Cymbella

Vol. 2. No. 2 (2016)

Revista de investigación y difusión sobre algas

Complete chloroplast genome for *Caulerpa racemosa* and compartive analyses of siphonous green seaweeds plastomes. Contribuciones a la ficología. Dr. Rafael Riosmena R.

Sociedad Mexicana de Ficología
www.somfico.org

DIRECTORIO

CRÉDITO DE FOTO DE LA PORTADA

Costra de una Rhodophyta costrosa con un erizo.
Ambiente intermareal en Akumal, Quintana Roo.
Foto de A. Sentíes.

CINTILLO LEGAL

Cymbella Revista de investigación y difusión sobre algas. Vol. 2, Núm. 2, mayo-agosto de 2016, es una publicación cuatrimestral editada por la Universidad Nacional Autónoma de México, Ciudad Universitaria, Delegación Coyoacán, C.P. 04510, México D.F. a través del Laboratorio de Algas Continentales. Ecología y Taxonomía de la Facultad de Ciencias, Circuito exterior s/n, Ciudad Universitaria, Col. Copilco, Del. Coyoacán, C.P. 04510, Ciudad de México, Tel. (55) 56225430, http:// cymbella.mx/, enm@ciencias.unam.mx. Editor responsable: Dr. Eberto Novelo Maldonado. Reserva de Derechos al Uso Exclusivo: 04-2016-112410454200-203. ISSN: en trámite. ISSN: en trámite. Responsable de la última actualización de este número, Laboratorio de Algas Continentales. Ecología y Taxonomía de la Facultad de Ciencias, Dr. Eberto Novelo Maldonado, Circuito exterior s/n, Ciudad Universitaria, Col. Copilco, Del. Coyoacán, C.P. 04510, Ciudad de México, fecha de la última modificación, 14 de diciembre de 2016.

Complete chloroplast genome for *Caulerpa racemosa* (Bryopsidales, Chlorophyta) and comparative analyses of siphonous green seaweed plastomes

Genoma completo del cloroplasto de *Caulerpa racemosa* (Bryopsidales, Chlorophyta) y un análisis comparativo de los plastomas de algas marinas sifonáceas.

Daryl W. Lam[1], Juan M. Lopez-Bautista[1*]

[1]Department of Biological Sciences, The University of Alabama, Tuscaloosa, Alabama, United States of America
*E-mail: jlopez@ua.edu

ABSTRACT

The green algal order Bryopsidales is mostly comprised of large conspicuous siphonous seaweeds. From this green algal order, a *de novo* chloroplast genome was sequenced for *Caulerpa racemosa*. The plastid genome was circular and lacked the inverted repeat commonly found in vascular green plants. The *C. racemosa* genome was 176,522 base pairs long and represents the largest plastid genome currently known for the Bryopsidales. Comparative genomic and phylogenetic analyses were performed with the addition of previously published bryopsidalean plastome data. Overall, the genome contained a similar gene complement to other bryopsidalean species. However, *C. racemosa* was missing the *ycf*47 gene that encodes for P-P-bond-hydrolysis-driven amino acid involved in protein translocation across the thylakoid membrane. Phylogenomic analysis, based on a 50-gene dataset, supported the current taxonomy for *C. racemosa* and was similar to previously published Bryopsidales phylogenies. Specifically, the monophyly of suborders Bryopsidineae and Halimedineae was strongly supported. Mauve based synteny analyses suggested several genomic rearrangement events in both suborders. Overall, there were more rearrangements in the Halimedineae as compared to the Bryopsidineae. When compared to other species of this order, the larger genome size of *C. racemosa* was due to more abundant and longer introns, more intergenic space, and the presence of large open reading frames (ORFs). Several of these relatively large ORFs are potentially from a horizontal gene transfer event from bacteria. Bacterial-related ORFs include several methyl-transferases, restriction endonuclases, and a DNA polymerase. Many of the aforementioned bacterial genes were found to be present in other green algal species and may represent pleisomorphic horizontal gene transfer events from bacterial to the plastid genomes to these siphonous marine plants.

Keywords: Bryopsidales, chloroplast genome, genetic analyses, plastomes, siphonous algae

RESUMEN

El orden de algas verdes Bryopsidales está compuesto principalmente por organismos sifonales conspicuos y grandes. De este orden de algas verdes fue secuenciado un genoma de cloroplasto *de novo* de *Caulerpa racemosa*. El genoma plastidial fue circular y careció de la repetición invertida que se ha encontrado comúnmente en las plantas verdes vasculares. El genoma de *C. racemosa* fue de 176,522 pares de bases de longitud y representa el genoma más grande de plástidos actualmente conocido

para las Bryopsidales. Se realizaron análisis genómicos comparativos y filogenéticos con la adición de los datos previamente publicados de plastomas bryopsidaleanos. En general, el genoma contuvo un complemento de genes similar a otras especies del grupo. Sin embargo, a *C. racemosa* le faltó el gen *ycf*47 que codifica la unión P-P inducida por hidrólisis de aminoácidos implicada en la translocación de proteínas través de la membrana del tilacoide. Los análisis filogenómicos, basado en un conjunto de datos de 50 genes, apoyaron la taxonomía actual de *C. racemosa* y fue similar a filogenias de Bryopsidales previamente publicadas. Específicamente, la monofilia de subórdenes Bryopsidineae y Halimedineae fue fuertemente soportada. Los análisis de Mauve basados en sintenia sugieren varios eventos de reordenamiento genómico en ambos subórdenes. En general, hubo más reordenamientos en Halimedineae en comparación con el Bryopsidineae. Cuando se compara con otras especies de este orden, el gran genoma de C*. racemosa* se distingue por intrones más abundantes y más largos, más espacio intergénico y la presencia de grandes arreglos de lectura abiertos (ORF). Varios de estos relativamente grandes ORFs son potencialmente de un evento de transferencia horizontal de genes de bacterias. ORFs relacionados con las bacterianas incluyen varias metil-transferasas, endonucleasas de restricción y un ADN polimerasa. Muchos de los genes bacterianos mencionados se encontraron que están presente en otras especies de algas verdes y puede representar eventos de transferencia horizontal de genes pleisomórficos de las bacterias a los genomas de plástidos de algas marinas sifonáceas.
Palabras clave: algas sifonales, análisis genéticos, Bryopsidales, genoma del cloroplasto, plastomas

INTRODUCTION

The division Chlorophyta is a species rich group strictly comprised of green algal organisms that contain chlorophylls a and b (Graham *et al.* 2009). The Chlorophyta lineage is sister to the Streptophyta lineage, a plant division that includes some green algae, mosses, liverworts, hornworts, and vascular plants (Lewis & McCourt 2004; Leliaert *et al.* 2012). The split between the two lineages has been estimated to have occurred ~936 million years ago (Becker 2013; Parfrey *et al.* 2011). Although not as well studied as the Streptophyta, chlorophytan species exhibit a vast array of morphological and ecological diversity. They exhibit a wide array of morphologies, a few examples include: single celled organisms that form palmelloid colonies (Zechman *et al.* 2010), flagellated free-living organisms (Pröschold *et al.* 2001), multicellular macrophytic blades (Melton *et al.* 2015), filamentous thalli with a single nucleus per cell (Rindi & Lopez-Bautista 2009), filamentous thalli with multiple nuclei per cell (Leliaert *et al.* 2007), calcified uninuclear siphons (Olsen *et al.* 1994), and multinuclear siphons (Lam & Zechman 2006). The Chlorophyta are ubiquitous and can be found in marine, freshwater, and terrestrial habitats (Graham *et al.* 2009). Recent systematic assessments of this division based on phylogenomic analyses inferred a large well-supported monophyletic group, which included taxa in the classes Chlorodendrophyceae, Chlorophyceae, Pedinophyceae, Trebouxiophyceae, and Ulvophyceae and was branded as the "core Chlorophyta" clade (Fučíková *et al.* 2014).

Within the "core Chlorophyta", the order Byopsidales (class Ulvophyceae) is comprised of 564 marine species and a single freshwater species all with siphonous thalli (Guiry & Guiry 2015). In other words, the vegetative thallus is a single undivided cell, which contains multiple nuclei that arose from mitosis without cytokinesis (Leliaert *et al.* 2015). A systematic revision of the Bryopsidales erected two suborders, the Bryopsidineae (including the genera *Bryopsis*, *Codium*, and *Derbesia*) and the Halimedineae (*Caulerpa*, *Halimeda*, and *Udotea*), and was based on morphological attributes within the grouping (Hillis-Colinvaux 1984). Based on Hillis-Colinvaux's assessment (Hillis-Colinvaux 1984) there are several important morphological features that distinguish the two suborders. The Halimedineae have two types of plastids (heteroplasty), amyloplasts and chloroplast, while Bryopsidineae species have only chloroplasts (homoplasty). Sexual reproduction characteristics also differ in the two suborders. The Bryopsidineae reproduce with septa separating the reproductive cells from the rest of the thallus, while the Halimedineae lack septa and reproduce in a holocarpic manner where most of the vegetative thallus is filled with gametes (Hillis-Colinvaux 1984). After the release of gametes, the greater majority of the protoplasm is lost with an empty "ghost thallus" remaining (Kooistra 2002). It should be noted that *Caulerpella* is an exception to this rule as the genus has been reported to be nonholocarpic (Fama *et al.* 2002). Molecular phylogenetic analyses based on the large subunit of the plastid-encoded gene RuBisCO supported the monophyly of both suborders and placed the cryptic genus *Pseudocodium*, which superficially looks like *Codium* but exhibits heteroplasty with Halimedineae taxa (Lam & Zechman 2006). Subsequent phylogenies using the same

molecular marker inferred that the marine genera *Avrainvillea* and *Cladocephalus* formed a robustly supported clade along with the freshwater species *Dichotomosiphon tuberosus* (Curtis *et al.* 2008). A time-calibrated multilocus (five plastid markers and one nuclear locus) phylogeny inferred the separation of the two suborders to have occurred in the Early Paleozoic, approximately 465 million years ago (Verbruggen *et al.* 2009).

Clearly, molecular based systematic analyses have increased our understanding of bryopsidalean evolutionary processes. Recent technological advances in high throughput DNA sequencing currently offer a wealth of gene data at a fraction of the cost of Sanger-based sequencing methods. Because chloroplasts are abundant in most macrophytes and each individual chloroplast contains multiple copies of genomic DNA, plastid genomes (plastomes) are a common target for botanical based studies (Lutz *et al.* 2011). The Bryopsidales are not an exception to this trend, with three completed chloroplast genomes currently published (Lü *et al.* 2011; Leliaert & Lopez-Bautista 2015). Within the suborder Halimedineae, the complete chloroplast genome (105,200 base pairs (bp)) of *Tydemania expeditionis* was recently assembled and annotated (Leliaert & Lopez-Bautista 2015). For the suborder Bryopsidineae two species of *Bryopsi*s have been assembled and annotated. *B. hypnoides* (153,429 bp) and *B. plumosa* (106,859 bp) have published plastomes (Lü *et al.* 2011; Leliaert & Lopez-Bautista 2015). All three plastomes were published as circular contigs (Lü *et al.* 2011; Leliaert & Lopez-Bautista 2015).

In order to infer the evolutionary history of this order of green algae, we have sequenced, *de novo* assembled, and performed annotations for the chloroplast genome of one bryopsidalean species, *Caulerpa racemosa*. Although there have been chloroplast genome analyses for C. *sertularoides* (Lehman & Manhart 1997) through Southern hybridization/analysis and restriction fragments analysis and partial plastid genome data (~ 30,000 bp, 23 genes) for *C. filiformis, C. racemosa* represent the first completely sequenced genome for the family Caulerpaceae. Here we present, a phylogenomic inference based on protein-coding plastid genes. In addition, we infer genomic rearrangement events for all taxa in the order with currently completed plastomes.

MATERIALS AND METHODS

Field collection and DNA extraction/sequencing

C. racemosa (voucher UNA00072801) was collected from La Parguera, Isla Magueyes, Enrique, Puerto Rico, USA on November 11, 2014. DWL identified these samples based on the two taxonomic keys (Littler & Littler 2000; Littler *et al.* 2008). DNA was extracted from silica-gel-dried *C. racemosa* using the E.Z.N.A. Plant DNA Extraction Kit (Omega Biotek Norcross, GA USA) following the protcol therein. For this species, a paired-end 101 bp library was constructed using the standard Illumina Truseq adapter. MiSeq DNA sequencing was performed at Cold Spring Harbor Laboratory (Cold Spring Harbor, NY, USA) on a multiplexed sequencing run that contained two other algal species. This run generated 2,862,510 paired-end reads (2 x 101 bp).

Genome assembly

The dataset was downloaded to the Alabama Supercomputer Center's Dense Memory Cluster (DMC). Fastq formatted files were quality checked using the software packages fastQValidator (https://github.com/statgen/fastQValidator) and fastqc (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/). The A5 Assembly pipeline (Tritt *et al.* 2012) was used to remove ambiguous and low quality portions of the raw read data, correct for sequencing errors, assembly, scaffolding, and reassembly using 120 gigabytes of RAM and a single processing core. In order to identify the chloroplast genome from the rest of the data, all contigs from the *de novo* assembly were compared against a custom local dataset comprised of protein coding sequences (CDS) and ribosomal RNA (rRNA) regions from previously published bryopsidalean chloroplast genomes: *Bryopsis hypnoides* (GenBank accession number NC_013359.1), *Bryopsis plumosa* (NC_026795.1), and *Tydemania expeditionis* (NC_026796.1) via MegaBLAST (Morgulis *et al.* 2008) with an E-value threshold of $< 1x10^{-10}$ through Geneious v 7.9 (Biomatters, http://www.geneious.com) using ten processing threads. The linear contigs had identical sequence data (~100 bp) on both ends. Subsequently, one of the identical sequence ends was removed and the contig was circularized in Geneious. In order to edit the nucleotide data for the contigs, the original raw read data was trimmed of low-quality bases and adapters sequences by Trim Galore! (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Trimmed reads were mapped on to the chloroplast contig in Geneious and a majority rule consensus sequences was created.

Gene annotation

Annotations were made in Geneious by mapping the aforementioned previously published gene regions to the *de novo* plastome. For protein-coding

genes open reading frames (ORFs) were identified in Geneious using a bacterial/plant plastid genetic code. Matching annotations were double-checked by BLASTx (Gish & States 1993) against the non-redundant NCBI database. Relatively large ORFs (> 700 bps) that did not match any of the aforementioned gene annotations were are also queried against the NCBI non-redundant database for potential gene homology to other organisms in GenBank. Intron-exon boundaries for protein coding genes were identified by translational alignment in Geneious. For rRNA genes these boundaries were inferred by MAFFT (Katoh & Standley 2013) alignments as implemented in Geneious. Transfer RNA (tRNA) gene predictions were made using tRNAscan-SE (Lowe & Eddy 1997) with the following parameters: search mode set to "Organellar", searching with Cove only (cutoff score = 15), Covariance model tRNA2.cm, max intron + var. length = 40, and pseudogene checking disabled. The *C. racemosa* plastome is available for download as NCBI/GenBank accession KT946602.

Phylogenomic analyses

Maximum likelihood tree based on *atp*A, *atp*B, *atp*E, *atp*F, *atp*H, *atp*I, *clp*P, *inf*A, *pet*A, *pet*B, *pet*G, *psa*A, *psa*B, *psa*C, *psa*J, *psb*A, *psb*B, *psb*C, *psb*D, *psb*E, *psb*F, *psb*H, *psb*I, *psb*J, *psb*K, *psb*L, *psb*N, *psb*T, *rbc*L, *rpl*2, *rpl*5, *rpl*14, *rpl*16, *rpl*20, *rpl*23, *rpl*36, *rps*3, *rps*4, *rps*7, *rps*8, *rps*9, *rps*11, *rps*12, *rps*14, *rps*18, *rps*19, *tuf*A, *ycf*3, *ycf*4, and *ycf*12 genes with a GTR + G model implemented in RAxML (Stamatakis 2014). Poorly aligned regions were masked using Gblocks (Talavera & Castresana 2007). Nodal support values were based on 1000 bootstrap replicates. Prasinophyte taxa were used to root this phylogeny.

From Treebase (http://treebase.org/), the nucleotide alignment from Fučíková *et al.* (2014) was downloaded (study 16203 id M24024). The concatenated dataset was separated into individual alignments by gene. Transcriptome sequence data for the taxon *Codium decorticatum* was removed from this dataset because its nucleotide similarity was nearly identical to published sequences for *Bryopsis plumosa* (Leliaert & Lopez-Bautista 2015). We added corresponding gene data from the following ulvophycean species: *Bryopsis plumosa* (GenBank NC_026795.1), *Caulerpa racemosa* (this study), *Tydemania expeditionis* (NC_026796.1), and *Ulva* sp. (GenBank KP720616.1). As in Fučíková *et al.* (2014), the *psa*M, *rpl*32, and the highly variable *ycf*1 gene were not included in the concatenated dataset. The single gene datasets were realigned with MAFFT and concatenated into a single alignment using Sequence Matrix (Vaidya *et al.* 2011). Hypervariable regions of the alignment were removed from the dataset using the Gblocks (Talavera & Castresana 2007) web server http://molevol.cmima.csic.es/castresana/Gblocks_server.html) allowing for smaller final blocks, gap positions within the final blocks, less strict flanking positions and many contiguous non-conserved positions. Phylogenetic analyses were performed in RAxML version 8.0.24 (Stamatakis 2014) with a GTR+G model with rapid bootstrapping (1000 replicates) and subsequent maximum likelihood search on the Alabama supercomputer DMC with 10 cores and 25 gigabytes of RAM.

Analysis of genomic rearrangements

The *C. racemosa* plastome was aligned using the progressive Mauve algorithm (Darling *et al.* 2004) using the full alignment option and automated calculation of locally co-linear block scores.

RESULTS AND DISCUSSION

For *C. racemosa*, a total of 2,322,843 reads were mapped to the 176,522 bp circular contig with an average coverage per site at 1701.3x and had an overall nucleotide similarity of 99.3% (Fig. 1). There were 76 protein coding genes, 27 tRNA genes, 3 rRNA genes, 18 introns, and 15 ORFs (≥ 700 bps in length) that were annotated to the *C. racemosa* plastid genome. This is currently the largest known plastid genome for the order Bryopsidales. *C. racemosa* was missing the *ycf*47 gene that was found in the bryopsidalean species *Bryopsis plumosa* and *Tydemania expeditionis*. However, it should be noted that the overall gene content for *C. racemosa* is quite similar to that of the previously published bryopsidalean plastomes (Lü *et al.* 2011; Leliaert & Lopez-Bautista 2015). *C. racemosa* contained the *til*S pseudogene at positions 62,978 to 62,908 (reverse orientation). Likewise, the *til*S pseudogene was found in both the *Bryopsis* and *Tydemania* plastomes (Leliaert & Lopez-Bautista 2015).

Since the overall gene content for *C. racemosa* among Bryopsidales is quite similar, the differences in genome size are partially due to variance of intergenetic space and introns. *C. racemosa* has longer intergenetic regions as compared to *Bryopsis plumosa* and *Tydemania expeditionis*. One example of intronic differences is exemplified in the alignment of rrl rRNA gene. The *C. racemosa* rrl gene contained six introns that increased the gene length from ~3 kbp to over 9 kbp when compared to *T. expeditionis*.

In the *C. racemosa* plastome, ORFs (≥ 700 bps in

Figure 1. Gene map of the C*aulerpa racemosa* chloroplast genome using OGDRAW. Genes oriented in the clockwise direction are annotated inside of the circle, while genes in the counterclockwise direction are on the outside of the circle. Genes have been color-coded based on functional categories in the legend (bottom left).

length) are numerous (15 ORFs total). ORFs 4, 5, and 8 when queried against the NCBI nr database were found to be significantly similar (E-values of 6.00 x 10^{-53}, 2.00 x 10^{-155}, and 9.00 x 10^{-53}, respectively) to bacterial DNA methylase proteins. Also in the *C. racemosa* ORF 6 was significantly similar (E-value < 0.00) to a type I bacterial restriction endonuclease. Restriction endonucleases are used in bacteria as a means of defense against foreign DNA (Arber & Linn 1969). These restriction endonucleases digest DNA at specific restriction sites (Arber & Linn 1969). Meanwhile, the bacteria's host genome is protected biochemically by DNA methylation. It is possible that *C. racemosa*'s plastids seem to exhibit a similar defense mechanism against foreign DNA invasion as many prokaryotic organisms. However, this hypothesis still needs further investigation as expression-based evidence (transcriptomics) is necessary to support

this theory. That being said, the most likely means of acquirement is via horizontal gene transfer from bacteria to the host's plastid genome. Multiple bacterial communities have been found inside of the *Caulerpa* thallus (Delbridge *et al.* 2004). Furthermore, ORFs of bacterial origin have been found in the plastomes of *Bryopsis* and *Tydemania* from previously published studies (Leliaert & Lopez-Bautista 2015), suggesting that transfer of bacterial genes into bryopsidalean plastome is fairly common in the order Bryopsidales. However, plastomes for more Bryopsidales species are necessary to elucidate the evolution of bacterial ORFs in these genomes.

Phylogenomic analyses

The 50-gene nucleotide alignment of chlorophytan algae was analyzed via maximum likelihood and the resulting phylogeny is depicted as Fig. 2. This tree was rooted with Prasinophytes (*sensu lato*) taxa. The green algal class Chlorophyceae formed a separate robustly supported (100% bs) clade. However, the monophyly of the classes Trebouxiophyceae and Ulvophyceae were unresolved. In contrast, the order Bryopsidales formed a robustly supported claded (100% bs). Within the order Bryopsidales, the suborders Bryopsidineae and Halmedineae formed separate robustly supported clades (100% bs and 100% bs, respectively), thus supporting the systematic/taxonomic assessment of Hillis-Colinvaux (1984) based on purely morphological observations. In general, the presented phylogenomic analysis is quite similar to previously published trees (Fučíková *et al.* 2014 ; Lam

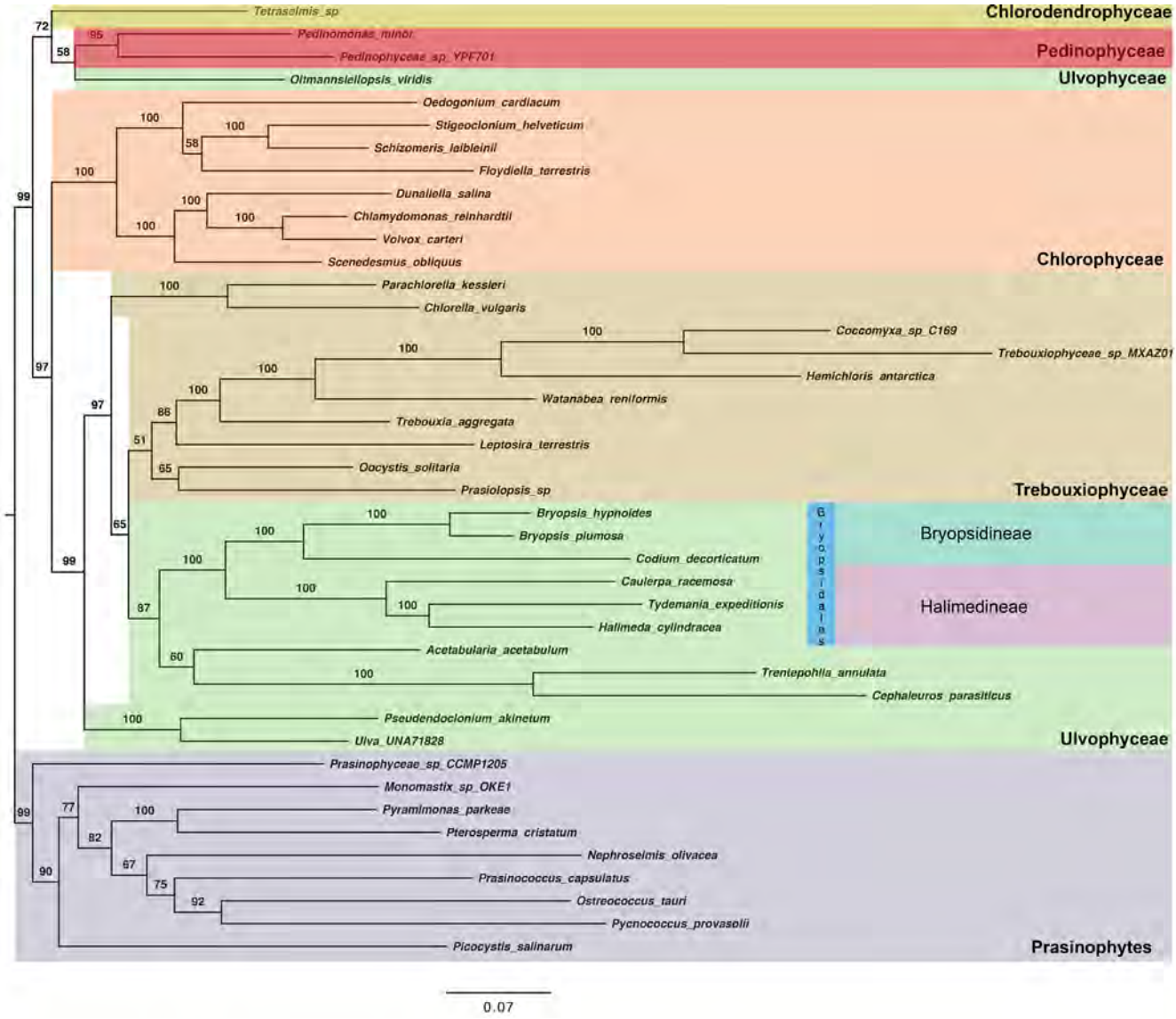

Figure 2. Phylogeny resulted from 50-gene nucleotide alignment of chlorophytan algae using maximum likelihood.

& Zechman 2006; Lieliaert & Lopez-Bautista 2015; Melton *et al.* 2015). This being said, more plastid gene data from a more diverse set of chlorophytan species is required before a robustly supported tree for this division of green algae can be inferred.

Chloroplast genome rearrangements

Based on the inferred monophyly of both suborders in the 50-gene phylogenomic analysis, separate Mauve alignment analyses were conducted for the Bryopsidineae and Halimedineae. These analyses included all currently completed chloroplast genome data for the order Bryopsidales (Lü *et al.* 2011; Leliaert & Lopez-Bautista 2015; and this study). Fig. 3 represents the plastid genomes of the two members of the suborder Bryopsidineae where colored locally collinear blocks (LCB) represent homologous sequence regions that do not contain major rearrangements. These results were quite similar to the results of Leliaert and Lopez-Bautista (2015). The number of genomic rearrangement in the suborder Bryopsidineae is fewer than the rearrangement events inferred for the suborder Halimedineae (Fig. 4). However, this might be an artifact of taxon sampling as Fig. 3 has two species from the same family/genus (Bryopsidaceae/*Bryopsis*), while Fig. 4 has two species from two different families (Caulerpaceae and Udoteaceae).

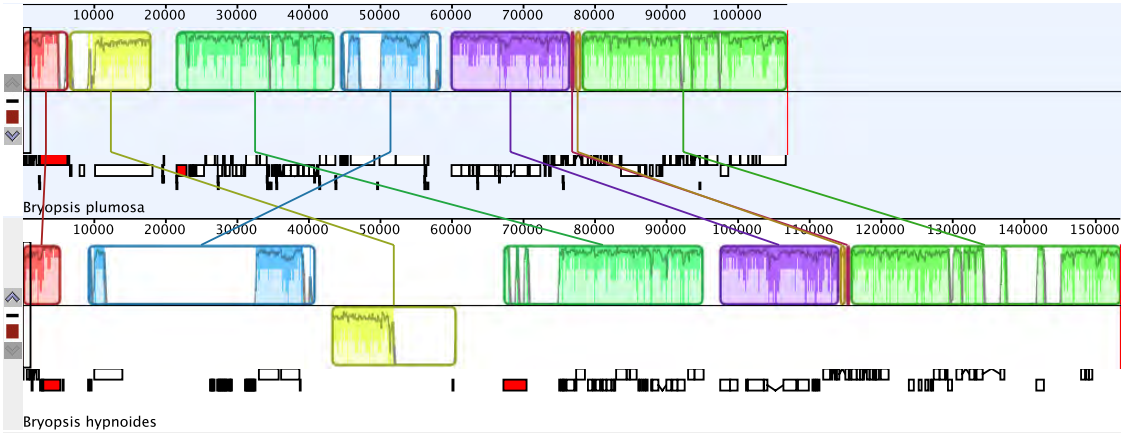

Fig. 3. Whole plastid genome alignments for the Bryopsidineae. The Mauve algorithm (Darling *et al.* 2004) was implemented on the plastomes of *Bryopsis plumosa* (Leliaert & Lopez-Bautista 2015), and *Bryopsis hypnoides* (Lü *et al.* 2011). Corresponding colored boxes represent local collinear blocks (LCB) and represent regions of homology. Inside of each LCB a sequence similarity profile is displayed. Inverted LCB are presented as blocks below the centerline.

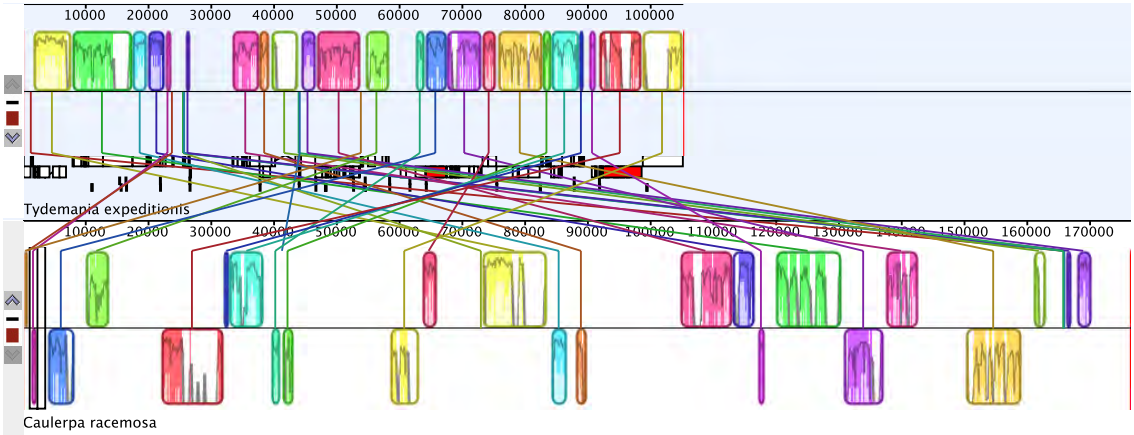

Fig. 4. Whole plastid genome alignments for the Halimedineae. The Mauve algorithm (Darling *et al.* 2004) was implemented on the plastomes of *Caulerpa racemosa* (this study) and *Tydemania expeditionis* (Leliaert & Lopez-Bautista 2015). Corresponding colored boxes represent local collinear blocks (LCB) and represent regions of homology. Inside of each LCB a sequence similarity profile is displayed. Inverted LCB are presented as blocks below the centerline.

CONCLUSIONS

The circular plastid genome of *C. racemosa* (176,522 bp) is currently the largest completely sequenced plastome for the Bryopsidales. For comparison, the complete chloroplast genome *Tydemania expeditionis* is 105,200 bp in length. Completed plastomes of *Bryopsis hypnoides* (153,429 bp) and *B. plumosa* (106,859 bp) have also been published (Lü *et al.* 2011; Leliaert & Lopez-Bautista 2015). Overall the genome contained similar gene complements to previously published bryopsidalean plastomes. The larger genome size of *C. racemosa* was due to differences in number and size of introns, intergenetic spacers, and ORFs. This species had ORFs that were most likely transferred horizontally from bacterial communities. In addition, *C. racemosa* had ORFs that were significantly similar to bacterial methyl-transferases and a restriction endonuclease genes. Phylogenomic inferences based on a 50-gene dataset supported the current taxonomy and previously published phylogenetic studies. Mauve based synteny analyses suggested several genomic rearrangement events for taxa in both suborders, however there were more extensive rearrangement events in the Halimedineae. Although these genomes are a large step toward the elucidation of the chlorophytan evolution, more plastid genomes from more green algal species are necessary in order to fully unravel the history of this green algal division.

ACKNOWLEDGEMENTS

We thank Trey Melton, Ana Tronholm, and David Ward for sample collection. Jeremy McDoniell, Hendrick Adams, and Dustin Steffan assisted DWL with DNA extractions and laboratory protocols. We acknowledge the Alabama Supercomputer Cluster for analysis time on its high performance supercomputer that was used for both short-read assembly and phylogenomic inference. To the College of Arts & Sciences and the Department of Biological Sciences of The University of Alabama for their support. The National Science Foundation through Assembling The Tree of Life for Green Algae GRAToL (DEB 1036495).

REFERENCES

Arber, W., & S. Linn. 1969. DNA modification and restriction. *Annual Review of Biochemistry* 38:467-500.

Becker, B. 2013. Snow ball earth and the split of Streptophyta and Chlorophyta. *Trends Plant Sciences*. 18: 180-183.

Curtis, N.E., C.J. Dawes & S.K. Pierce. 2008. Phylogenetic analysis of the large subunit RUBISCO gene supports the exclusion of *Avrainvillea* and *Cladocephalus* from the Udoteaceae (Bryopsidales, Chlorophyta). *Journal of Phycology* 44: 761-767.

Darling, A.C.E., B.Mau, F.R. Blattner & N.T. Perna. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research.* 14:1394-1403.

Delbridge, L., J. Coulburn, W. Fagerberg & L. Tisa. 2004. Community profiles of bacterial endosymbionts in four species of *Caulerpa*. *Symbiosis* 37:335-334.

Fama, F., B. Wysor, W.H.C.F. Kooistra& G.C. Zuccarello. 2002. Molecular phylogeny of the genus *Caulerpa* (Caulerpales, Chlorophyta) inferred from chloroplast *tuf*A gene. *Journal of Phycology* 38:1040-1050.

Fučíková, K., F. Leliaert, E.D. Cooper, P. Škaloud, S. DiHondt, O. De Clerck, C.F. Gurfel, L.A. Lewis, P.O. Lewis, J.M. Lopez-Bautista, C. Delwiche & H. Verbruggen. 2014. New phylogenetic hypotheses for the core Chlorophyta based on chloroplast sequence data. *Frontiers in Ecology and Evolution 2*. Articule 63. doi: 10.3389/fevo.2014.00063.

Gish, W. & D.J. States.1993. Identification of protein coding regions by database similarity search. *Nature Genetics* 3:266-272.

Graham, L.E., J.M. Graham & L.W. Wilcox. 2009. *Algae*. 2nd ed. Pearson Education, Inc., San Francisco.

Guiry, M.D. & G.M. Guiry. 2015. AlgaeBase. World-wide electronic publication, National University of Ireland, Galway. http://www.algaebase.org

Hillis-Colinvaux, L. 1984. Systematics of the Siphonales. In: D.E.G. Irvine, D.M. John. Eds. *Systematics of the green algae*. Academic Press Inc., London, pp. 271-296.

Katoh, K., & D.M. Standley. 2013. MAFFT multiple sequence alignment software version 7: improvement in performance and usability. *Molecular Biology and Evolution* 30: 772-780.

Kooistra, W.H.C.F. 2002. Molecular phylogenies of the Udoteaceae (Bryopsidales, Chlorophyta reveal nonomonophyly for *Udotea*, *Penicillus*, and *Chlorodesmis*. *Phycologia* 41: 453-462.

Lam, D.W., & F. Zechman. 2006. Phylogenetic analyses of the Bryopsidales (Ulvophyceae, Chlorophyta) based on RuBisCO large subunit gene sequences. *Journal of Phycology* 42: 669-678.

Lehman, R.L. & J.R. Manhart. 1997. A preliminary comparison of restriction fragment patterns in the genus *Caulerpa* (Chlorophyta) and the unique structure of the chloroplast genome of *Caulerpa sertularioides. Journal of Phycology* 33: 1055-1062.

Lelieart, F., O. De Clerck, H. Verbruggen, C. Boedeker & E. Coppejans. 2007. Molecular phylogeny of the Siphonocladales (Chlorophyta: Cladophorophyceae). *Molecular Phylogenetics and Evolution* 44: 1237-1256.

Leliaert, F., D.R. Smith, H. Moreau, M.D. Herron, H. Verbruggen, C.F. Delwiche & O. DeClerck. 2012. Phylogeny

and molecular evolution of the green algae. *CRC Critical Reviews in Plant Sciences* 31: 1-46.

Leliaert, F. & J.M. Lopez-Bautista. 2015. The chloroplast genomes of *Bryopsis plumosa* and *Tydemania expeditionis* (Bryopsidales, Chlorophyta): compact genomes and genes of bacterial origin. *BMC Genomics* 16: 204. doi: 10.1186/s12864-015-1418-3.

Leliaert, F., J.M. Lopez-Bautista & O. De Clerck. 2015. Class Ulvophyceae K.R. Mattox and K.D. Stewart. In: W. Frey. Ed. *Syllabus of Plant Families – A. Engler's Syllabus der Pflanzenfamilien* Part 2/1: *Photoautotrophic eukaryotic Algae*. Schweizerbart, Stuttgart, pp. 247-281

Lewis, L.A. & R. McCourt. 2004. Green algae and the origin of land plants. *American Journal of Botany* 91: 1535-1556.

Littler, D.S. & M.M. Littler. 2000. *Caribbean reef plants: an identification guide to the reef plants of the Caribbean, Bahamas, Florida, and Gulf of Mexico*. D.C. OffShore Graphics, Inc. Washington.

Littler, D.S., Littler, M.M., & M.D. Hanisak.2008. *Submersed plants of the Indian River Lagoon: a floristic inventory and field guide*. Washington, Offshore Graphics, Inc.

Lowe, T.M. & S.R. Eddy. 1997. t-RNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* 25: 955-964.

Lü, F., W. Xu, T. Chao, G. Wang, J. Niu, G. Pan & S. Hu. 2011. The *Bryopsis hypnoides* plastid genome: multimeric forms and complete nucleotide sequence. *PLoS One* 6(2): e14663. doi: 10.1371/journal.pone.0014663

Lutz, K.A., W. Wang, A. Zdepski & T.P. Michael. 2011. Isolation and analysis of high quality nuclear DNA with reduced organellar DNA for plant genome sequencing and resequencing. *BMC Biotechnology* 11:54. doi: 10.1186/1472-6750-11-54

Melton, J.T., F. Leliaert, A. Tronholm, & J.M. Lopez-Bautista. 2015. The complete chloroplast and mitochondrial genomes of the green macroalga *Ulva* sp. UNA00071828 (Ulvophyceae, Chlorophyta). *PLoS One* 10(4): e0121020. doi: 10.1371/journal.pone.0121020

Morgulis, A., G. Coulouris, Y. Rayteselis, T.L. Madden, R. Agarwala & A.A. Schäffer. 2008. Database indexing for production MegaBLAST searches. *Bioinformatics* 24(16):1757-64. doi: 10.1093/bioinformatics/btn322

Olsen, J.L., W.T. Stam, S. Berger & D. Menzel. 1994. 18S rDNA and evolution in the Dasycladales (Chlorophyta): modern living fossils. *Journal of Phycology* 30: 729-744.

Parfrey, L.W., D.J.G. Lahr, A.H. Knoll & L.A. Katz. 2011. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proceedings of the National Academy of Sciences of the United States of America* 108: 13624-13629.

Pröschold, T., B. Marin, U.G. Schlösser & M. Melkonian. 2001. Molecular phylogeny and taxonomic revision of *Chlamydomonas* (Chlorophyta). I. emendation of *Clamydomonas* Ehrenberg and *Chloromonas Gobi*, and description of *Oogamochlamys* gen. nov. and *Lobochlamys* gen. nov. *Protist* 152: 265-300.

Rindi, F., Lam, D.W., & J.L. Lopez-Bautista. 2009. Phylogenetic relationships and species circumscription in *Trentepohlia* and *Printzina* (Trentepohliales, Chlorophyta). *Molecular Phylogenetics and Evolution* 52: 329-339.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312-1313.

Talavera, G. & J. Castresana. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*. 56: 564-577.

Tritt, A., J.A. Eisen, M.T. Facciotti & A. Darling. 2012. An integrated pipeline for *de novo* assembly of microbial genomes. *PLoS One* 7(9): e42304. doi: 10.1371/journal.pone.0042304

Vaidya, G., D.J. Lohman & R. Meier. 2011. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics* 27: 171-180.

Verbruggen, H., M. Ashworth, S.T. LoDuca, C. Vlaeminck, E. Cocquyt, T. Sauvage, F. Zechman, D. Littler, M. Littler, F. Leliaert & O. DeClerck. 2009. A multi-locus time-calibrated phylogeny of the siphonous green algae. *Molecular Phylogenetics and Evolution* 50: 642-653.

Zechman, F.W., H. Verbruggen, F. Leliaert, M., Ashworth, M.A. Buchheim, M.W. Fawley, H. Spalding, C. Pueschel, J. Buchheim, B. Verghese & M.D. Hanisak. 2010. An unrecognized ancient lineage of green plants persists in deep marine waters. *Journal of Phycology* 46: 1288-1295.

Recibido: 21 septiembre
Revisado: 24 octubre
Corregido: 10 noviembre
Aceptado: 11 de noviembre
Revisores: 2 revisores anónimos

Table 1. BLASTx simlarities of ORFs in the plastid genomes of *Caulerpa racemosa* to known proteins.

	Bacterium	Protein	E-value	Similarity
ORF 1	*Methanosarcina siciliae*	group II intron reverse transcriptase/maturase	2.00E-19	31%
ORF 2	*Halomona*s sp.	group II intron reverse transcriptase/maturase	5.00E-70	31%
ORF 3	No significant hit	-	-	-
ORF 4	*Campylobacter concisus*	Adenine-specific DNA methylase	6.00E-53	37%
ORF 5	*Sulfurospirillum multivoran*s	DNA adenine methylase (dam)	2.00E-155	76%
ORF 6	*Cycloclasticus* sp.	Type I restriction enzyme R protein N terminus	0.00E+00	52%
ORF 7	No significant hit	-	-	-
ORF 8	*Prevotella conceptionensis*	DNA methyltransferase	9.00E-63	41%
ORF 9	*Desulfovibrio inopinatus*	DNA polymerase family A	1.00E-18	31%
ORF 10	No significant hit	-	-	-
ORF 11	*Hydrogenobacter thermophilus*	DNA polymerase family A	1.00E-41	27%
ORF 12	No significant hit	-	-	-
ORF 13	*Estrella lausannensis*	hypothetical protein (unknown function)	9.00E+00	33%
ORF 14	No significant hit	-	-	-
ORF 15	Delta-proteobacteria	hypothetical protein (unknown function)	6.5	34%

Comité Editorial