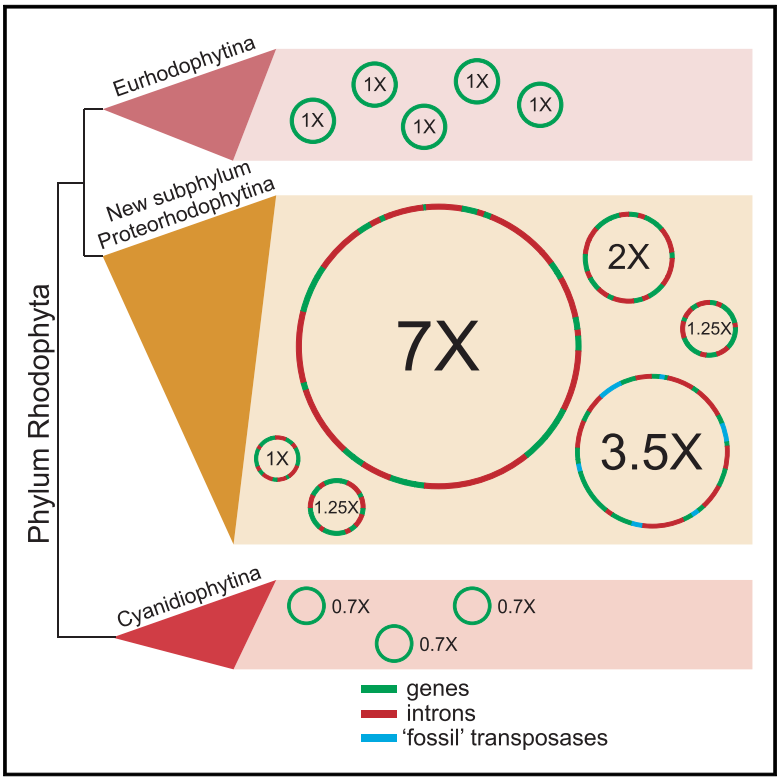
Report

Current Biology

The New Red Algal Subphylum Proteorhodophytina Comprises the Largest and Most Divergent Plastid Genomes Known

Graphical Abstract



Highlights

- Proteorhodophytina is a new red algal subphylum
- Proteorhodophytina comprises the largest plastid genome sequenced
- Proteorhodophytina comprises the most intron-rich plastid genome known
- Secondary red plastids evolved before the diversification of mesophilic red algae

Authors

Sergio A. Muñoz-Gómez, Fabián G. Mejía-Franco, Keira Durnin, Morgan Colp, Cameron J. Grisdale, John M. Archibald, Claudio H. Slamovits

Correspondence

claudio.slamovits@dal.ca

In Brief

Muñoz-Gómez et al. explored the diversity of red algal plastid genomes, which led them to discover the largest and most intron-rich plastid genomes known as well as to resolve the deep phylogeny of the red algae and shed light on the origin of secondary red algal plastids.

Muñoz-Gómez et al., 2017, Current Biology *27*, 1677–1684
June 5, 2017
http://dx.doi.org/10.1016/j.cub.2017.04.054

Current Biology
Report

The New Red Algal Subphylum Proteorhodophytina Comprises the Largest and Most Divergent Plastid Genomes Known

Sergio A. Muñoz-Gómez,[1] Fabián G. Mejía-Franco,[2] Keira Durnin,[1] Morgan Colp,[1] Cameron J. Grisdale,[1] John M. Archibald,[1,3] and Claudio H. Slamovits[1,3,4,*]
[1]Centre for Comparative Genomics and Evolutionary Bioinformatics, Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, NS B3H 4R2, Canada
[2]Departamento de Biología, Universidad CES, Medellín, Antioquia 050022, Colombia
[3]Canadian Institute for Advanced Research, Toronto, ON M5G 1Z8, Canada
[4]Lead Contact
*Correspondence: claudio.slamovits@dal.ca
http://dx.doi.org/10.1016/j.cub.2017.04.054

SUMMARY

Red algal plastid genomes are often considered ancestral and evolutionarily stable, and thus more closely resembling the last common ancestral plastid genome of all photosynthetic eukaryotes [1, 2]. However, sampling of red algal diversity is still quite limited (e.g., [2–5]). We aimed to remedy this problem. To this end, we sequenced six new plastid genomes from four undersampled and phylogenetically disparate red algal classes (Porphyridiophyceae, Stylonematophyceae, Compsopogonophyceae, and Rhodellophyceae) and discovered an unprecedented degree of genomic diversity among them. These genomes are rich in introns, enlarged intergenic regions, and transposable elements (in the rhodellophycean *Bulboplastis apyrenoidosa*), and include the largest and most intron-rich plastid genomes ever sequenced (that of the rhodellophycean *Corynoplastis japonica*; 1.13 Mbp). Sophisticated phylogenetic analyses accounting for compositional heterogeneity show that these four "basal" red algal classes form a larger monophyletic group, Proteorhodophytina subphylum nov., and confidently resolve the large-scale relationships in the Rhodophyta. Our analyses also suggest that secondary red plastids originated before the diversification of all mesophilic red algae. Our genomic survey has challenged the current paradigmatic view of red algal plastid genomes as "living fossils" [1, 2, 6] by revealing an astonishing degree of divergence in size, organization, and non-coding DNA content. A closer look at red algae shows that they comprise the most ancestral (e.g., [2, 7, 8]) as well as some of the most divergent plastid genomes known.

RESULTS AND DISCUSSION

New Red Algal Genomes Reveal High Phylogenetic Divergence and Diversity

Red algae are direct descendants of the first eukaryote that became photosynthetic after enslaving and converting a cyanobacterium into a photosynthetic organelle. Among plastid diversity, red algal plastid genomes (ptDNAs) are considered ancestral and evolutionarily stable in terms of gene content and genomic organization [1, 6], and therefore most closely resembling the last common ancestral genome of all plastids (e.g., [2]). For instance, red algal plastid genomes contain the largest repertoire of plastid genes reported so far (i.e., 166–235 protein-coding genes [9]). This large gene set includes numerous genes that are absent in all other plastid genomes, and which mostly encode gene-expression regulatory proteins, biosynthetic enzymes, and membrane transporters [6]. Furthermore, red algal plastid genomes have a very compact genomic organization and their genes appear to be evolving relatively slowly [2]. Red algae, therefore, are of particular importance in helping us elucidate the early evolution of plastids and their genomes.

Unfortunately, the diversity of red algal plastid genomes remains undersampled (e.g., see [10] for sampling deficiencies of the red algal phylogeny that are addressed in this study). Sampling has largely focused on the red algal mesophilic seaweeds (classes Bangiophyceae and Florideophyceae) and the unicellular thermoacidophilic members of the Cyanidiophyceae (e.g., [2, 3, 5]). More than 80% of the publicly available red algal plastid genomes belong to these three well-known classes (see [9]). In order to improve our understanding of the diversity and evolution of red algae and their plastids, we sequenced six new plastid genomes from the four phylogenetically disparate classes of mesophilic non-seaweed red algae: the Porphyridiophyceae (*Flintiella sanguinaria*), Stylonematophyceae (*Bangiopsis subsimplex*), Compsopogonophyceae (*Boldia erythrosiphon* and *Rhodochaete parvula*), and Rhodellophyceae (*Corynoplastis japonica* and *Bulboplastis apyrenoidosa*) (Figure 1).

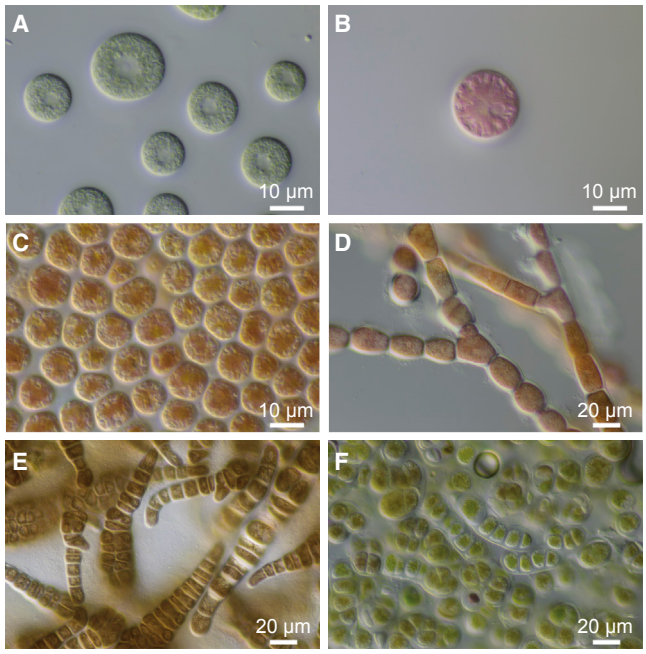


Figure 1. Mesophilic Non-seaweed Red Algae Belonging to the New Subphylum Proteorhodophytina, Whose Plastid Genomes Were Sequenced in This Study

(A) *Bulboplastis apyrenoidosa* NIES-2742 (Rhodellophyceae) as single unicells.

(B) *Corynoplastis japonica* NIES-2662 (Rhodellophyceae) as a single unicell.

(C) Palmelloid stage of *Flintiella sanguinaria* UTEX LB2060 (Porphyridiophyceae).

(D) Branched filaments of *Rhodochaete parvula* UTEX LB2715 (Compsopogonophyceae).

(E) Filaments of *Boldia erythrosiphon* UTEX LB2858 (Compsopogonophyceae).

(F) Filaments of *Bangiopsis subsimplex* UTEX LB2854 (Stylonematophyceae).

The scale bars represent 10 μm and 20 μm where indicated.

Table 1. Genomic Features and Statistics for the Newly Sequenced ptDNAs of Six Representatives of the Novel Red Algal Subphylum Proteorhodophytina

Species	Class	ptDNA Size (bp)	Non-genic DNA[a] (bp)	Intergenic DNA[b] (bp)	Intronic DNA[c] (bp)	Introns	Transposases[d]	Proteins[e]	tRNAs	IR	LSC (bp)	SSC (bp)
Bangiopsis subsimplex UTEX LB2854	Stylonematophyceae	205,002	58,249 (28.41%)	36,070 (17.59%)	22,179 (10.81%)	39	0	192	29	no	–	–
Boldia erythrosiphon UTEX LB2858	Compsopogonophyceae	226,658	88,798 (39.17%)	41,858 (18.46%)	46,940 (20.70%)	94	0	185	31	no	–	–
Rhodochaete parvula UTEX LB2715	Compsopogonophyceae	221,656	71,554 (32.28%)	36,790 (16.59%)	34,764 (15.68%)	72	0	190	31	yes	191,768 (86.51%)	20,150 (9.09%)
Flintiella sanguinaria UTEX LB2060	Porphyridiophyceae	370,672	225,396 (60.80%)	72,526 (19.56%)	152,870 (41.24%)	179	0	186	29	yes	343,401 (92.64%)	18,017 (4.86%)
Bulboplastis apyrenoidosa NIES-2742	Rhodellophyceae	610,063	485,888 (79.64%)	224,262 (36.76%)	261,721 (42.90%)	220	>31	163	30	yes	599,917 (98.33%)	402 (0.06%)
Corynoplastis japonica NIES-2662	Rhodellophyceae	1,127,474	991,673 (87.95%)	271,113 (24.04%)	720,560 (63.90%)	311	0	179	30	yes	1,116,741 (99.04%)	457 (0.04%)

LSC, large single copy; SSC, small single copy. See Table S1 for sequencing and assembly statistics for these ptDNAs.

[a]Includes intergenic and intronic DNA, as well as maturases and transposases.

[b]Includes intergenic maturases and transposases.

[c]Includes intronic maturases.

[d]Includes pseudogenized transposases.

[e]Excludes pseudogenes, maturases, and transposases.

Our genomic explorations into the mesophilic non-seaweed red algae have expanded the known range of ptDNA diversity among red algae (see Table 1 and Figures 2, S1; Figures S5–S8 on Mendeley Data, http://dx.doi.org/10.17632/hfhb433p9s.1)—some previous studies had hinted at the distinctiveness of the ptDNAs for some members from these groups [1, 4, 9, 11–13]. Whereas intergenic DNA tends to be relatively constant among these ptDNAs (~22%), most variation is caused by intronic DNA, which ranges from 10% to 64% of the total ptDNA size (Table 1). Unlike bangiophycean and florideophycean ptDNAs, which carry open reading frames (ORFs) of plasmid origin [14], we could not detect any plasmid ORFs on our new ptDNAs (data not shown), although characterization of plasmids for these red algal lineages remains to be done. The gene content of these classes, however, seems to have stabilized early in evolution and has not changed drastically relative to other red algae (Table 1; Table S2 on Mendeley Data, http://dx.doi.org/10.17632/txn2dnt5z6.1). Most of the mesophilic non-seaweed red algae have ptDNAs with a quadripartite organization defined by inverted repeats (IRs) that contain the rRNA gene operon (see Table 1). However, unlike the quadripartite organization of some red algal and most green plant and secondary red algal ptDNAs (e.g.,

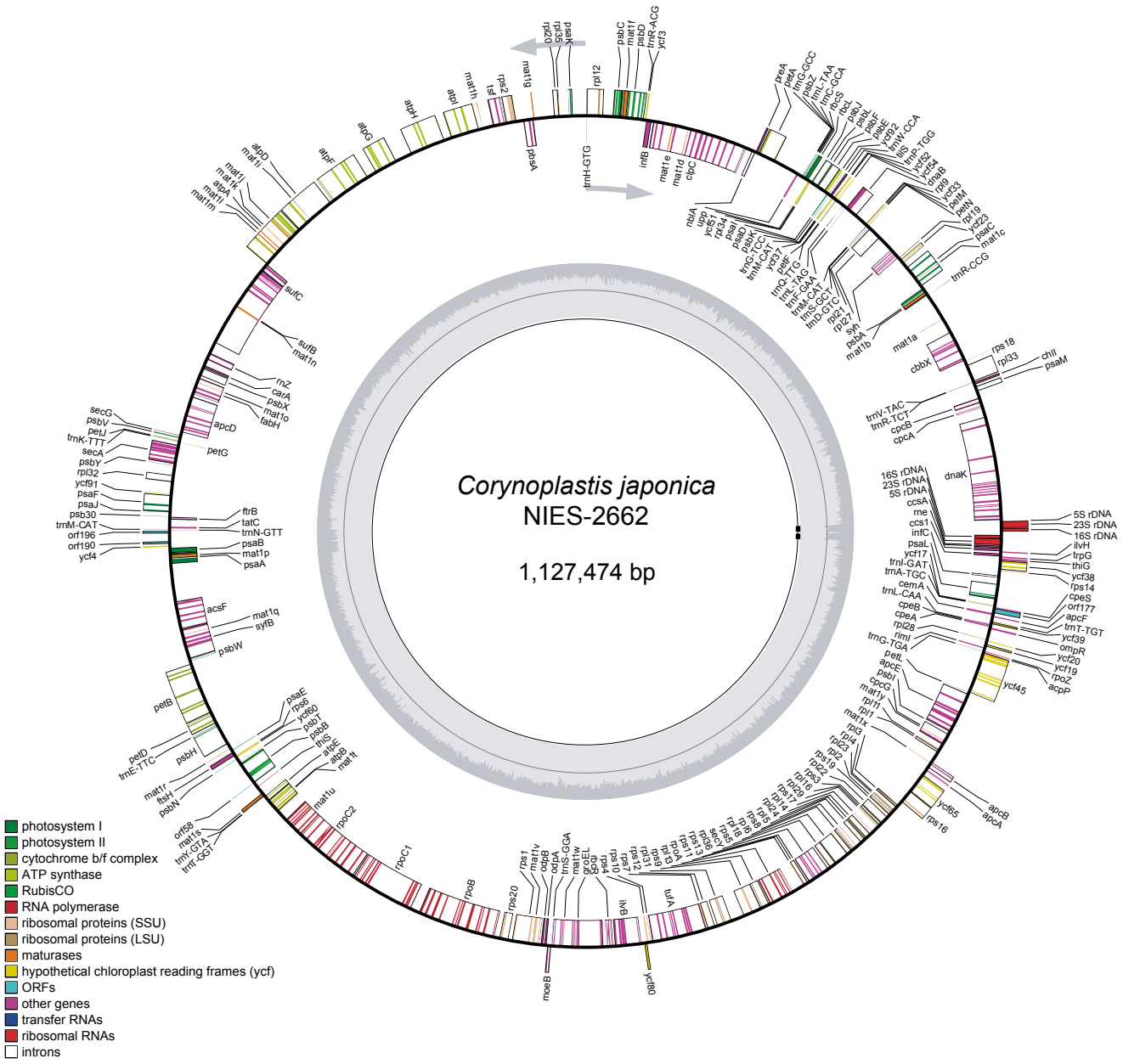


Figure 2. Plastid Genome Map of *Corynoplastis japonica* NIES-2662 Rhodellophyceae, which Represents the Largest Plastid Genome Sequenced to Date

Note the highly fragmented nature of its genes due to the massive proliferation of divergent group II introns (note that pseudogenized maturases are also shown). This picture resembles eukaryotic nuclear genes and makes rhodellophycean ptDNAs interesting models to understand the early invasion of eukaryotic genes by selfish introns. Inverted repeats containing the rRNA gene operon (marked as thick black lines in the innermost circle) flank a tiny SSC region of just 457 bp. See Figure S1 for the plastid genome map of the rhodellophycean *Bulboplastis apyrenoidosa* NIES-2742.

see [4, 15]), these IRs are divergent relative to the small single-copy (SSC) region (Figure S4 on Mendeley Data). Remarkably, the SSCs of rhodellophyceans are quite small, being only 402–457 bp long and containing no genes at all (Table 1). The high degree of divergence of these ptDNAs is also seen in their highly rearranged genomes (Figures S9–S11 on Mendeley Data). These results indicate that red algal ptDNAs encompass a greater degree of diversity than previously appreciated, challenging traditional views of their ancestral nature.

Massive Plastid Genome Expansion among Rhodellophyceans

All mesophilic non-seaweed red algae examined to date have plastid genomes that are larger than those of cyanidiophycean or seaweed red algae as a consequence of the proliferation of

divergent group II introns [4, 9, 13]. Prior to our study, the largest known red algal ptDNA belonged to the porphyridiophycean *Porphyridium sordidum* with a size of 259 kbp [9]. Outside the Rhodophyta, the largest ptDNAs reported are those of the green algae *Floydiella terrestris* (521 kbp [16]) and *Volvox carteri* (525 kbp [17]). It has also been suggested that the green seaweed *Acetabularia acetabulum* has a ptDNA >1 Mbp in size, although it has only been partially sequenced [18, 19]. Our expanded sample of red algal ptDNAs has uncovered the largest plastid genomes ever sequenced.

The rhodellophyceans *B. apyrenoidosa* (order Dixoniellales) [20] and *C. japonica* (order Rhodellales) [21] have reached unprecedented levels of genomic expansion, with genome sizes of 610 kbp and 1.12 Mbp, respectively. In the case of *C. japonica*, ptDNA expansion seems to have been driven by the massive proliferation of introns, which have come to occupy more than 63.9% of the total ptDNA size (Figure 2; Table 1). *C. japonica* has an impressive number of at least 311 introns in its genome (see Figure 2), surpassing the record previously held by euglenophyceans; the most intron-rich ptDNA documented had been that of *Euglena gracilis* strain Z, with ~150 introns [22–24]. *B. apyrenoidosa*, on the contrary, has enlarged intergenic regions that represent ~36% of its ptDNA size, although introns have also spread significantly (at least 220 introns) (Table 1). The cause for the expansion of intergenic spacers in the ptDNA of *B. apyrenoidosa* appears to be invasion by insertion sequences (ISs) of bacterial origin. *B. apyrenoidosa*'s ptDNA is exceptional in having as many as 55 predicted ISs (defined by complete, partial, or pseudogenized transposases; Figure S12 on Mendeley Data); previous reports are limited to only one or two examples of similarly acquired foreign genes (e.g., see [25–28]). Classification of these ISs suggests that these selfish DNA elements might have had at least two independent ancient origins, and that they subsequently spread (and were most likely inactivated through pseudogenization) within *B. apyrenoidosa*'s ptDNA (Figure S12 on Mendeley Data). The modes of expansion of rhodellophycean ptDNAs contrast with those in the green algae *F. terrestris* and *V. carteri*, whose intergenic regions have been populated by numerous short repeats (Table 1) [16, 17]. These examples further illustrate the extraordinary levels of genomic diversity hidden within the red algae.

A New Higher-Level Clade within the Rhodophyta

Previous studies delineated the major lineages within the red algae (now ranked as classes) but failed to resolve their deep interrelationships (e.g., [29–33]). It has therefore traditionally been assumed that the mesophilic non-seaweed red algae comprised a paraphyletic assemblage of independent lineages that successively diverged before the diversification of the red algal seaweeds (Eurhodophytina) and after their phylogenetic separation from the Cyanidiophytina [33–36]. A recent multi-gene analysis of 263 nuclear genes suggested a similar diversification pattern for the mesophilic non-seaweed red algae (i.e., Porphyridiophyceae, Stylonematophyceae, Compsopogonophyceae, and Rhodellophyceae), although support was equivocal [37]. It has been suggested that the lack of resolution at the red algal tree backbone is partly the result of a rapid early radiation of the mesophilic red algae [32, 36]. Phylogenetic analyses have been inconclusive regarding the deep phylogeny of the red algae, either because of poor taxon sampling or because of poor resolution at the deeper nodes (e.g., [4, 32, 37]).

We built a dataset of 170 plastid genome-encoded proteins to better understand the evolutionary history of red plastid genomes and the deep phylogeny of the red algae (see Supplemental Information). This dataset comprised ptDNA-encoded genes from 37 red algae, including our six newly sequenced "basal" red algae, and five green algae as an outgroup. We analyzed this dataset under maximum-likelihood and Bayesian frameworks and used sophisticated models that account for across-site rate and amino acid compositional heterogeneities (C60 and PMSF on IQ-TREE, and CAT+GTR on PhyloBayes [38, 39]; H. Wang, S. Susko, B. Minh, and A. Roger, personal communication; http://www.iqtree.org/doc/Complex-Models#site-specific-frequency-models). Our results recovered a strongly supported clade for the mesophilic non-seaweed red algae, and confidently resolved the interrelationships between the four previously "basal" red algal classes (Figure 3). Interestingly, support for this deep node, as well as its internal nodes, increased as we used more sophisticated models of protein sequence evolution (Figures S13–S25 on Mendeley Data). In our trees, the classes Compsopogonophyceae and Porphyridiophyceae are reciprocally monophyletic, similar to the sister relationship between the classes Stylonematophyceae and Rhodellophyceae (Figure 3). After our analyses had been completed, a study focused solely on the red algae used 298 nuclear genes to recover the four "basal" non-cyanidiophycean red algal classes as a larger monophyletic group [40]. This result corroborates the higher-level clade recovered here based on plastid genomes. Although support was high for this new higher-level clade, Qiu et al. [40] could not confidently resolve the interrelationships among the classes from which it is comprised.

We propose that the new higher-level clade formed by these four classes be named Proteorhodophytina subphylum novus to reflect the wide diversity of body types, lifestyles, and plastid genomes (Figure 3; see STAR Methods for formal taxon description). The shared evolutionary history among the four classes comprising this new subphylum furthermore explains the evolutionary trend in non-coding DNA acquisition by the plastid genomes of this major red algal lineage (Table 1) as well as their unique ptDNA organization relative to other red algae (Figure S4 on Mendeley Data). The putatively ancient nature and wide diversity of the Proteorhodophytina also help to partly explain why some traditionally used characters (such as filamentous thalli, asexuality, pit plugs, pyrenoids, plastid morphology, Golgi associations, ecological habitat, etc.) are most taxonomically "useful" at a class level and most likely have multiple independent origins (or losses) among red algal classes (see Figure S2, which primarily relies on [34, 41–44]). This higher-level classification scheme for the phylum Rhodophyta updates, unifies, and grounds the recent proposals of Saunders and Hommersand [34], Yoon et al. [36], and Ruggiero et al. [45] on a robust phylogenetic framework.

A Deep Origin of Secondary Red Plastids among Red Algae

Red algal plastids are well known to have spread across the eukaryotic tree by secondary (i.e., eukaryote-eukaryote) endosymbiosis. As a consequence, the chlorophyll-*c*-containing algae

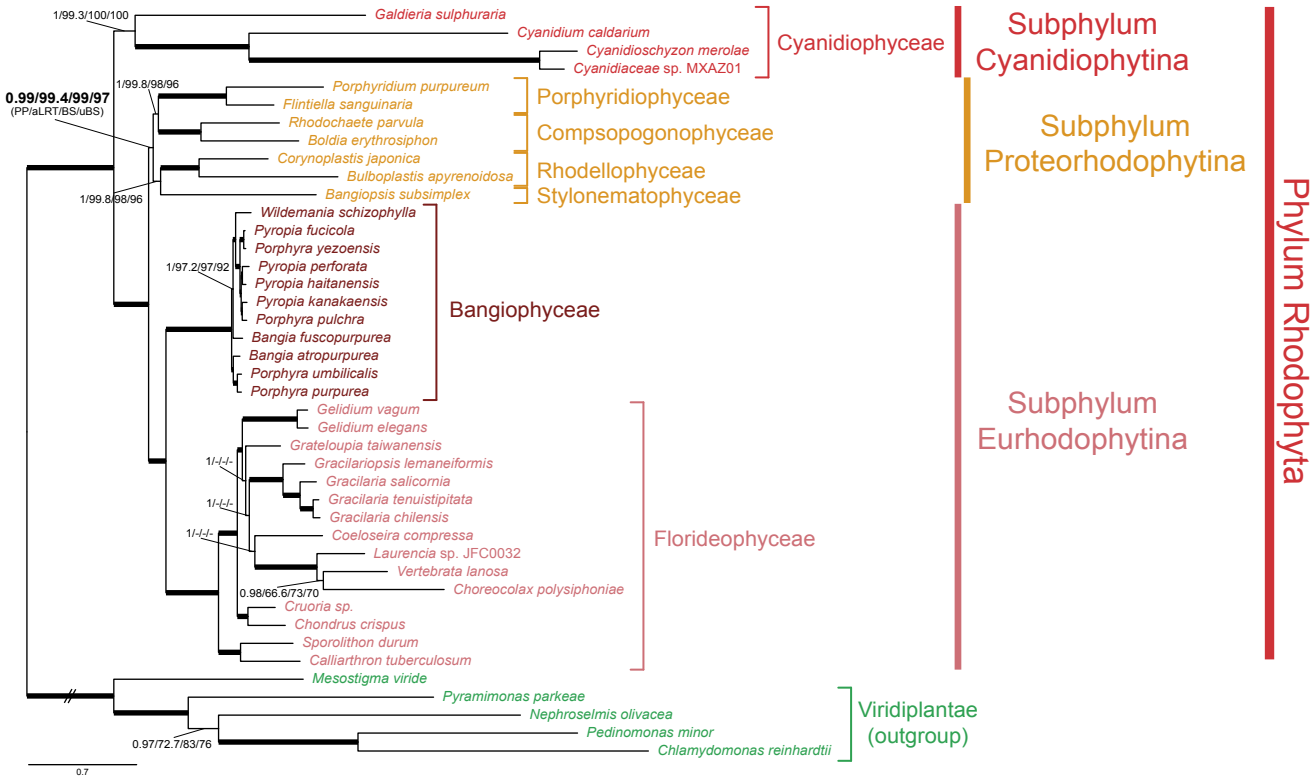


Figure 3. Phylogenetic Tree of Red Algal Plastid Genomes, which Resolved the Deep Phylogeny of the Phylum Rhodophyta and Established Its New Subphylum Proteorhodophytina

Consensus phylogram from two independent Markov chain Monte Carlo (MCMC) chains inferred under a Bayesian framework using PhyloBayes and the CAT+GTR+G4 model on a dataset of 170 plastid genome-encoded proteins. An identical topology is inferred under maximum likelihood with IQ-TREE using the models LG+C60+F+R6 and LG+PMSF+F+R6. Branch support is shown as Bayesian posterior probabilities/SH-aLRT support (%)/standard non-parametric bootstrap support (%)/ultrafast bootstrap support (%). Fully supported branches (1/100%/100%/100%) have been thickened. The most basal branch has been shortened 50% for illustration purposes (see hash marks). See Figure S2 for the phylogenetic distribution (on the topology here inferred) of taxonomic characters traditionally used for the red algae. See also Figure S3 for the phylogenetic placement of secondary red plastids in the phylogeny of the red algae.

evolved: cryptophytes, haptophytes, ochrophytes, dinophytes, and the chromerid algae (as well as some of their non-photosynthetic descendants, such as apicomplexans [46, 47]). Secondary red plastids have a monophyletic origin from within the red algae whether they were inherited vertically from a singular secondary endosymbiosis or further spread horizontally among other eukaryotes afterward. However, the precise red algal lineage that donated their plastid to the first secondary red alga (or red meta-alga) has remained elusive due to a lack of phylogenetic resolution and poor taxon sampling among the "basal" red algae (e.g., [31, 33, 48]). Recent phylogenetic trees addressing this question indicate that secondary red plastids are sister to the Eurhodophytina (the bangiophyceans and florideophyceans) or to a larger clade composed of both the Eurhodophytina and the Porphyridiophyceae (e.g., [4, 49]). Unfortunately, the lack of information for the classes Compsopogonophyceae, Stylonematophyceae, and Rhodellophyceae has precluded a clearer understanding of the origin of secondary red plastids. The establishment of a well-supported new red algal subphylum, the Proteorhodophytina, based on an expanded taxon sampling of "basal" red algae, can potentially shed light on the timing of the origin of secondary red plastids from within the red algae.

To revisit this question, we incorporated secondary red algal plastid genome sequences into our newly expanded plastid dataset (alveolate plastid genomes were excluded from our analyses due to their highly divergent nature). Our plastid protein trees suggest that secondary red plastids have an even deeper origin within the red algae. They branch as sister to all mesophilic red algae, i.e., the subphyla Eurhodophytina and the Proteorhodophytina (Figure S3). This phylogenetic position suggests that secondary red plastids could have evolved approximately ~1,500–1,194 million years ago [44, 50], although updated fossil-calibrated molecular chronograms are needed. The nature of the ancestor of secondary red plastids was most likely an ancestral mesophilic red algal unicell that lived in freshwater environments (based on the phylogenetic context of the secondary red algal lineage; e.g., see Figure S2).

Conclusions

This work (1) reports the largest plastid genome ever sequenced, (2) reports the most intron-rich plastid genome known, (3) reports the most extreme example of ancient transposition by insertion sequences in an organelle genome, (4) resolves the deep phylogeny of the red algae, (5) sheds light on the timing and origin of

secondary red plastids, and (6) reveals an unprecedented degree of diversity among red algal plastid genomes. It will be interesting to see whether future studies aimed at disentangling the ambiguous/conflicting signals in nuclear datasets corroborate the phylogenetic pattern based on plastid genomes shown here. Similarly, mitochondrial genomes will also prove helpful by providing an independent phylogenetic test. Additional sampling among the mesophilic non-seaweed red algae, especially the rhodellophyceans, will no doubt continue to push the boundaries of the wide spectrum of plastid genome diversity [51]. Our genomic survey has challenged the current paradigmatic view of red algal plastid genomes as "living fossils" that have retained an ancestral state by revealing an astonishing degree of divergence in size, organization, and non-coding DNA content. Mosaic evolution has produced an eclectic mixture of the most ancestral, as well as some of the most divergent, plastid genomes known among the red algae.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- METHOD DETAILS
  - Sequencing of red algal plastid genomes
  - Plastid genome assembly and circularization
  - Plastid genome annotation
  - Genomic analyses
  - Phylogenetic analyses
  - Taxon description
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and one table and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2017.04.054.

AUTHOR CONTRIBUTIONS

Conceptualization, S.A.M.-G., F.G.M.-F., and C.H.S.; Methodology, S.A.M.-G., J.M.A., and C.H.S.; Validation, S.A.M.-G., M.C., and C.J.G.; Formal Analysis, S.A.M.-G.; Investigation, S.A.M.-G., F.G.M.-F., K.D., M.C., and C.J.G.; Data Curation, S.A.M.-G., M.C., and C.J.G.; Writing – Original Draft, S.A.M.-G.; Writing – Review & Editing, S.A.M.-G., J.M.A., and C.H.S.; Visualization, S.A.M.-G.; Funding Acquisition, C.H.S.; Resources, C.H.S.; Supervision, J.M.A. and C.H.S.

ACKNOWLEDGMENTS

We thank Marlena Dlutek and Bruce Curtis for technical assistance. We are grateful to Sebastian Hess for sharing his expertise on light microscopy. This work was supported by a Killam predoctoral scholarship (to S.A.M.-G.), an Emerging Leaders in the Americas Program (ELAP) scholarship (to F.G.M.-F.), a Natural Sciences and Engineering Research Council of Canada (NSERC) postdoctoral fellowship (PDF 417789-2015 to C.J.G.), a NSERC undergraduate student research award (to M.C.), and a NSERC Discovery grant (RGPIN 05871-2014 to J.M.A.). J.M.A. and C.H.S. are fellows of the Integrated Microbial Biodiversity Program and the Canadian Institute for Advanced Research (CIFAR).

Received: March 27, 2017
Revised: April 24, 2017
Accepted: April 26, 2017
Published: May 18, 2017

REFERENCES

1. Lang, B.F., and Nedelcu, A.M. (2012). Plastid genomes of algae. In Genomics of Chloroplasts and Mitochondria Advances in Photosynthesis and Respiration, R. Bock, and V. Knoop, eds. (Springer), pp. 59–87.
2. Janouškovec, J., Liu, S.-L., Martone, P.T., Carré, W., Leblanc, C., Collén, J., and 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.
3. Wang, L., Mao, Y., Kong, F., Li, G., Ma, F., Zhang, B., Sun, P., Bi, G., Zhang, F., Xue, H., and Cao, M. (2013). Complete sequence and analysis of plastid genomes of two economically important red algae: *Pyropia haitanensis* and *Pyropia yezoensis*. PLoS ONE *8*, e65902.
4. Tajima, N., Sato, S., Maruyama, F., Kurokawa, K., Ohta, H., Tabata, S., Sekine, K., Moriyama, T., and Sato, N. (2014). Analysis of the complete plastid genome of the unicellular red alga *Porphyridium purpureum*. J. Plant Res. *127*, 389–397.
5. Verbruggen, H., and Costa, J.F. (2015). The plastid genome of the red alga *Laurencia*. J. Phycol. *51*, 586–589.
6. Kim, E., and Archibald, J.M. (2009). Diversity and evolution of plastids and their genomes. In The Chloroplast Plant Cell Monographs, A.S. Sandelius, and H. Aronsson, eds. (Springer), pp. 1–39.
7. Shivji, M.S. (1991). Organization of the chloroplast genome in the red alga *Porphyra yezoensis*. Curr. Genet. *19*, 49–54.
8. Reith, M., and Munholland, J. (1993). A high-resolution gene map of the chloroplast genome of the red alga *Porphyra purpurea*. Plant Cell *5*, 465–475.
9. Lee, J., Cho, C.H., Park, S.I., Choi, J.W., Song, H.S., West, J.A., Bhattacharya, D., and Yoon, H.S. (2016). Parallel evolution of highly conserved plastid genome architecture in red seaweeds and seed plants. BMC Biol. *14*, 75.
10. Verbruggen, H., Maggs, C.A., Saunders, G.W., Le Gall, L., Yoon, H.S., and De Clerck, O. (2010). Data mining approach identifies research priorities and data requirements for resolving the red algal tree of life. BMC Evol. Biol. *10*, 16.
11. Bernard, C., Thomas, J.C., Mazel, D., Mousseau, A., Castets, A.M., Tandeau de Marsac, N., and Dubacq, J.P. (1992). Characterization of the genes encoding phycoerythrin in the red alga *Rhodella violacea*: evidence for a splitting of the rpeB gene by an intron. Proc. Natl. Acad. Sci. USA *89*, 9564–9568.
12. Richaud, C., and Zabulon, G. (1997). The heme oxygenase gene (pbsA) in the red alga *Rhodella violacea* is discontinuous and transcriptionally activated during iron limitation. Proc. Natl. Acad. Sci. USA *94*, 11736–11741.
13. Perrineau, M.-M., Price, D.C., Mohr, G., and Bhattacharya, D. (2015). Recent mobility of plastid encoded group II introns and twintrons in five strains of the unicellular red alga *Porphyridium*. PeerJ *3*, e1017.
14. Lee, J., Kim, K.M., Yang, E.C., Miller, K.A., Boo, S.M., Bhattacharya, D., and Yoon, H.S. (2016). Reconstructing the complex evolutionary history of mobile plasmids in red algal genomes. Sci. Rep. *6*, 23744.
15. Douglas, S.E. (1994). Chloroplast origins and evolution. In The Molecular Biology of Cyanobacteria Advances in Photosynthesis, D.A. Bryant, ed. (Springer), pp. 91–118.
16. Brouard, J.-S., Otis, C., Lemieux, C., and Turmel, M. (2010). The exceptionally large chloroplast genome of the green alga *Floydiella terrestris* illuminates the evolutionary history of the Chlorophyceae. Genome Biol. Evol. *2*, 240–256.
17. Smith, D.R., and Lee, R.W. (2010). Low nucleotide diversity for the expanded organelle and nuclear genomes of *Volvox carteri* supports the mutational-hazard hypothesis. Mol. Biol. Evol. *27*, 2244–2256.

18. Palmer, J.D. (1985). Comparative organization of chloroplast genomes. Annu. Rev. Genet. *19*, 325–354.
19. de Vries, J., Habicht, J., Woehle, C., Huang, C., Christa, G., Wägele, H., Nickelsen, J., Martin, W.F., and Gould, S.B. (2013). Is ftsH the key to plastid longevity in sacoglossan slugs? Genome Biol. Evol. *5*, 2540–2548.
20. Kushibiki, A., Yokoyama, A., Iwataki, M., Yokoyama, J., West, J.A., and Hara, Y. (2012). New unicellular red alga, *Bulboplastis apyrenoidosa* gen. et sp. nov. (Rhodellophyceae, Rhodophyta) from the mangroves of Japan: phylogenetic and ultrastructural observations. Phycol. Res. *60*, 114–122.
21. Yokoyama, A., Scott, J.L., Zuccarello, G.C., Kajikawa, M., Hara, Y., and West, J.A. (2009). *Corynoplastis japonica* gen. et sp. nov. and Dixoniellales ord. nov. (Rhodellophyceae, Rhodophyta) based on morphological and molecular evidence. Phycol. Res. *57*, 278–289.
22. Hallick, R.B., Hong, L., Drager, R.G., Favreau, M.R., Monfort, A., Orsat, B., Spielmann, A., and Stutz, E. (1993). Complete sequence of *Euglena gracilis* chloroplast DNA. Nucleic Acids Res. *21*, 3537–3544.
23. Wiegert, K.E., Bennett, M.S., and Triemer, R.E. (2013). Tracing patterns of chloroplast evolution in euglenoids: contributions from *Colacium vesiculosum* and *Strombomonas acuminata* (Euglenophyta). J. Eukaryot. Microbiol. *60*, 214–221.
24. Bennett, M.S., and Triemer, R.E. (2015). Chloroplast genome evolution in the Euglenaceae. J. Eukaryot. Microbiol. *62*, 773–785.
25. Fan, W.-H., Woelfle, M.A., and Mosig, G. (1995). Two copies of a DNA element, 'Wendy', in the chloroplast chromosome of *Chlamydomonas reinhardtii* between rearranged gene clusters. Plant Mol. Biol. *29*, 63–80.
26. Brouard, J.-S., Otis, C., Lemieux, C., and Turmel, M. (2008). Chloroplast DNA sequence of the green alga *Oedogonium cardiacum* (Chlorophyceae): unique genome architecture, derived characters shared with the Chaetophorales and novel genes acquired through horizontal transfer. BMC Genomics *9*, 290.
27. Cattolico, R.A., Jacobs, M.A., Zhou, Y., Chang, J., Duplessis, M., Lybrand, T., McKay, J., Ong, H.C., Sims, E., and Rocap, G. (2008). Chloroplast genome sequencing analysis of *Heterosigma akashiwo* CCMP452 (West Atlantic) and NIES293 (West Pacific) strains. BMC Genomics *9*, 211.
28. Leliaert, F., and Lopez-Bautista, J.M. (2015). The chloroplast genomes of *Bryopsis plumosa* and *Tydemania expeditiones* (Bryopsidales, Chlorophyta): compact genomes and genes of bacterial origin. BMC Genomics *16*, 204.
29. Freshwater, D.W., Fredericq, S., Butler, B.S., Hommersand, M.H., and Chase, M.W. (1994). A gene phylogeny of the red algae (Rhodophyta) based on plastid rbcL. Proc. Natl. Acad. Sci. USA *91*, 7281–7285.
30. Ragan, M.A., Bird, C.J., Rice, E.L., Gutell, R.R., Murphy, C.A., and Singh, R.K. (1994). A molecular phylogeny of the marine red algae (Rhodophyta) based on the nuclear small-subunit rRNA gene. Proc. Natl. Acad. Sci. USA *91*, 7276–7280.
31. Oliveira, M.C., and Bhattacharya, D. (2000). Phylogeny of the Bangiophycidae (Rhodophyta) and the secondary endosymbiotic origin of algal plastids. Am. J. Bot. *87*, 482–492.
32. Yoon, H.S., Müller, K.M., Sheath, R.G., Ott, F.D., and Bhattacharya, D. (2006). Defining the major lineages of red algae (Rhodophyta). J. Phycol. *42*, 482–492.
33. Müller, K.M., Oliveira, M.C., Sheath, R.G., and Bhattacharya, D. (2001). Ribosomal DNA phylogeny of the Bangiophycidae (Rhodophyta) and the origin of secondary plastids. Am. J. Bot. *88*, 1390–1400.
34. Saunders, G.W., and Hommersand, M.H. (2004). Assessing red algal supraordinal diversity and taxonomy in the context of contemporary systematic data. Am. J. Bot. *91*, 1494–1507.
35. Maggs, C.A., Verbruggen, H., and De Clerck, O. (2007). Molecular systematics of red algae: building future structures on firm foundations. In Unravelling the Algae: The Past, Present, and Future of Algal Systematics, J. Brodie, and J. Lewis, eds. (CRC Press), pp. 103–121.
36. Yoon, H.S., Zuccarello, G.C., and Bhattacharya, D. (2010). Evolutionary history and taxonomy of red algae. In Red Algae in the Genomic Age: Cellular Origin, Life in Extreme Habitats and Astrobiology, J. Seckbach, and D.J. Chapman, eds. (Springer), pp. 25–42.
37. Burki, F., Kaplan, M., Tikhonenkov, D.V., Zlatogursky, V., Minh, B.Q., Radaykina, L.V., Smirnov, A., Mylnikov, A.P., and Keeling, P.J. (2016). Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. Proc. Biol. Sci. *283*, 20152802.
38. Lartillot, N., and Philippe, H. (2004). A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol. Biol. Evol. *21*, 1095–1109.
39. Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274.
40. Qiu, H., Yoon, H.S., and Bhattacharya, D. (2016). Red algal phylogenomics provides a robust framework for inferring evolution of key metabolic pathways. PLoS Curr. *8*, http://dx.doi.org/10.1371/currents.tol.7b037376e6d84a1be34af756a4d90846.
41. Cole, K.M., and Sheath, R.G. (1990). Biology of the Red Algae (Cambridge University Press).
42. Seckbach, J., and Chapman, D.J., eds. (2010). Red Algae in the Genomic Age: Cellular Origin, Life in Extreme Habitats and Astrobiology (Springer).
43. Graham, L.E., Graham, J.M., Wilcox, L.W., and Cook, M.E. (2016). Algae, *Version 3.3.1* (LJLM Press).
44. Yoon, H.S., Nelson, W., Lindstrom, S.C., Boo, S.M., Pueschel, C., Qiu, H., and Bhattacharya, D. (2016). Rhodophyta. In Handbook of the Protists, J.M. Archibald, A.G.B. Simpson, C.H. Slamovits, L. Margulis, M. Melkonian, D.J. Chapman, and J.O. Corliss, eds. (Springer), pp. 1–45.
45. Ruggiero, M.A., Gordon, D.P., Orrell, T.M., Bailly, N., Bourgoin, T., Brusca, R.C., Cavalier-Smith, T., Guiry, M.D., and Kirk, P.M. (2015). A higher level classification of all living organisms. PLoS ONE *10*, e0119248.
46. Cavalier-Smith, T. (2013). Symbiogenesis: mechanisms, evolutionary consequences, and systematic implications. Annu. Rev. Ecol. Evol. Syst. *44*, 145–172.
47. Keeling, P.J. (2013). The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. Annu. Rev. Plant Biol. *64*, 583–607.
48. Yoon, H.S., Hackett, J.D., Pinto, G., and Bhattacharya, D. (2002). The single, ancient origin of chromist plastids. Proc. Natl. Acad. Sci. USA *99*, 15507–15512.
49. Ševčíková, T., Horák, A., Klimeš, V., Zbránková, V., Demir-Hilton, E., Sudek, S., Jenkins, J., Schmutz, J., Přibyl, P., Fousek, J., et al. (2015). Updating algal evolutionary relationships through plastid genome sequencing: did alveolate plastids emerge through endosymbiosis of an ochrophyte? Sci. Rep. *5*, 10134.
50. Yang, E.C., Boo, S.M., Bhattacharya, D., Saunders, G.W., Knoll, A.H., Fredericq, S., Graf, L., and Yoon, H.S. (2016). Divergence time estimates and the evolution of major lineages in the florideophyte red algae. Sci. Rep. *6*, 21361.
51. Smith, D.R., and Keeling, P.J. (2015). Mitochondrial and plastid genome architecture: reoccurring themes, but significant differences at the extremes. Proc. Natl. Acad. Sci. USA *112*, 10177–10184.
52. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.
53. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. *19*, 455–477.
54. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.
55. Wyman, S.K., Jansen, R.K., and Boore, J.L. (2004). Automatic annotation of organellar genomes with DOGMA. Bioinformatics *20*, 3252–3255.
56. Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. *25*, 955–964.

57. Lagesen, K., Hallin, P., Rødland, E.A., Staerfeldt, H.-H., Rognes, T., and Ussery, D.W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. *35*, 3100–3108.

58. Lohse, M., Drechsel, O., Kahlau, S., and 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*, W575–W581.

59. Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., and Chandler, M. (2006). ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. *34*, D32–D36.

60. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647–1649.

61. Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.-A., and Barrell, B. (2000). Artemis: sequence visualization and annotation. Bioinformatics *16*, 944–945.

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

63. Volfovsky, N., Haas, B.J., and Salzberg, S.L. (2001). A clustering method for repeat analysis in DNA sequences. Genome Biol. *2*, RESEARCH0027.

64. Katoh, K., Kuma, K., Toh, H., and Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. *33*, 511–518.

65. Criscuolo, A., and Gribaldo, S. (2010). BMGE (block mapping and gathering with entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol. Biol. *10*, 210.

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

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

68. Minh, B.Q., Nguyen, M.A.T., and von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol. *30*, 1188–1195.

69. Lartillot, N., Rodrigue, N., Stubbs, D., and Richer, J. (2013). PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. Syst. Biol. *62*, 611–615.

70. Drew, K.M., and Ross, R. (1965). Some generic names in the Bangiophycidae. Taxon *14*, 93–99.

71. Schmitz, F. (1896). Bangiaceae. In Die natürlichen Pflanzenfamilien nebst ihren Gattungen und wichtigeren Arten insbesondere den Nutzpflanzen unter Mitwirkung zahlreicher hervorragender Fachgelehrten, Teil 1, A. Engler, and K. Prantl, eds. (Verlag von Wilhelm Engelmann), pp. 307–316.

72. Bornet, E. (1892). Les Algues de P.-K.-A. Schousboe (Emile Le Maout).

73. Agardh, C.A. (1824). Systema Algarum (Literis Berlingianis).

74. Bird, C.J., McLachlan, J., and de Oliveira, E.C. (1986). *Gracilaria chilensis* sp.nov. (Rhodophyta, Gigartinales), from Pacific South America. Can. J. Bot. *64*, 2928–2934.

75. De Luca, P., Taddei, R., and Varano, L. (1978). « Cyanidioschyzon merolae »: a new alga of thermal acidic environments. Webbia *33*, 37–44.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Bangiopsis subsimplex	UTEX	LB2854
Boldia erythrosiphon	UTEX	LB2858
Rhodochaete parvula	UTEX	LB2715
Flintiella sanguinaria	UTEX	LB2060
Bulboplastis apyrenoidosa	NIES	2742
Corynoplastis japonica	NIES	2662
Critical Commercial Assays		
ZR Fungal/Bacterial DNA MicroPrep Kit	Zymo Research	Cat#D6005
REPLI-g Mini Kit	QIAGEN	Cat#150023
NEBNext Ultra II DNA Library Prep Kit	New England Biolabs	Cat#E7645S
Nextera XT DNA Library Preparation Kit	Illumina	FC-131-1024
Deposited Data		
ptDNA sequence (*B. subsimplex* UTEX LB2854)	NCBI GenBank	KY709207
ptDNA sequence (*B. erythrosiphon* UTEX LB2858)	NCBI GenBank	KY709208
ptDNA sequence (*R. parvula* UTEX LB2715)	NCBI GenBank	KY709209
ptDNA sequence (*F. sanguinaria* UTEX LB2060)	NCBI GenBank	KY709210
ptDNA sequence (*B. apyrenoidosa* NIES-2742)	NCBI GenBank	KY709211
ptDNA sequence (*C. japonica* NIES-2662)	NCBI GenBank	KY709212
Reads (*B. subsimplex* UTEX LB2854)	NCBI SRA	SRR5278935
Reads (*B. erythrosiphon* UTEX LB2858)	NCBI SRA	SRR5278934
Reads (*R. parvula* UTEX LB2715)	NCBI SRA	SRR5278930
Reads (*F. sanguinaria* UTEX LB2060)	NCBI SRA	SRR5278931
Reads (*B. apyrenoidosa* NIES-2742)	NCBI SRA	SRR5309331
Reads (*B. apyrenoidosa* NIES-2742)	NCBI SRA	SRR5309330
Reads (*C. japonica* NIES-2662)	NCBI SRA	SRR5278932
Multi-gene dataset (42x170.phy) and associated phylogenetic trees	TreeBASE	http://purl.org/phylo/treebase/phylows/study/TB2:S20932
Multi-gene dataset (90x106.phy) and associated phylogenetic trees	TreeBASE	http://purl.org/phylo/treebase/phylows/study/TB2:S20932
Figures S4–S25	Mendeley Data	http://dx.doi.org/10.17632/hfhb433p9s.1
Table S2	Mendeley Data	http://dx.doi.org/10.17632/txn2dnt5z6.1
Oligonucleotides		
Fs-rRNA_F: TTTTCGCTCGCCGCTACTAA	This paper	N/A
Fs-rRNA_R: AACTCGCCTACGTGAAGGTG	This paper	N/A
Software and Algorithms		
Trimmomatic 0.32	[46]	http://www.usadellab.org/cms/?page=trimmomatic
SPAdes 3.6.2	[47]	http://cab.spbu.ru/software/spades/
Geneious 9.1.2	[49]	https://www.geneious.com/
DOGMA	[50]	https://dogma.ccbb.utexas.edu/
Artemis v16.0	[51]	http://www.sanger.ac.uk/science/tools/artemis
RNAweasel		http://megasun.bch.umontreal.ca/RNAweasel/
tRNAscan-SE 2.0	[52]	http://lowelab.ucsc.edu/tRNAscan-SE/
RNAmmer 1.2	[53]	http://www.cbs.dtu.dk/services/RNAmmer/
OGDraw v1.2	[54]	http://ogdraw.mpimp-golm.mpg.de/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ISsaga 2.0		http://issaga.biotoul.fr/
HMMER v3.1b2		http://hmmer.org/
MAUVE	[55]	http://darlinglab.org/mauve/mauve.html
RepeatFinder		http://www.geneious.com/plugins/repeat-finder
Phobos v3.3.11	C. Mayer	http://www.ruhr-uni-bochum.de/ecoevo/cm/cm_phobos.htm
MAFFT v7.205	[56]	http://mafft.cbrc.jp/alignment/software/
BMGE v1.12	[57]	ftp://ftp.pasteur.fr/pub/gensoft/projects/BMGE/
RAxML 8.2.6	[58]	http://sco.h-its.org/exelixis/web/software/raxml/index.html
FigTree v1.4.3		http://tree.bio.ed.ac.uk/software/figtree/
SequenceMatrix 1.8	[59]	http://gaurav.github.io/taxondna/
IQ-TREE 1.5.0	[33]	http://www.iqtree.org/
PhyloBayes MPI 1.7a	[32]	www.phylobayes.org/
BLAST	[48]	https://blast.ncbi.nlm.nih.gov/Blast.cgi

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Claudio H. Slamovits (claudio.slamovits@dal.ca).

METHOD DETAILS

Sequencing of red algal plastid genomes

We chose six red algal species based on availability in culture collections and phylogenetic diversity among the Rhodophyta. These are *Bangiopsis subsimplex* UTEX LB2854 (Stylonematophyceae), *Boldia erythrosiphon* UTEX LB2858 and *Rhodochaete parvula* UTEX LB2715 (Compsopogonophyceae), *Corynoplastis japonica* NIES-2662 and *Bulboplastis apyrenoidosa* NIES-2742 (Rhodellophyceae) and *Flintiella sanguinaria* UTEX LB2060 (Porphyridiophyceae). Culture maintenance was done in a Thermo Scientific Precision incubator at 23°C and a 12:12 hr (night:day) cycle. In order to obtain genomic DNA (gDNA) from *F. sanguinaria* UTEX LB2060 a whole genome amplification (WGA) was performed using the REPLI-g Mini Kit (QIAGEN). DNA extraction of total gDNA for all other species was carried out with the ZR Fungal/Bacterial DNA MicroPrep Kit (Zymo Research) using a BIO101/Savant FastPrep FP120 high-speed bead beater and 20 uL of proteinase K (20 mg/mL). Sequencing libraries were made using the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs) for all species except *F. sanguinaria* UTEX LB2060. A sequencing library for *F. sanguinaria* UTEX LB2060 was made using the Nextera XT DNA Library Preparation Kit (Illumina). DNA sequencing libraries were sequenced in an Illumina MiSeq instrument. Post-assembly Sanger sequencing using the primers Fs-rRNA_F and Fs-rRNA_R was done to confirm the variable 16S rRNA-23S rRNA intergenic region in *F. sanguinaria* UTEX LB2006.

Plastid genome assembly and circularization

Sequencing reads produced in an Illumina MiSeq were trimmed with Trimmomatic 0.32 using the options: HEADCROP:16 LEADING:30 TRAILING:30 MINLEN:36. Illumina adapters were similarly removed with Trimmomatic 0.32 [52] using the option ILLUMINACLIP. Trimmed reads (both paired and unpaired) were assembled with SPAdes 3.6.2 [53] using the option –careful. ptDNA contigs were retrieved based on BLAST searches [54] using red algal ptDNA-encoded proteins as queries against the assembled SPAdes scaffolds. Manual work was done in Geneious 9.1.2 [60] and consisted in mapping reads to the ends of the initially identified ptDNA contigs (by BLAST searches) in order to extend them, and then searching for contigs that overlapped these extensions. Doing this for every contig allowed us to locate adjacent contigs and orient them properly until the genome map (which was being drawn on paper) circularized itself. The high degree of divergence between potential repetitive sequences did not hinder automated assembly of the reads and facilitated the subsequent circularization of the genome (see Figure S1 for sequencing and assembly statistics).

Plastid genome annotation

An initial rough annotation that served as a guide was done using DOGMA (https://dogma.ccbb.utexas.edu; [55]). Most protein-coding genes and exons were manually annotated by BLAST searches [54] of predicted ORFs. Small external exons were found by (1) identifying terminal insertions (as short as six amino acids long) in reference red algal proteins relative to our partially annotated proteins through BLASTp alignments, (2) searching these short amino acid insertions against our new genomes using tBLASTn with

relaxed parameters (e.g., PAM30 matrix and word size = 2), and (3) manually checking the candidate exons for the right genomic location flanking the gene in question. Small internal exons were found by (1) searching a database of reference red algal proteins with candidate introns using BLASTx and relaxed parameters (see above), (2) retaining hits that came from the gene from which the intron was derived, (3) checking for the appropriate internal position and conservation of the short amino acid sequence, and (4) confirming that these candidate exons corresponded to deletions in our partially annotated proteins relative to the reference red algal proteins in tBLASTn alignments. Final polishing of the ptDNA annotations was done in Artemis v16.0 [61]. Intron prediction was carried out by defining exons through BLAST searches [54], and in combination with predictions from RNAweasel (http://megasun.bch.umontreal.ca/RNAweasel). tRNA genes were predicted with tRNAscan-SE 2.0 [56], whereas rRNA genes were predicted with RNAmmer 1.2 [57]. Circular maps for the ptDNAs were generated with OGDraw v1.2 (http://ogdraw.mpimp-golm.mpg.de/; [58]).

Genomic analyses

Insertion sequences and their ORFs in the ptDNA of *B. apyrenoidosa* were predicted and classified using the ISsaga server (https://www-is.biotoul.fr) which relied on BLAST searches against a curated dataset of IS transposases [59]. IS transposases were also manually checked using BLASTp [54] and Artemis v16.0 [61]. Group II intron maturases were identified with HMMER v3.1b2 (http://hmmer.org/) searches using the Pfam domain profiles RVT_1 (PF00078), RVT_N (PF13655), GIIM (PF08388), and Intron_maturas2 (PF01348) against databases of all ptDNA ORFs > 50 amino acids. Maturase candidates were manually corroborated by doing BLASTp searches against the nr database. Multiple whole genome alignments were done with MAUVE (http://darlinglab.org/mauve/mauve.html; [62]) in Geneious 9.1.2 [60] using default options that automatically calculate seed weights and the minimum LCB score (introns were removed from ptDNAs prior to alignment). Short dispersed repeats with a minimum length of 50 bp were searched with RepeatFinder [63], whereas short tandem repeats of 20-100 bp in size were searched with Phobos v3.3.11 (http://www.rub.de/spezzoo/cm/cm_phobos.htm).

Phylogenetic analyses

A dataset of 170 genes (38,348 sites; 15% missing data) comprising 37 red algae only, plus five green algae as outgroup, was assembled. A second dataset of 106 genes (23,193 sites; 8% missing data) and 91 taxa including secondary red plastids (cryptophytes, haptophytes and ochrophytes) was assembled as well. Homologs were first gathered after BLAST searches [54] using *Porphyridium purpureum* protein sequences as queries against a local database of plastid genomes. Retrieved amino acid sequences were aligned with MAFFT v7.205 [64] and the L-INS-i method. Multiple alignments were then trimmed using BMGE v1.12 [65] with default parameters, except for -m BLOSUM30 and -g 0.15. Single gene trees were built using RAxML 8.2.6 [66] and the model PROTGAMMALGF and then manually inspected to remove suspicious sequences and paralogs using FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/). Single gene alignments were concatenated in a supermatrix using SequenceMatrix 1.8 [67]. Maximum likelihood (ML) analyses and Bayesian (BY) inferences were carried out using IQ-TREE 1.5.0 [39, 68] and PhyloBayes MPI 1.7a [38, 69], respectively. Stationarity and convergence of the Bayesian MCMC chains were checked using the program Tracer v1.6 (http://beast.bio.ed.ac.uk/Tracer) and following the guidelines of the PhyloBayes manual.

Plastid genome-encoded markers were selected based on how well represented they are among the chosen taxa. For our 170-gene dataset, all genes that were present in at least 70% of the ingroup taxa (red algae) were retained. For our 106-gene dataset we were more stringent and all genes that were present in at least 80% of the ingroup taxa (red and secondary red algae) were retained. Genes included in the phylogenetic datasets are the following (names in bold are genes present in both datasets): *accA, accB, accD, acpP, acsF, apcA, apcB, apcD, apcE, apcF, argB,* ***atpA***, ***atpB***, ***atpD***, ***atpE***, ***atpF***, ***atpG***, ***atpH***, ***atpI***, *carA,* ***cbbX***, ***ccs1***, ***ccsA***, *cemA,* ***chlI***, ***clpC***, *cpcA, cpcB, cpcG, cpeA, cpeB,* ***dnaB***, ***dnaK***, *fabH, ftrB,* ***ftsH***, *gltB,* ***groEL***, ***ilvH***, *infB, infC, moeB, odpA, odpB, ompR, pbsA,* ***petA***, ***petB***, ***petD***, ***petF***, ***petG***, *petJ, pgmA, preA,* ***psaA***, ***psaB***, ***psaC***, ***psaD***, ***psaE***, ***psaF***, ***psaI***, *psaJ, psaK,* ***psaL***, ***psbA***, ***psbB***, ***psbC***, ***psbD***, ***psbE***, ***psbF***, ***psbH***, ***psbI***, ***psbJ***, ***psbK***, ***psbL***, ***psbN***, ***psbT***, ***psbV***, ***psbW***, ***psbY***, *psbZ,* ***rbcL***, ***rbcR***, ***rbcS***, *rne,* ***rpl1***, ***rpl11***, ***rpl12***, ***rpl13***, ***rpl14***, ***rpl16***, ***rpl18***, ***rpl19***, ***rpl2***, ***rpl20***, ***rpl21***, ***rpl22***, ***rpl23***, ***rpl24***, ***rpl27***, *rpl28, rpl29,* ***rpl3***, ***rpl31***, ***rpl33***, ***rpl36***, ***rpl4***, ***rpl5***, ***rpl6***, *rpl9,* ***rpoA***, ***rpoB***, ***rpoC1***, ***rpoC2***, *rpoZ, rps1,* ***rps10***, ***rps11***, ***rps12***, ***rps13***, ***rps14***, ***rps16***, ***rps17***, ***rps18***, ***rps19***, ***rps2***, ***rps20***, ***rps3***, ***rps4***, ***rps5***, *rps6,* ***rps7***, ***rps8***, ***rps9***, ***secA***, ***secY***, *syfB, syh,* ***tatC***, ***thiG***, *thiS, trpA, trpG, trxA,* ***tsf***, ***tufA***, *ycf12,* ***ycf16***, *ycf19, ycf23,* ***ycf24***, *ycf26, ycf29,* ***ycf3***, *ycf34,* ***ycf35***, *ycf36, ycf38,* ***ycf39***, ***ycf45***, ***ycf46***, *ycf52, ycf53, ycf54, ycf60, ycf62, ycf65, ycf80, ycf85, ycxr.*

Taxon description

Proteorhodophytina S. A. Muñoz-Gómez, F. G. Mejía-Franco, K. Durnin, M. Colp, C. J. Grisdale, J. M. Archibald & C. H. Slamovits, subphylum nov.

Description

Red algae with unicellular, pseudofilamentous or filamentous bodies inhabiting mesophilic freshwater and marine environments. Members of this group do not develop macroscopic, pseudoparenchymatous or parenchymatous seaweed-type thalli as those of the Eurhodophytina; variable plastid morphologies and organellar associations; plastid genomes with a large number of introns and a quadripartite organization where the rRNA operon-containing inverted repeats are divergent relative to the small single copy region. The name ‘‘Proteorhodophytina’’ makes reference to the Greek god of the sea Proteus, able to display many different

forms, and alludes to the vast ('protean') phenotypic and genotypic diversity exhibited by the members of this new subphylum. Phylogenetic (node-based) definition: the least inclusive clade containing *Porphyridium purpureum* Drew & Ross 1965 [70] (Porphyridiophyceae), *Bangiopsis subsimplex* Schmitz 1896 [71] (Stylonematophyceae), *Bulboplastis apyrenoidosa* Kushibiki, Yokoyama, Iwataki, Yokoyama, West & Hara 2012 [21] (Rhodellophyceae) and *Rhodochaete parvula* Thuret ex Bornet 1892 [72] (Compsopogonophyceae). The Proteorhodophytina does not include: *Porphyra purpurea* Agardh 1824 [73] (Bangiophyceae), *Gracilaria chilensis* Bird, McLachlan & Oliveira 1986 [74] (Florideophyceae), and *Cyanidioschyzon merolae* De Luca, Taddei & Varano 1978 [75] (Cyanidiophyceae).

DATA AND SOFTWARE AVAILABILITY

The ptDNA sequences of *B. subsimplex* UTEX LB2854, *B. erythrosiphon* UTEX LB2858 and *R. parvula* UTEX LB2715, *C. japonica* NIES-2662 and *B. apyrenoidosa* NIES-2742, and *F. sanguinaria* UTEX LB2060 were deposited in NCBI GenBank under the accession numbers KY709207, KY709208, KY709209, KY709210, KY709211 and KY709212. Raw sequencing reads were deposited on the NCBI SRA archive under the study accession number SRP100563. Multi-gene datasets as well as phylogenetic trees inferred in this study were deposited at TreeBASE (http://purl.org/phylo/treebase/phylows/study/TB2:S20932). Figures S4–S25 (http://dx.doi.org/10.17632/hfhb433p9s.1) and Table S2 (http://dx.doi.org/10.17632/txn2dnt5z6.1) were deposited in Mendeley Data.