Additional Supplemental Information (Figs. S4-S25)

The new red algal subphylum Proteorhodophytina comprises the largest and most divergent plastid genomes known


Figure legends

Figure S4. Phylogenetic distribution of general ptDNA organizations (i.e., orientations of the rRNA gene-containing inverted repeats relative to the SSC region) among the seven classes of the phylum Rhodophyta. Some members in most red algal classes (e.g., Cyanidiophyceae, Stylonematophyceae, Bangiophyceae, and Florideophyceae) seem to have independently lost one of the rRNA gene operons.

Figure S5. Plastid genome map of Bangiopsis subsimplex UTEX LB2854 (Stylonematophyceae). There is a unique rRNA gene operon (marked by a thick black line in the innermost circle). Pseudogenes, including fragmented maturases, are shown.

Figure S6. Plastid genome map of Boldia erythrosiphon UTEX LB2858 (Compsopogonophyceae). There is a unique rRNA gene operon (marked by a thick black line in the innermost circle). Pseudogenes, including fragmented maturases, are shown.

Figure S7. Plastid genome map of Rhodochaete parvula UTEX LB2715 (Compsopogonophyceae). Inverted repeats containing the rRNA gene operon are marked as thick black lines in the innermost circle. Pseudogenes, including fragmented maturases, are shown.

Figure S8. Plastid genome map of Flintiella sanguinaria UTEX LB2060 (Porphyridiophyceae). Inverted repeats containing the rRNA gene operon are marked as thick black lines in the innermost circle. Pseudogenes, including fragmented maturases, are shown.
Figure S9. MAUVE multiple genome alignment between the plastid genomes of *Gracilaria chilensis* (Florideophyceae), *Pyropia perforata* (Bangiophyceae), and *Galdieria sulphuraria* strain 074W (Cyanidiophyceae).

Figure S10. MAUVE multiple genome alignment between the plastid genomes of *Boldia erythrosiphon* UTEX LB2858 (Compsopogonophyceae), *Rhodochaete parvula* UTEX LB2715 (Compsopogonophyceae), and *Flintiella sanguinaria* UTEX LB2060 (Porphyridiophyceae).

Figure S11. MAUVE multiple genome alignment between the plastid genomes of *Bangiopsis subsimplex* UTEX LB2854 (Stylonematophyceae), *Bulboplastis apyrenoidosa* NIES 2742 (Rhodellophyceae), and *Corynoplastis japonica* NIES-2662.

Figure S12. Classification of IS transposases in the plastid genome of *Bulboplastis apyrenoidosa* NIES 2742 based on ISsaga.

Figure S13. Maximum likelihood (ML) phylogenetic tree of primary and secondary red plastids (91x106 dataset) inferred with IQTREE under the LG+F+R5 model. Branch support shown as: SH-aLRT (%) / Ultrafast bootstrap (%). Cyanidiophyceae (in red); Proteorhodophytina (in dark yellow); Bangiophyceae (in maroon); Florideophyceae (in pink); secondary red plastids (in brown).

Figure S14. Maximum likelihood (ML) phylogenetic tree of primary and secondary red plastids (91x106 dataset) inferred with IQTREE under the LG4X+F+R5 model. Branch support shown as: SH-aLRT (%) / Ultrafast bootstrap (%). Cyanidiophyceae (in red); Proteorhodophytina (in dark yellow); Bangiophyceae (in maroon); Florideophyceae (in pink); secondary red plastids (in brown).

Figure S15. Maximum likelihood (ML) phylogenetic tree of primary and secondary red plastids (91x106 dataset) inferred with IQTREE under the LG+C20+F+R5 model. Branch support shown as: SH-aLRT (%) / Ultrafast bootstrap (%). Cyanidiophyceae (in red); Proteorhodophytina (in dark yellow); Bangiophyceae (in maroon); Florideophyceae (in pink); secondary red plastids (in brown).
Figure S16. Maximum likelihood (ML) phylogenetic tree of primary and secondary red plastids (91x106 dataset) inferred with IQTREE under the LG+C60+F+R5 model. Branch support shown as: SH-aLRT (%) / Ultrafast bootstrap (%). Cyanidiophyceae (in red); Proteorhodophytina (in dark yellow); Bangiophyceae (in maroon); Florideophyceae (in pink); secondary red plastids (in brown).

Figure S17. Maximum likelihood (ML) phylogenetic tree of primary and secondary red plastids (91x106 dataset) inferred with IQTREE under the LG+C60+F+R5 model. Branch support shown as: SH-aLRT (%) / Ultrafast bootstrap (%). Cyanidiophyceae (in red); Proteorhodophytina (in dark yellow); Bangiophyceae (in maroon); Florideophyceae (in pink); secondary red plastids (in brown).

Figure S18. Maximum likelihood (ML) phylogenetic tree of primary and secondary red plastids (91x106 dataset) inferred with IQTREE under the LG+PMSF1+F+R5 model. Branch support shown as: SH-aLRT (%) / Standard non-parametric bootstrap / Ultrafast bootstrap (%). Cyanidiophyceae (in red); Proteorhodophytina (in dark yellow); Bangiophyceae (in maroon); Florideophyceae (in pink); secondary red plastids (in brown).

Figure S19. Maximum likelihood (ML) phylogenetic tree of red algal plastids (42x170 dataset) inferred with IQTREE under the LG+F+R5 model. Branch support shown as: SH-aLRT (%) / Ultrafast bootstrap (%). Cyanidiophyceae (in red); Proteorhodophytina (in dark yellow); Bangiophyceae (in maroon); Florideophyceae (in pink).

Figure S20. Maximum likelihood (ML) phylogenetic tree of red algal plastids (42x170 dataset) inferred with IQTREE under the LG4X+F+R5 model. Branch support shown as: SH-aLRT (%) / Ultrafast bootstrap (%). Cyanidiophyceae (in red); Proteorhodophytina (in dark yellow); Bangiophyceae (in maroon); Florideophyceae (in pink).

Figure S21. Maximum likelihood (ML) phylogenetic tree of red algal plastids (42x170 dataset) inferred with IQTREE under the LG+C20+F+R5 model. Branch support shown as: SH-aLRT (%) / Ultrafast bootstrap (%). Cyanidiophyceae (in red); Proteorhodophytina (in dark yellow); Bangiophyceae (in maroon); Florideophyceae (in pink).
Figure S22. Maximum likelihood (ML) phylogenetic tree of red algal plastids (42x170 dataset) inferred with IQTREE under the LG+C60+F+G4 model. Branch support shown as: SH-aLRT (%) / Ultrafast bootstrap (%). Cyanidiophyceae (in red); Proteorhodophytina (in dark yellow); Bangiophyceae (in maroon); Florideophyceae (in pink).

Figure S23. Maximum likelihood (ML) phylogenetic tree of red algal plastids (42x170 dataset) inferred with IQTREE under the LG+C60+F+R5 model. Branch support shown as: SH-aLRT (%) / Ultrafast bootstrap (%). Cyanidiophyceae (in red); Proteorhodophytina (in dark yellow); Bangiophyceae (in maroon); Florideophyceae (in pink).

Figure S24. Maximum likelihood (ML) phylogenetic tree of red algal plastids (42x170 dataset) inferred with IQTREE under the LG+C60+F+R6 model. Branch support shown as: SH-aLRT (%) / Ultrafast bootstrap (%). Cyanidiophyceae (in red); Proteorhodophytina (in dark yellow); Bangiophyceae (in maroon); Florideophyceae (in pink).

Figure S25. Maximum likelihood (ML) phylogenetic tree of red algal plastids (42x170 dataset) inferred with IQTREE under the LG+PMSF1+F+R6 model. Branch support shown as: SH-aLRT (%) / Standard non-parametric bootstrap / Ultrafast bootstrap (%). Cyanidiophyceae (in red); Proteorhodophytina (in dark yellow); Bangiophyceae (in maroon); Florideophyceae (in pink).
Bangiopsis subsimplex
UTEX LB2854
205,002 bp
Figure S7

Rhodochaete parvula
UTEX LB2715

221,656 bp
Figure S9

Gracilaria chilensis

Pyropia perforata

Galdiera sulphuraria strain 074W
Figure S11

Bangiopsis subsimplex UTEX LB2854

Bulboplástis apyrenoidosa NIES 2742

Corynoplástis japonica NIES-2662
ISsaga prediction and classification in *B. apyrenoidosa*’s ptDNA*

- IS607 (53.25%)
- IS200/IS605 (46.75%)

99 ORFs related to ISs were found in this ptDNA
- 1 Putative complete ORFs
- 32 Putative partial ORFs
- 0 Pseudogenes related to IS ORFs
- 66 Uncategorized ORFs

Predicted IS number: 77
Different ISs: 55

**IS prediction table**

<table>
<thead>
<tr>
<th>IS family</th>
<th>ORF distribution*</th>
<th>Different ISs</th>
<th>Total ISs</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS200/IS605 ssgr IS1341</td>
<td>0 / 3 / 0 / 25</td>
<td>22</td>
<td>28</td>
</tr>
<tr>
<td>IS607</td>
<td>1 / 29 / 0 / 33</td>
<td>25</td>
<td>41</td>
</tr>
<tr>
<td>IS200/IS205</td>
<td>0 / 0 / 0 / 8</td>
<td>8</td>
<td>8</td>
</tr>
</tbody>
</table>

*complete / partial / pseudogene / unknown

*These numbers are not (necessarily) correct and they are only a guide. Indeed, our manual inspection of transposase-related ORFs suggests that most of these are pseudogenized which made their precise count difficult.
Figure S23