

MITOGENOME ANNOUNCEMENT

 OPEN ACCESS

## The complete chloroplast genome of *Gracilariopsis lemaneiformis*, an important economic red alga of the family Gracilariaceae

Yurong Zhang<sup>a</sup>, Yuan-Ming Guo<sup>a</sup>, Tie-Jun Li<sup>a</sup>, Ching-Hung Chen<sup>b</sup>, Kang-Ning Shen<sup>c</sup> and Chung-Der Hsiao<sup>d</sup>

<sup>a</sup>Marine Fisheries Research Institute of Zhejiang Province, Key Lab of Mariculture and Enhancement of Zhejiang Province, Zhoushan, Zhejiang, China; <sup>b</sup>WeThink Biotech Inc., Taoyuan, Taiwan; <sup>c</sup>Center of Excellence for the Oceans, National Taiwan Ocean University, Keelung, Taiwan;

<sup>d</sup>Department of Bioscience Technology, Chung Yuan Christian University, Chung-Li, Taiwan

### ABSTRACT

The complete chloroplast DNA (cpDNA) of a famous red alga of the family Gracilariaceae, *Gracilariopsis lemaneiformis*, was deduced by using next-generation sequencing and *de novo* assembly technology. The complete cpDNA of *G. lemaneiformis* consists of 182 505 bp and encodes 230 unique genes consisting 204 protein-coding genes (PCGs), 21 transfer RNA genes, 3 ribosomal RNA genes, 1 transfer-messenger RNA genes and 1 non-coding RNA genes. Among 204 PCGs, *ccsA* gene is interrupted by a intron. Unlike the typical quadripartite structure (a pair of inverted repeats separated by the small single-copy and large single-copy units) of cpDNA in higher plants, the complete cpDNA of *G. lemaneiformis* is very compact, containing no inverted repeat and just one copy of rRNA gene cluster consisting of 16S, 23S and 5S rRNA genes. The genic regions account for 83.7% of whole cpDNA genome, and the G + C content of the cpDNA was 27.4%. The low G + C content of *G. lemaneiformis* cpDNA is largely contributed by high A + T content in the PCGs and non-coding regions. A phylogenetic analysis of the 15 complete cpDNA from rhodophyta shows that *G. lemaneiformis* is closely related to macroalga *Gracilaria salicornia*. The complete cpDNA of *G. lemaneiformis* provides essential and important DNA molecular data for further phylogenetic and evolutionary analysis for rhodophyta.

### ARTICLE HISTORY

Received 27 November 2015  
Accepted 22 December 2015

### KEYWORDS

Chloroplast genome;  
*Gracilariopsis lemaneiformis*;  
next-generation sequencing;  
rhodophyta

### Introduction

The red alga *Gracilariopsis lemaneiformis* is one of the most important economic seaweeds in China (Tseng 2001). It provides good food source for abalone aquaculture (Cruz-Rivera & Friedlander 2011) and also been used as a biofiltration material to clean waste water (Cahill et al. 2010). Single-rope floating raft cultivation of this species has been conducted since 1990s in the south of China such as Fujian and Guangdong provinces. It is the most well-known agarophyte mainly used for agar extraction due to its enormous yield in China, with the agar produced from it accounting for more than 90% of the total annual production (Xiu-Geng et al. 1999). *G. lemaneiformis* is also considered to be an ideal material for genetic research because its whole life cycle can be completed in laboratory conditions. Recently, genome survey sequencing has been conducted in *G. lemaneiformis*, providing lots of essential genetic background information of this species (Zhou et al. 2013). In this study, the complete chloroplast genome of *G. lemaneiformis* was sequenced and characterized for the first time, aiming at providing useful information for further genetic and phylogenetic studies in this alga.

Sample of *G. lemaneiformis* (voucher no. 475) was collected from Dalian, Liaoning Province of China. Genomic DNA was extracted following the modified CTAB DNA extraction

protocol (Attitalla 2011) and then subjected to build up genomic library and pair-end sequencing (2X300 bp) by MiSeq (Illumina Inc., San Diego, CA, USA). By using commercial software (Geneious V9, Biomatters Ltd., Auckland, New Zealand), about 9.7% (1 376 840 out of 14 239 248) raw reads were *de novo* assembled to produce circular form of complete chloroplast DNA (cpDNA) genome with high average coverage of 2332 X. The complete cpDNA of *G. lemaneiformis* consists of 182 505 bp (GenBank KU179794), showing 83% identity to macroalga *Gracilaria salicornia* (Campbell et al. 2014) by sequence alignment.

Annotation of the assembled cpDNA genome was performed with DOGMA (Wyman et al. 2004), cpGAVAS (Liu et al. 2012) and manual inspected to predict protein-coding genes(PCGs), transfer RNA (tRNA) genes, ribosomal RNA (rRNA) genes and other genes. The cpDNA contains 230 unique genes consisting of 21 tRNA, 3 rRNA (16S–23S–5S gene cluster), 1 transfer-messenger RNA (*ssrA* gene), 1 non-coding RNA (*rnpB* gene) and 204 PCGs. Unlike the typical quadripartite structure (a pair of inverted repeats separated by the small single-copy and large single-copy unit) of cpDNA in higher plants, the complete cpDNA of *G. lemaneiformis* is very compact, containing no inverted repeat and just one copy of rRNA gene cluster consisting of 16S, 23S and 5S rRNA

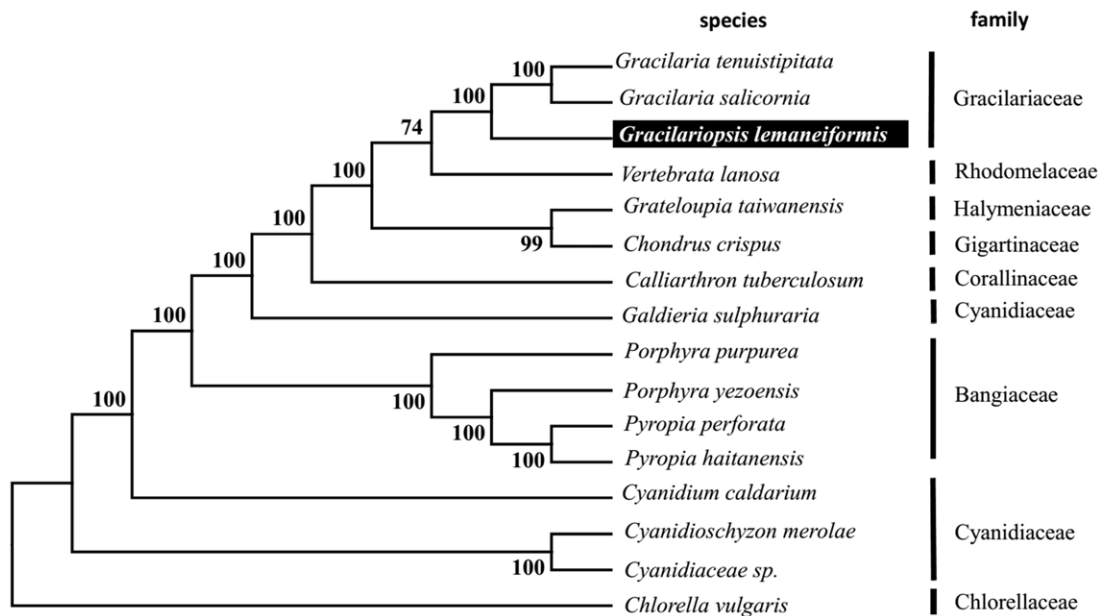


**Figure 1.** Molecular phylogeny of *Gracilariopsis lemaneiformis* and other related species in rhodophyta based on complete chloroplast genome. The complete chloroplast genome is downloaded from GenBank and the phylogenetic tree is constructed by Maximum likelihood method with 500 bootstrap replicates. The gene's accession number for tree construction is listed as follows: *Gracilaria tenuistipitata* (AY673996), *Gracilaria salicornia* (KF861575), *Gracilariopsis lemaneiformis* (KU179794), *Vertebrata lanosa* (KP308097), *Grateloupia taiwanensis* (NC\_021618), *Chondrus crispus* (HF562234), *Calliarthron tuberculosum* (KC153978), *Galdieria sulphuraria* (KJ700459), *Porphyra purpurea* (PPU38804), *Pyropia yezoensis* (KC517072), *Pyropia perforata* (KC904971), *Pyropia haitanensis* (KC464603), *Cyanidium caldarium* (AF022186), *Cyanidioschyzon merolae* (AB002583), *Cyanidiaceae sp.* (KJ569775) and *Chlorella vulgaris* (NC\_001865).

genes. This result is consistent with previous reported cpDNA genome in rhodophyta (Green 2011). There are only one PCG (*ccsA*, which is required for haeme attachment to chloroplast c-type cytochromes) containing single intron. The cpDNA consists of 83.7% genic regions, and the overall G + C content of the complete cpDNA is 27.4%. The G + C content is 48.9% for tRNA, 46.6% for rRNA and 28.9% for PCGs. Therefore, the low G + C content for the entire cpDNA is due to high A + T content in the PCGs and non-coding regions.

To validate the phylogenetic position of *G. lemaneiformis*, we used MAFFT (Katoh et al. 2002) to perform multiple sequence alignment and MEGA6 (Tamura et al. 2013) to construct a maximum-likelihood tree (with 500 bootstrap replicates) containing complete cpDNA of 15 algae species in rhodophyta. *Chlorella vulgaris* derived from Chlorophyta was used as out-group for tree rooting. Result shows *G. lemaneiformis* is closely related to *Gracilaria tenuistipitata* and *G. salicornia*, which are another two members in the family Gracilariaceae, with high bootstrap value supported (Figure 1). In conclusion, the complete cpDNA of *G. lemaneiformis* is decoded for the first time in this study and provides essential and important DNA molecular data for further phylogenetic and evolutionary analysis for rhodophyta.

## Declaration of interest

None of the authors report any conflict of interest. The authors alone are responsible for the content and writing of the paper. The research is supported from the grants from Natural Science Foundation of Zhejiang Province (LQ14B070002) and Zhejiang Province Science and Technology Department public technology research project of Agriculture (2015C32001).

## References

- Attitalla IH. 2011. Modified CTAB method for high quality genomic DNA extraction from medicinal plants. *Pak J Biol Sci.* 14:998–999.
- Cahill PL, Hurd CL, Lokman M. 2010. Keeping the water clean—seaweed biofiltration outperforms traditional bacterial biofilms in recirculating aquaculture. *Aquaculture* 306:153–159.
- Campbell MA, Presting G, Bennett MS, Sherwood AR. 2014. Highly conserved organellar genomes in the Gracilariaceae as inferred using new data from the Hawaiian invasive alga *Gracilaria salicornia* (Rhodophyta). *Phycologia* 53:109–116.
- Cruz-Rivera E, Friedlander M. 2011. Feeding preferences of mesograzers on aquacultured *Gracilaria* and sympatric algae. *Aquaculture* 322:218–222.
- Green BR. 2011. Chloroplast genomes of photosynthetic eukaryotes. *Plant J.* 66:34–44.
- Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30:3059–3066.
- Liu C, Shi L, Zhu Y, Chen H, Zhang J, Lin X, Guan X. 2012. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genomics* 13:715.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol.* 30:2725–2729.
- Tseng C. 2001. Algal biotechnology industries and research activities in China. *J Appl Phycol.* 13:375–380.
- Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20:3252–3255.
- Xiu-Geng F, Ying B, Shan L. 1999. Seaweed cultivation: traditional way and its reformation. *Chinese J Oceanol Limnol.* 17:193–199.
- Zhou W, Hu Y, Sui Z, Fu F, Wang J, Chang L, Guo W, Li B. 2013. Genome survey sequencing and genetic background characterization of *Gracilariopsis lemaneiformis* (Rhodophyta) based on next-generation sequencing. *PLoS One* 8:e69909.