involving other red algal groups: *Kumanoa americana* (Lee et al. 2016) and *Sheathia arcuata* (Nan et al. 2017).

This investigation is a first phylogenomic approach for Batrachospermales based on cp genomes of seven representatives of the order, including a wide range of vegetative morphology, reproductive characters, and taxonomic position. We aimed to (i) describe the characteristics of cp genomes of key genera in the order and compare them to cp genomes of other red algal groups; (ii) infer the phylogenetic relationships among these members in order to clarify the infra-ordinal classification.

MATERIAL AND METHODS

Taxon sampling and species identification. Seven species of the freshwater red algal order Batrachospermales, including a relatively wide range of vegetative and reproductive morphologies (Table S1 in the Supporting Information), were sequenced. All specimens were desiccated in silica gel, except for *Paralemanea* sp., which was conserved as a dried herbarium specimen. Voucher specimens are lodged at the following herbaria: Tunghai University (THU)—*Batrachospermum macrosporum*, *Kumanoa mahlacensis,* and *Sheatia arcuata*; Ghent University (GENT)—*Paralemanea* sp.; and São Paulo State University (SJRP)—*Batrachospermum viride-brasiliense*, *Kumanoa ambigua,* and *Sirodotia delicatula*. Specimens were identified based on current diagnostic characters using recent taxonomic literature (Vis and Sheath 1992, Kumano 2002, Entwisle et al. 2009, Vis et al. 2012, Johnston et al. 2014). In addition, taxa were checked for sequence identity for the gene encoding the Rubisco large subunit (*rbc*L) against sequences in the GenBank (Benson et al. 2013) database.

Sequencing and assembling of cp DNA. Total genomic DNA was extracted using a CTAB DNA extraction protocol with minor modifications (Doyle and Doyle 1987). Extracted DNA was quantified and checked for purity at A260/280 nm (Nanodrop, Thermo Fisher Scientific, Waltham, Massachusetts, USA) prior to storage at −20°C. Sequencing of the genomic DNA was performed either on an Illumina Next Seq or an IonTorrent NGS platform (Table S1). For the first sequencing run on Illumina, libraries of 350 nt fragments were prepared from DNA extracts of each sample using a TruSeq Nano LT kit. Each library was given a unique barcode and sequenced on the Illumina HiSeq 2000 platform. For subsequent runs, libraries of 500 nt fragments were prepared using a KapaBiosystemS DNA Library Preparation Kit (KK8232) and sequenced on either a HiSeq 2500 or a NextSeq 500. Ion Torrent PGM™ sequencing was performed using a 318 chip Emulsion PCR in the Ion OneTouch 2 400 kit, with samples checked on a Bioanalyser. For library prep, the New England BioLabs NEBNext® Fast DNA Fragmentation kit with insert size of 400 nt was used.

Low-quality ends of the reads (Phred score <30) were trimmed with fastx-toolkit (https://github.com/agordon/fastx_toolkit, last accessed June 06, 2017). For Illumina libraries, trimmed reads shorter than 35 bp were discarded, while for IonTorrent libraries, trimmed reads shorter than 50 bp were discarded. To assemble the cp genomes, de novo assemblies were performed on the trimmed reads using four different assemblers: CLC Genomics Workbench version 7.5.1

(http://www.clcbio.com, last accessed on June 06, 2017), using an automatic word size of 63 and standard parameters; Geneious 8.0.5 (Biomatters, www.geneious.com, last accessed on June 06, 2017), using medium sensitivity parameters; SOAPdenovo v. 2.223 (Luo et al. 2012), using a range of kmers (23–113); and SPAdes v. 3.6.2, using kmers 21, 33, 55, and 77 (Bankevich et al. 2012). cp contigs were identified from the total assemblies after a sequence similarity search against a local database of Rhodophyta organellar genes. Despite using similar algorithms, different assemblers can result in the assembly of cp contigs with identical sequences, but different lengths: Employing different assemblers can be a successful strategy to obtain a full-length chloroplast genome where a single assembly fails. Circular mapping contigs were obtained by scaffolding the organellar contigs with identical sequences in Geneious with stringent parameters (minimum overlap identity 95%, maximum gap size 0, maximum mismatches 10).

Annotations of cp genomes. Annotation of the cp genomes was carried out with two complementary approaches: (i) Sequences were uploaded and processed with the free online tools MFannot using standard parameters (Mega Sun Inc., Saint Louis, MO, USA; http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl, last accessed March 22, 2017) and DOGMA (Wyman et al. 2004), using 80% identity cut-off for RNAs, 60% identify cut-off for protein-coding genes, and a BLAST E-value lower than $1e^{-5}$. (ii) Open reading frames (ORFs) longer than 100 aa were predicted with Geneious. Similarities to protein sequences were identified with sequence similarity searches against the nonredundant protein database at NCBI and against a local database of Rhodophyta organellar proteins. The intron class (group I or group II) was identified by comparing the RNA secondary structures predicted with MFold (Zuker 2003) and the models by Michel et al. (1989) and Michel and Westhof (1990). Positions of the start and stop codons of protein-coding genes were refined manually. Repetitive elements on the chloroplast genomes were predicted with inverted, tandem, and palindrome algorithms from the EMBOSS package using standard parameters (Rice et al. 2000). Sequences identified as inverted repeats were discarded if shorter than 15 nt, while tandem repeats were discarded if the sequence identity within each repeat was lower than 70%. The final annotations were imported into Geneious for comparison of gene content and genome structure. Cp genomes were oriented to have their first basepair downstream of the *psbD* gene, encoding the photosystem protein D2. Genome maps displayed in Figures S1–S7 in the Supporting Information were generated with OGDRAW (Lohse et al. 2013).

Phylogenomic analyses. For phylogenomic analyses, the cp genomes of 27 representative red algae were used, seven of them newly sequenced in this study. The corresponding GenBank accession numbers are *Batrachospermum viride-brasiliense* (this study, MG252484), *B. macrosporum* (this study , MG252483), *Calliarthron tuberculosum* (NC_021075), *Chondrus crispus* (NC_020795), *Galaxaura rugosa* (NC_031657), *Gracilaria salicornia* (NC_023785), *G. tenuistipitata* (NC_006137), *G. chilensis* (NC_029860), *G. lemaneiformis* (KP330491), *Grateloupia taiwanensis* (NC_021618), *Helminthocladia australis* (NC_031658), *Kumanoa ambigua* (this study, MG252485), *K. americana* (NC_031178), *K. mahlacensis* (this study, MG252486), *Laurencia* sp. (LN833431), *Liagora brachyclada* (NC_031667), *Liagoropsis maxima* (NC_031662), *Nemalion* sp. (LT622871), *Paralemanea* sp. (this study, MG252487), *Porphyra pulchra* (NC_029861), *Scinaia undulata* (NC_031664), *Sheathia arcuata* (this study, MG252488), *Sirodotia delicatula* (this study, MG252489), *Sporolithon durum* (NC_029857), *Thorea hispida* (NC_031171), *Vertebrata lanosa* (NC_026523), *Yamadaella caenomyce* (NC_031666). Alignments of individual genes were

performed at the nucleotide level and manually concatenated using Geneious. For phylogenomic analyses, we used only genes present in more than 10 taxa, totaling 177 genes (Table S2 in the Supporting Information). The final alignment was 132,691 nt positions in length. The nucleotidic alignments were translated to the corresponding amino acidic sequences. Maximum likelihood (ML) phylogenetic trees inferred from nucleotide and amino acid sequences were obtained with RAxML (Stamatakis 2014), available on the CIPRES Science Gateway V3.3 portal (https://www.phylo.org/), using the following parameters: 1,000 bootstraps, CATGTR substitution model, and DAYHOFF protein substitution matrix.

Whole-genome alignments and analysis of genome rearrangements. Synteny between cp genomes was carried out through a whole-genome alignment with the progressive-Mauve 2.3.1 algorithm (Darling et al. 2010) implemented in Geneious using the full alignment option, automatically calculated seed weights, and automated calculation of locally collinear block (LCB) scores. To estimate the minimal number of rearrangements among the selected cp genomes, double-cut and join (DCJ) genome distances were calculated with the MAUVE plugin implemented in Geneious.

RESULTS

Assembly, metrics, and organization of cp DNA. All the Batrachospermales cp genomes sequenced mapped as circular molecules (Figs. S1–S7). The assembled genomes had the following characteristics (Table 1 and Table S2, Figs. S1–S7): average coverage ranging from 176× (*Kumanoa mahlacensis*) to 1,984× (*Sirodotia delicatula*), lengths from 171,722 nt (*Batrachospermum viride-brasiliense*; Fig. S2) to 185,555 nt (*S. delicatula*; Fig. S7), number of protein-coding genes from 164 (*B. macrosporum* and *S. delicatula*; Table S2, Figs. S1 and S7) to 172 (*B. viride-brasiliense*; Table S2, Fig. S2), G+C contents between 28.0% (*B. macrosporum*) and 30.5% (*Paralemanea* sp.), number of tRNA ranging from 28 (*K. ambigua*, *K. mahlacensis*, and *Paralemanea* sp.) to 37 (*B. macrosporum*), noncoding DNA between 14.2% (*Paralemanea* sp.) and 18.2% (*K. ambigua*), and repetitive elements ranging from 1.3% (*Paralemanea* sp.) to 3.0% (*B. macrosporum*) of the assembled genome.

Batrachospermales cp genomes were compact, and their intergenic regions accounted for less than 20% of their length (Table 1). Short dispersed repetitive elements (palindromes and inverted repeats) accounted for 1%–3% of Batraschorpermales cp genomes (Table 1). The sequenced cp genomes were gene-dense (high gene content for the genome size, and reduced intergenic regions) and encoding for 219 (*K. ambigua*) to 225 genes (*B. viride-brasiliense*), whereas 205 genes were shared among all the cp genomes (Table S2). They harbored three rRNA genes (*rnl–rns–rnz* gene cluster) and one noncoding RNA (*rnpB*). Among all the genes predicted in the sequenced cp genomes, only one intron (group II intron) was present in a conserved position in the light-independent protochlorophyllide oxidoreductase subunit B gene (*chl*B). The repertoire of Batrachospermales cp genes was largely shared with most of the cp genomes of Florideophyceae sequenced so far, and 134 genes were in common with 20 additional Florideophyceae representatives of the major lineages in this clade and the Bangiophyceae *Porphyra purpurea* (Table S2).

In terms of gene loss, the most remarkable finding was the lack of *apc*F in all sequenced taxa of Batrachospermales (Table S2). In addition, the following genes that are usually found in other Florideophyceae were absent in the freshwater orders Batrachospermales and Thoreales (Table S2): the transcriptional regulatory protein *omp*R and two conserved genes of unknown function of the *ycf* family (*ycf*34 and *ycf*46). In contrast, the gene *ycf*27, encoding a putative transcriptional regulator, that is present in only four other taxa of Florideophyceae, was also found in these two freshwater orders.

Rearrangements in the Florideophyceae cp genome. Whole-genome alignments revealed the absence of rearrangements among the sequenced Batrachospermales cp genomes, as indicated by DCJ values (Table S3 in the Supporting Information). A single collinear block was predicted with the MAUVE algorithm for whole-genome alignments (Fig. 1). When additional 20 cp genomes representative of the major clades in Florideophyceae and the Bangiophyceae *Porphyra purpurea* were added to the analysis, a maximum of eight DCJ rearrangements could be counted among *P. purpurea* and representatives

TABLE 1. Overview of genome metrics of the Batrachospermales plastid genomes sequenced in this study.

| Species	GC%	Genome size (nt)	# protein-coding genes	# YCFs	# ORFs	# tRNAs	# Introns*	% noncoding DNA	% repetitive elements	Coverage (mean; max; min)
Batrachospermum viride-brasiliense	28.2	171,722	172	23	5	30	1	18.0	1.5	961; 5,343; 1
B. macrosporum	28.0	179,687	164	23	8	37	1	14.7	3.0	478; 1,394; 41
Kumanoa ambigua	28.1	183,003	165	26	9	28	1	18.2	2.9	345; 844; 4
K. mahlacensis	29.8	181,361	166	27	7	28	1	15.8	1.6	176; 2,198; 1
Paralemanea sp.	30.5	180,393	167	28	8	28	1	14.2	1.3	364; 1,934; 1
Sheathia arcuata	29.0	182,807	166	27	8	30	1	15.2	1.9	1,708; 3,536; 1
Sirodotia delicatula	29.1	185,555	164	27	8	30	1	17.7	2.8	1,984; 3,485; 460

*All group II introns.
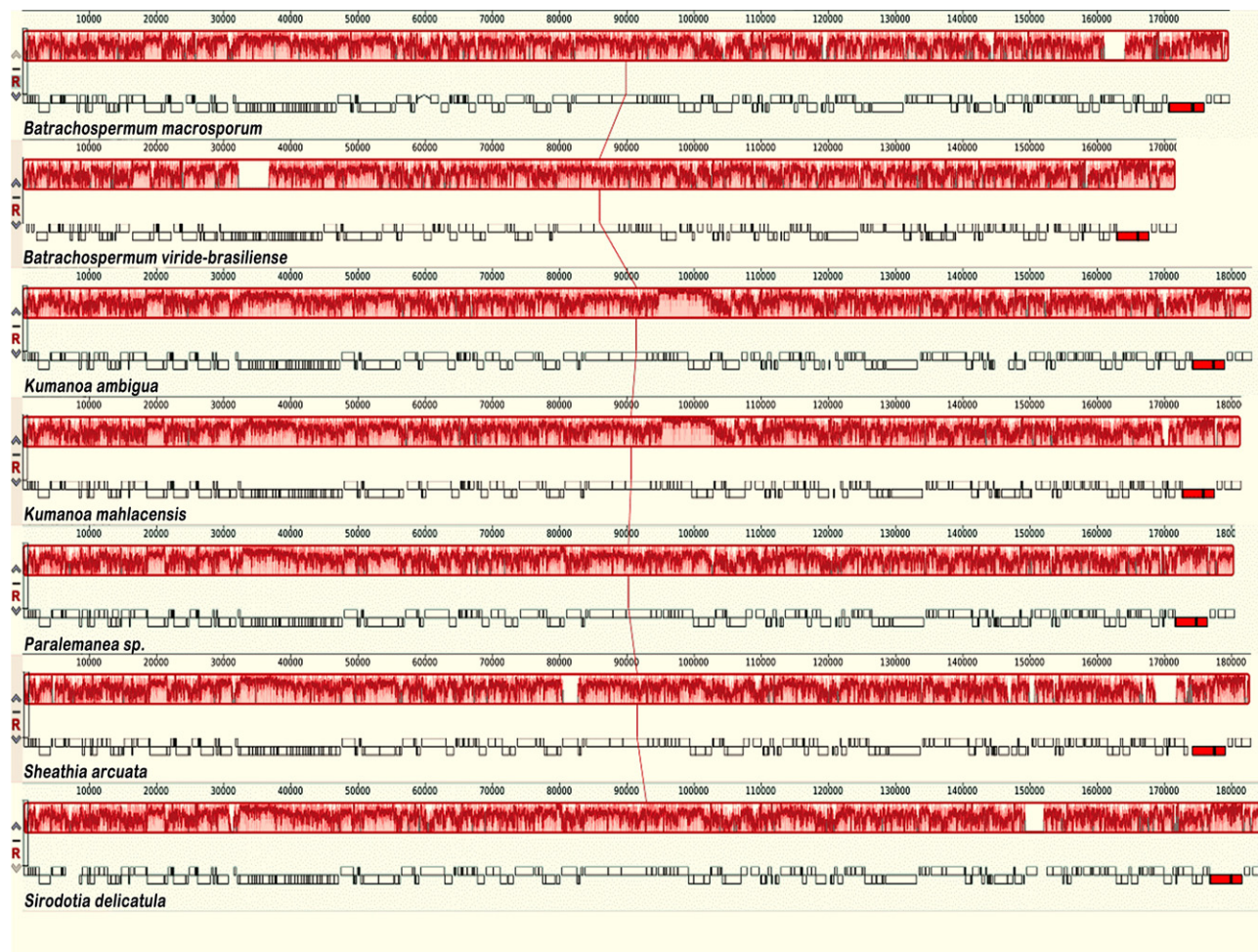
Fig. 1. Cp genome alignment of the Batrachospermales plastid genomes sequenced. Whole-genome MAUVE alignments showing the conserved structure and the collinearity between the Batrachospermales plastid genomes sequenced in this study. A unique locally collinear block (LCB) was recovered, and it is indicated by a corresponding box. Within each LCB, a sequence similarity profile is reported. Annotations are reported below the LCB: protein-coding genes and tRNAs are represented by white boxes, while rRNAs genes are represented by filled (shaded) boxes. The position of the boxes above or below the line refers to the orientation of the gene. [Color figure can be viewed at wileyonlinelibrary.com]

of the other orders, resulting in 17 collinear blocks (Fig. 2B). Within the cp genomes of the representative taxa of Florideophyceae analyzed, the maximum number of DCJ rearrangements was three, observed among the members of the Gracilariales, Halymeniales, Gigartinales, Ceramiales, Sporolithales, Corallinales, and Thoreales (Table S3).

Phylogeny. The ML phylogenetic tree for the cp genome dataset (Fig. 2A) resulted in excellent bootstrap support (100%) at all nodes for the subclass and order-level groups and mixed support for relationships of other levels. Nemaliophycidae formed a distinct clade separated from the other subclasses (Corallinophycidade and Rhodymeniophycidade). Within the Nemaliophycidae, Nemaliales was well resolved as a distinct clade from the one formed by Batrachospermales and Thoreales. Among the freshwater orders, Thoreales appeared as the earliest branching lineage. Within the Batrachospermales,

two groups were recovered, both with full support, one formed by *Paralemanea, Sheathia,* and *Sirodotia,* and another one formed by *Batrachospermum* spp. and *Kumanoa* spp. While *B. macrosporum* was characterized by a long branch, the three *Kumanoa* species were closely related taxa.

DISCUSSION

Most features of the cp genomes of the Batrachospermales species sequenced in this study are similar to those reported for other species of Florideophyceae (Janouškovec et al. 2013, Costa et al. 2016, Díaz-Tapia et al. 2017). Batrachospermales cp genomes are large, gene-dense and have highly conserved gene order. However, some relevant new features were revealed in our study. Genome sizes in members of the Batrachospermales (171,722–185,555 nt) are near the lowest
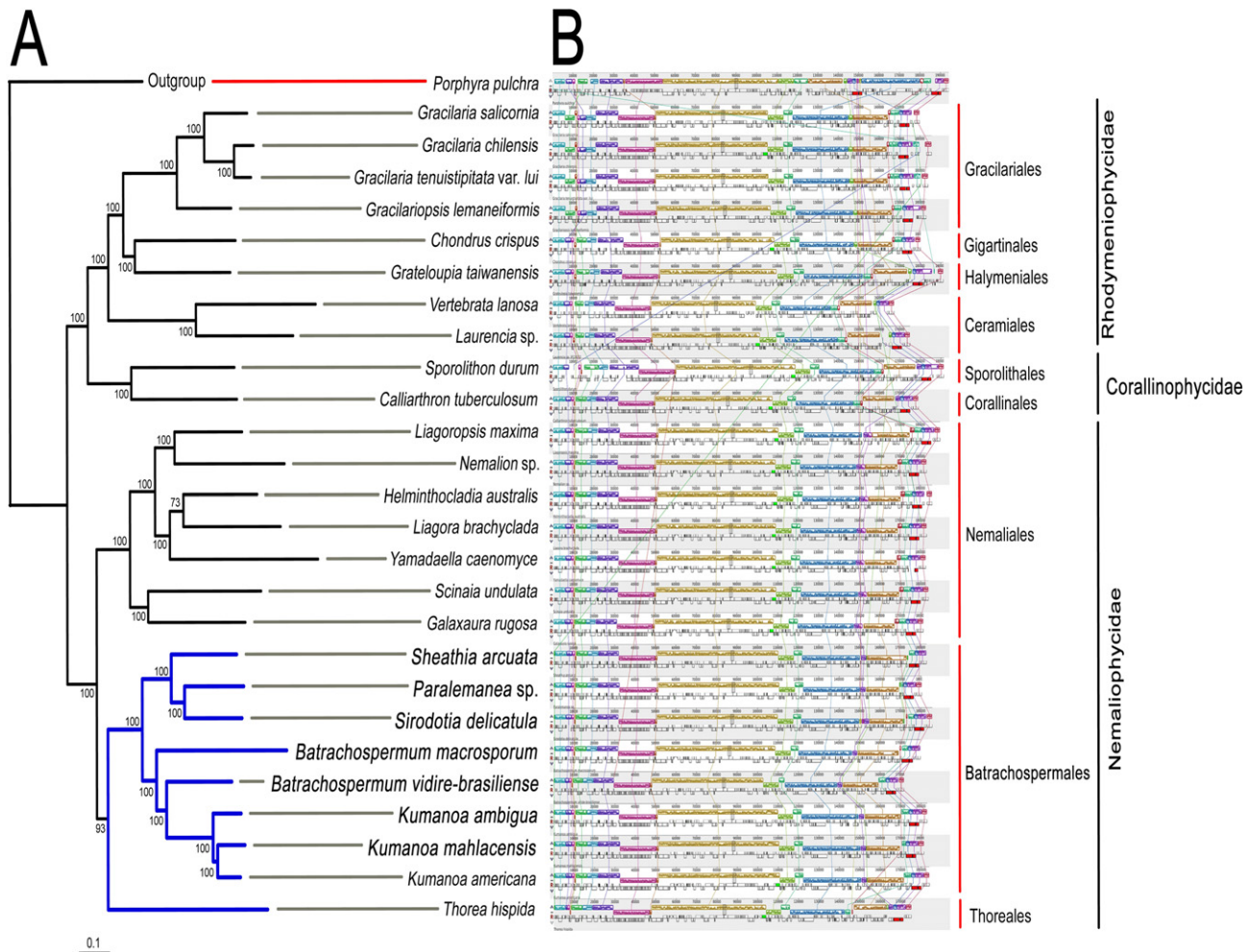
FIG. 2. Phylogenetic relationships and plastid genome rearrangements between the major representatives of Florideophyceae. (A) Phylogenetic ML tree of major orders of Florideophyceae inferred from the concatenated alignment of 177 chloroplast genes. The tree was rooted on *Porphyra pulcra*. Freshwater orders are represented by thicker lines. (B) Whole-genome MAUVE alignments showing the conserved structure and the collinearity between the red algae plastid genomes analyzed. Locally collinear blocks (LCBs) are indicated by corresponding boxes. Within each LCB, a sequence similarity profile is reported. Annotations are reported below the LCBs: protein-coding genes and tRNAs are represented by white boxes, while rRNAs genes are represented by filled (shaded) boxes. The position of the boxes above or below the line refers to the orientation of the gene. [Color figure can be viewed at wileyonlinelibrary.com]

ranges reported for the Florideophyceae (171,392–194,153 nt; Costa et al. 2016, Lee et al. 2016 and references therein). The differences in cp genome size within the order are high (~13,800 bp) in comparison with other orders where more than one member was sequenced: Ceramiales, Gelidiales, Gracilariales, Hildenbrandiales, and Nemaliales (4,000–8,000 nt; Costa et al. 2016, Lee et al. 2016, Díaz-Tapia et al. 2017). In terms of protein-coding genes, members of the Batrachospermales have the lowest gene content (ranging from 164 to 172) among the Florideophyceae (184–235; Costa et al. 2016, Lee et al. 2016), with only the exception of *K. americana* (198 genes; Lee et al. 2016). Cp genome organization within red algae was shown to be very conserved in comparison with the complex history of rearrangements in green algal cp

(Janouškovec et al. 2013, Lam et al. 2015, Leliaert and Lopez-Bautista 2015, Melton et al. 2015, Turmel et al. 2015, Costa et al. 2016). Organization of cp genomes in the Batrachospermales both from our data and a previous study (Lee et al. 2016) follows this pattern. Unlike the typical chloroplast genome quadripartite structure in higher plants, with a large and a small single-copy region separated by two inverted repeats (Green 2011), the cp genome of Batrachospermales did not contain long inverted repeats, similar to other sequenced Florideophyceae (Verbruggen and Costa 2015, Costa et al. 2016). Several short inverted repeats of 15–30 bp were present at the 3′ terminus of genes in members of the Batrachospermales, and they could serve as RNA-processing signals (Stern and Gruissem 1987, Jiao et al. 2004).

Lee et al. (2016) analyzed a representative set of red algal cp genomes to generate a highly resolved multigene tree for the phylum Rhodophyta. They found that most florideophycean species, except Hildenbrandiophycidae, had three different cp genome architectures, named R1 (Rhodophyta-type 1), R2, and R3 type. Differences between the R1- and R2-type plastid genome architectures encompassed one inversion between the rDNA operons, whereas two inversions were found between R1 and R3 types, and three inversions were found between the R2 and R3 types (Lee et al. 2016). R1 and R2 types co-occurred in multiple places in the tree, even within highly supported monophyletic clades. With particular interest to this study, they found that within the well-supported Nemaliophycidae, the freshwater order Thoreales (*Thorea*) had a R2 type, whereas a R1 type was present in the Batrachospermales (*Kumanoa*). In addition, they reported that the cp genome of Nemaliophycidae had unique characteristics among the Florideophyceae, in that they are the only members of this class with *chl*L and *chl*N genes, encoding the light-independent protochlorophyllide oxidoreductase subunits L and N, respectively.

In terms of gene loss, it is important to note the lack of *apc*F in all sequenced taxa of Batrachospermales, which encodes the allophycocyanin beta subunit, a phycobilin component of the light-harvesting proteins characteristic of red algae (Apt and Grossman 1993). Costa et al. (2016) also reported the lack of a gene (*pbs*A) involved in the production of phycobilins in the Nemaliales taxa, which has also been lost in several Rhodomelaceae taxa. The loss of the *apc*F gene has never been reported in florideophyte cp: Its lack in the Batrachospermales can be interpreted as either an actual loss or a gene could have been transferred to the nucleus. Chang et al. (2015) reported an *apc*F mutant, where the function of *apc*F was possibly replaced by another allophycocyanin (*apc*B). Thus, an alternative hypothesis is that *apc*F could have been replaced by another type of allophycocyanin. Indeed, 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. At this stage, neither of these possibilities can be confirmed without more representative genomic data available.

Our cp phylogenomics approach recovered part of the phylogeny within the order. Two groups were recovered, and the genus *Kumanoa* (the most species rich in the order and with three cp genomes described) confirmed as a clade. The single family scheme proposed within the Batrachospermales by Entwisle et al. (2009) could not be refined due to the relatively limited set of taxa studied. Associations with morphological traits within the order were not possible at this stage, since genera with similar vegetative morphology, for example, *Batrachospermum,*

Kumanoa, Sirodotia, and *Sheathia,* were placed in distinct clades. For reproductive characters, no clear grouping were evident, with taxa like *Batrachospermum macrosporum* and *Sheathia arcuata,* which share the presence of straight and long carpogonial branches and stalked carposporophytes (Necchi 2016), positioned in different clades. A more representative set of taxa should be analyzed to propose a sound and universal infra-ordinal classification for the Batrachospermales.

The phylogenetic relationships between the two freshwater orders within the subclass Nemaliophycidae remain poorly understood. Yang et al. (2016) showed these two orders as part of a large clade, with Batrachospermales diverging earlier (~330 mya) from a highly diverse group including the Thoreales and several orders including freshwater and marine representatives (e.g., Acrochaetiales, Balbianiales, and Nemaliales). These results strongly contradict the statements by Nan et al. (2017) that Batrachospermales and Thoreales are derived from the marine relative *Palmaria palmata* (Palmariales) ~415–484 mya. We therefore advocate that the inclusion of more taxa in Nemaliophycidae is critical to properly approach the issue in terms of the evolutionary transition from marine to freshwater habitats in the freshwater lineages including Batrachospermales, Thoreales, and Balbianiales.

CONCLUSIONS

Our study contributes toward the knowledge of cp genome evolution in Batrachospermales by the analysis of seven species. The phylogeny based on concatenated nucleotide genome data roughly supports the current taxonomic system for the order, but a more representative number of taxa should be studied to refine the infra-ordinal classification. We demonstrate that all cp genomes are large, gene-dense and have highly conserved gene order, with only a few cases of gene loss. Comparative analyses confirm previous studies of members of Florideophyceae that indicate cp DNA in the Batrachospermales is highly conserved, with little variation in gene composition. However, some relevant new features were revealed in our study: Genome sizes in members of the Batrachospermales are near the lowest limit reported for the Florideophyceae; the differences in cp genome size within the order are high in comparison with other orders where more than one member was sequenced (Ceramiales, Gelidiales, Gracilariales, Hildenbrandiales, and Nemaliales); members of the Batrachospermales have the lowest number of protein-coding genes (ranging from 164 to 172) among the Florideophyceae. We also reinforce that the relationship between the freshwater orders Batrachospermales and Thoreales within the Nemaliophycidae remains poorly understood.

Apt, K. E. & Grossman, A. R. 1993. The phycobilissome β subunit gene of allophycocyanin is located on the plastid genome in *Aglaothamnion neglectum* (Rhodophyta) and cotranscribed with an underscribed open reading frame. *J. Phycol.* 29:716–8.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19:455–77.

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. 2013. GenBank. *Nucleic Acids Res.* 41:36–42.

Butterfield, N. 2000. *Bangiomorpha pubescens* n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. *Paleobiology* 26:386–404.

Chang, L., Liu, X., Li, Y., Liu, C. C., Yang, F., Zhao, J. & Sui, S. F. 2015. Structural organization of an intact phycobilissome and its association with photosystem II. *Cell Res.* 25:726–37.

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 Evol. Biol.* 16:205.

Darling, A. E., Mau, B. & Perna, N. T. 2010. ProgressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* 5:e11147.

Díaz-Tapia, P., Maggs, C. A., West, J. A. & Verbruggen, H. 2017. Analysis of chloroplast genomes and a supermatrix inform reclassification of the Rhodomelaceae (Rhodophyta). *J. Phycol.* 53:920–37.

Doyle, J. J. & Doyle, J. L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19:11–5.

Entwisle, T. J., Vis, M. L., Chiasson, W. B., Necchi, O. Jr & Sherwood, A. R. 2009. Systematics of the batrachospermales (Rhodophyta): a synthesis. *J. Phycol.* 45:704–15.

Garbary, D. J. & Gabrielson, P. W. 1990. Taxonomy and evolution. *In* Cole, K. M. & Sheath, R. G. [Eds.] *Biology of the Red Algae.* Cambridge University Press, New York, pp. 477–98.

Glazer, A. N., Chan, C. F. & West, J. A. 1997. An unusual phycocyanobilin-containing phycoerythrin of several bluish-colored, acrochaetioid, freshwater red algal species. *J. Phycol.* 33:617–24.

Green, B. R. 2011. Chloroplast genomes of photosynthetic eukaryotes. *Plant J.* 66:34–44.

Guiry, M. D. & Guiry, G. M. 2017. *AlgaeBase.* World-wide electronic publication, National University of Ireland, Galway. Available at: http://www.algaebase.org (accessed 03 March 2017).

Gurgel, C. F. D. & Lopez-Bautista, J. 2007. *Red Algae. In Encyclopedia of Life Sciences.* John Wiley & Sons, Chichester.

Janouškovec, J., Shao-Lun, L., Martone, P. T., Carré, W., Leblanc, C., Collén, J. & Keeling, P. J. 2013. Evolution of red algal plastid genomes: ancient architectures, introns, horizontal gene transfer, and taxonomic utility of plastid markers. *PLoS ONE* 8:e59001.

Jiao, H. S., Hicks, A., Simpson, C. & Stern, D. B. 2004. Short dispersed repeats in the *Chlamydomonas* chloroplast genome are collocated with sites for mRNA 3′ end formation. *Curr. Genet.* 45:311–22.

Johnston, E. T. J., Lim, P. E., Buhari, N., Keil, E. J., Djawad, M. I. & Vis, M. L. 2014. Diversity of freshwater red algae (Rhodophyta) in Malaysia and Indonesia from morphological and molecular data. *Phycologia* 53:329–41.

Kumano, S. 2002. *Freshwater Red Algae of the World.* Biopress, Bristol, 375 pp.

Lam, D. W., Verbruggen, H., Saunders, G. W. & Vis, M. L. 2015. Multigene phylogeny of the red algal subclass Nemaliophycidae. *Mol. Phylogenet. Evol.* 94:730–6.

Lee, J., Cho, C. H., Park, S. I., Choi, J. W., Song, H. S., West, J. A., Battacharya, D., et al. 2016. Parallel evolution of highly conserved plastid genome architecture in red seaweeds and seed plants. *BMC Biol.* 14:75.

Leliaert, F. & Lopez-Bautista, J. M. 2015. The chloroplast genomes of *Bryopsis plumosa* and *Tydemania expeditiones* (Bryopsidales, Chlorophyta): compact genomes and genes of bacterial origin. *BMC Genom.* 16:204.

Leliaert, F., Tronholm, A., Lemieux, C., Turmel, M., DePriest, M., Battacharya, D., Karol, K. G., et al. 2016. Chloroplast phylogenomic analyses reveal the deepest-branching lineage of the Chlorophyta, Palmophyllophyceae class. nov.. *Sci. Rep.* 6:25367.

Lohse, M., Drechsel, O., Kahlau, S. & Bock, R. 2013. OrganellarGenomeDRAW–a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res.* 41:1–7.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., et al. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1:18.

Melton, J. T., Leliaert, F., Tronholm, A. & Lopez-Bautista, J. M. 2015. The complete chloroplast and mitochondrial genomes of the green macroalga *Ulva* sp. UNA00071828 (Ulvophyceae, Chlorophyta). *PLoS ONE* 12:e0184196.

Michel, F., Kazuhiko, U. & Haruo, O. 1989. Comparative and functional anatomy of group II catalytic introns- a review. *Gene* 82:5–30.

Michel, F. & Westhof, E. 1990. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* 216:585–610.

Muller, K. M., Sherwood, A. R., Pueschel, C. M., Gutell, R. R. & Sheath, R. G. 2002. A proposal for a new red algal order, the Thoreales. *J. Phycol.* 38:807–20.

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. *Curr. Biol.* 27:1677–84.

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. *Sci. Rep.* 7:2934.

Necchi Jr., O. 2016. Red algae (Rhodophyta) in rivers. *In* Necchi Jr., O. [Ed.] *River Algae.* Springer, Cham, pp. 65–91.

Pueschel, C. M. & Cole, K. M. 1982. Rhodophycean pit plugs: an ultrastructural survey with taxonomic implications. *Am. J. Bot.* 69:703.

Rice, P., Longden, I. & Bleasby, A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* 16:276–7.

Salomaki, E. D., Nickles, K. R. & Lane, C. E. 2015. The ghost plastid of *Choreocolax polysiphoniae.* *J. Phycol.* 51:217–21.

Saunders, G. W. & Hommersand, M. H. 2004. Assessing red algal supraordinal diversity and taxonomy in the context of contemporary systematic data. *Am. J. Bot.* 91:1494–507.

Sheath, R. G., Muller, K. M., Vis, M. L. & Entwisle, T. J. 1996. A re-examination of the morphology, ultrastructure and

classification of genera in the Lemaneaceae (Batrachos-permales, Rhodophyta). *Phycol. Res.* 44:233–46.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–3.

Stern, D. B. & Gruissem, W. 1987. Control of plastid gene expression: 3′ inverted repeats act as mRNA processing and stabilizing elements, but do not terminate transcription. *Cell* 51:1145–57.

Turmel, M., Otis, C. & Lemieux, C. 2015. Dynamic evolution of the chloroplast genome in the green algal classes Pedinophyceae and Trebouxiophyceae. *Genome Biol. Evol.* 7:2062–82.

Verbruggen, H. & Costa, J. F. 2015. The plastid genome of the red alga *Laurencia*. *J. Phycol.* 51:586–9.

Vis, M. L., Necchi, O., Chiasson, W. B. & Entwisle, T. J. 2012. Molecular phylogeny of the genus *Kumanoa* (Batrachospermales, Rhodophyta). *J. Phycol.* 48:750–8.

Vis, M. L. & Sheath, R. G. 1992. Systematics of the freshwater red algal family Lemaneaceae in North America. *Phycologia* 31:164–79.

Wyman, S. K., Jansen, R. K. & Boore, J. L. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20:3252–5.

Yang, E. C., Boo, S. M., Battacharya, D., Saundres, G. W., Knoll, A. H., Fredericq, S., Graf, L., et al. 2016. Divergence time estimates and the evolution of major lineages in the florideophyte red algae. *Sci. Rep.* 6:21361.

Yoon, H. S., Müller, K. M., Sheath, R. G., Ott, F. D. & Bhattacharya, D. 2006. Defining the major lineages of red algae (Rhodophyta). *J. Phycol.* 42:482–92.

Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31:3406–15.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web site:

Figure S1. Gene map of the chloroplast genome of *Batrachospermum macrosporum*. Genes shown on the outside of the circle are transcribed counterclockwise. Annotated genes are colored according to the functional categories shown in the legend bottom left.

Figure S2. Gene map of the chloroplast genome of *Batrachospermum viride-brasiliense,* Genes shown on the outside of the circle are transcribed counterclockwise. Annotated genes are colored according to the functional categories shown in the legend bottom left.

Figure S3. Gene map of the chloroplast genome of *Kumanoa ambigua*. Genes shown on the outside of the circle are transcribed counterclockwise. Annotated genes are colored according to the functional categories shown in the legend bottom left.

Figure S4. Gene map of the chloroplast genome of *Kumanoa mahlacensis*. Genes shown on the outside of the circle are transcribed counterclockwise. Annotated genes are colored according to the functional categories shown in the legend bottom left.

Figure S5. Gene map of the chloroplast genome of *Paralemanea* sp. Genes shown on the outside of the circle are transcribed counterclockwise. Annotated genes are colored according to the functional categories shown in the legend bottom left.

Figure S6. Gene map of the chloroplast genome of *Sheatia arcuata*. Genes shown on the outside of the circle are transcribed counterclockwise. Annotated genes are colored according to the functional categories shown in the legend bottom left.

Figure S7. Gene map of the chloroplast genome of *Sirodotia delicatula*. Genes shown on the outside of the circle are transcribed counterclockwise. Annotated genes are colored according to the functional categories shown in the legend bottom left.

Table S1. Collection details and sequencing methods of Batrachospermales species analyzed in this study.

Table S2. Comparison of chloroplast gene content in the plastid genomes included in the phylogenetic analysis.

Table S3. DCJ values for plastid genomes analyzed, calculated as rearrangement distances based on N-way LCBs.