

# Complete Sequences of the Mitochondrial DNA of the Wild *Gracilariopsis lemaneiformis* and Two Mutagenic Cultivated Breeds (Gracilariaceae, Rhodophyta)

Lei Zhang<sup>1</sup>, Xumin Wang<sup>2</sup>, Hao Qian<sup>1</sup>, Shan Chi<sup>1</sup>, Cui Liu<sup>1</sup>, Tao Liu<sup>1\*</sup>

**1** College of Marine Life Sciences, Ocean University of China, Qingdao, Shandong, People's Republic of China, **2** Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, People's Republic of China

## Abstract

The complete mitochondrial DNA (mtDNA) of *Gracilariopsis lemaneiformis* was sequenced (25883 bp) and mapped to a circular model. The A+T composition was 72.5%. Forty six genes and two potentially functional open reading frames were identified. They include 24 protein-coding genes, 2 rRNA genes, 20 tRNA genes and 2 ORFs (*orf60*, *orf142*). There is considerable sequence synteny across the five red algal mtDNAs falling into Florideophyceae including *Gr. lemaneiformis* in this study and previously sequenced species. A long stem-loop and a hairpin structure were identified in intergenic regions of mt genome of *Gr. lemaneiformis*, which are believed to be involved with transcription and replication. In addition, the mtDNAs of two mutagenic cultivated breeds ("981" and "07-2") were also sequenced. Compared with the mtDNA of wild *Gr. lemaneiformis*, the genome size and gene length and order of three strains were completely identical except nine base mutations including eight in the protein-coding genes and one in the tRNA gene. None of the base mutations caused frameshift or a premature stop codon in the mtDNA genes. Phylogenetic analyses based on mitochondrial protein-coding genes and rRNA genes demonstrated *Gracilariopsis andersonii* had closer phylogenetic relationship with its parasite *Gracilariophila oryzoides* than *Gracilariopsis lemaneiformis* which was from the same genus of *Gracilariopsis*.

**Citation:** Zhang L, Wang X, Qian H, Chi S, Liu C, et al. (2012) Complete Sequences of the Mitochondrial DNA of the Wild *Gracilariopsis lemaneiformis* and Two Mutagenic Cultivated Breeds (Gracilariaceae, Rhodophyta). PLoS ONE 7(6): e40241. doi:10.1371/journal.pone.0040241

**Editor:** Jan Aerts, Leuven University, Belgium

**Received:** March 15, 2012; **Accepted:** June 3, 2012; **Published:** June 29, 2012

**Copyright:** © 2012 Zhang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by Province Science and Technology in the Guangdong Project (2010B060200010, 2010B020201015) and the National Public Benefit Research Foundation of China (201203063). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: liutao@ouc.edu.cn

## Introduction

*Gracilariopsis* is an important economical marine red algae and widely used in food industry, agar extraction, etc [1,2]. *Gracilariopsis lemaneiformis* has been cultivated on large scales in both the southern and the northern parts of China for food and agar production, playing an effective role against coastal eutrophication [3]. For the past few years, *Gracilariopsis* cultivation has become the third largest seaweed cultivation industry only after *Laminaria* and *Porphyra* in China. Among *Gracilariopsis* seaweeds, the *Gr. lemaneiformis* is one of the most important cultivated breeds, because of its high yields and commercially valuable extracts.

Mutation is one of the important means of breeding new varieties of *Gr. lemaneiformis*. Chemical mutagen treatments (EMS: ethyl methanesulphonate and N-methyl-N-nitrosoguanidine) could increase the frequency of mutations in *Gracilaria* species [4]. In China, two important commercial cultivated breeds of *Gr. lemaneiformis* ("981" and "07-2") are both mutant stains. The strain "981" was bred from the wild *Gr. lemaneiformis* after mutation by N-methyl-N'-nitro-N-nitrosoguanidine (MNNG) and selected with hydroxyproline (HYP) protocol. The strain "981" has the characteristics of high temperature tolerance, fast growth rate and high levels of content and quality of agar [5]. "981" has been extensively cultivated as a source of commercial agar in the coastal areas of Fujian and Guangdong Province of China. The strain

"07-2" was obtained from the original "981" by the use of MNNG and selected with HYP protocol. The new strain ("07-2") has excellent features of fast growing and strong resistance to bad environment [6].

Mitochondria are thought to be derived from eubacterial endosymbionts [7], because they have biochemical machineries to replicate and transcribe of their own genomes. Mitochondrial DNA has the features of compact size, mostly maternal inheritance and fast evolutionary rate, it is a good molecular marker for evolutionary and population studies [8,9]. The *cox1* and *cox2-3* genes were good mtDNA taxonomic barcodes often used for species identification [10]. The whole mitochondrial genome makes it possible for the combination of multiple protein-coding genes, providing strong and precise phylogenetic analysis. However, limited information of complete mt genome is available. Since the first report on the complete mt genome of *Chondrus crispus* [11], there are only 6 red algal complete mt genomes deposited in the GenBank [12–14]. But for the seaweeds of brown algae (Phaeophyceae) and green algae (Chlorophyta), the complete mitochondrial data has reached 21 and 14 species respectively, more than that of red algal species. Here, we sequenced the complete mitochondrial genomes of wild *Gr. lemaneiformis* and two mutation strains. The comparison of the genome sequence and architecture reveals the influence of mutation to mtDNA. This work will contribute to further study about molecular phylogenetic

analysis of red algae and germplasm improvement for the economic *Gracilariopsis* species.

## Results

### Genome Organization and Comparison

The mitochondrial DNA of *Gr. lemaneiformis* is a circular molecule of 25883 nucleotides (GenBank accession number: JQ071938), with an overall A+T content of 72.5%. This base composition is consistent with previously reports of red algal mitochondrial genomes [11–14]. Forty six genes including 24 protein-coding genes, 2 ribosomal RNA (rRNA) genes and 20 transfer RNA (tRNA) genes were identified. No intron is inserted in any of these genes and all genes are encoded on both heavy strand (H-strand) and light strand (L-strand) with approximately the same encoding proportion on each strand. The genes are encoded in two opposite major transcriptional directions, suggesting the existence of two main transcriptional units (Figure 1). Compared to the six known red algal mitochondrial genomes, the protein-coding gene content and order in all seven mt genomes are virtually identical. Besides, there is considerable sequence synteny across the five mitochondrial genomes which fall in Florideophyceae including *Gr. lemaneiformis* in this study and previously sequenced: *Gracilariopsis andersonii*, *Gracilariophila oryzoides*, *Chondrus crispus* and *Plocamiocolax pulvinata* [14] (Table 1).

Compared to the mt genome of *Gr. andersonii* that from the same genus of *Gracilariopsis*, both genomes code for 26S rRNA gene (LSU) and 16S rRNA gene (SSU). The protein-coding gene content and order are almost identical with the following exceptions: the *yml39* gene is present in the *Gr. lemaneiformis* genome between *cox3* and *cob* gene; a fragment of nearly 2000 bp contains three predict ORFs are characteristic in intergenic region of *Gr. andersonii* mt genome; the gene *rps11* appears in both mt genomes but in opposite directions; the gene length of *rps12* in *Gr. andersonii* is 129 bp shorter than that of *Gr. lemaneiformis* for a premature stop codon. In addition, in mt genome of *Gr. lemaneiformis*, a fragment of ~600 bp containing a predicted ORF (*orf60*) is inserted into intergenic region between *sdhB* and *nad6* gene; another fragment of ~150 bp containing a stem-loop structure is present between *atp6* and *secY* gene (trn-Ala and trn-Asn, Figure 1).

### Gene Content and Codon Usage

The mtDNA of *Gr. lemaneiformis* codes for three subunits of the cytochrome oxidase (*cox1-3*), apocytochrome b (*cob*), seven subunits of the NADH dehydrogenase complex (*nad1-6*, *nad4L*), four ATPase subunits (*atp6*, *atp8*, *atp9*, *yml39*), five ribosomal proteins (*rps3*, *rps11*, *rps12*, *rpl16*, *rpl20*) and three succinate dehydrogenase complex (*sdhB*, *sdhC*, *sdhD*). Start and stop codons of all protein-coding genes were determined based on alignments with the corresponding genes and proteins of other Florideophyceae species. As shown in Table 2, the codon ATG was used as start codon for all genes, the stop codons include two types: TAA and TAG.

Twenty tRNA genes were determined and show standard cloverleaf secondary structures with the length ranging from 70–88 bp. All tRNA genes are dispersed throughout the *Gr. lemaneiformis* on both strands (Figure 1). Like all the other algal mt genomes published, the genome did not contain all the tRNAs needed to complete translation alone.

### Secondary Structures

At the demarcation point of the two opposite transcriptional units, a long and stable stem-loop (63 nt) was identified in

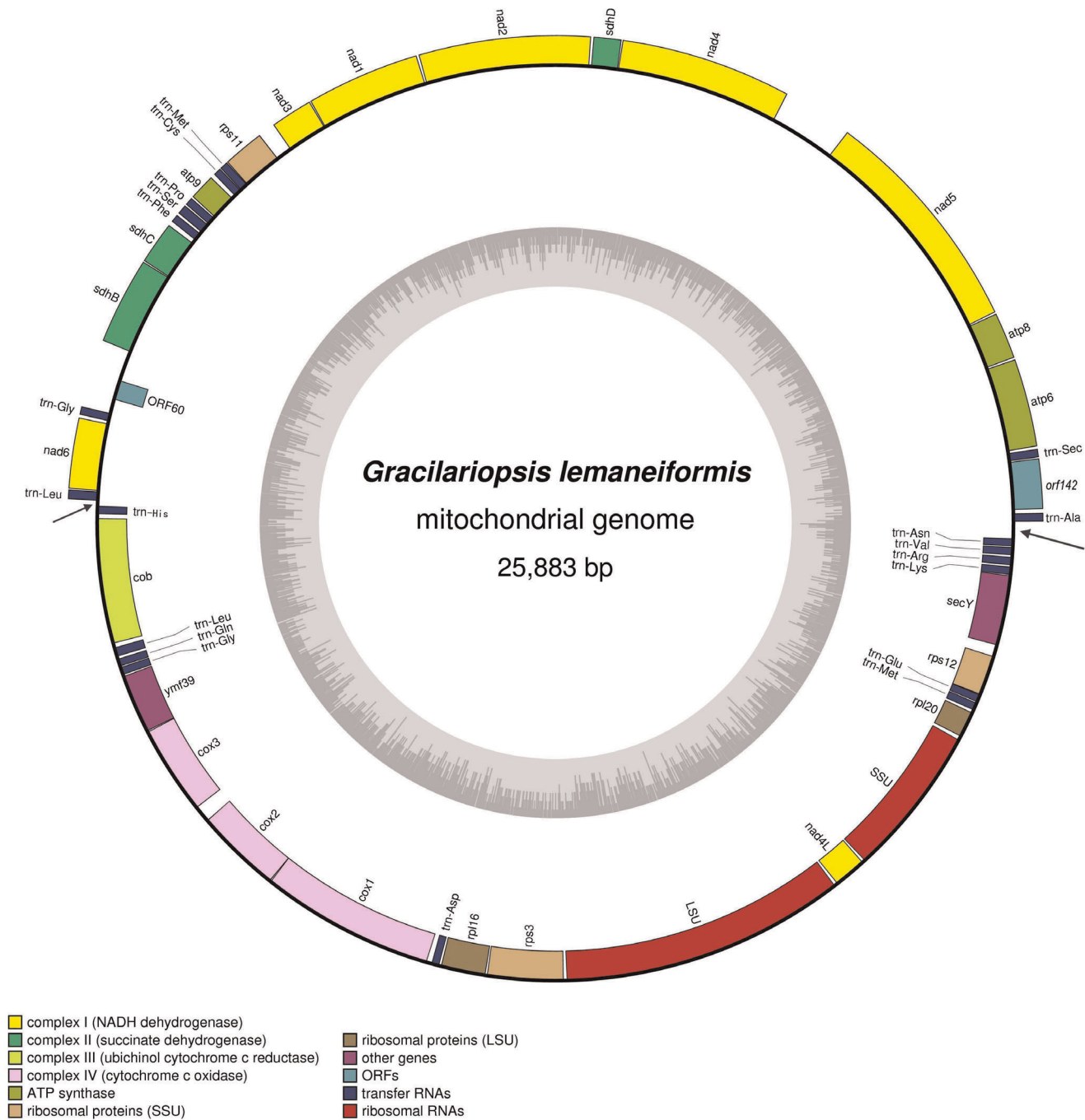
intergenic region of mt genome of *Gr. lemaneiformis*. Nearly diametrically opposite on the circular mt genome, another transcriptional demarcation point between trn-Leu and trn-His (Figure 1), a hairpin structure (21 nt) was found. Both secondary structures were complete inverted repeat sequences. The stem-loop and hairpin structures are likely to cause the termination of transcription, moreover, inverted repeat sequences provide the same binding site of enzyme at different directions. So, these structures could be involved in DNA transcription. Such strict stem-loop structures are also present in intergenic regions of *Chondrus crispus* (58 nt) and *Plocamiocolax pulvinata* (51 nt) mt genomes [11,14], and also locate at the demarcation point of different transcriptional units. There was no visible sequence homology between these secondary structures even with similar sequence length. However, all of these three inverted repeat sequences contain polymers of A and T (n = 6, 8, Figure 2). The polymers with low free energy tend to unlink easily and may be a recognition site of some enzymes. So the stem-loop structure is putative origin of replication of mtDNA.

### Comparison of mtDNA of Wild and Two Mutation Strains

The complete mt genomes of two mutation strains (“981”, “07-2”) were sequenced using the same method as the wild *Gr. lemaneiformis*. The mitochondrial genome size and gene order of the two mutation strains were the same as the wild *Gr. lemaneiformis*. Only nine base mutations were detected and all base mutations located at gene coding regions. Of which, eight were present in the protein-coding genes and one in the tRNA gene. The protein-coding genes contained base mutations were *atp6*, *nad5*, *rps11*, *sdhC* and *cob* genes. However, none of these base mutations caused frameshifts or a premature stop codon in the mtDNA genes. The only one mutation present in tRNA gene did not make any change to the structure and function of *tRNA<sup>Pro</sup>*. Among the eight base mutations present in the protein-coding genes, five of them are samesense mutations and the other three are nonsynonymous mutations (Table 3).

### Phylogenetic Analyses

Three data partitions (19 protein-coding genes, LSU and SSU) from seven red algal mt genomes were used to construct phylogenetic trees. The aligned sequence length of protein-coding genes, LSU and SSU were 15915 bp, 2882 bp and 1573 bp respectively. Phylogenetic trees with bootstrap values are present in Figure 3. Both MP and ML analyses yielded the same trees for each data partition and phylogenetic trees of three data partitions were similar with each other, in which *Cyanidioschyzon merolae* and *Porphyra purpurea* formed single clade as they belong to Cyanidiphyceae and Bangiophyceae respectively; the remaining five species falling into the same class (Florideophyceae) clustered together (Figure 3). Among three *Gracilariaceae* species, phylogenetic analyses demonstrated *Gr. andersonii* had closer relationship with *Gracilariophila oryzoides*, even though *Gr. andersonii* and *Gr. lemaneiformis* belong to the same genus of *Gracilariopsis*. The topological differences were mainly occurred within Florideophyceae clade. Within this clade, the branching order of *Plocamiocolax pulvinata* and *Chondrus crispus* were different among the phylogenetic trees based on different data partitions of mtDNA: *Plocamiocolax pulvinata* clustered out ahead of *Chondrus crispus* when using the data partition of protein-coding genes, whereas the branch order was opposite when using the data partition of LSU and SSU (Figure 3).



**Figure 1. Gene map of *Gracilariopsis lemaneiformis* mitochondrial genome.** Genes outside the map are transcribed counter clockwise and those inside the map are transcribed clockwise. Genes belonging to different functional groups are color coded. The black arrows indicate the positions of stem-loop and hairpin structures.  
doi:10.1371/journal.pone.0040241.g001

**Discussion**

To date, the complete mitochondrial genomes of seven red algae (including this work) have been reported. According to the results of previous studies and our research, the mtDNA sizes of red algae are ranging from 25.1 kb to 36.7 kb, and the A+T content of these mitochondrial genomes are 72.0–76.1% [11–14]. However, for the seaweeds, the mtDNA sizes of brown and green algae are in the range of 31.6–58.5 kb and 12.9–95.8 kb respectively. It is obvious that there exists great variance between

mt genomes of seaweeds, including the kinds of protein-coding genes, species of tRNAs and rRNA genes. Moreover, the mtDNAs of green algae have special model of linear molecule and circular molecule [15,16]. In addition, the mtDNAs of brown and red algae were both circular molecule and encoded on H-strand and L-strand, but the encoding proportion of two strands are very different. For red algae, the encoding proportion of each strand is nearly the same; while the L-strand of brown algae just encodes several ribosomal protein genes, the other protein-coding genes

**Table 1.** The comparisons of protein-coding genes and rRNA genes of five mitochondrial genomes from Florideophyceae.

Genes	<i>Chondrus crispus</i>	<i>Plocamiocolax pulvinata</i>	<i>Gracilariopsis andersonii</i>	<i>Gracilariophila oryzoides</i>	<i>Gracilariopsis lemaneiformis</i>
Protein-coding genes					
<i>atp6</i>	+	+	+	+	+
<i>atp8</i>	+	–	+	–	+
<i>atp9</i>	+	+	+	+	+
<i>cob</i>	+	+	+	+	+
<i>cox1</i>	+	+	+	+	+
<i>cox2</i>	+	+	+	+	+
<i>cox3</i>	+	+	+	+	+
<i>nad1</i>	+	+	+	+	+
<i>nad2</i>	+	+	+	+	+
<i>nad3</i>	+	+	+	+	+
<i>nad4</i>	+	+	+	+	+
<i>nad4L</i>	+	+	+	+	+
<i>nad5</i>	+	+	+	+	+
<i>nad6</i>	+	+	+	+	+
<i>sdhB</i>	+	+	+	+	+
<i>sdhC</i>	+	+	+	–	+
<i>sdhD</i>	+	+	+	+	+
<i>secY</i>	+	+	+	+	+
<i>ymf39</i>	+	+	–	+	+
<i>rpl16</i>	+	+	+	+	+
<i>rpl20</i>	+	+	+	+	+
<i>rps3</i>	+	+	+	+	+
<i>rps11</i>	+	+	+	+	+
<i>rps12</i>	+	+	–	+	+
Ribosomal RNA genes					
rRNA SSU	+	+	+	+	+
rRNA LSU	+	+	+	+	+

doi:10.1371/journal.pone.0040241.t001

and all tRNAs genes are encoded on H-strand, indicating a clear strand-specific bias in encoding genes. As noted by Gray, the mitochondrial genomes of red algae are incredibly well conserved in both content and genome architecture [17]. In despite of the discrepant mt genome size of red algae, the A+T content remains at a relatively stable level. Especially for the five species from Florideophyceae, the number and order of mtDNA genes are almost the same (Table 1).

According to previous reports, the mitochondrial genome size had undergone a substantial expansion from the green algae *Chara vulgaris* to land plants [18], which was mainly accounted for the enlargement of intergenic regions and duplications, and also get DNA from the chloroplast and the nucleus [19–22]. However, such phenomenon has not been observed in the mt genomes of brown and red seaweeds except for few species have few introns [14]. For the seven known mt genomes of red algae, compared to the primitive red alga *Cyanidioschyzon merolae* from Cyanidiophyceae, the genome sizes of other six species from Bangiophyceae and Florideophyceae show a process of decreasing [11–14], the mitochondrial genomes of red algae have a trend to become smaller and more stable.

### Codon Usage

For the 24 protein-coding genes of mt genome of *Gr. lemaneiformis*, there are just two types of stop codon, TAA and TAG. As reported before, special in the mitochondrial genome, the modified genetic code TGA specifies tryptophan instead of the stop codon [23]. These modifications are considered to be correlated to an A+T pressure [24]. Like the green-plant *Acanthamoeba castellanii* which displays a mitochondrial A+T content of 71% use a modified genetic code, whereas another green alga *Prototheca wickerhamii* has the similar A+T content of 74% but with no usage of TGA as stop codon nor tryptophan [25,26]. This situation of a deviant genetic code was observed in our research, and the codon TGA distributed over nearly all the genes of complete mt genome of *Gr. lemaneiformis*. For the known mt genomes of green and brown algae, we chose the gene *cox1* for a simple search, the results indicated no modified genetic code TGA was used in brown algae and only one was present in green algae. However, all the seven red algal mt genomes contain different numbers of modified genetic code TGA, it can be said this modified genetic code has the highest frequency in red algae. Moreover, such phenomenon was also observed in *cox1* genes of other red algal genera of *Eucheuma* and *Kappaphycus* [10].

**Table 2.** Characteristics of the mitochondrial protein-coding genes of *Gracilariopsis lemaneiformis*.

Gene	Size (bp)	Codon			Strand
		Amino acid	Start	Stop	
<i>apt6</i>	762	253	ATG	TAA	H
<i>atp8</i>	405	134	ATG	TAA	H
<i>nad5</i>	1992	663	ATG	TAA	H
<i>nad4</i>	1476	491	ATG	TAG	H
<i>sdhD</i>	240	79	ATG	TAA	H
<i>nad2</i>	1476	491	ATG	TAA	H
<i>nad1</i>	981	326	ATG	TAA	H
<i>nad3</i>	366	121	ATG	TAA	H
<i>rps11</i>	360	119	ATG	TAA	H
<i>atp9</i>	231	76	ATG	TAA	H
<i>sdhC</i>	372	123	ATG	TAA	H
<i>sdhB</i>	753	250	ATG	TAA	H
<i>nad6</i>	609	202	ATG	TAG	H
<i>cob</i>	1143	380	ATG	TAA	L
<i>ymf-39</i>	543	180	ATG	TAA	L
<i>cox3</i>	819	272	ATG	TAA	L
<i>cox2</i>	792	263	ATG	TAA	L
<i>cox1</i>	1596	531	ATG	TAA	L
<i>rpl16</i>	411	136	ATG	TAA	L
<i>rps3</i>	696	231	ATG	TAA	L
<i>nad4L</i>	306	101	ATG	TAA	L
<i>rpl20</i>	243	80	ATG	TAG	L
<i>rps12</i>	369	122	ATG	TAG	L
<i>secY</i>	735	244	ATG	TAA	L

doi:10.1371/journal.pone.0040241.t002

### Influence of Mutation

Colour mutants are relatively widespread in red algae [27]. Moreover, chemical mutagen treatments can increase the frequency of mutations [4]. For *Gr. lemaneiformis*, spontaneous colour mutants have been observed in the nature environment. Early in 2002, Sui found that the absorption spectra of phycoerythrins (PE) changed significantly between *Gracilaria lemaneiformis* Greville and its pigmental mutants [28]. The two mutants ("981" and "07-2") used in this study were the products mutated by chemical mutagenesis and selected the strains with excellent characteristics. We sequenced the complete mt genomes of the two mutants, the comparison of mtDNA between the mutants and wild *Gr. lemaneiformis* revealed that there was no

**Table 3.** Comparison of mtDNA base mutations between wild *Gr. lemaneiformis* and two mutation strains.

Gene	Position	wild	"981"	"07-2"	Strand
<i>apt6</i>	732	ATT (Ile)	ATT (Ile)	ATC (Ile)	H
<i>nad5</i>	1092	CAG (Gln)	CAA (Gln)	CAG (Gln)	H
<i>rps11</i>	188	TGT (Cys)	TGT (Cys)	TAT (Tyr)	H
<i>tRNA<sup>Pro</sup></i>	27	A (trn-Pro)	G (trn-Pro)	A (trn-Pro)	H
<i>sdhC</i>	93	TTC (Phe)	TTC (Phe)	TTT (Phe)	H
<i>cob</i>	756	CAT (His)	CAT (His)	CAC (His)	L
<i>cob</i>	30	TCT (Ser)	TCC (Ser)	TCT (Ser)	L
<i>cob</i>	25–26	CTT (Leu)	ACT (Thr)	CTT (Leu)	L

doi:10.1371/journal.pone.0040241.t003

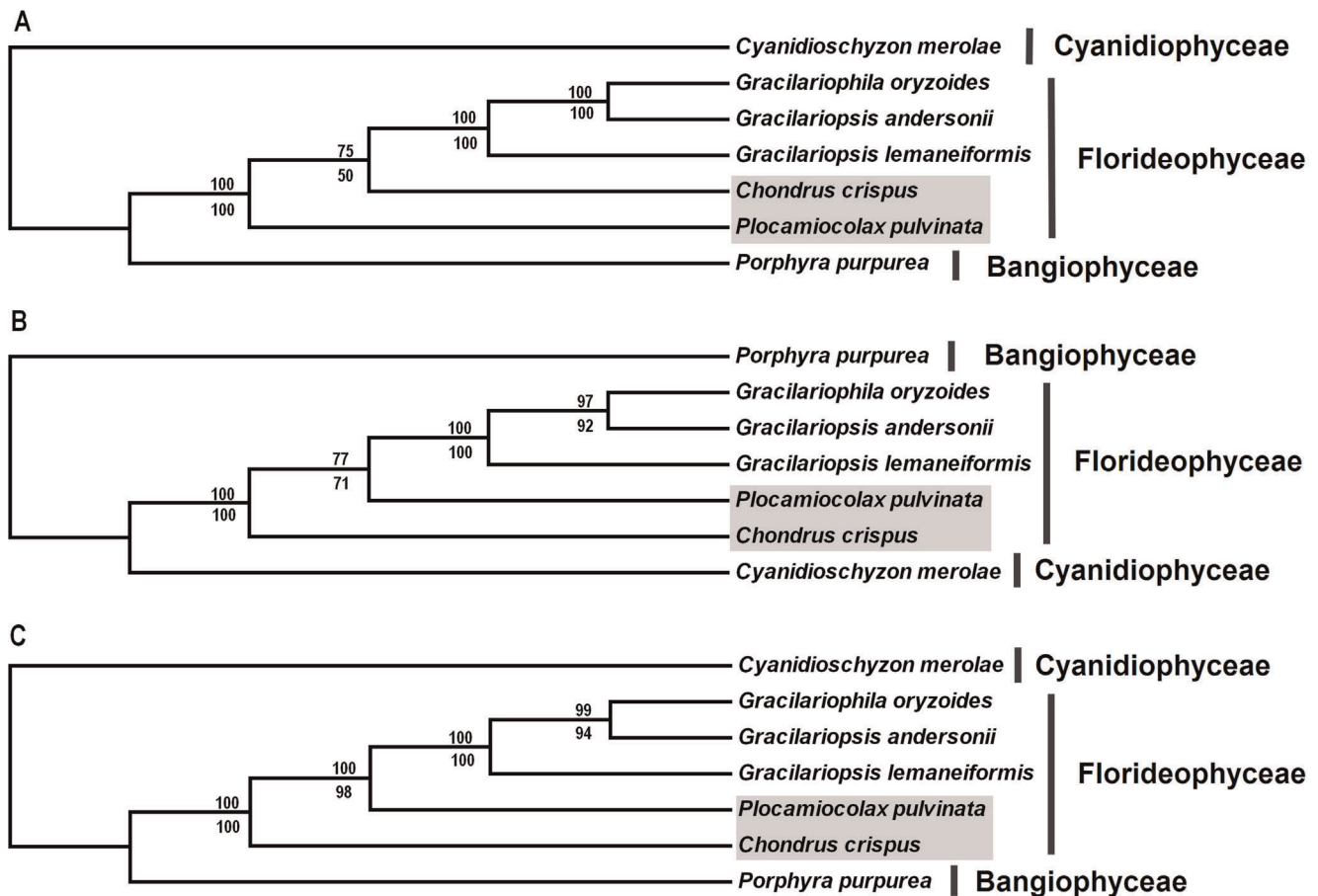
obvious variance except several amino acids changing caused by base mutations. Among the eight base mutations present in the protein-coding genes, five of them are same-sense mutations and cause no impact to the gene structure and gene function. The other three are nonsynonymous mutations, but there is no significant clue for the influence on the protein-coding genes except an amino acid change. These results suggest that the mutation does not cause the large-scale rearrangements and obvious insertion or deletion to *Gr. lemaneiformis* mt genome. To further understand the exact gene in relevant to the new characteristics particular for mutants, more efforts need to be done.

### Origin of Replication

As reported before, the mitochondrial genome of mammalian has the displacement-loop (D-loop) which also known as the control region, including transcriptional promoters for both strands and the heavy strand replication origin [29,30]. This noncoding region attracted many studies for its supposedly rapid rate of evolution [31]. In our research, a length of 63 nt complete inverted repeat sequence was detected in intergenic region of mt genome of *Gr. lemaneiformis*, and the stem-loop structure located at the demarcation point of two transcriptional units. This inverted repeat sequence seems like the D-loop structure of animal mt genome that can cause the termination of transcription, so it was presumed to be the origin of replication and in relevant with transcription. Similar structures were found in *C. crispus* and *P. pulvinata* mt genomes, and in *P. pulvinata* it was considered to be the origin of replication [14]. However, these strict secondary structures were only found in three mt genomes of red algae, and no similar sequence was found throughout the all known mt genomes of brown algae. Interestingly, even *Gr. andersonii* and *Gr. lemaneiformis* both belong



**Figure 2.** The comparison of sequences of stem-loop and hairpin structure from different mt genomes. **1:** *Gracilariopsis lemaneiformis* (126 bp). **2:** *Plocamioocolax pulvinata* (102 bp). **3:** *Chondrus crispus* (116 bp). **4:** the hairpin structure of *Gr. lemaneiformis* (42 bp). The dotted line indicates the middle point from where both sides are complete inverted repeat sequences. The shadow indicates polymers of A and T (n = 6, 8). doi:10.1371/journal.pone.0040241.g002



**Figure 3. ML and MP trees of seven red algae based on different data partitions of mtDNA. A:** 19 protein-coding genes. **B:** LSU. **C:** SSU. The ML and MP trees have the same topology and only one is shown. Numbers above and below nodes are ML and MP bootstrap support values respectively. The shadow indicates topological difference between trees based on different data partitions. doi:10.1371/journal.pone.0040241.g003

to the genus of *Gracilariopsis*, no such structure was found in the mt genome of *Gr. andersonii*.

In our research, the complete mitochondrial sequence was obtained by PCR strategy and sequenced by dideoxy chain termination method, the instability of PCR reaction may cause the deletion or base mutation of DNA fragments, special for the region include secondary structure. Frequently, normal rTaq enzyme can easily cross the folded hairpin structure and result in the loss of partial PCR fragment. Therefore, the accuracy and integrity of sequence must be considered when choosing the strategy of PCR to obtain the complete mitochondrial genome.

## Materials and Methods

### Ethics Statement

As a normal kind of economical red algae, no specific permits are required for the studies on *Gr. lemaneiformis* so far. The wild *Gr. lemaneiformis* was collected from Shandong Province of China in May 2011, the two mutation strains (“981”, “07-2”) were collected from Fujian Province of China in April 2011. All materials were collected in the coastal area and the location is neither privately-owned nor protected places.

### DNA Isolation, Genome Sequencing

Total DNA was extracted from fresh thalli with modified CTAB method [32]. mtDNA of *Gr. lemaneiformis* was amplified by

polymerase chain reaction (PCR) using primers (Table S1) designed in Primer Premier 5.0 from the conserved regions of the known red algal mitochondrial genomes [11–14]. The PCR reaction conditions include an initial denaturation cycle of 94°C for 3 min followed by 30 cycles of 94°C for 30 s, annealing temperature depending on primer sequences for 45 s, and 72°C for 1 min. A final extension step was performed at 72°C for 15 min followed by a 4°C hold. For large fragments, the extension time depend on the sequence length (approximately 1 min for 1,000 bp). PCR products were purified and sent for sequencing on an ABI 3730 automated sequencer (Applied Biosystems). Individual sequences were edited and assembled into contigs using DNASTar (DNASTAR, Inc., Madison, USA).

### Genome Annotation

Genes were identified using the BlastN and BlastX algorithms to compare the predicted ORFs to the NCBI GenBank database [33]. Ribosomal RNA (rRNA) genes and coding regions were determined by sequence alignment with the known mitochondrial genes of *Gr. andersonii* and *Gracilariophila oryzoides* [14]. Transfer RNA (tRNA) genes were identified with tRNAscan-SE 1.21 software [34]. Mitochondrial genome map of *Gr. lemaneiformis* was produced by using OGDRAW software [35]. Sequence alignment and base composition were conducted using MEGA 4.0 software [36].

## Phylogenetic Analyses

Phylogenetic analyses of the red algal species were performed based on the following data set: 19 protein-coding genes (*atp6*, *atp9*, *cob*, *cox1*, *cox2*, *cox3*, *nad1*, *nad2*, *nad3*, *nad4L*, *nad4*, *nad5*, *nad6*, *rpl16*, *rps11*, *rps12*, *rps3*, *sdhB*, *sdhD*), 26S rRNA gene (LSU) and 16S rRNA gene (SSU). All sequences are commonly present in all seven mt genomes publicly available in the GenBank database. Sequences were aligned using MEGA 4.0 [36] software and edited manually. Maximum parsimony (MP) analyses were performed with PAUP\*4.0b10 [37]. Heuristic searches were conducted with 1000 random addition replicates and tree bisection reconnection (TBR) branch swapping with “multrees” option. For maximum likelihood (ML) analyses, RA×ML version 7.0 [38] was used with searches relied on the general time-reversible (GTR) and gamma models. The local bootstrap probability of each branch was calculated by 1000 replicates. MEGA 4.0 was used for displaying and printing phylogenetic trees.

## References

- Santelices B, Doty MS (1989) A review of *Gracilaria* faming. *Aquaculture* 78: 59–133.
- Xia BM, Zhang JF (1999) *Flora Algarum Marinarum Sinicarum: Rhodophyta*. Science Press. 55–58.
- Xu JT, Gao KS (2007) Growth, pigments, UV-absorbing compounds and agar yield of the economic red seaweed *Gracilaria lemaneiformis* (Rhodophyta) grown at different depths in the coastal waters of the South China Sea. *J Appl Phycol* 20: 231–236.
- Zhang X, Van Der Meer JP (1987) A study on heterosis in diploid gametophytes of the marine red alga *Gracilaria tikvahiae*. *Bot Mar* 30: 309–314.
- Meng L, Xu D, Chen WZ, Zhang XC (2009) Selection and characterization of a new strain of *Gracilaria lemaneiformis*. *Periodical of Ocean University of China* 39: 94–98.
- Chen WZ, Xu D, Wang LG, Men L, Du H, et al. (2009) Preliminary Study on economic characteristics and agar characteristics of two new strains of *Gracilaria lemaneiformis*. *Periodical of Ocean University of China* 39: 437–442.
- Gray MW (1992) The endosymbiont hypothesis revisited. *Int Rev Cytol* 141: 233–357.
- Wilson AC, Cann RL, Carr SM, George M, Gyllensten UB, et al. (1985) Mitochondrial DNA and two perspectives on evolutionary genetics. *Biol J Linn Soci* 26: 375–400.
- Brown WM, George M JR, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. *Pro Nat Acad Sci U S A* 76: 1967–1971.
- Conklin KY, Kurihara A, Sherwood AR (2009) A molecular method for identification of the morphologically plastic invasive algal genera *Eucheuma* and *Kappaphycus* (Rhodophyta, Gigartinales) in Hawaii. *J Appl Phycol* 21: 691–699.
- Leblanc C, Boyen C, Richard O, Bonnard G, Grienerberger JM, et al. (1995) Complete sequence of the mitochondrial DNA of the rhodophyte *Chondrus crispus* (Gigartinales): Gene content and genome organization. *J Mol Biol* 250: 484–495.
- Burger G, Saint-Louis D, Gray MW, Lang BF (1999) Complete sequence of the mitochondrial DNA of the red alga *Porphyra purpurea*: Cyanobacterial introns and shared ancestry of red and green algae. *Plant Cell* 11: 1675–1694.
- Ohta N, Sato N, Kuroiwa T (1998) Structure and organization of the mitochondrial genome of the unicellular red alga *Cyanidioschyzon merolae* deduced from the complete nucleotide sequence. *Nucleic Acids Research* 26: 5190–5198.
- Hancock L, Goff L, Lane C (2010) Red algae lose key mitochondrial genes in response to becoming parasitic. *Genome Biol Evol* 2: 897–910.
- Nosek J, Tomaska L, Kucejova B (2004) The chromosome end replication: lessons from mitochondrial genetics. *J Appl Biomed* 2: 71–79.
- Smith DR, Hua J, Lee RW (2010) Evolution of linear mitochondrial DNA in three known lineages of *Polytomella*. *Curr Genet* 56: 427–438.
- Gray MW, Lang BF, Burger G (2004) Mitochondria of protists. *Annu Rev Genet* 38: 477–524.
- Turmel M, Otis C, Lemieux C (2003) The mitochondrial genome of *Chara vulgaris*: Insights into the mitochondrial DNA architecture of the last common ancestor of green algae and land plants. *The Plant Cell* 15: 1888–1903.
- Turmel M, Otis C, Lemieux C (2002) The chloroplast and mitochondrial genome sequences of the charophyte *Chaetosphaeridium globosum*: Insights into the timing of the events that restructured organelle DNAs within the green algal lineage that led to land plants. *Proc Natl Acad Sci USA* 99: 11275–11280.
- Unsold M, Marienfeldt JR, Brandt P, Brennicke A (1997) The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nat Genet* 15: 57–61.
- Kubo T, Nishizawa S, Sugawara A, Itchoda N, Estiati A, et al. (2000) The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for tRNA(Cys)(GCA). *Nucleic Acids Res* 28: 2571–2576.
- Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, et al. (2002) The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: Frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Mol Genet Genomics* 268: 434–445.
- Boyen C, Leblanc C, Bonnard G, Grienerberger JM, Kloreg B (1994) Nucleotide sequence of the *cox3* gene from *Chondrus crispus*: evidence that UGA encodes tryptophan and evolutionary implications. *Nucl Acids Res* 22: 1400–1403.
- Osawa S, Jukes TH, Watanabe K, Muto A (1992) Recent evidence for evolution of genetic code. *Microbiol Rev* 56: 229–264.
- Burger G, Plante I, Lonergan KM, Gray MW (1995) The mitochondrial DNA of the amoeboid protozoan, *Acanthamoeba castellanii*: complete sequence, gene content, and genome organization. *J Mol Biol* 245: 522–537.
- Wolff G, Plante I, Lang BF, Kück U, Burger G (1994) Complete sequence of the mitochondrial DNA of the chlorophyte alga *Prototheca wickerhamii*. *J Mol Biol* 237: 75–86.
- Van Der Meer JP (1990) Genetics. In: Cole KM, Sheath RG, editors. *Biology of the Red Algae*. Cambridge: University Press. 103–119.
- Sui ZH, Zhang XC, Cheng XJ (2002) Comparison of phycobiliproteins from *Gracilaria lemaneiformis* (Rhodophyceae) and its pigment mutants in spectral and molecular respects. *Acta Botanica Sinica* 44: 557–561.
- Clayton DA (1982) Replication of animal mitochondrial DNA. *Cell* 28: 693–705.
- Chang DD, Clayton DA (1986) Identification of primary transcription start sites of mouse mitochondrial DNA: accurate in vitro initiation of both heavy- and light-strand transcriptions. *Mol Cell Biol* 6: 1446–1453.
- Hoelzel AR, Hancock JM, Dover GA (1991) Evolution of the cetacean mitochondrial D-Loop region. *Mol Biol Evol* 8: 475–493.
- Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* 12: 13–15.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389–3402.
- Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25: 955–964.
- Lohse M, Dreshsil O, Bock R (2007) OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr Genet* 52: 267–264.
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 24: 1596–1599.
- Swofford DL (1998) PAUP\*. *Phylogenetic Analysis Using Parsimony* (\*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
- Stamatakis A (2006) RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690.

## Supporting Information

**Table S1** Primers sequences designed to sequence three strains of *Gracilariopsis lemaneiformis*. (DOC)

## Acknowledgements

We would like to thank Tianxiang Gao for critical review of the manuscript. We are indebted to Yan Sun for assistance with experiments.

## Author Contributions

Conceived and designed the experiments: TL LZ. Performed the experiments: LZ. Analyzed the data: LZ HQ. Contributed reagents/materials/analysis tools: XMW HQ SC CL TL. Wrote the paper: LZ TL.