Highly Conserved Mitochondrial Genomes among Multicellular Red Algae of the Florideophyceae

Eun Chan Yang1,2,†, Kyeong Mi Kim3,†, Su Yeon Kim4, JunMo Lee4, Ga Hun Boo5, Jung-Hyun Lee6, Wendy A. Nelson7,8, Gangman Yi9, William E. Schmidt10, Suzanne Fredericq10, Sung Min Boo5, Debashish Bhattacharya11, and Hwan Su Yoon4,*

1Marine Ecosystem Research Division, Korea Institute of Ocean Science & Technology, Ansan, Korea
2Department of Marine Biology, Korea University of Science and Technology, Daejeon, Korea
3Bioresource Systematics Department, National Marine Biodiversity Institute of Korea, Seocheon, Chungnam, Korea
4Department of Biological Sciences, Sungkyunkwan University, Suwon, Korea
5Department of Biology, Chungnam National University, Daejeon, Korea
6Marine Biotechnology Research Division, Korea Institute of Ocean Science & Technology, Ansan, Korea
7National Institute for Water and Atmospheric Research, Wellington, New Zealand
8School of Biological Sciences, University of Auckland, Auckland, New Zealand
9Department of Computer Science and Engineering, Gangneung-Wonju National University, Wonju, Korea
10Biology Department, University of Louisiana at Lafayette
11Department of Ecology, Evolution and Natural Resources and Department of Marine and Coastal Sciences, Rutgers University
*Corresponding author: E-mail: hsyoon2011@skku.edu.
†These authors contributed equally to this work.

Accepted: July 28, 2015

Data deposition: Newly determined and annotated 11 mitochondrial sequences of red algae have been deposited at GenBank under the accessions KF290995, KF649303–5, and KJ398158–64.

Abstract

Two red algal classes, the Florideophyceae (approximately 7,100 spp.) and Bangiophyceae (approximately 193 spp.), comprise 98% of red algal diversity in marine and freshwater habitats. These two classes form well-supported monophyletic groups in most phylogenetic analyses. Nonetheless, the interordinal relationships remain largely unresolved, in particular in the largest subclass Rhodymeniophycidae that includes 70% of all species. To elucidate red algal phylogenetic relationships and study organelle evolution, we determined the sequence of 11 mitochondrial genomes (mtDNA) from 5 florideophycean subclasses. These mtDNAs were combined with existing data, resulting in a database of 25 florideophytes and 12 bangiophytes (including cyanidiophycean species). A concatenated alignment of mt proteins was used to resolve ordinal relationships in the Rhodymeniophycidae. Red algal mtDNA genome comparisons showed 47 instances of gene rearrangement including 12 that distinguish Bangiophyceae from Hildenbrandiophycidae, and 5 that distinguish Hildenbrandiophycidae from Nemaliophycidae. These organelle data support a rapid radiation and surprisingly high conservation of mtDNA gene syntheny among the morphologically divergent multicellular lineages of Rhodymeniophycidae. In contrast, we find extensive mitochondrial gene rearrangements when comparing Bangiophyceae and Florideophyceae and multiple examples of gene loss among the different red algal lineages.

Key words: Florideophyceae, mitochondrial genome, ordinal relationship, red algal evolution, rhodophytes, Rhodymeniophycidae.

Introduction

Mitochondria (mt) supply chemical energy to cells and have a long evolutionary history that extends to the origin of eukaryotes (Gray et al. 1999; Burger et al. 2003; Koonin 2010). Most mt genomes are characterized by a compact structure (exceptions include plants that contain exceptionally large mt genomes) and maternal inheritance. Due to their ubiquitous distribution and conserved functions, complete mtDNA data...
Mitochondrial Genome Evolution of Red Algae

have been used in diverse studies including population genetics, phylogenetics, and organellar genome evolution (Unseld et al. 1997; Turmel et al. 2002; Ogihara et al. 2005; Ratnasingham and Hebert 2007; Zink and Barrowclough 2008; Morin et al. 2010). For example, analysis of mtDNA data from land plants revealed gene replacements through horizontal gene transfer and endosymbiotic gene transfer to the nucleus (Adams and Palmer 2003; Bergthorsson et al. 2003, 2004). Similar to plants, Rhodophyta (red algal) mtDNA data have been used in systematic and population studies as molecular markers (Saunders 2005; Yang et al. 2008; Freshwater et al. 2010).

Red algae belong to an anciently diverged group (Supergroup Plantae or Archaeplastida) whose ancestor hosted the plastid primary endosymbiosis (Bhattacharya et al. 2004). This phylum is composed of seven classes (i.e., Bangiophyceae, Compsopogonophyceae, Cyanidiophyceae, Florideophyceae, Porphyridiophyceae, Rhodellophyceae, and Stylocladophyceae) with the most species-rich class Florideophyceae containing the subclasses Hildenbrandiophyceidae, Nemaliophyceidae, Corallinophyceidae, Ahnfeltiophyceidae, and Rhodymeniophyceidae (Saunders and Hommersand 2004; Yoon et al. 2006; Le Gall and Saunders 2007; Le Gall et al. 2010). A putative member of the Bangiophyceae played a central role in the eukaryotic tree of life as a donor through a single or more likely multiple secondary plastid endosymbiotes that gave rise to chlorophyll-c containing groups, such as diatoms, dinoflagellates, haptophytes, and cryptophytes (Yoon et al. 2004; Chan et al. 2011; Bhattacharya et al. 2013). Despite the importance of red algae in eukaryote evolution, only a few representative genomes (nuclear and organellar) have been characterized thus far (Ohta et al. 2003; Hancock et al. 2010; Smith et al. 2012; Janouškovec et al. 2013). Although the complete mtDNA from 26 red algal species has been determined (two Cyanidiophyceae [Leblanc et al. 1995; Jain et al. 2015], ten Bangiophyceae [Burger et al. 1999; Mao et al. 2012; Smith et al. 2012; Hwang et al. et al. 2013; Hughey et al. 2014; Kong et al. 2014; Silva and Hughey 2014; Bangia fuscoruppurea, Bi G, Cao M, Mao Y, unpublished data, NC_026905), and 14 Florideophyceae species [Leblanc et al. 1995; Hancock et al. 2010; Zhang et al. 2012; Campbell et al. 2014; DePriest et al. 2014; Kim, Yang, Yi, et al. 2014; Kim, Yang, Boo, et al. 2014; Tablizo and Lluisma 2014; Yang, Kim, Boo, et al. 2014; Yang, Kim KM, Kim SY, et al. 2014; Kim et al. 2015; Lee et al. 2015]), the broader evolutionary history of this genome has been challenging to reconstruct because of insufficient taxon sampling and the high level of structural variation seen in the available mt genomes. Apart from a recent report of the corallinophycean (nongeniculate) species Sporolithon durum (Kim et al. 2015), the remaining mt genome data are primarily from the Rhodymeniophyceidae.

Understanding mtDNA evolution in Rhodophyta, in particular as it relates to the sister taxa Bangiophyceae and Florideophyceae that comprise greater than 98% of red algal diversity (Guiry MD and Guiry GM 2015) is of great interest to understand mitochondrial evolution in general. The Bangiophyceae contains approximately 193 reported species, all with simple thalli, including the economically important Porphyra and Pyropia (source of sushi wrap, kim or nori) from which mtDNA from ten species has been determined (Mao et al. 2012; Smith et al. 2012; Hwang et al. 2013; Hughey et al. 2014). The Florideophyceae is the largest class of red algae containing greater than 95% (6,760 spp.) of reported species diversity in the phylum Rhodophyta [approximately 7,100 species (Guiry MD and Guiry GM 2015)]. Phylogenetic relationships of the five subclasses within the Florideophyceae are well-resolved based on molecular and ultrastructural data (e.g., Saunders and Hommersand 2004; Le Gall and Saunders 2007). The Hildenbrandiophyceidae, the earliest diverging subclass (which likely lacks a gametophyte stage) has a sister relationship to the remaining subclasses (Pueschel and Cole 1982; Le Gall and Saunders 2007). Of the remaining well-supported subclasses, the Rhodymeniophyceidae is the largest, containing ≥5,000 species (70% of red algae). In spite of the strongly supported monophyly of the subclass, relationships among the 12 rhodymeniophycidan orders remain unresolved (Saunders et al. 2004; Le Gall and Saunders 2007; Verbruggen et al. 2010).

Here, we report 11 mtDNA sequences, 3 from Ahnfeltia plicata (Ahnfeltiophyceae), Hildenbrandia rubra (Hildenbrandiophyceidae), Palmaria palmata (Nemaliophyceae), and the remaining 8 from the following rhodymeniophycidan species: Asparagopsis taxiformis (Bonnemaisoniales), Ceramium japonicum (Ceramiaceae), Gelidium elegans (Gelidiaceae), Plocamium cartilagineum (Plocamiaceae), Riquetopsis sp. (Pyyssoneliales), Schizymenia dubyi (Nemastomatales), Schimmellmannia schousboei (Acrosymphytales), and Sebdenia flabellata (Sebdeniales).

The aims of the present work were 1) to elucidate phylogenetic relationships at higher taxonomic levels and identify major evolutionary transitions in the Rhodophyta using a database of 37 mt genomes; 2) to compare the coding sequences (CDs) content and synteny between the Bangiophyceae and Florideophyceae, and among the five florideophycean subclasses; and 3) to determine the extent of mtDNA gene synteny and loss within the Rhodymeniophyceidae.

Materials and Methods

Taxon Sampling, DNA Isolation, Genome Sequencing, and Annotation

The red algal species used in this study were chosen to sample all 5 subclasses of the Florideophyceae, as well as 12 orders of the Rhodymeniophyceidae (table 1). We collected nine red algal species:
### Table 1
Red Algal Taxonomy and General Features of the Mitochondrial Genome Used in the Present Study

<table>
<thead>
<tr>
<th>Taxa</th>
<th>Size (bp)</th>
<th>GC Content (%)</th>
<th>Number of Gene Accession</th>
<th>CDS</th>
<th>tRNA</th>
<th>rRNA</th>
<th>Intron</th>
<th>Intronic ORF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cyanidiophyceae</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyanidioschyzon merolae</td>
<td>32,211</td>
<td>36.5</td>
<td></td>
<td>35</td>
<td>25</td>
<td>3</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Galdieria sulphuraria</td>
<td>21,428</td>
<td>44.0</td>
<td></td>
<td>19</td>
<td>7</td>
<td>2</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Bangiophyceae</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bangia fuscopurpurea</td>
<td>43,517</td>
<td>33.0</td>
<td></td>
<td>31</td>
<td>23</td>
<td>2</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Porphyra purpurea</td>
<td>36,753</td>
<td>33.5</td>
<td></td>
<td>29</td>
<td>24</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Porphyra umbilicalis</td>
<td>29,123</td>
<td>31.9</td>
<td></td>
<td>25</td>
<td>24</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Pyropia fucicola</td>
<td>35,035</td>
<td>32.5</td>
<td></td>
<td>31</td>
<td>23</td>
<td>2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Pyropia haitanensis</td>
<td>37,023</td>
<td>30.7</td>
<td></td>
<td>24</td>
<td>24</td>
<td>2</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Pyropia kanakaensis</td>
<td>39,300</td>
<td>30.0</td>
<td></td>
<td>30</td>
<td>25</td>
<td>2</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Pyropia perforata</td>
<td>40,042</td>
<td>31.8</td>
<td></td>
<td>24</td>
<td>23</td>
<td>2</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Pyropia tenera</td>
<td>42,269</td>
<td>32.8</td>
<td></td>
<td>25</td>
<td>23</td>
<td>2</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Pyropia yezoensis</td>
<td>41,688</td>
<td>32.7</td>
<td></td>
<td>27</td>
<td>27</td>
<td>2</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td><strong>Average:</strong></td>
<td>37,391</td>
<td>32.2</td>
<td></td>
<td>27.2</td>
<td>23.9</td>
<td>2</td>
<td>3.8</td>
<td>3.9</td>
</tr>
<tr>
<td><strong>Florideophyceae</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Hildenbrandiophycidae</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hildenbrandia rubra</td>
<td>33,066</td>
<td>36.4</td>
<td></td>
<td>25</td>
<td>18</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td><strong>Nemaliophycidae</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palmaria palmata</td>
<td>29,735</td>
<td>32.2</td>
<td></td>
<td>26</td>
<td>22</td>
<td>2</td>
<td>2</td>
<td>—</td>
</tr>
<tr>
<td><strong>Corallinophycidae</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sporolithon durum</td>
<td>27,202</td>
<td>28.4</td>
<td></td>
<td>23</td>
<td>20</td>
<td>2</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td><strong>Ahnfeltiophycidae</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ahnfeltia plicata</td>
<td>32,878</td>
<td>33.4</td>
<td></td>
<td>24</td>
<td>24</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td><strong>Rhodymeniophycidae</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asparagopsis taxiformis</td>
<td>26,097</td>
<td>26.7</td>
<td></td>
<td>25</td>
<td>24</td>
<td>2</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>Ceramium japonicum</td>
<td>26,110</td>
<td>28.5</td>
<td></td>
<td>22</td>
<td>24</td>
<td>2</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Chondrus crispus</td>
<td>25,836</td>
<td>27.9</td>
<td></td>
<td>29</td>
<td>24</td>
<td>4</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>Gelidium elegans</td>
<td>24,922</td>
<td>29.5</td>
<td></td>
<td>23</td>
<td>19</td>
<td>2</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>Gelidium vagum</td>
<td>24,901</td>
<td>30.5</td>
<td></td>
<td>23</td>
<td>18</td>
<td>2</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>Gracilaria chilensis</td>
<td>26,898</td>
<td>27.6</td>
<td></td>
<td>25</td>
<td>25</td>
<td>2</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>Gracilaria salicornia</td>
<td>25,727</td>
<td>28.4</td>
<td></td>
<td>25</td>
<td>20</td>
<td>2</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>Gracilariotheca oryzoides</td>
<td>25,161</td>
<td>28.1</td>
<td></td>
<td>23</td>
<td>20</td>
<td>2</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>Gracilariospora andersonii</td>
<td>27,036</td>
<td>28.0</td>
<td></td>
<td>27</td>
<td>19</td>
<td>2</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>Gracilariospora chorda</td>
<td>26,543</td>
<td>27.6</td>
<td></td>
<td>26</td>
<td>24</td>
<td>2</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Gracilariospora lemaneiformis</td>
<td>25,883</td>
<td>27.5</td>
<td></td>
<td>26</td>
<td>21</td>
<td>2</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>Grateloupin angusta</td>
<td>27,943</td>
<td>30.2</td>
<td></td>
<td>25</td>
<td>19</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Grateloupin taiwanensis</td>
<td>28,906</td>
<td>31.4</td>
<td></td>
<td>25</td>
<td>24</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Kappaphycus striatus</td>
<td>25,242</td>
<td>30.1</td>
<td></td>
<td>24</td>
<td>24</td>
<td>2</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>Plocamiocholax pulvinata</td>
<td>25,894</td>
<td>26.0</td>
<td></td>
<td>25</td>
<td>20</td>
<td>2</td>
<td>2</td>
<td>—</td>
</tr>
<tr>
<td>Plocamium cartilagineum</td>
<td>26,187</td>
<td>23.6</td>
<td></td>
<td>24</td>
<td>25</td>
<td>2</td>
<td>2</td>
<td>—</td>
</tr>
<tr>
<td>Rhodymenia pseudopalmata</td>
<td>26,166</td>
<td>29.5</td>
<td></td>
<td>24</td>
<td>21</td>
<td>2</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>Riquetophycus sp.</td>
<td>26,261</td>
<td>25.7</td>
<td></td>
<td>24</td>
<td>22</td>
<td>2</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>Schimmelpinnia schousboe</td>
<td>25,816</td>
<td>26.7</td>
<td></td>
<td>23</td>
<td>21</td>
<td>2</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>Schizymenia dubyi</td>
<td>26,347</td>
<td>25.8</td>
<td></td>
<td>25</td>
<td>24</td>
<td>2</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>Sebdenia flabellata</td>
<td>25,887</td>
<td>28.5</td>
<td></td>
<td>23</td>
<td>21</td>
<td>2</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td><strong>Average:</strong></td>
<td>26,866</td>
<td>28.7</td>
<td></td>
<td>24.6</td>
<td>21.7</td>
<td>2.1</td>
<td>1.3</td>
<td>0.08</td>
</tr>
</tbody>
</table>

*Number of protein-coding genes exclude intronic-orf.

**Newly determined mitochondrial genomes.**
Phylogenetic Analyses

To reduce tree artifacts due to high DNA divergence, amino acid sequences were used to reconstruct phylogenies. The inferred amino acid sequences of every protein-coding gene from 37 mtDNAs were aligned using the T-Coffee package (under the default setting of ‘-mode fmcoffee’) and then refined manually. This concatenated alignment comprised 6,436 amino acids (data set c24p, 24 CDS; gene no. 1–24 in supplementary table S3, Supplementary Material online). The alignment and tree files are available from TreeBase (http://treebase.org/treebase-web/search/study/summary.html?id=15150&x-access-code=7bddd870b7e5e820b3737098d36e3c4c&agreement=ok), the RedTol website (dbdata.rutgers.edu/redseq), and from the corresponding author upon request.

The phylogeny and statistical support for monophyletic groups were inferred using the maximum-likelihood (ML and ML bootstrap analysis) method with RAxML v.8.1.17 (Stamatakis 2006) under the specified models. We used the best fit independent amino acid model (amino acid substitution matrix with empirical frequencies [I] + gamma distributed rate heterogeneity [G]) for each gene partition (24 genes) of the concatenated data set to reduce the heterogeneity effect among genes on the phylogeny. The specific amino acid substitution matrix for each partition was selected by using the protein model selection procedure provided in the RAxML program website (http://sco.h-its.org/exelixis/resource/download/software/ProteinModelSelection.pl). This script automatically determines the best-scoring model from all available (18 amino acid substitution matrices with/without amino acid frequencies estimation) to use with RAxML. The best-fit models for the genes were as follows: MTZOA for atp6 (ATP synthase F0 subunit 6), atp9, cob (apo-cytochrome b), cox2 (cytochrome c oxidase subunit 2), cox3, nad1 (NADH dehydrogenase subunit 1), nad2, nad3, nad4, nad5, JTTDCMUT for atp8, rpl16 (ribosomal protein subunit L16), rpl20, rps3 (ribosomal protein subunit S3), rps11, rps12, CPREV for cox1, nad4L, nad6, sdh3 (succinate-cytochrome c oxidoreductase subunit 3), FLU for sdh2, and JTT for sdh4, secY, ymf39. The independent evolutionary models for the partitions of each gene were unlinked by ‘-q’ options. The ‘-f a’ option was used in RAxML for the simultaneous search for the best likelihood tree and rapid bootstrap analysis with ‘-# 1,000’ (1,000 bootstrap replications) with default options of ‘-i’ (automatically optimized SPR branch rearrangement for heuristic search) and ‘-c’ (25 distinct rate categories).

To test concordance of monophyletic groups between different data sets, we prepared 23 sub-data sets from the c24p (table S3) based on three criteria. First, based on the mean amino acid (AA) P-distance of each gene alignment that was calculated with PAUP* v.4.0b10 (Swofford 2002), 24 CDSS were sorted by mean values (ranging from 0.046 [atp9] to 0.621 [rpl20]) and then the genes with high mean P-distances
were incrementally removed for the sub-data sets by 0.05 intervals (PV1–8). Second, 24 CDSS were ordered by AA length of gene (ranging from 156 [atp8] to 684 [nad5]), and the genes were incrementally removed for the sub-data sets by 50 AA length intervals (PL1–8). Third, 24 CDSS were sorted by translation counts as listed in supplementary table S3, Supplementary Material online. The electron transport, oxidative phosphorylation protein (PE), and ribosomal protein (PR) genes were grouped as sub-data sets that include PE (atp-cob-cocx-nad-sdh-secY-ymf39), PE1–5 (atp6-9, cob-cox1-3, nad1-6, sdh2-4, and secY-ymf39), and PR (rpl-rps). We compared the ML bootstrap values among the full c24p and 23 sub-data sets (100 resampling for bootstrap analysis; supplementary table S2 and fig. S2, Supplementary Material online).

Bayesian Inference (BI) was done using MrBayes v.3.2.4 (Ronquist and Huelsenbeck 2003) using the best-fit model for the data set of the c24p. Preliminary searches under the mixed aamodel (prset aamodel=mixed;) suggested the JONES amino acid substitution model as the best-fit for BI. The JONES + G model was used as a fixed model [prset aamodel=fixed(jones)] for all partitions. All model parameters were unlinked for each partition by using the program settings [prset applyto=(all) ratepr=variable; lset applyto=(all) rates=gamma ngammacat=4; unlink revmet=(all) statefreq=(all) shape=(all)]. Twenty million Metropolis-coupled Markov Chain Monte Carlo (MCMC) simulations were completed with the following parameters: Two independent runs with different random start points, one cold chain and five heated chains for each run, and tree sampling at every 1,000th generation. The Bayesian Posterior Probabilities (BPP) of the monophyletic nodes were calculated using trees (total 2,002) post burn-in; that is, 10 million generations.

Alternative tree topologies were evaluated using the Approximately Unbiased (AU) and Shimodaira Hasegawa (SH) tests (Shimodaira 2002) implemented in the CONSEL v.1.20 package (Shimodaira and Hasegawa 2001). All 24 CDS combined data (c24p) were used for the paired-site test with the best phylogeny and alternative phylogenies. We selected five alternative relationships of rhodophyceanic orders based on sub-data set analyses that had more than 50% of nodes supported with greater than 80% maximum-likelihood bootstrap (MLB) (supplementary fig. S2, Supplementary Material online). The test was performed with 100,000 bootstrap replicates under the same amino acid model settings and partitions as used in the best ML search. We repeated the test three times to confirm the reproducibility of the analysis.

Results and Discussion

Characteristics of 11 Novel mt Genomes

We determined 11 mt genomes from representatives of the 4 florideophycean subclasses; that is, Ahnfeltiophycidae, Hildenbrandiophycidae, Nemaliophycidae, and Rhodymeniophycidae (table 1 and supplementary fig. S1, Supplementary Material online). These data were combined with existing mt genomes from 10 bangiophycean, 2 cyanidiophycean, and 14 florideophycean taxa, including 6 genomes recently determined by our group (Kim, Yang, Yi et al. 2014; Kim, Yang, Boo, et al. 2014; Yang, Kim, Boo, et al. 2014; Yang, Kim KM, Kim SY, et al. 2014; Kim et al. 2015; Lee et al. 2015). This resulted in a database of 37 mt genomes from red algae for the phylogenetic and comparative analyses.

The newly determined mt genomes have a compact gene arrangement with 44–52 predicted genes (i.e., 22–26 protein-coding, 2 rRNAs, and 18–25 tRNAs; supplementary fig. S1, Supplementary Material online), which are shared among most red algae (Hancock et al. 2010; Zhang et al. 2012; Yang, Kim KM, Kim SY, et al. 2014). Differences were found in the absence of a few genes including atp8, nad4L, rpl20, sdh3, sdh4, secY, and ymf39, for example, the lack of atp8 in Ahnfeltia, Hildenbrandia, and Plocamium (see supplementary table S3, Supplementary Material online). The rfl (large subunit rRNA) of Ahnfeltia, Hildenbrandia, and Palmaria includes an intron (with/without an intronic ORF); the cox1 gene contains an intron with an ORF in Ahnfeltia and Hildenbrandia and without an ORF in Plocamium; and tml in all species includes an intron except in Ceramium, Gelidium, and Hildenbrandia (supplementary fig. S1 and table S3, Supplementary Material online). The mt genomes of the Bangiophyceae were larger in size (i.e., min. = 29.1 kb, max. = 43.5 kb, average = 37.4 kb) and had a higher GC-content (min. = 30%, max. = 33.5%, average = 32.2%) than those of the Florideophyceae (size: min. = 24.9 kb, max. = 33.1 kb, average = 26.9 kb; GC-content: min. = 23.6%, max. = 36.4%, average = 28.7%). Within the Florideophyceae, H. rubra contained the largest mt genome and the highest GC content (33.1 kb, 36.4% GC; see table 1).

Mitochondrial Genome-Based Red Algal Phylogeny

Phylogenetic analysis of concatenated mitochondrial data (c24p) from 37 red algal species strongly supports the currently accepted relationships of Rhodophyta (fig. 1). These include a close relationship of Bangiophyceae and Florideophyceae (Müller et al. 2001; Saunders and Hommersand 2004), the recognition of each of the five subclasses, and intersubclass relationships within the Florideophyceae (Le Gall and Saunders 2007; Verbruggen et al. 2010).

The most notable finding of the phylogenetic analysis was the resolution of ordinal relationships among 21 species of Rhodymeniophycidae (fig. 1). The subclass Rhodymeniophycidae is the most morphologically diverse of the red algae and includes greater than 5,015 reported species under the current 12-order taxonomic system (Guiry MD and Guiry GM 2015). The lineages in this subclass likely underwent a rapid radiation as indicated by the mostly short and dense
intermediate nodes among the orders, with only a few ordinal relationships resolved with weak to moderate statistical support (Le Gall and Saunders 2007; Verbruggen et al. 2010). Our mtDNA phylogeny nonetheless provides novel insights into the relationships of some of the rhodymeniophycidan groups. Within the subclass, the Bonnemaisoniales (A. taxiformis) diverged first and is sister to the remaining 11 orders with high support (i.e., 100% MLB and 1.0 BPP) in the c24p data set (node-a) and in the sub-data sets (15 sub-data sets support node-a with >91% MLB; supplementary fig. S2, Supplementary Material online). The Peyssonneliales (Riquetophycus sp.) and Gigartinales (Chondrus and Kappaphycus) grouped together with strong support (node-b in fig. 1, 100% MLB and 1.0 BPP) using c24p and 15 sub-data sets (>91% MLB in supplementary fig. S2, Supplementary Material online) in this restricted taxa data set of Gigartinales sensu lato. However, the monophyly of node-c (11 orders) was less well supported, that is, 75% MLB and 1.0 BPP in the c24p. The topology of sub-data sets (i.e., the sister relationship of Asparagopsis and Kappaphycus as shown in fig. 1 as alt-1) was not rejected in both AU and SH tests (supplementary fig. S3C, Supplementary Material online).

A larger number of taxa are likely required to resolve the phylogenetic relationships of the Bonnemaisoniales, Gigartinales, and Peyssonneliales (Krayesky et al. 2009).

After the divergence of the Gelidiales (Gelidium spp.; 94% MLB and 1.0 BPP; node-d), two monophyletic clades (node-e and node-h) were found, that is, node-e: Plocamiales (Plocamiocolax pulvinata and P. cartilagineum) + Ceramiales (C. japonicum) (node-f) + Acrosymphytales (S. schousboei) clade (58% MLB and 0.59 BPP) and node-h: Rhodymeniales (Rhodymenia pseudopalmata) + Sebdeniales (Se. flabellata) (node-k) + Halymeniales (Grateloupia spp.) (node-l) + Gracilariales (Gracilaria spp., Gracilariopsis, and Gracilariotheca onychoides) + Nemastomatales (Schiz. dubyi) (node-i) clade (67% MLB and 0.99 BPP). Interordinal relationships of each clad are mostly congruent with results from prior studies (Le Gall and Saunders 2007; Le Gall et al. 2010; Verbruggen et al. 2010). However, interordinal relationships of the Acrosymphytales, Ceramiales, and Plocamiales were unresolved and an alternative topology (alt-2 in fig. 1) was not rejected by the tree topology tests (supplementary fig. S3B, Supplementary Material online). Because the Ceramiales is the largest group of red algae including greater than 2,655 species (Guiry MD and Guiry GM 2015), additional mt genome data from representative taxa of the order (such as an early branched Spyridia) may provide better resolution of this node.

Mt genome data included in this analysis resolved the relationship of 16 of the 25 florideophycean orders (Guiry MD and Guiry GM 2015). It has generally been difficult to resolve class-level relationships among red algae (i.e., Compsopogonophyceae, Porphyridiophyceae, and Stylonematophyceae) and intraclass relationships (i.e., Nemaliophycidae and Rhodymeniophycidae) using multigene data (Saunders and Hommersand 2004; Yoon et al. 2006; Le Gall and Saunders 2007; Verbruggen et al. 2010; Scott et al. 2011). However, our analysis using mtDNA data (data set c24p and 23 sub-data sets) provides moderate to strong support for intraclass relationships in the Rhodymeniophycidae. Based on these results, we expect that mt genome data from major representatives could be a promising tool to further resolve the relationships of the red algae.

Mitochondrial Genome Evolution in Red Algae

To infer mt genome evolution in the red algae, we compared gene synteny from 35 genomes (excluding the highly divergent extremophiles Cyanidioschyzon merolae and Gallidaria sulphuraria) that represent two classes (Bangiophyceae and Florideophyceae), all 5 of the florideophycean subclasses, and all 12 rhodymeniophycidan orders (fig. 2). These data (supplementary fig. S1, Supplementary Material online) encompass the phylogenetic breadth of the Florideophyceae.

The overall structure of mt genomes was mapped onto the best phylogeny (fig. 1) to infer the ancestral genome content, patterns of gene acquisition and loss, and types of rearrangements (i.e., numbers on fig. 2). Because of a conserved linear arrangement of three adjacent genes around mt rRNAs (nrs [small subunit rRNA]-nad4L-rrl, approximately 4,300 bp in length excluding intron regions), we used it as 5′- and 3′-anchor points of the mtDNA maps (NR segment in fig. 2). We infer that a total of 51 mitochondrial genes were present in the hypothetical last common ancestor of Bangiophyceae and Florideophyceae (Ra in fig. 2) including 24 protein-coding genes, 2 rRNAs, and 25 tRNAs (supplementary table S4, Supplementary Material online).

The gene content of Bangiophyceae and Florideophyceae is nearly identical in terms of protein-coding genes (supplementary table S3, Supplementary Material online). However, two major syntenic differences involving inversion/translocation events were found between Bangiophyceae (e.g., Porphyra) and Hildenbrandiophyceae (Hildenbrandia) (see no. 1–12 in fig. 2) and between Hildenbrandiophyceae and Nemaliophyceae (Palmaria) (see no. 14, 15, 17–19). The majority of rearrangements occurred through gene inversions (12 inversions of 17 rearrangements, fig. 2) between Bangiophyceae and Hildenbrandiophyceae, and between Hildenbrandiophyceae and Nemaliophyceae. Rearrangements of a single gene were more frequent than simultaneous multigene rearrangements (i.e., more than two genes involved; i.e., genes no. 5, 7, 10, and 14).

The trnl gene (without intron; no. 6) exists between two large conserved regions, that is, nad2-sdh4-nad4-nad5-atp8-atp6 (ANS segment; approximately 6,500 bp) and ymf39-cox3-cox2-cox1 (no. 7 CY segment; approximately 4,000 bp).
in length when the intron and intronic ORF are excluded) in the bangiophycean lineage. In the Florideophyceae, the CY segment is inverted (5'-cox1-cox2-cox3-ymf39-3'), whereas the ANS segment is retained with the trnI-intron in all species (no. 21) but not in Hildenbrandia (no. 6; 5 bp intergenic region between nad4 and nad5), Gelidium (no. 28), and Ceramium (no. 34). This indicates that the trnI gene was present in the common ancestor of Bangiophyceae and Florideophyceae.

**FIG. 1.**—Phylogeny of the red algae (Rhodophyta) based on mt genome data. Tree constructed using the ML method based on 24 concatenated genes (6,345 amino acids from 24 protein coding). The support values for each node are calculated from MLB and BPP. Asterisks after species names indicate newly determined mt genomes, followed by red algal order, subclass, and class system. Alternative tree branch position of species (alt) indicated by dot line with arrow, that is, alt-1 for Asparagopsis and alt-2 for Schimmelmania. Alternative topologies are available in supplementary figure S3, Supplementary Material online.
Mitochondrial genome evolution in red algae. Mitochondrial gene order of 27 red algal species is illustrated along the best-supported phylogeny and associated taxonomy. Three classes (BANGIOPHYCEAE, CYANIDIOPHYCEAE, and FLORIDEOPHYCEAE) and five subclasses (Ahnfeltiophycidae, Corallinophycidae, Hildenbrandiophycidae, Nemaliophycidae, and Rhodymeniophycidae) of the
(Ra in fig. 2, supplementary table S4, Supplementary Material online). In the Bangiophyceae, the ANS and CY segments were present in all bangiophycean species with/without intron and intronic-ORF in cox1. However, in the Florideophyceae, the ANS segments acquired an intron in trnl and a reduced intron and intronic-ORF in cox1 of the CY segment. The trnl has gained an intron and it is inserted in the ANS segment (i.e., between nad4 and nad5) in the florideophycean ancestor (Rd). Thus, the trnl-intron was lost independently in Hildenbrandia, Gelidium, and Ceramium. This trnl-mediated intron acquisition is the first report of such an event in red algae; however, acquisition of heterogeneous DNA (protein-coding genes and introns) has been reported in the Viridiplantae, that is, in green algae and land plants (Palmer et al. 2000; Berghthorsson et al. 2004; Smith et al. 2013).

Since the split of the hildenbrandiophycidan ancestor (i.e., node Rd in fig. 2), gene synteny has been highly conserved, with changes only occurring in tRNAs and in intronic regions. Lineage-specific CDS gene losses were found in florideophycean species (red numbers in fig. 2). For example, there were losses of atp8 (no. 16), rpl20 (no. 20), trnl (no. 6), and ymf39 (no. 13) in Hildenbrandia, whereas losses of rps11 (no. 22) and rpl20 (no. 23) occurred in Sporolithon. In contrast to the conserved CDS gene content (except the rps11 inversion in Gracilariosis andersonii; no. 42), there are several types of variation in tRNAs (i.e., gene duplication and insertion/deletion, see fig. 2 and supplementary table S3, Supplementary Material online) not only between the subclasses but also within the Rhodymeniophycidae in tmnH (H—H gray regions around tmnL-nad6-trnG) and/or tmY and one of each tmnR and tmnS (Y-S-R-Y gray region between tRNA and tmN).

In this study, the new florideophycean mt genome data surprisingly show that overall gene synteny is conserved among subclasses, except for the Hildenbrandiaceae. Based on the fossil record and molecular clock studies, the florideophycean ancestor (Rd) originated approximately 900 Ma and the nemaliophycidan ancestor (Ra) appeared at least 550 Ma (Corallinocystidae fossils) (Butterfield 2001; Xiao et al. 2004; Yoon et al. 2004). Despite the long evolutionary history of the florideophycean red algae, most are characterized by a highly conserved mtDNA synteny (fig. 2).

**Reduction of Protein-Coding Genes**

Based on our current collection of mt genomes, at least 51 genes (excluding hypothetical proteins) were present in the common ancestor (Ra) of Bangiophyceae and Florideophyceae (supplementary table S4, Supplementary Material online). We inferred a total of 29 lineage-specific gene losses during red algal evolution (no. 6, 13, 16, 20, 22–41, 43–47 in fig. 2). We marked these gene losses on the map of percent AT-content versus gene length, which shows a negative correlation between these values (fig. 3). Surprisingly, genes that were absent in some species were grouped in the top-left corner. These genes are small in size (<800 bp in mean length) and have a high AT content (>73% mean AT content). The rpl20 (relatively short with the highest AT content) was lost in 18 species among 35 investigated red algae (except Cyanidiophyceae), and this was followed by gene losses of ymf39 (6 species), atp8 (5 species), sdh3 (3 species), sdh4 and nad4L (2 species), and rps11 and secY (1 species). Gene loss appears however to be uncorrelated with phylogenetic relationships (see, red number on fig. 2). For example, within the strongly supported clade of the Halymeniales, Sebdeniales, and Rhodymeniales (100% MLB), rpl20 was lost only in R. pseudopalma (no. 47), whereas it was retained in only three species. Interestingly, 9/10 bangiophycean species lost the rpl20 gene, except for Porphyra umbilicalis. It is reasonable to infer the rpl20 data as representing four independent losses in Porphyra purpurae, B. fusco-purpurea, the ancestor of six Pyropia species, and Wildelmania schizophylla, respectively, as opposed to one loss in the ancestor of the Bangiophyceae (Rb) with a regain of rpl20 because it is extremely improbable to regain a gene in the correct reading frame after its loss.

There are three genes located near the boundary of the gene loss; that is, nad6, rps3 and rpl16, which are small in size (612, 691, and 419 bp mean length, respectively) and have a
high AT-content (73.6%, 74.1%, and 73.5%, respectively; fig. 3) that are retained in all 35 bangiophycean and florideophycean red algal mt genomes. Based on the absence of the ribosomal and succinate dehydrogenase genes (rpl, rps, and sdh) in land plants (Adams et al. 2002; Notsu et al. 2002; Adams and Palmer 2003), and the trend of gene loss in red algae, rpl16 and rps3 may be candidates for future gene loss (or perhaps already lost in some yet to be studied rhodophytes) than nad6. Meanwhile, many lost genes (rpl20, rps11, sdh4, secY, and ymf39) and candidates for loss (rpl16, rps2, rps3, and sdh2) were lost in G. sulphuraria but retained in Cy. merolae (Cyanidiophyceae; fig. 2, supplementary table S3, Supplementary Material online). The pattern that high AT-content genes are more frequently lost than low AT-content ones in red algae is an intriguing one, for which we currently have no explanation for the mechanisms. It may be a phenomenon restricted to this phylum, and further study is needed.

The lost genes have most likely been replaced by copies that were transferred to the nucleus (Lang et al. 1999). In red algae, G. sulphuraria is the only species with completed nuclear (Schönknecht et al. 2013) and mt genomes, and the mt genome has eight CDS losses (fig. 2, supplementary table S3, Supplementary Material online). We found one of the eight CDSs, mitochondrial-derived sdh2 gene, in the nuclear genome of G. sulphuraria (XM_005707240.1). G. sulphuraria succinate dehydrogenase [ubiquinone] iron-sulfur protein [Gasu_20110] mRNA, complete CDS, 275 AA;
E value = 5e-105; 93% Query coverage, 59% identity) by using tBLASTn search with xdh2 of *Cy. merolae* (gi|8954361|ref|NP_059350.1; 265 AA) on the National Center for Biotechnology Information. However, other genes showed matches with corresponding plastid encoded genes (*rpl16, rps3, rps11, rps12*) or very short regions of similarity to nuclear genes (*rpl20, shd4, ymf39*). Additional genomes from Florideophyceae and mt-driven gene transfer to the nucleus are needed to test the genome reduction and gene transfer scenario presented here.

Among all reported mt genomes shown in figure 2, group II introns are present in all species except *Ceramium, Cyanidoschyzon, Galdieria, and Gelidium* (Dai et al. 2003; Mao et al. 2012; Smith et al. 2012; Yang, Kim, Boo, et al. 2014). It is interesting, however, that only three genes (*rrl, cox1, and trnI*) contain introns with or without intronic-ORFs (Leblanc et al. 1995; Hancock et al. 2010; Hwang et al. 2013; Kim, Yang, Boo, et al. 2014; Yang, Kim, Boo, et al. 2014). In the bangiophycean species (*Bangia, Porphyra, Pyropia, and Wildemania*), one or two copies of *rrl*-introns are present (Smith et al. 2012; Hwang et al. 2013), whereas these tend to be lost in most florideophycean taxa (i.e., only present in *Hildenbrandia, Palmaria*, and *Ahnfeltia*). A similar pattern is found for the *cox1* intron, where 1–3 copies are present in bangiophycean taxa but sporadically present in the Florideophyceae (supplementary table S3, Supplementary Material online). In contrast, the florideophycean lineage acquired the *trnI*-intron without the intronic-ORF after divergence of the Hildenbrandiophyceae.

The distribution of introns in red algae together with the 11 novel mt genomes suggests that the majority of red algal introns (*rrl* and *cox1*) originated prior to the split of the two major classes (*Ra in fig. 2; the last common ancestor of Bangiophyceae and Florideophyceae*). Subsequent intronic-ORF loss resulted in pseudogenization of intronic-ORFs, followed by intron loss. Intron gain, however, may occur with an ORF-lacking intron, as seen for the *trnI*-intron, which is an apparent case of a recent intron gain.

**Conclusions**

Our study elucidates mt genome evolution in the Florideophyceae through analysis of this genome in *Ahnfeltia, Hildenbrandia, Palmaria*, and diverse rhodymeniophycean species. The phylogeny based on concatenated amino acid mt genome data supports the current taxonomic system for Bangiophyceae and Florideophyceae. We demonstrate that all Florideophyceae mt genomes have compact and conserved architectures but they also contain several unique features, such as gene rearrangements that discriminate lineages and the gain of group II introns in *trnI* genes. Comparative analyses suggest that the mtDNA in Florideophyceae (except the Hildenbrandiophyceae), although representing a wide variety of morphologically divergent taxa, is surprisingly highly conserved with little variation in gene composition and with shared synteny across the class. This result is consistent with a rapid radiation of the class followed by strong constraints on mitochondrial gene content and genome evolution.

**Supplementary Material**

Supplementary figures S1–S3 and tables S1–S4 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

**Acknowledgments**

The authors thank to Ms Sun Jeong Lim for data analysis and Dr Ju Yun Bae and Dr Joon Park (Life Technologies) for technical support. This work was supported by grants from the Basic Science Research Program through the National Research Foundation of Korea (NRF) founded by the Ministry Science, ICT and Future Planning (NRF-2014R1A1A01006512) to E.C.Y., the Korean Rural Development Administration Next-generation BioGreen21 (PJ011121), the National Research Foundation of Korea (MEST: 2014R1A2A2A01003588), and Marine Biotechnology Program (PJ700620) funded by Ministry of Oceans and Fisheries, Korea to H.S.Y., National Science Foundation RedToL (DEB 0936884, 0936855, 1317114) to D.B., S.F., and H.S.Y., Korea Research Foundation (2012-0704) to S.M.B., a Marine and Extreme Genome Research Center Program grant of the Ministry of Oceans and Fisheries to J.H.L., a GoMRH (CVC) award to S.F., and NIWA Core funding to W.A.N. The authors declare that they have no competing interests.

**Literature Cited**


Mitochondrial Genome Evolution of Red Algae


Associate editor: John Archibald