

RESEARCH ARTICLE

Open Access



# Complete chloroplast genome of *Gracilaria firma* (Gracilariaceae, Rhodophyta), with discussion on the use of chloroplast phylogenomics in the subclass Rhodymeniophycidae

Poh-Kheng Ng<sup>1</sup>, Showe-Mei Lin<sup>1\*</sup>, Phaik-Eem Lim<sup>2\*</sup>, Li-Chia Liu<sup>1</sup>, Chien-Ming Chen<sup>3</sup> and Tun-Wen Pai<sup>3</sup>

## Abstract

**Background:** The chloroplast genome of *Gracilaria firma* was sequenced in view of its role as an economically important marine crop with wide industrial applications. To date, there are only 15 chloroplast genomes published for the Florideophyceae. Apart from presenting the complete chloroplast genome of *G. firma*, this study also assessed the utility of genome-scale data to address the phylogenetic relationships within the subclass Rhodymeniophycidae. The synteny and genome structure of the chloroplast genomes across the taxa of Eurhodophytina was also examined.

**Results:** The chloroplast genome of *Gracilaria firma* maps as a circular molecule of 187,001 bp and contains 252 genes, which are distributed on both strands and consist of 35 RNA genes (3 rRNAs, 30 tRNAs, tmRNA and a ribonuclease P RNA component) and 217 protein-coding genes, including the unidentified open reading frames. The chloroplast genome of *G. firma* is by far the largest reported for Gracilariaceae, featuring a unique intergenic region of about 7000 bp with discontinuous vestiges of red algal plasmid DNA sequences interspersed between the *nbIA* and *cpeB* genes. This chloroplast genome shows similar gene content and order to other Florideophycean taxa. Phylogenomic analyses based on the concatenated amino acid sequences of 146 protein-coding genes confirmed the monophyly of the classes Bangiophyceae and Florideophyceae with full nodal support. Relationships within the subclass Rhodymeniophycidae in Florideophyceae received moderate to strong nodal support, and the monotypic family of Gracilariales were resolved with maximum support.

**Conclusions:** Chloroplast genomes hold substantial information that can be tapped for resolving the phylogenetic relationships of difficult regions in the Rhodymeniophycidae, which are perceived to have experienced rapid radiation and thus received low nodal support, as exemplified in this study. The present study shows that chloroplast genome of *G. firma* could serve as a key link to the full resolution of *Gracilaria sensu lato* complex and recognition of *Hydropuntia* as a genus distinct from *Gracilaria sensu stricto*.

**Keywords:** Chloroplast genome, *Gracilaria firma*, Gracilariaceae, Conserved synteny, Phylogenomic analyses, Red algal plasmid remnants

\* Correspondence: linsm@ntou.edu.tw; phaikem@um.edu.my

<sup>1</sup>Institute of Marine Biology, National Taiwan Ocean University, Keelung 20244, Taiwan

<sup>2</sup>Institute of Ocean and Earth Sciences, University of Malaya, Kuala Lumpur 50603, Malaysia

Full list of author information is available at the end of the article



## Background

Rhodophyta is a monophyletic phylum currently divided into seven classes, including Bangiophyceae, Compsopogonophyceae, Cyanidiophyceae, Florideophyceae, Porphyridiophyceae, Rhodellophyceae and Stylonematophyceae [1, 2]. The Florideophyceae which accommodates more than 6700 species of red algae [3] has only chloroplast genomes of 15 species published to date. Despite the sporadic studies [4–6] on the phylogenetic relationships among the Florideophyceae inferred using chloroplast genome data, the whole chloroplast genome has been demonstrated to be a promising resource to resolve the red algal relationships, attributable to the conserved nature of the slowly-evolving genome [4]. In line with that, emergence of the next generation high-throughput sequencing technologies and the associated exponential decline in the cost of sequencing would open up the opportunity for more whole genome sequencing projects targeted to gain more novel insights into the evolutionary relationships within the red algal lineage.

*Gracilaria firma* Chang et Xia is an agar-producing seaweed distributed in the tropical and subtropical regions in the western Pacific [7–9]. It is an economically important marine crop that has been cultivated on commercial scale in several countries including Taiwan [10], Vietnam [11] and the Philippines [12]. The seaweed exemplifies superior growth and agar quality among other investigated Gracilarioid agarophytes [13]. Apart from serving as the source of agar which seeks versatile applications in food and pharmaceutical industries, *G. firma* is also harvested and made into local delicacies for human consumption [12]. Some regional abalone farms preferred using *G. firma* as the natural feed [10].

This study presents the complete chloroplast genome of *G. firma*, adding to the number of chloroplast genome available for the genus *Gracilaria*, one of the largest red algal genera which encompassed more than a hundred species [3] and has undergone numerous taxonomic revisions with contradicting conclusions [14–16]. Analysis of the synteny and genome structure concurred that the red algal chloroplast genomes are very compact and conserved across the subphylum Eurhodophytina, with different orders exhibiting syntenies that discriminate lineages. Phylogenomic analyses including most of the recently available (as of the date of writing this manuscript in June 2016) taxa across eight orders of Florideophyceae were conducted. The relevance of chloroplast genomic data in addressing the phylogenetic relationships at different hierarchical levels in Rhodymeniophycidae was also discussed.

## Methods

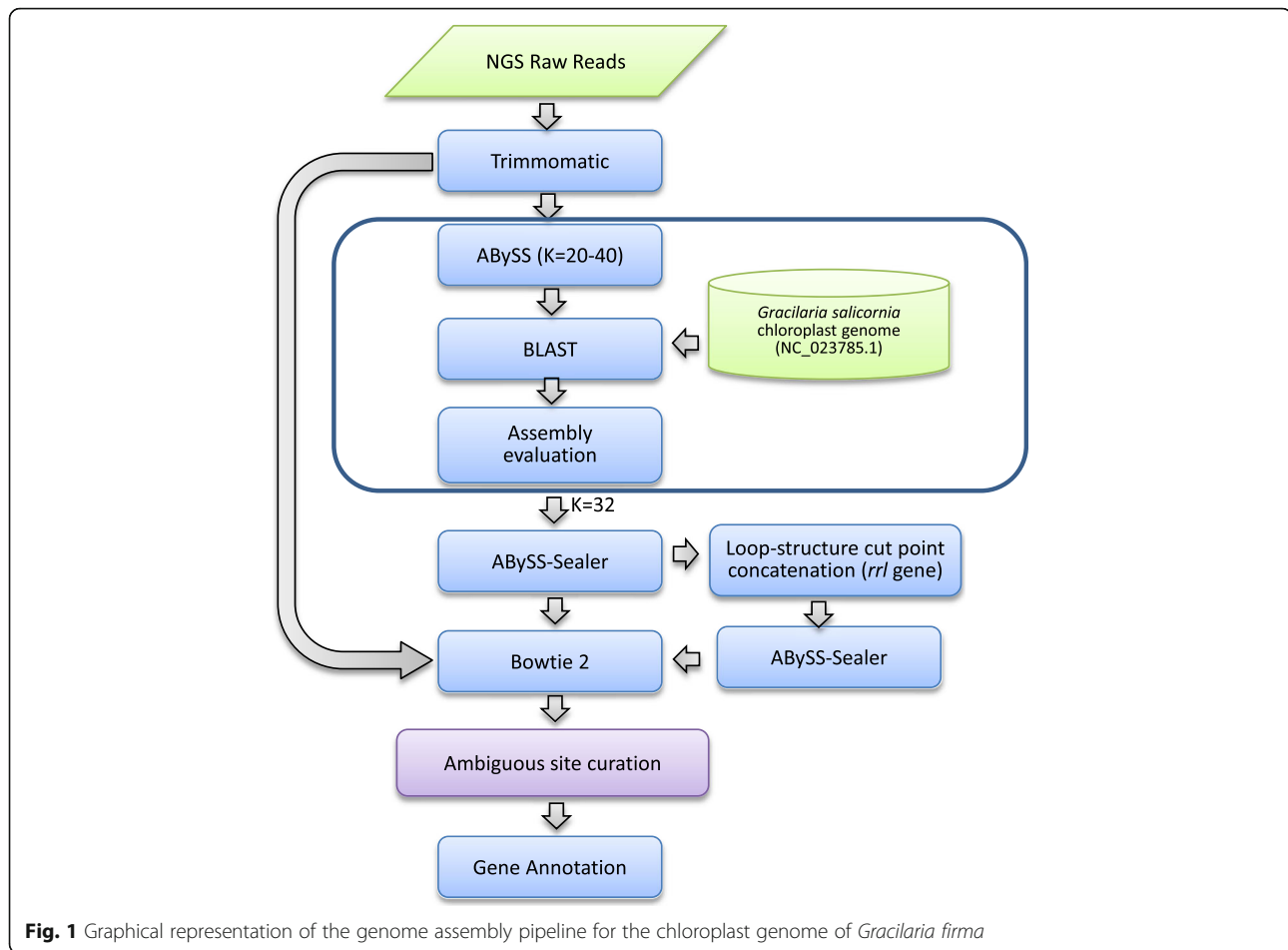
### Taxon sampling and sequencing

The plant material of *G. firma* was procured from a cultivation farm operated in Kouhu Township, Yunlin

County, Taiwan. Genomic DNA was extracted from fresh thallus of *G. firma* using the DNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instruction and sent to a company for library prep and sequencing (ScienceVision Sdn Bhd, Selangor, Malaysia). A corresponding voucher specimen is deposited in the Herbarium of the National Taiwan Ocean University, Taiwan (NTOU) under the accession number NTOU-KH-5i2016-Gf. The library was prepared using a Nextera XT kit (Illumina, San Diego, CA, USA) and sequenced with 250 bp pair-end reads on the MiSeq sequencing platform (Illumina, San Diego, CA, USA).

### De novo assembly and annotation of the *Gracilaria firma* chloroplast genome

The genome assembly pipeline is shown in Fig. 1. The sequenced NGS raw reads were stored in Fastq format with a total of 7,962,495 read pairs, and trimmed using Trimmomatic v0.36 [17] with the following parameter settings: LEADING:15, TRAILING:15, SLIDINGWINDOW: 4:15 and MINLEN:36. A total of 7,490,913 read pairs (94.1%) survived filtering. The genome was first assembled using the ABySS *de novo* assembler v1.9.0 [18] with different k-mer parameters varying from 20 to 40, and the best assembly results were determined based on the information content estimates (total number of gene BLAST hits against the reference genome of *G. salicornia* (NC\_023785)). A single contig of 186,354 bp with 150 mapped genes, assembled using the k-mer size of 32, was identified as the probable candidate of the complete chloroplast genome. Reads were also mapped back to the contig using Bowtie 2 [19] to validate the assembly. The contig was further refined using the Sealer tool [20] implemented in ABySS to verify the gaps represented by character *N* within the original ABySS output. Manual curation of the ambiguous sites reported from ABySS Sealer was performed via the visualization tool IGV v2.3 [21] based on the reference mapping results. As circularity was not inferred in the refined contig, the gap in the *rri* gene was closed by applying the ABySS Sealer and manual ambiguous site curation to the concatenation of the end segments of the contig and a segment of *N* sequences (50 bases). The final genome was validated by mapping the reads against the refined contig, where 1,140,489 of the 14,981,826 filtered reads mapped to the chloroplast genome with a mean coverage of 996×. An independent genome assembly was also conducted with the IDBA-UD *de novo* assembler v1.0.9 [22] using  $\text{mink} = 35$  and  $\text{maxk} = 100$ , coupled with the reference-guided ARC assembler v1.1.3 [23] using the reference genome of *G. tenuistipitata* (AY673996) to result in a single contig of 187,249 bp. One of the self-similar ends encoding the *groEL* gene was manually trimmed to yield a contig representative of the circular



chloroplast genome of *G. firma*. Both assembly approaches yielded single contig with identical sequences.

A combination of automated pipelines and manual verification was used to annotate the chloroplast genome of *G. firma*. Initial gene calling was accomplished using DOGMA [24] with a 60% cutoff for protein-coding genes, 80% for RNAs, and an *e*-value cut-off of  $1e-5$  for BLAST hits. The rRNA genes were determined using RNAmmer 1.2 server [25] with the 'kingdom of input sequences' selected as 'Bacteria'. The tRNAs were identified using tRNAscan-SE v1.21 [26] with default parameters and the source identified as 'Mito/Chloroplast'. Intron and tmRNA were searched using ARAGORN [27], while the ribonuclease P gene (*rnpB*) was detected using RNAweasel [28]. Open reading frames (ORFs) longer than 25 amino acids within the intergenic regions were searched using the NCBI's ORF finder (<https://www.ncbi.nlm.nih.gov/orf-finder>), with the genetic code option of 'bacterial, archaeal and plant plastid'. BLASTp homology searches against all the non-redundant protein sequences from GenBank were used to determine the start and stop codon positions for each protein-coding gene, including those detected by DOGMA and the small or missing genes recovered from

ORF finder. The circular genome map was generated using OGDraw [29]. The *G. firma* chloroplast genome sequence and annotation was deposited into GenBank under accession number KX601051.

#### Synten analyses on the red algal chloroplast genomes

Whole-genome alignments were generated using the progressiveMauve aligner implemented in Mauve v20150226 [30] under default settings. The synteny within the Gracilariales was assessed from the alignment of the chloroplast genomes of *G. firma* and related species from the same order, with which the genome sequences were reoriented to have the beginning of the *psaM* gene as the first position for alignment. In view of the variable genome structure as well as the lack of coherence in the designation of the first gene for chloroplast genomes across the Eurhodoptytina, the synteny among the Florideophyceae and Bangiophyceae was assessed by visual inspection of the whole-genome alignments variously generated with the chloroplast genome of *G. firma* serving as the reference to detect the conserved segments of sequence free from any internal rearrangements, also known as the locally collinear blocks (LCBs).

### Phylogenomic analyses

Two datasets were assembled for the phylogenomic analyses based on the chloroplast genome data of 21 taxa of Rhodophyta, including the data newly obtained for *G. firma* (Table 1). Analyses on the protein-coding genes were performed with amino acid sequences to avoid phylogenetic artefacts caused by synonymous substitutions-induced convergent base composition [31]. The first dataset included 79 protein-coding genes present in all taxa, selected based on the recommendation in [4]: *acsF*, *apcA*, *apcB*, *apcD*, *apcE*, *apcF*, *atpA*, *atpB*, *atpF*, *atpG*, *atpI*, *carA*, *ccsA*, *chlI*, *clpC*, *cpcA*, *cpcB*, *cpcG*, *dnaK*, *frbB*, *ftsH*, *groEL*, *infC*, *pdhA*, *pdhB*, *petA*, *petB*, *petF*, *petJ*, *preA*, *psaA*, *psaB*, *psaF*, *psbA*, *psbB*, *psbC*, *psbD*, *psbV*, *rbcR*, *rpl1*, *rpl2*, *rpl3*, *rpl4*, *rpl5*, *rpl6*, *rpl11*, *rpl12*, *rpl12*, *rpl16*, *rpl18*, *rpl19*, *rpl21*, *rpl22*, *rpl23*, *rpoB*, *rpoC1*, *rpoC2*, *rps2*, *rps3*, *rps4*, *rps5*, *rps7*, *rps10*, *rps11*, *rps12*, *rps13*, *rps14*, *secY*, *sufB*, *sufC*, *tatC*, *thiG*, *trpA*, *trxA*, *tsf*, *tufA*, *ycf4*, *ycf39* and *ycf65*. These genes are considered as the most promising candidate

genes for phylogenetic, barcoding and population studies, as they demonstrated low non-synonymous substitution rates across distantly related red algae. In line with the aim of phylogenomics to reduce random or sampling error with the use of large dataset [32], a larger concatenated dataset representing the expansion of the first dataset with the protein-coding genes present in all examined taxa were also analyzed. The second dataset consisted of 146 protein-coding genes, including the following in addition to the first dataset: *acpP*, *atpD*, *atpE*, *atpH*, *cbbX*, *cemA*, *gltB*, *ilvB*, *ilvH*, *nblA*, *ntcA*, *petD*, *petG*, *psaD*, *psaE*, *psaI*, *psaJ*, *psaK*, *psaL*, *psaM*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbN*, *psbT*, *psbW*, *psbX*, *psbY*, *psbZ*, *rbcL*, *rbcS*, *rpl13*, *rpl14*, *rpl20*, *rpl24*, *rpl27*, *rpl28*, *rpl29*, *rpl31*, *rpl32*, *rpl33*, *rpl34*, *rpl35*, *rpl36*, *rpoA*, *rpoZ*, *rps6*, *rps8*, *rps9*, *rps16*, *rps17*, *rps18*, *rps19*, *rps20*, *secA*, *tilS*, *trpG*, *ycf3*, *ycf19*, *ycf53*, *ycf54*, *ycf60* and *ycf80*.

Amino acid sequences from each individual gene were aligned using MAFFT v7.222 [33] with the 'L-INS-I'

**Table 1** Red algal taxa analyzed in this study and their chloroplast genome composition

Species and GenBank accession number	Reference	Size (bp)	GC (%)	Protein-coding genes	tRNAs	rRNAs
Cyanidiophyceae						
<i>Cyanidioschyzon merolae</i> (AB002583)	[40]	149,987	37.6	207	31	3
<i>Cyanidium caldarium</i> (AF022186)	[55]	164,921	32.7	201	30	3
Porphyridiophyceae						
<i>Porphyridium purpureum</i> (AP012987)	[56]	217,694	30.3	224	29	6
Bangiophyceae						
<i>Porphyra umbilicalis</i> (JQ408795)	[57]	189,933	32.9	209	37	6
<i>Pyropia haitanensis</i> (KC464603)	[44]	195,597	33.1	213	37	6
<i>Wildemanina schizophylla</i> (KR028420)	[58]	193,008	34.4	211	34	4
<i>Bangia atropurpurea</i> (KR028420)	Unpublished	189,505	32.4	206	35	4
Florideophyceae						
Corallinophycidae						
<i>Calliarthron tuberculosum</i> (KC153978)	[4]	178,981	29.2	201	31	3
<i>Sporolithon durum</i> (KT266785)	[6]	191,465	29.3	202	30	3
Rhodymeniophycidae						
<i>Chondrus crispus</i> (HF562234)	[4]	180,086	28.7	204	30	3
<i>Gelidium elegans</i> (KT266786)	[6]	174,748	30.2	203	30	3
<i>Gelidium vagum</i> (KT266787)	[6]	179,853	29.9	201	30	3
<i>Gracilaria firma</i> (KX601051)	This study	187,001	28.1	219	30	3
<i>Gracilaria salicornia</i> (KF861575)	[41]	179,757	28.8	204	31	3
<i>Gracilaria chilensis</i> (KT266788)	[6]	185,637	29.3	203	30	3
<i>Gracilaria tenuistipitata</i> var. <i>liui</i> (AY673996)	[38]	183,885	29.2	207	31	3
<i>Gracilariopsis lemaneiformis</i> (KU179794)	[59]	182,505	27.4	204	31	3
<i>Grateloupia taiwanensis</i> (KC894740)	[42]	191,270	30.6	235	31	3
<i>Coeloseira compressa</i> (KU053957)	[60]	176,291	29.0	201	29	3
<i>Laurencia snackeyi</i> (LN833431)	[39]	174,935	30	200	29	3
<i>Vertebrata lanosa</i> (KP308097)	[43]	167,158	30	192	27	3

strategy. Each individual alignment was subjected to trimming using the program Gblocks v0.91b available at [http://phylogeny.lirmm.fr/phylo.cgi/one\\_task.cgi?task\\_type=gblocks](http://phylogeny.lirmm.fr/phylo.cgi/one_task.cgi?task_type=gblocks) [34] for the removal of ambiguous regions which were poorly aligned or contained gaps under the setting for a more stringent selection that does not allow many contiguous non-conserved positions. The alignments of each individual gene were variously concatenated using Bioedit v7.2.5 [35] to result in two datasets comprised of 20,033 and 27,205 positions each. Maximum likelihood (ML) tree search were implemented in PhyML v3.0 [36], based on the MtZoa + G + I + F and cpREV + G + I + F model automatically selected by the program for the 79-gene and 146-gene datasets respectively. Branch support for both datasets were evaluated using the SH-like approximate Likelihood Ratio Test (SH-aLRT) implemented in PhyML. Only the 146-gene dataset was subjected to bootstrap analysis with 1000 bootstrap replicates. Bayesian inference (BI) was conducted with MrBayes v3.2.6 [37], using the cpREV + G + I + F model on both datasets as the best-fitting MtZoa + G + I + F model deduced for the 79-gene dataset was not implemented in MrBayes, with two parallel independent runs, each of which consisted of one cold chain and three hot chains of Markov chain Monte Carlo iterations for one million generations. The trees were sampled every 100th generation. Convergence of the runs to the stationary distribution was determined by looking at the standard deviation of split frequencies (always less than 0.01) and by the convergence of the parameter values in the two independent runs. The first 25% of the total number of the trees were discarded as burn-in, and the remaining trees were used to calculate a 50% majority rule tree and to determine the posterior probabilities for all datasets. Members of the Cyanidiales, *Cyanidioschyzon merolae* and *Cyanidium caldarium* were designated as the outgroup taxa based on the global phylogenetic searches recovered for the red algae [2].

## Results and discussion

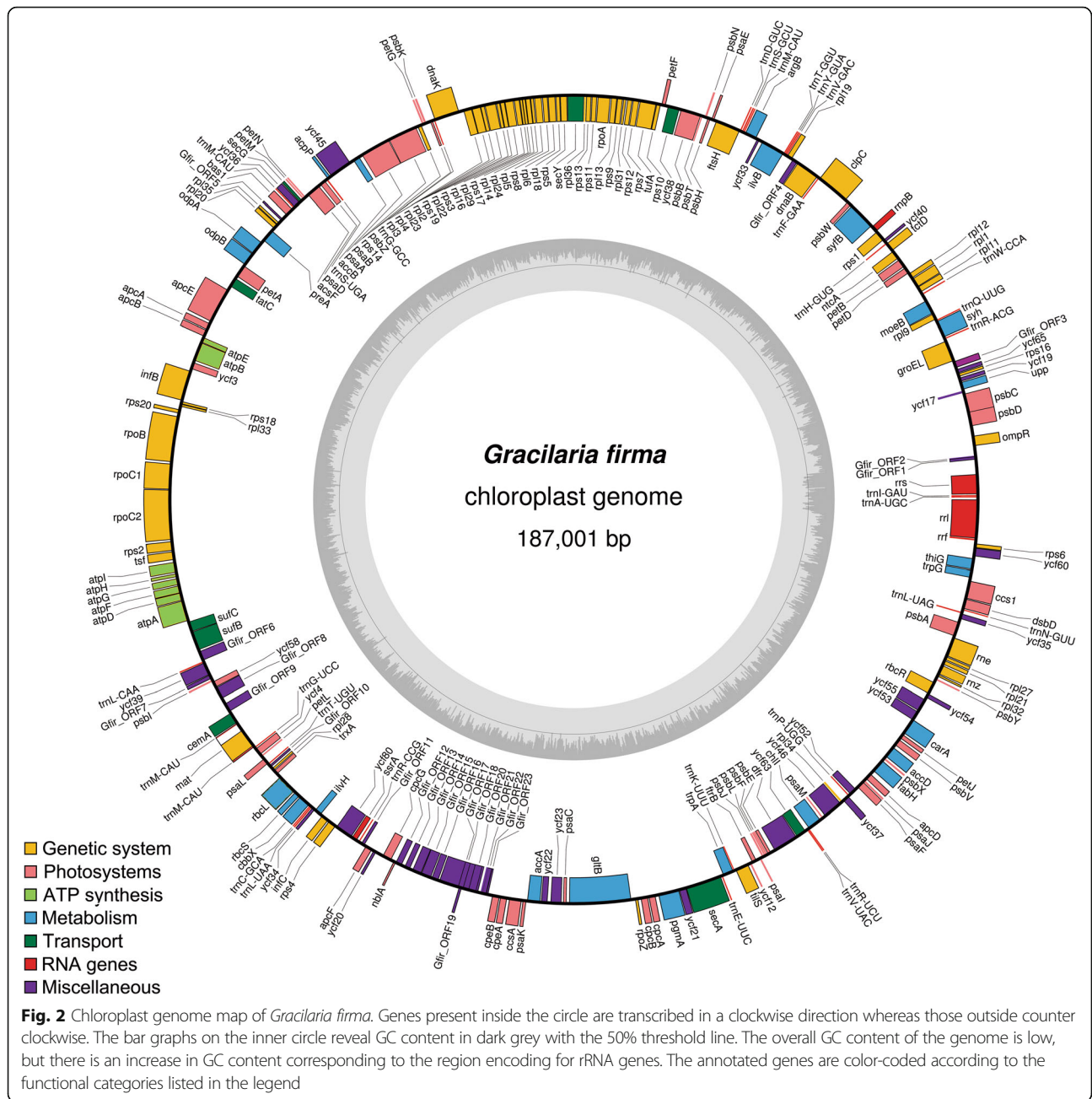
### Characteristics of the chloroplast genome of *Gracilaria firma*

The complete circular chloroplast genome of *G. firma* is 187,001 bp in length. *Gracilaria firma* has the largest chloroplast genome reported within the order Gracilariales thus far, with the chloroplast genomes of *G. salicornia*, *G. chilensis*, *G. tenuistipitata* var. *liui* and *Gracilariopsis lemaneiformis* each reporting the size of 179, 185, 183 and 183 kb. The genome has an overall GC content of 28.1% (Fig. 2). In contrast to the tRNA and rRNA genes which are richer in GC content (51.3 and 44.8% each on average), the protein-coding regions and intergenic spacers accounted for the generally low overall GC content (28.5 and 16.5% each on average). The genome size and base composition of *G. firma* is comparable to those reported for other chloroplast

genomes in red algae and followed the typical AT-rich trend (Table 1). The standard start codon ATG was used in 90.8% of the ORFs of the chloroplast genome. Alternative start codons were also used to initiate translation, including ATC for *Gfir\_ORF15*; GTG for *rps8*, *petM*, *rbcS*, *infC* and *psaF*; TTG for *ompR*, *ycf65*, *trxA*, *ycf20*, *Gfir\_ORF13*; ATA for *tctD*, *rpl22*, *infB*, *Gfir\_ORF11*, *Gfir\_ORF17* and *Gfir\_ORF23*; and ATT for *Gfir\_ORF3*, *Gfir\_21* and *chlL*. TAA was the most commonly used stop codon (72.4%), followed by TAG (15.7%) and TGA (12.0%). Despite the rather large variation in genome size, the chloroplast genome of *G. firma* demonstrated an extensive conservation of synteny to other members of Gracilariales, as evidenced by the single LCB resulted from the chloroplast genome alignment (Fig. 3).

Genetic information is densely packed in the chloroplast genome of *G. firma*, with the combined coding regions spanning 83.5% of the genomic sequence. The intergenic spacers were 122 bp long on average and overlapping ORFs between seven gene pairs were observed, including *psbC-psbD* (7 bp), *rpl24-rpl14* (1 bp), *rpl14-rps17* (4 bp), *rps17-rpl29* (4 bp), *rpl23-rpl4* (28 bp), *rps18-rpl33* (4 bp) and *atpF-atpD* (4 bp). Such overlapping is not unprecedented in the red algal chloroplast genome. For instance, the overlapping of *psbC-psbD* had been reported in *Gracilariopsis lemaneiformis*, *G. tenuistipitata* var. *liui* and *Laurencia snackeyi* [5, 38, 39]. *Cyanidioschyzon merolae* even showed up to 40% of gene overlapping in the chloroplast genome, including the *rps14-rps17* gene pair [40].

A total of 252 coding regions were predicted for the genome, including three rRNA genes (*rrs*, *rnl* and *rnf*), 30 tRNA genes, a sufficient set for the chloroplast protein synthesis machinery, one tmRNA (*ssra*, tag peptide: AKNNILNLSKQLVVCV) gene, one ribonuclease P RNA component (*rnpB*), and 217 protein-coding genes including unidentified ORFs (Fig. 2, Table 2). As with the chloroplast genome of many other Florideophycean algae, that of *G. firma* featured only a single copy of rRNA operon composed of three rRNA genes and two tRNA genes (*trnI* and *trnA*) sandwiched between the *rrs* and *rnl* genes. Of the 30 tRNA in the chloroplast genome of *G. firma*, three encoded for arginine, two for valine, two for threonine, three for methionine, two for serine, two for glycine, and three for leucine. The other amino acids were encoded by single tRNA gene. One of the tRNA genes encoding methionine had a group II intron encoding a maturase. The presence of this group II intron in the *trnM* gene is not uncommon, as it is regarded as a signature characteristic of most florideophytes [4]. However, the intron can be easily overlooked during annotation [38]. Of the 217 protein-coding genes annotated on the *G. firma* chloroplast genome, 193 are widely conserved and named genes, including the

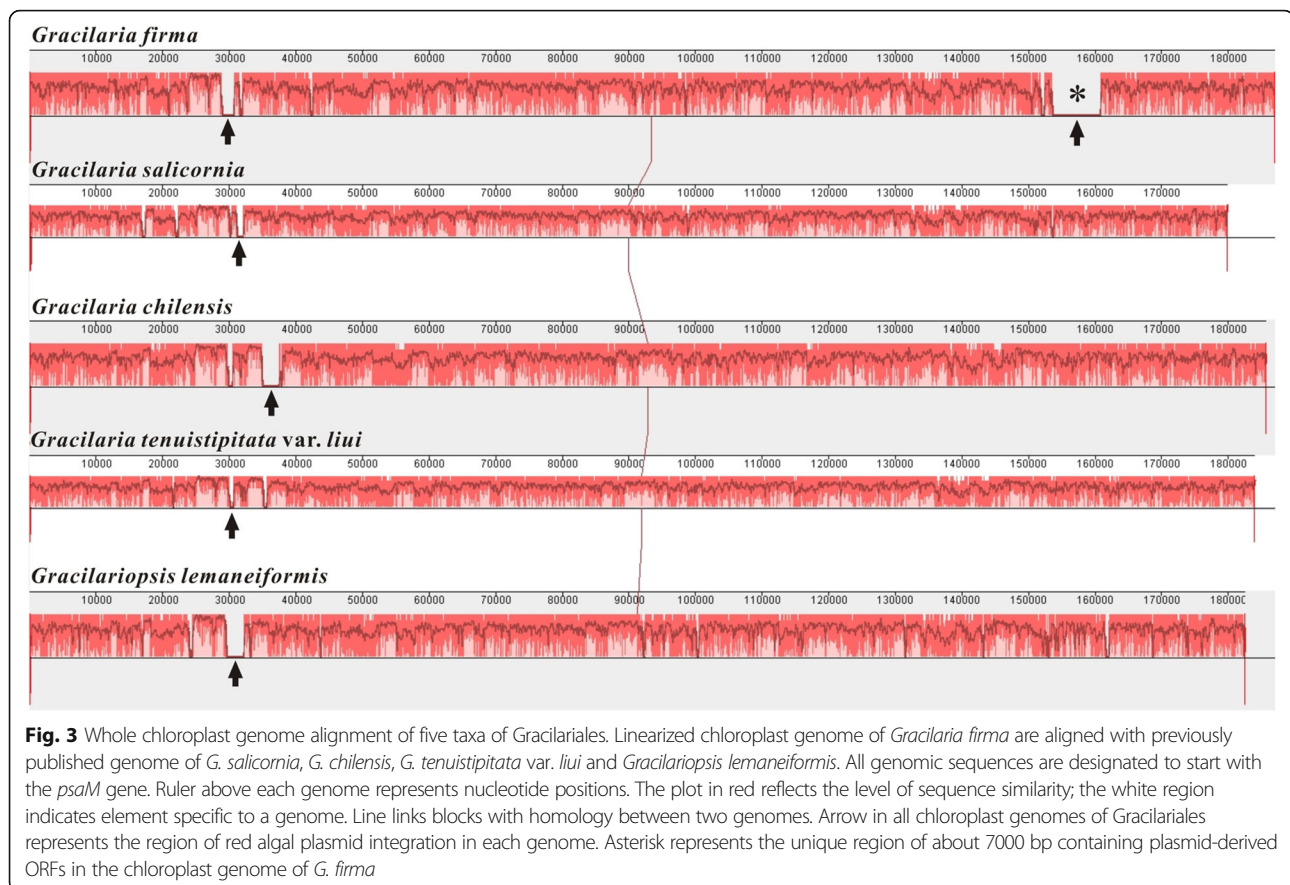


conserved hypothetical genes. None of the protein-coding genes contains introns. Twenty three ORFs were annotated in the chloroplast genome of *G. firma*, of which 14 of them showed significant homology to the ORFs of red algal plasmids against BLASTp search with the threshold *e*-value of  $1e-5$  (Additional file 1), eight of them showed significant similarity with ORFs from other red algae (*e*-value  $<1e-5$ ), and one of them showed remote similarity to the *pbsA* gene (*e*-value = 0.7). The gene content of the chloroplast genome of *G. firma* is very similar to that of other florideophyte chloroplast genomes, with suites of genes involved in photosynthesis,

electron transport, translation, biosynthesis of amino acids, fatty acids, pigments and other cellular processes, apart from the lack of the *pbsA* gene which is also missing in *G. chilensis* [6].

**Red algal plasmid-derived regions in the chloroplast genome of *Gracilaria firma***

A total of 14 ORFs homologous to red algal plasmids (*Gfir\_ORF1*, *Gfir\_ORF2*, *Gfir\_ORF12*, *Gfir\_ORF13*, *Gfir\_ORF14*, *Gfir\_ORF15*, *Gfir\_ORF16*, *Gfir\_ORF17*, *Gfir\_ORF18*, *Gfir\_ORF19*, *Gfir\_ORF20*, *Gfir\_ORF21*, *Gfir\_ORF22* and *Gfir\_ORF23*) are interspersed in the



*rrs-ompR* and *nbla-cpeB* intergenic regions of the *G. firma* chloroplast genome, with *Gfir\_ORF1* and *Gfir\_ORF2* found in the former spacer, and the remaining in the latter (Fig. 4). These ORFs showed moderate to strong similarity to ORFs found in the plasmids of *G. chilensis* (Gch7220 and Gch3937), *G. robusta* (Gro4970) and *Gp. lemaneiformis* (Gle4293) (Additional file 1). ORF5 in Gch7220 is almost similar to ORF2 in Gch3937 with 99% amino acid identity. ORF1 in Gle4293 is similar to ORF1 and ORF4 in Gch7220, ORF1 in Gch3937 and ORF1 in Gro4970 with 34–38% amino acid identity, all of which constituting part of the protein family DUF1368 with unknown function specific to red algae.

*Gfir\_ORF1* and *Gfir\_ORF2* are two overlapping ORFs which appear to be truncated versions of single ORF (ORF5/ORF2) found in the Gch7220 or Gch3937 plasmids of *G. chilensis*, corresponding to amino acids 1 to 36 and 44 to 101, respectively (Additional file 1). *Gfir\_ORF20* which shares 76 and 74% amino acid identity with ORF5 from Gch7220 and ORF2 from Gch3937 at 199 amino acids is slightly longer than the comparable ORF from *G. chilensis* plasmids (Additional file 1). The high sequence identity may indicate recent horizontal transfer of this ORF from *G. chilensis*-specific plasmid to *G. firma* chloroplast. *Gfir\_ORF18* and *Gfir\_ORF19*

display considerable amino acid similarity with ORF4 (61% amino acid identity corresponding to amino acids 17 to 91) and ORF7 (64% amino acid identity corresponding to amino acids 23 to 61) of Gro4970 plasmid respectively. The remaining ORFs are homologous to all five ORFs in the Gle4293 plasmid of *Gp. lemaneiformis*. *Gfir\_ORF17*, *Gfir\_ORF16*, *Gfir\_ORF15*, *Gfir\_ORF14* and *Gfir\_ORF13* showed similarity to ORF1, ORF2, ORF3, ORF4 and ORF5 of Gle4293 plasmid (Fig. 4), with each ORF corresponding to amino acids 57 to 417 (53% identity), 16 to 132 (49% identity), 17 to 262 (53% identity), 1 to 168 (59% identity) and 17 to 149 (58% identity). Apart from *Gfir\_ORF17*, *Gfir\_ORF12*, *Gfir\_ORF21*, *Gfir\_ORF22* and *Gfir\_ORF23* are also truncated versions of the ORF1 from Gle4293 plasmid, each corresponding to amino acids 1 to 161 (42% identity), 375 to 418 (59% identity), 275 to 359 (64% identity) and 226 to 274 (61% identity). The integration of whole plasmid into the chloroplast of *G. firma* may have occurred, considering the presence of degenerated fragments homologous to all five ORFs from Gle4293 plasmid in a contiguous manner in the intergenic region between the *nbla* and *cpeB* genes of the *G. firma* chloroplast genome.

Whole-genome alignment of the chloroplast genome of all five members of Gracilariales revealed a common

**Table 2** Functional classification of *Gracilaria firma* chloroplast genes

Classification	Genes
Genetic system	
Maintenance	<i>dnaB, rne, rnz, mat</i>
RNA polymerase	<i>rpoA, rpoB, rpoC1, rpoC2, rpoZ</i>
Transcription factors	<i>ompR, rbcR, tctD, ntcA</i>
Translation	<i>infB, infC, tsf, tufA</i>
Ribosomal proteins	<i>rpl1, rpl2, rpl3, rpl4, rpl5, rpl6, rpl9, rpl11, rpl12, rpl13, rpl14, rpl16, rpl18, rpl19, rpl20, rpl21, rpl22, rpl23, rpl24, rpl27, rpl28, rpl29, rpl31, rpl32, rpl33, rpl34, rpl35, rpl36, rps1, rps2, rps3, rps4, rps5, rps6, rps7, rps8, rps9, rps10, rps11, rps12, rps13, rps14, rps16, rps17, rps18, rps19, rps20</i>
tRNA processing	<i>tisL</i>
Protein quality control	<i>clpC, dnaK, ftsH, groEL</i>
Photosystems	
Phycobilisomes	<i>apcA, apcB, apcD, apcE, apcF, cpcA, cpcB, cpcG, cpeA, cpeB, nbla, ycf58</i>
Photosystem I	<i>psaA, psaB, psaC, psaD, psaE, psaF, psal, psaj, psak, psal, psam, ycf3, ycf4</i>
Photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbN, psbT, psbV, psbW, psbX, psbY, psbZ, ycf12</i>
Cytochrome complex	<i>petA, petB, petD, petF, petG, petJ, petL, petM, petN, ccsA, ccs1</i>
Redox system	<i>acsF, bas1, ftrB, dsbD, trxA</i>
ATP synthesis	
ATP synthase	<i>atpA, atpB, atpD, atpE, atpF, atpG, atpH, atpI</i>
Metabolism	
Carbohydrates	<i>odpA, odpB, rbcl, rbcS, cbbX, pgmA</i>
Lipids	<i>accA, accB, accD, acpP, fabH</i>
Nucleotides	<i>carA, upp</i>
Amino acids	<i>argB, gltB, syh, ilvB, ilvH, trpA, trpG, syfB</i>
Cofactors	<i>chll, moeB, preA, thiG</i>
Transport	
Transport	<i>cemA, secA, secG, secY, tatC, sufB, sufC, ycf38, ycf63</i>
Unknown	
Conserved ORFs	<i>dfr, ycf17, ycf19, ycf20, ycf21, ycf22, ycf23, ycf33, ycf34, ycf35, ycf36, ycf37, ycf39, ycf40, ycf45, ycf46, ycf52, ycf53, ycf54, ycf55, ycf60, ycf65, ycf80</i>
Unique ORFs	<i>Gfir_ORF1, Gfir_ORF2, Gfir_ORF3, ..., Gfir_ORF23</i>
RNA genes	
rRNAs	<i>rrf, rrs, rrl</i>
tRNAs	

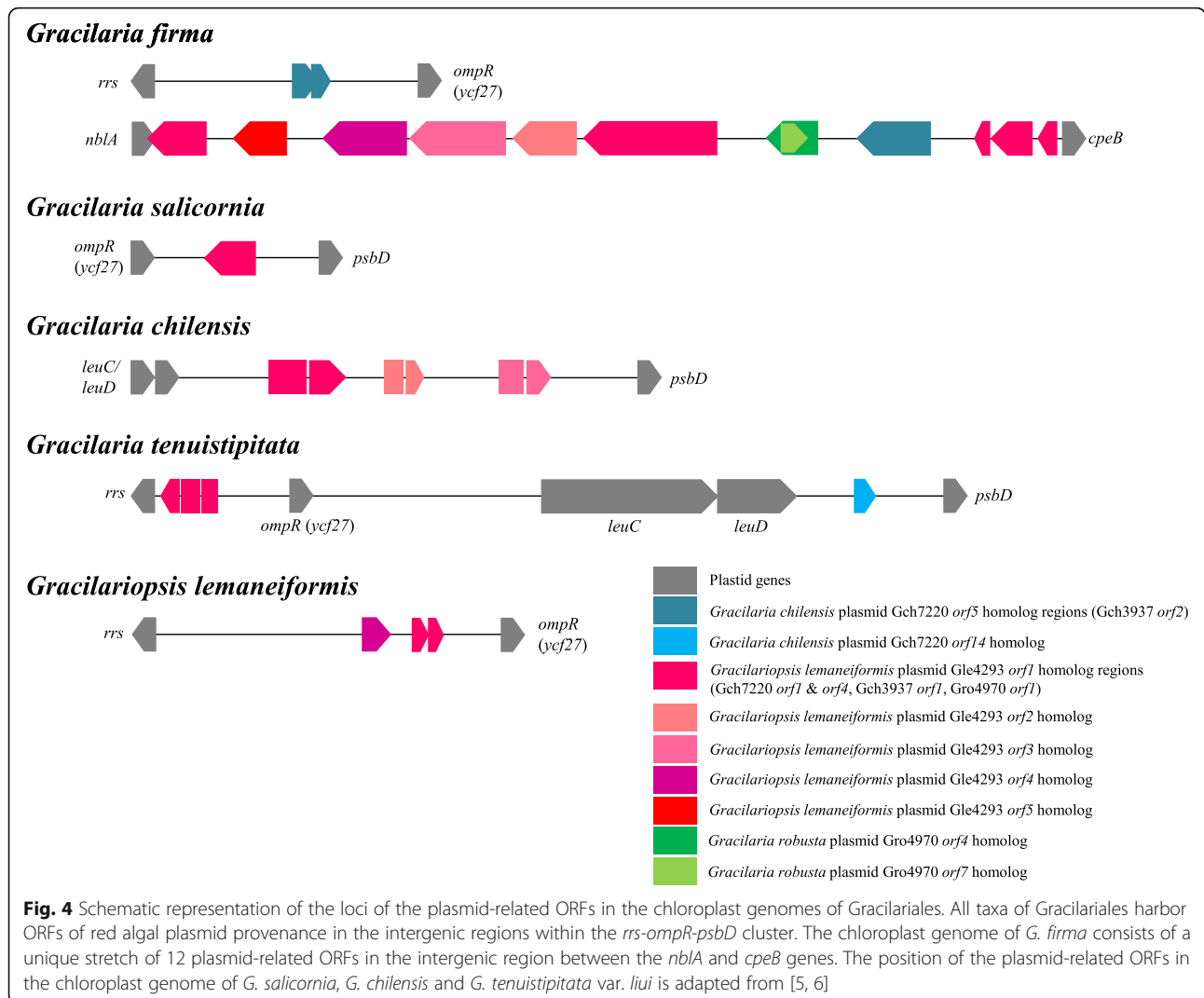
**Table 2** Functional classification of *Gracilaria firma* chloroplast genes (Continued)

	<i>trnA, trnC, trnD, trnE, trnF, trnG, trnH, trnI, trnK, trnL, trnL, trnL, trnM, trnM, trnM, trnN, trnP, trnQ, trnR, trnR, trnR, trnS, trnS, trnT, trnT, trnV, trnV, trnW, trnY</i>
Miscellaneous RNAs	<i>ssrA, rmpB</i>

syntenic break at the position between 30,000 and 40,000, which corresponds to the intergenic regions within the *rrs-ompR-psbD* cluster (Fig. 3). This region represents the “hotspot” for plasmid integration in the chloroplast genome of Gracilariales, as ORFs homologous to red algal plasmids were found in this region based on the previous studies [5, 6] (Fig. 4). The intergenic region between *ompR* and *psbD* in the chloroplast genome of *G. salicornia* has an ORF that shares strong similarity with ORF1 of Gch7220 plasmid [5, 6]. And the corresponding spacer region in the chloroplast genome of *G. chilensis* has a gene pair *leuC-leuD* of bacterial origin and three plasmid-derived regions which are homologous to the ORF1, ORF2 and ORF3 from Gle4293 plasmid respectively [6]. In the chloroplast genome of *G. tenuistipitata* var. *liui*, the intergenic region between *rrs* and *ompR* in the chloroplast genome was reported to have a plasmid-derived region homologous to ORF1 of Gle4293 plasmid [6], and the region between *ompR* and *psbD* has a gene pair *leuC-leuD* and an ORF similar to ORF14 of Gch7220 plasmid [5]. Three ORFs homologous to the ORF5 and ORF1 of Gle4392 plasmid were found in the spacer region between *rrs* and *ompR* in the chloroplast genome of *Gp. lemaneiformis*. A notable feature unique to the chloroplast genome of *G. firma* is the presence of a relatively large syntenic break with 12 red algal plasmid homologs spanning about 7 kb at position ~153,000 between *nbla* and *cpeB* (asterisk in Fig. 3). This region was not observed in any other examined taxa, and likely to have contributed to the expansion of chloroplast genome size in *G. firma*.

Despite the caveat that the sequenced library was not searched to confirm the existence of extrachromosomal plasmid in *G. firma*, we preclude the possibility of mapping artefact in the assembly for the expanded chloroplast genome on two grounds: (1) single contigs with identical sequence were obtained from two independent assemblies using different *de novo* and reference-guided assemblers, and (2) the presence of red algal plasmid remnants in the chloroplast genome has been reported in red algal taxa with or without naturally occurring extrachromosomal plasmid [4, 5, 38, 41, 42]. Previous studies have verified the occurrence of plasmid-derived regions in the chloroplast genome of *Gp. lemaneiformis* [5], and confirmed the consistency of the plasmid-





derived sequences within individuals and populations of *Gelidium elegans*, *Porphyra pulchra* and *Sporolithon durum* [6], using customized primer pairs for PCR. The copy number and position of the homologous plasmid-derived ORFs in chloroplast genomes is inconsistent with the red algal phylogenetic relationships [6]. Incorporation of foreign genetic materials into the maternally inherited organelles may have rendered their fixation in a population. The plasmid may have recombined randomly during the integration into the chloroplast genome and led to gene truncation and subsequent gene loss, such that vestiges of red algal plasmid were perceived only in certain but not all red algal species. Such species-specific integration of plasmid-derived ORFs is consistent with the evolution of mobile genetic elements. The plasmids are thought to be analogous to transposable elements with mobility that can contribute to the gain or loss among closely related genomes [6]. However, the actual role of red algal plasmids remains

elusive. While recognizing the possible role of red algal plasmids in mediating gene transfer between foreign DNA and organelles, a study considered the plasmids as parasitic elements that spread plasmid-derived DNA regions in different organelles [6]. It was suggested that the ubiquitous red algal plasmid remnants may be associated with their function in ancient horizontal gene transfer among nucleus, chloroplast and mitochondrion genomes [5].

#### Evolution of chloroplast genome in Eurhodophytina

The gene content and gene order of 21 chloroplast genomes representing the classes Bangiophyceae and Florideophyceae was compared to infer the evolution of chloroplast genome in red algae. The highly divergent unicellular thermophiles of the classes Cyanidiophyceae and Porphyridiophyceae were not included in the genome alignment. Only one representative from each genus of Bangiophyceae was included in the analyses to

emphasize the resolution of the relationship within Florideophyceae using chloroplast genome data. The parasitic red alga *Choreocolax* was also excluded from the genome alignment as it experienced many rearrangements and losses despite sharing regions of synteny with other Florideophycean taxa [43].

A total of 229 chloroplast genes were inferred to present in the hypothetical last common ancestor of the Bangiophyceae and Florideophyceae, including 194 protein-coding genes, 3 rRNA genes, and 35 tRNA genes, based on the currently available chloroplast genomes examined in this study (Table 3). Comparison of the whole-genome

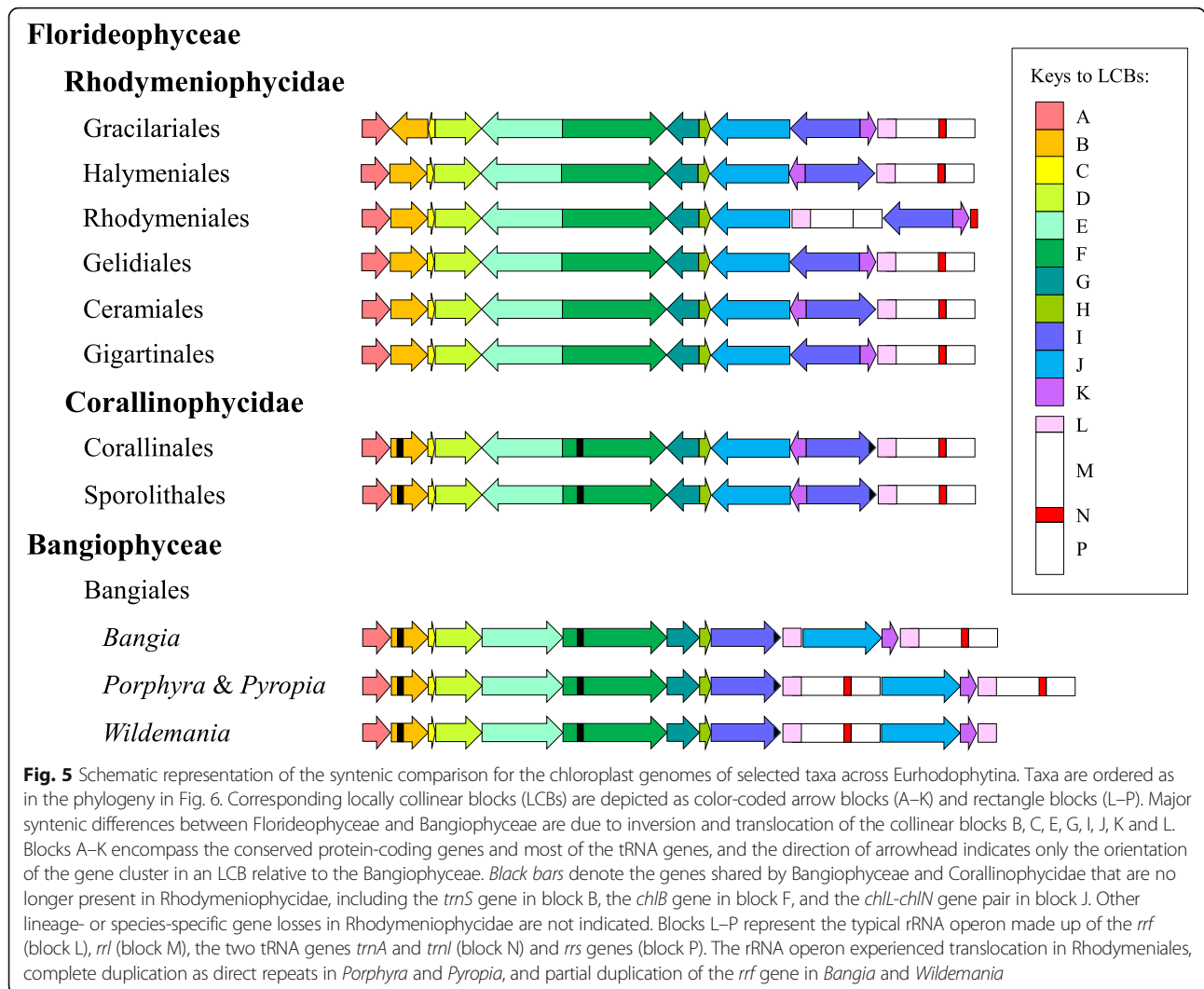
alignments revealed a largely identical gene content in the chloroplast genome of Bangiophyceae and Florideophyceae with a few exceptions. The retention of such large number of chloroplast gene repertoires known in photosynthetic eukaryotes including the complete tRNA set and the genes encoding transcriptional regulators, and the position of genes on the same strand in operon-like structures, suggest ancient gene content in the red algal chloroplast genomes [38, 44]. Apart from the *glnB* gene flanked by the *rpl33* and *rps20* genes, the representatives of Bangiophyceae also featured three additional tRNA genes not found in the Florideophycean representatives, including a *trnA* gene in the tRNA gene cluster of *trnM-trnS-trnD*, a *trnS* gene positioned between *ftsH* and *psaE*, and a *trnL* gene between *fabH* and *psbX*. The *chlB*, *chlL* and *chlN* genes, along with the *trnS* gene positioned upstream of the *moeB* gene, which were initially considered exclusive to the Bangiophycean species [38] before more chloroplast genomes of Florideophyceae are available, were also found in the representatives of Corallinophycidae but not Rhodymeniophycidae (black bars in the arrow blocks B, F and J in Fig. 5). The absence of those genes encoding redundant tRNA variants and subunits of the chloroplast light-independent protochlorophyllide reductase was considered as outright gene loss rather than gene transfer from chloroplast to nucleus given that all the genes involved have been completely lost in many other photosynthetic eukaryotes [4]. Additional taxa sampling to include the other subclasses of Florideophyceae would reveal whether such pattern of gene loss is characteristic to the Florideophyceae. Nevertheless, the presence or absence of certain genes in the chloroplast genome should not be taken as reliable phylogenetic characters because gene transfer to the nucleus or gene loss has occurred numerous times independently across various eukaryotic lineages [45]. Lineage- or species-specific loss of a few genes including *dfr*, *dnaB*, *petL*, *pbsA*, *psb30*, *rpl9*, *syfB*, *syh*, *secG*, and *thiS* were also observed. A few instances are the loss of *dfr* in *Chondrus crispus*, *Vertebrata lanosa*, and *Laurencia snackeyi*, the independent loss of *syh*, *dnaB*, *syfB* and *ycf21* in *Sporolithon durum*, and loss of *pbsA* in *G. firma* and *G. chilensis*.

Fifteen LCBs were identified chloroplast genomes, with blocks A–K comprising the protein-coding and tRNA genes, and blocks L–P representing the rRNA operon (Fig. 5). Blocks E, F, I and J encompassed a large portion of the conserved gene repertoire of red algal chloroplast genomes. Some blocks comprised of only one or two genes, including blocks L, M, N and P, as a result of single gene translocation (Table 3). Most of the genomic rearrangements involved simultaneous inversion or translocation of gene cluster in the collinear blocks B, C, E, G, I, J, and K. Major syntenic differences observed between Bangiophyceae and Corallinophycidae

**Table 3** Gene cluster content of Eurhodophytina

LCB	Conserved genes present in LCB
A	<i>orf243</i> [ <i>ycf27</i> or <i>ompR</i> ], <i>psbD</i> , <i>psbC</i> , <i>orf198</i> [ <i>upp</i> ], <i>ycf19</i> , <i>rps16</i> , <i>ycf65</i> , <i>groEL</i> , <i>trnR</i> (ACG), <i>syh</i> , <i>trnQ</i> (UUG), <i>trnR</i> (CCU)
B	<i>psbW</i> , <i>syfB</i> , <i>rps1</i> , <i>orf71</i> [ <i>ycf40</i> or <i>thiS</i> ], <i>trnH</i> (GUG), <i>ycf29</i> [ <i>tctD</i> ], <i>ycf28</i> [ <i>ntcA</i> ], <i>petB</i> , <i>petD</i> , <i>rpl12</i> , <i>rpl1</i> , <i>rpl11</i> , <i>trnW</i> (CCA), <i>orf277</i> [ <i>moeB</i> ], <i>trnS</i> (GGA), <i>rpl9</i>
C	<i>dnaB</i> , <i>trnF</i> (GAA), <i>clpC</i>
D	<i>rpl19</i> , <i>trnV</i> (GAC), <i>trnY</i> (GUA), <i>trnT</i> (GGU), <i>ilvB</i> , <i>ycf33</i> , <i>argB</i> , <i>trnM</i> (CAU), <i>trnA</i> (GGC), <i>trnS</i> (GCU), <i>trnD</i> (GUC), <i>ftsH</i> , <i>trnS</i> (CGA), <i>psaE</i> , <i>psaH</i> , <i>psaN</i> , <i>psaT</i> , <i>psbB</i> , <i>ycf38</i> , <i>petF</i>
E	<i>rps14</i> , <i>petG</i> , <i>psbK</i> , <i>psbZ</i> , <i>trnG</i> (GCC), <i>dnaK</i> , <i>rpl3</i> , <i>rpl4</i> , <i>rpl23</i> , <i>rpl2</i> , <i>rps19</i> , <i>rpl22</i> , <i>rps3</i> , <i>rpl16</i> , <i>rpl29</i> , <i>rps17</i> , <i>rpl14</i> , <i>rpl24</i> , <i>rpl5</i> , <i>rps8</i> , <i>rpl6</i> , <i>rpl18</i> , <i>rps5</i> , <i>secY</i> , <i>rpl36</i> , <i>rps13</i> , <i>rps11</i> , <i>rpoA</i> , <i>rpl13</i> , <i>rps9</i> , <i>rpl31</i> , <i>rps12</i> , <i>rps7</i> , <i>tufA</i> , <i>rps10</i>
F	<i>psaB</i> , <i>psaA</i> , <i>accB</i> , <i>orf565</i> [ <i>ycf45</i> ], <i>acpP</i> , <i>trnS</i> (UGA), <i>psaD</i> , <i>chlB</i> , <i>ycf59</i> [ <i>ascF</i> ], <i>petN</i> , <i>orf71</i> [ <i>secG</i> or <i>ycf47</i> ], <i>ycf36</i> , <i>trnM</i> (CAU), <i>orf199</i> [ <i>bas1</i> ], <i>pbsA</i> , <i>rpl35</i> , <i>rpl20</i> , <i>preA</i> , <i>odpA</i> [ <i>pdhA</i> ], <i>odpB</i> [ <i>pdhB</i> ], <i>petA</i> , <i>tatC</i> , <i>apcE</i> , <i>apcA</i> , <i>apcB</i> , <i>atpE</i> , <i>atpB</i> , <i>ycf3</i> , <i>infB</i> , <i>rps18</i> , <i>rpl33</i> , <i>rps20</i> , <i>rpoB</i> , <i>rpoC1</i> , <i>rpoC2</i> , <i>rps2</i> , <i>tsf</i> , <i>atpI</i> , <i>atpH</i> , <i>atpG</i> , <i>atpF</i> , <i>atpD</i> , <i>atpA</i> , <i>ycf16</i> [ <i>sufC</i> ], <i>ycf24</i> [ <i>sufB</i> ], <i>trnL</i> (CAA), <i>ycf39</i> , <i>psbI</i> , <i>orf149</i> [ <i>ycf58</i> ], <i>cemA</i> , <i>trnM</i> (CAU)
G	<i>infC</i> , <i>ilvH</i> , <i>trnL</i> (UAA), <i>trnC</i> (GCA), <i>cbbX</i> , <i>rbcS</i> , <i>rbcl</i> , <i>trxA</i> , <i>rpl28</i> , <i>trnT</i> (UGU), <i>psaL</i> , <i>ycf7</i> [ <i>petL</i> ], <i>ycf4</i> , <i>trnG</i> (UCC)
H	<i>rps4</i> , <i>orf450</i> [ <i>ycf80</i> ], <i>trnR</i> (CCG), <i>apcF</i> , <i>ycf20</i>
I	<i>trnL</i> (UAG), <i>ycf35</i> , <i>psbA</i> , <i>rne</i> , <i>rpl27</i> , <i>rpl21</i> , <i>orf263</i> [ <i>ycf56</i> ], <i>rpl32</i> , <i>ycf32</i> [ <i>psbY</i> ], <i>rbcR</i> , <i>orf108</i> [ <i>ycf54</i> ], <i>orf320</i> [ <i>ycf55</i> ], <i>orf238</i> [ <i>ycf53</i> ], <i>carA</i> , <i>petJ</i> , <i>psbV</i> , <i>accD</i> , <i>psbX</i> , <i>trnL</i> (GAG), <i>fabH</i> , <i>apcD</i> , <i>psaJ</i> , <i>psaF</i> , <i>orf174</i> [ <i>ycf52</i> ], <i>trnP</i> (UGG), <i>ycf37</i> , <i>rpl34</i> , <i>ycf46</i> , <i>chlN</i> , <i>chlL</i>
J	<i>psaM</i> , <i>chlI</i> , <i>trnR</i> (UCU), <i>trnV</i> (UAC), <i>orf263a</i> [ <i>ycf63</i> ], <i>ycf26</i> [ <i>dfr</i> ], <i>psbE</i> , <i>psbF</i> , <i>psbL</i> , <i>psbJ</i> , <i>psal</i> , <i>frtB</i> , <i>ycf12</i> [ <i>psb30</i> ], <i>orf327</i> [ <i>ycf62</i> or <i>tif5</i> ], <i>trnK</i> (UUU), <i>trpA</i> , <i>trnE</i> (UUC), <i>secA</i> , <i>ycf21</i> , <i>pgmA</i> , <i>cpcA</i> , <i>cpcB</i> , <i>ycf61</i> [ <i>rpoZ</i> ], <i>gltB</i> , <i>psaC</i> , <i>ycf23</i> , <i>ycf22</i> , <i>accA</i> , <i>psaK</i> , <i>ccsA</i> , <i>cpeA</i> , <i>cpeB</i> , <i>ycf18</i> [ <i>nblA</i> ], <i>cpcG</i>
K	<i>trnN</i> (GUU), <i>orf240</i> [ <i>dsbD</i> or <i>ccdA</i> ], <i>ccs1</i> , <i>trpG</i> , <i>thiG</i> , <i>orf203</i> [ <i>ycf60</i> ], <i>rps6</i>
L	<i>Rrf</i>
M	<i>Rrl</i>
N	<i>trnA</i> (UGC), <i>trnI</i> (GAC)
P	<i>rrs</i>

The gene content inferred to present in the hypothetical last common ancestor of Bangiophyceae and Florideophyceae are listed in conserved clusters identified using Mauve. The alphabetical code of each conserved gene cluster or locally collinear block (LCB) corresponds to that in Fig. 5. The genes in each cluster are listed in the order they present in the Bangiophyceae in Fig. 5. Alternative names applied for the homologous genes in the Florideophyceae are indicated in square brackets



were the inversion of collinear blocks E, G, J and K, and the translocation of block I from the original position immediately downstream of block H exhibited by Bangiophyceae, to a novel position between block K and the rRNA operon. The gene order has been highly conserved since the split from Bangiophyceae and a high degree of synteny was observed in Florideophyceae. Representatives of the Corallinophycidae, Ceramiales and Halymeniales exhibited similar gene order across the examined Florideophycean lineages, with some gene losses in the Rhodymeniophycidae as mentioned above. An inversion of the collinear block (I + K) occurred in the Gracilariales, Gigartinales, Gelidiales and Rhodymeniales with respect to its orientation in the Corallinophycidae, Ceramiales and Halymeniales. In addition, *C. compressa* of the Rhodymeniales featured a unique translocation of the three rRNA genes (blocks L, M and P) from the original position downstream of block K as in the Gracilariales, Gigartinales and Gelidiales, to a

position flanked by block J upstream and block I downstream. The Gracilariales exclusively exhibited inversion of the blocks B and C not seen in other orders.

The Florideophycean taxa possessed only a single copy of rRNA operon typically comprised of the *rrf*, *rrl*, *trnA*, *trnI* and *rrs* genes, except for the rRNA operon of *Coeloseira compressa* which comprised only three rRNA genes. The representatives of Bangiophyceae showed more variations in the number of the rRNA genes. *Porphyra umbilicalis* and *Pyropia haitanensis* possessed two copies of rRNA operon each made up of the *rrf-rrl-trnA-trnI-rrs* gene cluster as direct repeats, with one copy of the operon positioned between blocks I and J, and another downstream of block K. Hughey et al. [46] reported that *Pyropia perforata* lacked the second copy of rRNA operon in other species of *Porphyra* and *Pyropia*. Instead of having two copies of rRNA operon like most species of *Porphyra* and *Pyropia*, both *Bangia atropurpurea* and *Wildemanina schizophylla* have an

additional *rrf* gene on top of the single copy of rRNA operon possessed.

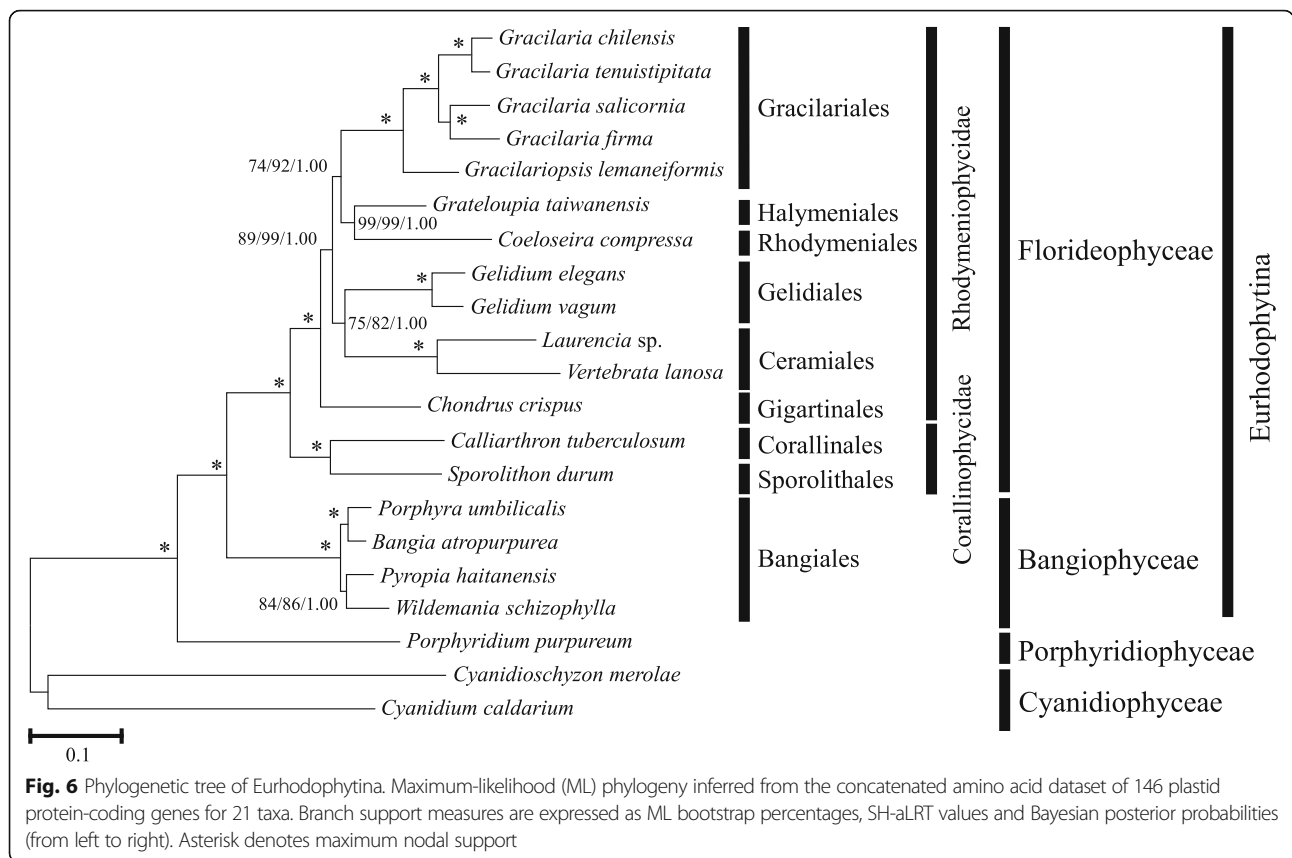
Inclusion of more red algal taxa of the Bangiophyceae and Florideophyceae for the comparison of genomic synteny has resulted in the observation of more rearrangement patterns across the subphylum Eurhodophytina, with an increase in the number of conserved gene cluster from 11 in previous studies [4, 38] to 15 in present study. The chloroplast genomes of the Bangiales which had four genera examined in the present study exhibited identical genome structure when the rRNA operons were not taken into consideration. The red algal chloroplast genomes are highly conserved in gene content and order, considering the relatively minimal number of gene lost and extent of genomic rearrangement across Eurhodophytina over a substantial evolutionary distance since the divergence of Bangiophyceae and Florideophyceae that has occurred at least 940 million years ago [5, 47]. However, the gene order pattern was not reflective of the ordinal relationships inferred from the phylogenomic analyses. Parallel evolution in gene order is observed in Florideophyceae such that identical gene order pattern arose independently several times in different lineages.

#### Phylogenetic relationships inferred using the chloroplast genome

The phylogenies inferred using different phylogenetic approaches (ML vs BI) and gene sampling (79 vs 146 concatenated protein-coding genes) on 21 taxa recovered topologies which are largely congruent, with the exception in the Bayesian phylogeny inferred using the 79-gene dataset, where *Coeloseira compressa* was positioned basal to the Ceramiales. Additional gene sampling improved the branch support for the relationships between Gelidiales and Ceramiales with the SH-aLRT value increased from 40 in the 79-gene dataset to 82 in the 146-gene dataset. Only the maximum likelihood (ML) phylogeny inferred from the concatenated dataset containing 146 protein-coding genes of the chloroplast genomes present in all 21 examined taxa was presented (Fig. 6), with the posterior probabilities recovered from the Bayesian analysis appended on the nodes of the ML phylogeny along with the SH-aLRT and bootstrap branch supports. Consistent with previous studies [47–49], Fig. 6 resolved the subphylum Eurhodophytina as two fully supported monophyletic groups corresponding to the classes Bangiophyceae and Florideophyceae. The subclass Corallinophycidae represented by *Calliarthron tuberosum* (Corallinales) and *Sporolithon durum* (Sporolithales) was positioned basal to the subclass Rhodymeniophycidae with maximum nodal support. *Chondrus crispus* (Gigartinales) diverged first within the subclass Rhodymeniophycidae with strong nodal support (BS = 89%, SH-aLRT = 99; PP = 1.00). The remaining taxa of Rhodymeniophycidae were split into

two groups, of which the basal group encompassed the Gelidiales (*Gelidium*) and Ceramiales (*Vertebrata lanosa* and *Laurencia snackeyi*), and another group consisted of Gracilariales (*Gracilaria salicornia*, *G. firma*, *G. chilensis*, *G. tenuistipitata* var. *liui* and *Gracilariopsis lemaneiformis*), Halymeniales (*Grateloupia taiwanensis*) and Rhodymeniales (*Coeloseira compressa*). The interordinal relationships within the Rhodymeniophycidae received moderate to strong nodal support (BS > 70%; SH-aLRT > 80; PP = 1.00). The relationships within the monotypic family of Gracilariales were fully supported with *Gp. lemaneiformis* being positioned basal to the remaining *Gracilaria* species.

The relationships among most of the orders in the subclass Rhodymeniophycidae have been identified as one of the evolutionary lineages in the red algal phylogeny by Verbruggen et al. [50]. These relationships require considerable analyses by the inclusion of more markers and wider taxon sampling to increase the informative characters for better phylogenetic resolution. The poorly supported interordinal relationships within the subclass is largely attributed to the rapid radiation of lineages (i.e. the Gigartinales *sensu lato*) at the base of the subclass. We refrain from drawing concluding remarks on the phylogenetic relationships of the Rhodymeniophycidae with notion that the mere addition of one taxon from previously studied Gracilariales in the underrepresented taxon sampling will not result in significant changes to the interordinal relationships within the subclass. Relative phylogenetic affinities of the taxa (or the representative orders) examined in this study based on the analyses using chloroplast genome data in this study were compared with those attained from previous work based on different sets of markers over wide sampling [47–50]. Interordinal relationships within the subclass Rhodymeniophycidae recovered in this study are congruent with the results inferred using the mitochondrial genome [48] and also multiple markers of different genomic origins in some studies (*rbcL*, *psaA*, *psbA*, EF2, SSU, LSU and *cox1* [47]; SSU, LSU, EF2, *rbcL* and CO1-5P [49]). These results, however, are in contrast with that inferred using DNA data of 14 markers (EF2, 23S rDNA, 28S rDNA, 16S rDNA, 18S rDNA, *cox1*, *psaA*, *psaB*, *psbA*, *psbC*, *psbD*, *rbcL*, *rbcS* and *tufA*) belonging to all three genomic compartments mainly mined from GenBank, of which ten markers were poorly represented [50]. The first divergence of *Chondrus crispus* in Gigartinales from other orders received strong nodal support in this study and moderate to strong support in previous studies [47–49]. The placement of the family accommodating this taxon in the most recently derived order in Florideophyceae was poorly supported in [50]. In contrast to the poorly supported grouping of Gracilariales with Ceramiales and Gelidiales in [50], this study



recovered a moderately supported relationships among the Gracilariales, Halymeniales and Rhodymeniales (BS = 74, PP = 1.00). Consistent relationships among Gracilariales, Halymeniales and Rhodymeniales with moderate nodal support (BS = 67, PP = 1.00) was also recovered in the phylogeny inferred using mitochondrial genome and wider taxon sampling [48], but the relationships among these three orders were poorly supported in phylogenies constructed using multigene data over broad taxonomic breadth of Rhodymeniophycidae [47, 49]. While large multi-locus datasets for a broad taxonomic breadth are considered to be the preferred solution to resolve the relationships in the Rhodymeniophycidae [50], genome-scale data based on limited taxon sampling as in this study can still recover similar phylogenetic inference for the common taxa used in different studies. This suggests that the use of chloroplast genome data could help to improve the support for the interordinal relationships previously identified using multiple markers on Florideophyceae-wide sampling.

Classification in the family Gracilariales has always been in a state of flux over the years, with more than hundred species passing under the generic names of *Gracilaria* and *Hydropuntia* based on the morphological and anatomical features [14–16]. It was not until the rather comprehensive molecular study based on the *rbcL*

gene had successfully delineated the *Gracilaria sensu lato* into three groups which correspond to a new lineage and the previously defined genera of *Gracilaria* and *Hydropuntia*, that the systematics within Gracilariaceae was considered stabilized [14]. *Hydropuntia* was maintained as a genus distinct from *Gracilaria* based on the well-circumscribed reproductive features, despite the lineage being supported by high Bayesian posterior probabilities and low ML bootstrap percentages in the molecular phylogeny [14]. It differs from the *Gracilaria sensu stricto* which features cystocarps composed of gonimoblasts and carposporangia arranged in longer chains that are dichotomously or irregularly branching and tubular filaments connecting the gonimoblasts to the pericarp, by possessing cystocarps composed of gonimoblasts cells in short chains terminating in apical carposporangia that are arranged radially, and tubular filaments connecting the gonimoblasts to the base of the cystocarp. However, a recent study on the phylogeny of Gracilariaceae using the *rbcL*, *cox1* and UPA indicated the non-monophyly of *Hydropuntia* and suggested its reduction to *Gracilaria* [16]. The taxa of Gracilariaceae examined in present study are representative of the three previously delineated groups, with *G. salicornia* as *Gracilaria sensu stricto* [14], *G. firma* as *Hydropuntia* [51], and *G. tenuistipitata* and *G. chilensis* as the new lineage

[14]. Despite the limited taxon sampling, the fully resolved phylogeny of Gracilariaceae inferred using the concatenated dataset of 146 chloroplast protein-coding genes in this study implied the potential of chloroplast genome in resolving the *Gracilaria sensu stricto* conundrum. Inclusion of more samples of *Gracilaria sensu lato* for analyses may eventually lead to the recognition of *Hydropuntia* as a genus distinct from *Gracilaria sensu stricto*.

The interpretation and accuracy of the branch support measured as ML bootstrap replicate proportions and Bayesian posterior probabilities is still an issue of dispute, especially when the deep relationships involving rapid radiation of lineages are considered, although the likelihood-based phylogenetic reconstruction in the non-parametric maximum-likelihood and Bayesian frameworks have established themselves as the methods of choice [52]. The very short internal branches separating the major lineages of Gracilariales, Halymeniales, Rhodymeniales, Gelidiales and Ceramiales with long terminal branches (Fig. 6) suggest that these lineages of the subclass Rhodymeniophycidae underwent a rapid radiation [50]. The branching order is difficult to resolve when there is a rapid accumulation of many lineages in short period of time and when sequences become saturated, where the phylogenies typically exhibit very short internal branches with low ML bootstrap but high Bayesian posterior probabilities supports. It has been noted that the ML bootstrapping is computational intensive and often underestimates the branch support, whereas the Bayesian estimators inflate the confidence of the corresponding branches [53]. The values estimated by SH-aLRT, an alternative non-parametric approach to the conventional bootstrapping, which corroborated most of the relationships supported by high Bayesian posterior probabilities in this study, are likely to be another measure to support the data accurately. The branch supports derived from SH-aLRT are considered to be more consistent and conservative than conventional bootstrapping [54]. In addition, Bayesian posterior probabilities may be more indicative of the evolutionary relationships compared to the ML bootstrap replicate proportions which have been typically considered conservative when the full alignment was analyzed, as the deep relationships received increased ML bootstrap supports parallel to the high Bayesian posterior probabilities when analyses were conducted on the progressively more conservative alignments generated by site-stripping [49].

## Conclusions

This study presents the chloroplast genome of *Gracilaria firma* and identified unique red algal plasmid DNA remnants in the genome. Despite some lineage- and species-specific gene losses in the Florideophyceae, the chloroplast genomes across the subphylum Eurhodophytina are highly

conserved in synteny and genome architecture. The chloroplast genomes hold substantial information that can be tapped for improving the resolution of the phylogenetic relationships at all taxonomic hierarchical levels. However, additional study including improved taxon sampling, additional sequence data and further exploration of analyses options such as data partitions and evolutionary model selections is warranted to resolve the relationships within the Rhodymeniophycidae [47–50].

## Additional file

**Additional file 1:** Unique ORFs of *Gracilaria firma* homologous to red algal plasmids. The *G. firma* ORFs showed significant similarity ( $e$ -value <  $1e-5$ ) to the plasmids of *Gracilariopsis lemaneiformis*, *G. chilensis* and *G. robusta*, with the percentage identity and similarity significance expressed after the plasmid homolog. (DOCX 14 kb)

## Acknowledgements

We are grateful to Miss Yealing-Tay from BioEasy Sdn. Bhd. for her technical advice on the analysis of the sequenced reads. We also thank two anonymous reviewers for their constructive and valuable comments that helped to improve the previous version of this manuscript.

## Funding

Funding was provided by the Taiwan Ministry of Science and Technology (102-2628-B-019-002-MY3 and 104-2811-B-019-004 to SML) and the Ministry of Higher Education, Malaysia, MoHE-HIR Grant (H-50001-00-A000025 to PEL).

## Availability of data and materials

The alignments used for the phylogenomic analyses are available from the corresponding authors upon request.

## Authors' contributions

PKN and LCL obtained the samples and performed the experiment. PKN, CMC and TWP analyzed the data. PKN, SML and PEL conceived and wrote the manuscript. All authors read and approved the final manuscript.

## Competing interests

The authors declare that they have no competing interests.

## Consent for publication

Not applicable.

## Ethics approval and consent to participate

Not applicable.

## Author details

<sup>1</sup>Institute of Marine Biology, National Taiwan Ocean University, Keelung 20244, Taiwan. <sup>2</sup>Institute of Ocean and Earth Sciences, University of Malaya, Kuala Lumpur 50603, Malaysia. <sup>3</sup>Department of Computer Science and Engineering, National Taiwan Ocean University, Keelung 20244, Taiwan.

Received: 18 October 2016 Accepted: 22 December 2016

Published online: 06 January 2017

## References

- Saunders GW, Hommersand MH. Assessing red algal supraordinal diversity and taxonomy in the context of contemporary systematic data. *Am J Bot*. 2004;91:1494–507.
- Yoon HS, Müller KM, Sheath RG, Ott FD, Bhattacharya D. Defining the major lineages of red algae (Rhodophyta). *J Phycol*. 2006;42:482–92.
- Guiry MD, Guiry GM. Algaebase. World-wide electronic publication, National University of Ireland, Galway. 2016. <http://www.algaebase.org>. Accessed 15 Jun 2016.
- Janouškovec J, Liu SL, Martone PT, Carré W, Leblanc C, Collén J, et al. Evolution of red algal plastid genomes: ancient architectures, introns,

- horizontal gene transfer, and taxonomic utility of plastid markers. *PLoS One*. 2013;8:e59001.
5. Du QW, Bi GQ, Mao YX, Sui ZH. The complete chloroplast genome of *Gracilariopsis lemaneiformis* (Rhodophyta) gives new insight into the evolution of family Gracilariaceae. *J Phycol*. 2016;52(3):441–50.
  6. Lee J, Kim KM, Yang EC, Miller KA, Boo SM, Bhattacharya D, et al. Reconstructing the complex evolutionary history of mobile plasmids in red algal genomes. *Sci Rep*. 2016;6:23744.
  7. Chang CF, Xia BM. Taxonomic studies on *Gracilaria* from China. *Stud Mar Sinica*. 1976;11:91–163.
  8. Terada R, Baba M, Yamamoto H. New record of *Gracilaria firma* Chang et Xia (Rhodophyta) from Okinawa, Japan. *Phycol Res*. 2000;48(4):291–4.
  9. Lin SM. Marine benthic macroalgal flora of Taiwan: Part I Order Gracilariales (Rhodophyta). 1st ed. Keelung: National Taiwan Ocean University; 2009.
  10. Ajisaka T, Chiang YM. Recent status of *Gracilaria* cultivation in Taiwan. *Hydrobiologia*. 1993;260/261:335–8.
  11. Titlyanov EA, Titlyanova TV, Oham VH. Stocks and the use of economic marine macrophytes of Vietnam. *Russ J Mar Biol*. 2012;38:285–98.
  12. Trono GC. Diversity of the seaweed flora of the Philippines and its utilization. *Hydrobiologia*. 1999;398/399:1–6.
  13. Arano KG, Trono GC, Montano NE, Hurtado AQ, Villanueva RD. Growth, agar yield and quality of selected agarophyte species from the Philippines. *Bot Mar*. 2000;43(6):517–24.
  14. Gurgel CFD, Fredericq S. Systematics of the Gracilariaceae (Gracilariales, Rhodophyta): a critical assessment based on *rbcl* sequence analyses. *J Phycol*. 2004;40:138–59.
  15. Bird CJ, Rice EL, Murphy CA, Ragan MA. Phylogenetic relationships in the Gracilariales (Rhodophyta) as determined by 18S rDNA sequences. *Phycologia*. 1992;31(6):510–22.
  16. Lyra GM, Costa ES, Jesus PB, Matos JC, Caires TA, Oliveira MC, et al. Phylogeny of Gracilariaceae (Rhodophyta): evidence from plastid and mitochondrial nucleotide sequences. *J Phycol*. 2015;51(2):356–66.
  17. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
  18. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I. ABySS: a parallel assembler for short read sequence data. *Genome Res*. 2009;19:1117–23.
  19. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
  20. Paulino D, Warren RL, Vandervalk BP, Raymond A, Jackman SD, Birol I. Sealer: a scalable gap-closing application for finishing draft genomes. *BMC Bioinformatics*. 2015;16:230.
  21. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013;14:178–92.
  22. Peng Y, Leung HC, Yiu SM, Chin FY. IDBA-UD: a de novo assembler for single-cell and metagenomics sequencing data with highly uneven depth. *Bioinformatics*. 2012;28:1420–8.
  23. Hunter SS, Lyon RT, Sarver BAJ, Hardwick K, Forney LJ, Settles ML. Assembly by reduced complexity (ARC): a hybrid approach for targeted assembly of homologous sequences. *bioRxiv*. 2015. <http://dx.doi.org/10.1101/014662>.
  24. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*. 2004;20:3252–325.
  25. Lagesen K, Hallin P, Rødland EA, Stærfeldt H, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res*. 2007;35:3100–8.
  26. Lowe TM, Eddy SR. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. 1997;25:955–64.
  27. Laslett D, Canback B. ARAGORN, a program for the detection of transfer RNA and transfer-messenger RNA genes in nucleotide sequences. *Nucleic Acids Res*. 2004;32:11–6.
  28. Lang BF, Laforest M, Burger G. Mitochondrial introns: A critical view. *Trends Genet*. 2007;23:119–25.
  29. Lohse M, Drechsel O, Bock R. OrganellarGenomeDRAW (OGDRAW) – a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr Genet*. 2007;52:267–74.
  30. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res*. 2004;14(7):1394–403.
  31. Cox CJ, Li B, Foster PG, Embley TM, Civián P. Conflicting phylogenies for early land plants are caused by composition biases among synonymous substitutions. *Syst Biol*. 2014;63(2):272–9.
  32. Philippe H, Brunkmann H, Lavrov DV, Littlewood DTJ, Manuel M, Wörheide G, et al. Resolving difficult phylogenetic questions: Why more sequences are not enough. *PLoS Biol*. 2009;9(3):e1000602.
  33. Katoh K, Standley DM. MAFFT Multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772–80.
  34. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res*. 2008;31:W465–9.
  35. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser*. 1999;41:95–8.
  36. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol*. 2010;59(3):307–21.
  37. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012;61:1–4.
  38. Hagopian JC, Reis M, Kitajima JP, Bhattacharya D, Oliveira MC. Comparative analysis of the complete plastid genome sequence of the red alga *Gracilaria tenuistipitata* var. *liui* provides insights into the evolution of Rhodoplasts and their relationship to other plastids. *J Mol Evol*. 2004;59:464–77.
  39. Verbruggen H, Costa JF. The plastid genome of the red alga *Laurencia*. *J Phycol*. 2015;51:586–9.
  40. Ohta N, Matsuzaki M, Misumi O, Miyagishima SY, Nozaki H, Tanaka K, et al. Complete sequence and analysis of the plastid genome of the unicellular red alga *Cyanidioschyzon merolae*. *DNA Res*. 2003;10:67–77.
  41. Campbell MA, Presting G, Bennett MS, Sherwood AR. Highly conserved organellar genomes in the Gracilariales as inferred using new data from the Hawaiian invasive alga *Gracilaria salicornia* (Rhodophyta). *Phycologia*. 2014;53:109–16.
  42. DePriest MS, Bhattacharya D, López-Bautista JM. The plastid genome of the red macroalga *Grateloupia taiwanensis* (Halymeniaceae). *PLoS One*. 2013;8:e68246.
  43. Salomaki ED, Nickles KR, Lane CE. The ghost plastid of *Choreocolax polysiphoniae*. *J Phycol*. 2015;51:217–21.
  44. Wang L, Mao YX, Kong FN, Li GY, Ma F, Zhang BL, et al. Complete sequence and analysis of plastid genome of two economically important red alga-*Pyropia haitanensis* and *Pyropia yezoensis*. *PLoS One*. 2013;8:e65902.
  45. Lang BF, Nedelcu A. Plastid Genomes of Algae. In: Bock R, Knoop V, editors. *Genomics of chloroplasts and mitochondria*. Netherlands: Springer; 2012. p. 59–87.
  46. Hughey JR, Gabrielson PW, Rohmer L, Tortolani J, Silva M, Miller KA, et al. Minimally destructive sampling of type specimens of *Pyropia* (Bangiales, Rhodophyta) recovers complete plastid and mitochondrial genomes. *Sci Rep*. 2014;4:5113.
  47. Yang EC, Boo SM, Bhattacharya D, Saunders GW, Knoll AH, Fredericq S, et al. Divergence time estimates and the evolution of major lineages in the florideophyte red algae. *Sci Rep*. 2016;6:21361.
  48. Yang EC, Kim KM, Kim SY, Lee J, Boo GH, Lee J, et al. Highly conserved mitochondrial genomes among multicellular red algae of the Florideophyceae. *Genome Biol Evol*. 2015;7(8):2394–406.
  49. Saunders GW, Filloramo G, Dixon K, Le Gall L, Maggs CA, Kraft GT. Multigene analyses resolve early diverging lineages in the Rhodymeniophycidae (Florideophyceae, Rhodophyta). *J Phycol*. 2016. doi:10.1111/jpy.12426.
  50. Verbruggen H, Maggs CA, Saunders GW, Le Gall L, Yoon HS, De Clerck O. Data mining approach identifies research priorities and data requirements for resolving the red algal tree of life. *BMC Evol Biol*. 2010;10:16.
  51. Le NH, Lin SM. The discontinuous geographic distribution of *Gracilaria firma* (Gracilariaceae, Rhodophyta) from the Gulf of Tonkin to the Gulf of Thailand along the coastlines of Vietnam. *Vietnamese J Biotechnol*. 2005;3(3):373–80.
  52. Verbruggen H, Theriot EC. Building trees of algae: some advances in phylogenetic and evolutionary analysis. *Eur J Phycol*. 2008;43(3):229–52.
  53. Wrobel B. Statistical measures of uncertainty for branches in phylogenetic trees inferred from molecular sequences by using model-based methods. *J Appl Genet*. 2008;49:49–67.
  54. Simmons MP, Norton AP. Divergent maximum-likelihood-branch-support values for polytomies. *Mol Phylogenet Evol*. 2014;73:87–96.
  55. Glöckner G, Rosenthal A, Valentin K. The structure and gene repertoire of an ancient red algal plastid genome. *J Mol Evol*. 2000;51:382–90.
  56. Tajima N, Sato S, Maruyama F, Kurokawa K, Ohta H, Tabata S, et al. Analysis of the complete plastid genome of the unicellular red alga *Porphyridium purpureum*. *J Plant Res*. 2014;127:389–97.

57. Smith DR, Hua J, Lee RW, Keeling PJ. Relative rates of evolution among the three genetic compartments of the red alga *Porphyra* differ from those of green plants and do not correlate with genome architecture. *Mol Phylogenet Evol.* 2012;65:339–44.
58. Hughey JR. Genomic and phylogenetic analysis of the complete plastid genome of the California endemic seaweed *Wildemania schizophylla*. *Madrono.* 2016;63(1):34–8.
59. Zhang Y, Guo Y-M, Li T-J, Chen C-H, Shen K-N, Hsiao C-D. The complete chloroplast genome of *Gracilariopsis lemaneiformis*, an important economic red alga of the family Gracilariaceae. *Mitochondrial DNA B Resour.* 2016;1(1):2–3.
60. Kilpatrick ZM, Hughey JR. Mitochondrial and plastid genome analysis of the marine red alga *Coeloseira compressa* (Champiaceae, Rhodophyta). *Mitochondrial DNA B Resour.* 2016;1(1):456–8.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

