MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE
KYIV NATIONAL UNIVERSITY OF TECHNOLOGIES AND DESIGN
Faculty of Chemical and Biopharmaceutical Technologies
Department of Biotechnology, Leather and Fur

QUALIFICATION THESIS

on the topic **Phylogenetic analysis of mitochondrial genomes of *Gracilaria* and *Gracilariopsis***

First (Bachelor's) level of higher education
Specialty 162 "Biotechnology and Bioengineering"
Educational and professional program "Biotechnology"

Completed: student of group BEBT-20
Zexuan ZHANG

Scientific supervisor
Liubov ZELENA, Ph.D., As. prof.

Reviewer
Tetiana SHCHERBATIUK, Dr. Sc., Prof.

Kyiv 2024

KYIV NATIONAL UNIVERSITY OF TECHNOLOGIES AND DESIGN

Faculty: Chemical and Biopharmaceutical Technologies
Department: Biotechnology, Leather and Fur
First (Bachelor's) level of higher education
Specialty: 162 Biotechnology and Bioengineering
Educational and professional program Biotechnology

APPROVE
Head of Department of Biotechnology,
Leather and Fur, Professor,
Doctor of Technical Science
Olena MOKROUSOVA

_____
«___»_____2024

ASSIGNMENTS
FOR THE QUALIFICATION THESIS
Zexuan Zhang

1. Thesis topic **Phylogenetic analysis of mitochondrial genomes of *Gracilaria* and *Gracilariopsis***
scientific supervisor Liubov Zelena, PhD
approved by the order of KNUTD "__"_____20__, №_____

2. Initial data for work: assignments for qualification thesis, scientific literature on the topic of qualification thesis, materials of Pre-graduation practice

3. Content of the thesis (list of questions to be developed): literature review; object, purpose, and methods of the study; experimental part; conclusions

4. Date of issuance of the assignments_____

EXECUTION SCHEDULE

№	The name of the stages of the qualification thesis	Terms of performance of stage	Note on performance
1	Introduction	From 01 April 2024 to 11 April 2024	
2	Chapter 1. Literature review	From 06 April 2024 to 20 April 2024	
3	Chapter 2. Object, purpose, and methods of the study	From 21 April 2024 to 30 April 2024	
4	Chapter 3. Experimental part	From 01 May 2024 to 10 May 2024	
5	Conclusions	From 07 May 2024 to 12 May 2024	
6	Draw up a bachelor's thesis (final version)	From 20 May 2024 to 25 May 2024	
7	Submission of qualification work to the supervisor for feedback	From 25 May 2024 to 27 May 2024	
8	Submission of bachelor's thesis to the department for review (14 days before the defense)	27 May 2024	
9	Checking the bachelor's thesis for signs of plagiarism (10 days before the defense)	31 May 2024	
10	Submission of bachelor's thesis for approval by the head of the department (from 7 days before the defense)	03 June 2024	

I am familiar with the task:

Student _____Zexuan ZHANG

Scientific supervisor _____Liubov ZELENA

SUMMARY

Zexuan Zhang. Phylogenetic analysis of mitochondrial genomes of *Gracilaria* and *Gracilariopsis*. Manuscript.

Qualification thesis on the specialty 162 «Biotechnology and Bioengineering». – Kyiv National University of Technologies and Design, Kyiv, 2024.

In this study, complete genome sequences of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* were obtained by GenBank database. The mitochondrial genomes of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* were compared and analyzed by Geneious, Mauve, tRNAscan-SE and other software. The total length of mitochondrial genome of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* was 25,829bp and 25,883bp, respectively. *Gracilaria chouae* encoded 50 genes, including 24 protein coding genes, 2 rRNAs, 23 tRNAs and orf148. *Gracilariopsis lemaneiformis* encoded 48 genes. It included 24 protein-coding genes, 2 rRNAs, 20 tRNAs, and 2 ORFs (orf60, orf143). The proportions of A, T, C and G were similar, and the content of G+C was respectively 28.8% and 27.5%. The lengths of the two RRNA genes are different, the types and numbers of the 24 protein-coding genes are the same, and the lengths of most of the protein-coding genes (17) are completely the same. *atp8, nad5, nad2, sdhC, rpl16, rpl20* and *secY* are different in length. *trnH* gene is rearranged in mitochondrial genomes of *Gracilaria chouae* and *Gracilariopsis lemaneiformis*. The relevant parameters of mitochondrial genome codon are different between them. T The ENC values of mitochondrial genome of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* were 50.07 and 48.85, respectively, both of which were greater than 35, indicating that the codon use of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* were weak. The mean values of CBI and RSCU of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* were different.

Keywords: Gracilaria chouae; Gracilariopsis lemaneiformis; comparative genome

TABLE OF CONTENTS

INTRODUCTION

Gracilaria chouae belongs to the Rhodophyta, Florideophyceae, Gigartianles, Gracilariaceae, *Gracilaria*. It is endemic species of China. It is mainly distributed in the coastal areas south of Zhejiang Province. As an important raw material for extracting agar, it also has a variety of edible, medicinal and other economic values.

Gracilariopsis lemaneiformis belongs to Rhodophyta, Florideophyceae, Gigartianles, Gracilariaceae, *Gracilariopsis*. Is distributed in coastal areas of China and is widely distributed in many coastal areas around the world. In addition to rich in agar, it has important value in food, medicine and maintaining Marine ecological balance.

In this study, complete genome sequences of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* were obtained by GenBank database. The mitochondrial genomes of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* were compared and analyzed by Geneious, Mauve, tRNAscan-SE and other software. The total length of mitochondrial genome of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* was 25,829bp and 25,883bp, respectively. *Gracilaria chouae* encoded 50 genes, including 24 protein coding genes, 2 rRNAs, 23 tRNAs and orf148. *Gracilariopsis lemaneiformis* encoded 48 genes. It included 24 protein-coding genes, 2 rRNAs, 20 tRNAs, and 2 ORFs (orf60, orf143). The proportions of A, T, C and G were similar, and the content of G+C was respectively 28.8% and 27.5%. The types of tRNA in mitochondrial genomes of the two are not exactly the same, and the secondary structure of the two rRNA genes is shamrock-shaped, and most of them do not have extra rings. The lengths of the two RRNA genes are different, the types and numbers of the 24 protein-coding genes are the same, and the lengths of most of the protein-coding genes (17) are completely the same. *atp8, nad5, nad2, sdhC, rpl16, rpl20* and *secY* are different in length. *trnH* gene is rearranged in mitochondrial genomes of *Gracilaria chouae* and *Gracilariopsis lemaneiformis*. It can be used as a molecular marker between the two genera. The results of collinearity analysis showed

that the gene arrangement of *Gracilaria* was generally conservative. The relevant parameters of mitochondrial genome codon are different between them. T The ENC values of mitochondrial genome of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* were 50.07 and 48.85, respectively, both of which were greater than 35, indicating that the codon use of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* were weak. The mean values of CBI and RSCU of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* were different.

From a phylogenetic perspective, we compared *Gracilaria chouae* and *Gracilariopsis lemaneiformis* with 20 protein-coding genes shared by 11 species of *Gracilaria*, *Gracilriopsis* and *Corallina officinalis*. The phylogenetic tree was constructed by BI and ML models. This suggests that *Gracilaria* and *Gracilriopsis* species evolved independently in the course of evolution, and *Gracilaria* and *Gracilriopsis* are two separate genera.

The relevance of the topic is Comparative analysis based on mitochondrial genome.

The purpose of the study is the Understanding of taxonomic problems of Gracilaria.

The objectives of the study is the comparative analysis of Gracilaria and Gracilaria from the perspective of mitochondrial genome provides theoretical evidence for the classification of Gracilaria family

The subject of the study: *Gracilaria chouae* and *Gracilariopsis lemaneiformis*.

Research methods: Consult literature and use bioinformatics software.

The scientific novelty: It is analyzed from the perspective of mitochondrial genome

The practical significance of the results obtained is *Gracilaria* and *Gracilriopsis* are two separate genera.

CHAPTER 1
LITERATURE REVIEW

1.1 Overview of Gracilaria chouae

1.1.1 Classification and natural distribution of Gracilaria chouae

Gracilaria chouae belongs to Rhodophyta, Florideophyceae, Gigartianles, Gracilariaceae, *Gracilaria* [1]. *Gracilaria chouae* is a subtropical seaweed endemic to China. It is mainly distributed in the coastal area south of Zhejiang Province.

1.1.2 Morphological characteristics of Gracilaria chouae

Gracilaria chouae is transparent purplish-red, light red when fresh, slightly darker after drying, the meat is thick and juicy, brittle, and very easy to break. The algal body is erect, solitary or clumpy, cylindrical, with a small disc-like fixator at the base. The common height is 15~20cm, the maximum is 45cm, the width is 2~3mm, the main branch is cylindrical, the branches are 2~4 times, irregularly alternate, partial or bifurcated, the base of the branches is wider, the tip is thin [5].

According to the cross-sectional diagram of the interior of the algal body, the central part of the algal body is the pulp, which is composed of large irregular round to oval parenchyma cells, and the outer part is a cortex composed of one or two layers of nearly round inner cortex cells. The outermost epidermal layer cells contain chromatosomes. Conspicuous protruding surface, conical or hemispherical [2]

1.1.3 Economic value of Gracilaria chouae

Gracilaria chouae can be used as a Marine vegetable because of its brittleness, thickness and juiciness and rich protein and polysaccharide. It also has certain medicinal value and can be used as raw material for aquatic health products. In addition, *Gracilaria chouae* can also have the characteristics of fast growth, high yield, simple cultivation method, etc. It is suitable for the extraction of agarice as a raw material, and is also a high-quality feed for abalone culture [3]. When it is mixed with cultured animals such as fish, ginseng and shellfish, it can reduce the concentration of nitrogen and phosphorus in the water, increase the dissolved oxygen,

inhibit the growth of microalgae and restore the aquaculture water environment. *Gracilaria chouae* can also provide necessary shelter conditions for cultured animals and increase the yield of cultured animals [4]. *Gracilaria chouae* as a large red algae, can absorb nitrogen, phosphorus, carbon, etc., in the growth process, which can improve water quality and delay water eutrophication, and has wide application value [28].

1.2 Overview of Gracilariopsis lemaneiformis

1.2.1 Classification and natural distribution of Gracilariopsis lemaneiformis

Gracilariopsis lemaneiformis belongs to Rhodophyta, Florideophyceae, Gigartianles, Gracilariaceae, *Gracilariopsis*. *Gracilariopsis lemaneiformi*s is a kind of red algae Marine plant. It is distributed all over the coast of our country and widely distributed in many sea areas worldwide, such as the coast of Japan, Canada and the United States, Venezuela and South Africa [7]. It grows on rocks or sand from intertidal to low tidal zones and is half-buried on sand-covered rocks.

1.2.2 Morphological characteristics of Gracilariopsis lemaneiformis

Gracilariopsis lemaneiformis frond is erect, linear, cylindrical and clumpy, with a disc-like, large and flat fixator at the base, 10~50 cm high and up to 1 m high. *Gracilariopsis lemaneiformis* is mostly purple brown, fresh *Gracilariopsis lemaneiformis* frond is purple red, sometimes slightly green or yellow, soft bone, after drying brown, algae body wrinkled. Generally with a more obvious trunk, diameter 1~2 mm, branching 1~2 times, alternate or partial in all directions, branch diameter 0.5~1 mm. Mature algal bodies have more elongated branches, especially the upper branches are often exposed into whiplash, and the branches of immature algal bodies are densely covered with short branchlets, the base of the branch is generally not shrinkable, sometimes slightly shrinkable, the base of the branchlets is often slightly wider, the diameter of the branch is 0.5~2 mm, the end of the branch is gradually pointed, and the branch itself is often broken due to damage, and new branches are regenerated at the cross section [1].

The inner part of the algal body is the pulp composed of large parenchyma cells, and the outer part is 2~5 layers of gradually smaller cortical cells. The tetrasporangium is purplish red, scattered on the surface of the algal body, buried in the cortical cells, cross-shaped division, sperm capsules are born in shallow pits or in the sunken part of the reproductive nest, pale yellow. The capsule fruit is spherical or hemispherical [1].

1.2.3 Economic value of Gracilariopsis lemaneiformis

Asparagus is rich in gelatine and can be used to produce gelatine for use as biological experimental medium. Dragon's asparagus has a cold taste, has the effect of softening firm phlegm, clearing heat and reducing water, and has certain medicinal value [6]. As a dual-purpose seaweed for food and medicine, dragon's asparagus has good nutritional value. In the food industry, dragon's asparagus has developed a variety of foods and achieved good economic effects. In 2000, there was a practice of processing asparagus into spicy, sour and sweet flavor food through alkali treatment, softening, seasoning, sterilization and other treatment technologies [25]. In addition, the dragon beard vegetable was treated and fermented by microbial fermentation, and the dragon beard vegetable was made into flavored seaweed sauce [26]. In addition, asparagus is used to make yogurt, and the dietary fiber extracted from asparagus can be made into a delicate dietary fiber yogurt, which is more conducive to intestinal health [27]. At the same time, the cultivation of asparagus can effectively consume nitrogen and phosphorus and CO2 in the air, which has important environmental benefits in maintaining the Marine ecological balance [8]. Asparagus also contains phycoglobin and contains 0.0315% cholesterol [1].

1.3 Overview of mitochondria

1.3.1 Origin of mitochondria

There are two hypotheses about the origin of mitochondrion: endosymbiosis theory and non-endosymbiosis theory. Nowadays, endosymbiosis theory is widely

accepted, believing that mitochondrion originated from an endosymbiosis event, and its ancestors are a kind of Gram-negative bacteria that can perform tricarboxylic acid cycle and electron transfer, called proto-mitochondrion. It is also the ancestor of alpha-proteobacterial, an aerobic bacterium that is a branch of the phylum proteobacteria and closely related to Rickettsia. The core evidence for this conclusion is that, based on molecular data accumulated over the past few decades, the mitochondrial genome clearly has remnants of the true bacterial genome [10].

A number of molecular biology and bioinformatics studies based on mitochondrial genomes have supported the following conclusions: First, there are significant similarities between mitochondrial genomes and bacterial genomes. Specifically, the eukaryotic mitochondrial genome can be said to be a simplified version of the bacterial genome in essence, and its similarities are reflected in many aspects, including DNA configuration similar to bacterial circular DNA, and certain similarities with bacteria in base ratio, gene structure characteristics, mRNA and rRNA sedimentation coefficients, etc. Second, mitochondria have a separate and complete protein synthesis system, which in many ways is more similar to the bacterial protein synthesis system. For example, protein synthesis in mitochondria begins with N-formylmethionine, unlike that in eukaryotic cells; Mitochondria, chloroplasts and prokaryotes contain 5SrRNA, while many eukaryotic ribosomes contain 5.8SrRNA. In addition, protein synthesis factors in mitochondria have recognition specificity for prokaryotic ribosomes, but not for cytoplasmic ribosomes. These findings further support the close association of mitochondrial genomes with bacterial genomes. Finally, in terms of genetic code, the genetic code of mitochondria is more similar to that of Proteobacteria [11]. It has been suggested that this symbiosis occurred about 1.7 billion years ago, at about the same time as the evolutionary divergence that gave rise to eukaryotes and archaea [34].

1.3.2 Characteristics of mitochondrial genome

The mitochondrial matrix contains circular DNA molecules that can direct the synthesis of some of its own proteins, which is the mitochondrial genome. Mitochondria have their own set of genetic systems, relatively independent of the nuclear chromosome genome. Some male sterile plants have been shown to have genes in their mitochondria.

Since its formation, mitochondria have embarked on a long road of evolution, which is not only reflected in the structure, but also significantly reflected in its genome level. In the long period of time, the mitochondrial genome is constantly exchanging genes with the nucleus to realize the transfer of genetic information. It is worth noting that the mitochondrial genome is relatively limited in size and contains a small number of genes, significantly smaller than the size of the bacterial genome. Given the idea of endosymbiosis that mitochondria originated from phagocytic bacteria, their genomes should have been similar in size to bacteria. To explain this difference, the researchers hypothesized that some of the genes in the original mitochondria were lost during evolution, while most of the remaining genes were transferred to the nucleus of the host cell. As a result, more than 98% of the proteins expressed within mitochondria today are actually encoded by nuclear genes.

In addition, mitochondrial genomes are typically concentrated in a single molecule of mtDNA, which is highly utilized. The genes of the mitochondrial genome are so tightly arranged that in some regions they overlap (the last base of the previous gene is connected to the first base of the next gene). These introns in mtDNA may play a specific role in the processing and translation of gene transcripts.

1.3.3 Applications of mitochondrial genomes

Because of its small size, lack of recombination and simple structure, mitochondrial genome is easy to be sequenced and analyzed in depth. And because of the rapid rate of evolution of mitochondrial DNA (mtDNA) and the maternal-inherited nature of most species, it has become an indispensable and

commonly used tool in the fields of phylogenetics and comparative genetics, for example as a molecular marker.

These advantages of mtDNA allow us to gain a deeper understanding of its structure, function, and genetic properties, and to further explore the interactions between mitochondrial and nuclear genomes. With the continuous progress and popularization of DNA sequencing technology, the research on mtDNA is increasingly in-depth, and its sequence analysis will be more widely applied in evolutionary genetics and molecular systematics, providing us with a powerful tool to reveal the mystery of life evolution [20].

1.4 Research progress of mitochondrial genome of red algae

Algae as a type of eukaryotic organism containing plastids, have extremely rich diversity, covering over 6000 unicellular and multicellular groups. They are widely distributed and occupy an important position in the Earth's ecosystem. Since the origin of mitochondria, through the initial endosymbiosis of plastids, three major algal groups have been formed: glaucophyte, red algae, and green algae. Subsequent secondary or even multiple endosymbiosis have further promoted the diversified development of algal populations, resulting in a variety of algal groups.

Although various algae exhibit certain phylogenetic relationships at the level of systematic evolution, their mitochondrial genomes exhibit significant differences. Generally speaking, mitochondrial DNA is composed of a chromosome with multiple copies, and this structure is related to the genetic sequences of the ancestors of α-Proteobacteria maintain a high degree of similarity. These characteristics provide valuable clues for the in-depth study of algal biology [10].

The morphology of the red algae community exhibits extremely high diversity, covering about 760 genera and a total of more than 4410 species, with multicellular species dominating and relatively few single celled species. Red algae are mainly distributed in oceans around the world, especially in temperate and tropical nearshore waters, but there are also a few species that can grow in freshwater. These species not

only possess rich biodiversity, but also have significant economic and ecological significance.

The plastids of red algae have unique structural characteristics, namely non stacked thylakoids. In addition, red algae also use their unique phycoerythrin as an auxiliary photosynthetic pigment, which provides them with unique advantages in photosynthesis. Red algae lack centrioles and flagella throughout their entire life cycle, which makes them particularly unique in biological classification and phylogenetic research [10].

The mitochondrial genome of red algae is relatively small compared to other biological groups, ranging in size from 25.2 to 41.7 kbp, with relatively low gene content. Moreover, its genome configuration is circular, and the gene arrangement is compact, with only 3.6% to 11.8% of the spacer region. There are multiple gene overlap regions in the mitochondrial genome of red algae, and their gene coding sequence is also extremely conservative. Red algae have three succinate dehydrogenase genes encoded by mtDNA, namely *sdh*2, *sdh*3, and *sdh*4.

In terms of the coding of tRNA genes, the number of red algae groups is relatively moderate, ranging from 18 to 27. Most red algae choose TGA as the codon for encoding tryptophan, which reflects the uniqueness of red algae in genetic information coding.

Rhodophyta species exhibit significant diversity in introns. Although protein coding genes generally contain introns, in red algae, these introns are mainly concentrated in *trn*I and *rnl* genes. For example, *Chondrus crispus* and all seaweed species carry type II introns. It is worth noting that *Pyropia yezoensis* has the highest number of introns, up to five, while *Porphyra umbilicalis* and *Chondrus crispus* each have only one type II intron. However, studies have shown that *Gracilaria lemaneiformis does* not contain introns [12].

Of particular concern is the significant similarity between the two introns in the *RNL* gene of *P.purpurea* and the intron in the *Calathrix sp.* of cyanobacteria, suggesting a recent gene level transfer event between bacteria and mitochondria. In the

mitochondrial genome of *G. lemanei formis*, the spacer region of the *ATP6* and *secY* genes contains a 126bp reverse repeat sequence, which is rich in poly A and T and can form a stable stem ring structure, possibly serving as the starting site for mtDNA replication. This special structure was first discovered in the mitochondrial genome of Gracilaria species [18].

1.5 Research progress of mitochondrial genome acquisition methods (sequencing methods)

1.5.1 Traditional mitochondrial genome sequencing methods

The traditional mitochondrial genome sequencing methods rely on physical isolation of mitochondrial DNA and clone library sequencing technology. This method first involves the separation and purification steps of mitochondrial DNA, usually using cesium chloride density gradient centrifugation, alkaline lysis, or differential centrifugation to achieve purification. Subsequently, these DNA fragments were segmented to adapt to the effective sequencing length of the sequencer. Common fragmentation methods include ultrasonic fragmentation and restriction endonuclease cleavage. Subsequently, these short fragments of DNA were cloned onto suitable plasmid vectors and sequenced using a sequencer. Finally, the mitochondrial whole genome sequence was obtained through sequence splicing technology.

Although this method has the advantages of obtaining high-purity mtDNA and ensuring sequence accuracy, its operation process is lengthy and complex, involving multiple time-consuming steps, with a low success rate and high cost. In addition, the isolation and purification of mitochondrial DNA also require fresh tissue materials, which may be limited in practice. Therefore, this method is currently less commonly used for full-length sequencing of mitochondrial genomes.

1.5.2 Homologous PCR amplification direct sequencing method

The core principle of homologous PCR sequencing of genomes is to design PCR primers on the total DNA template by utilizing information from organelle genomes of species that have been obtained and are closely related to the species being sequenced.

The design of these primers needs to ensure coverage of the entire organelle genome, followed by amplification of specific organelle DNA fragments using PCR technology and cloning onto appropriate vectors for sequencing. Finally, through sequence splicing technology, we can obtain complete organelle genome sequences.

In the process of primer design, it is usually chosen to design in the conserved region of the organelle genome, and a pair of primers are designed every approximately 1000 bp to ensure that there is an overlap of 50 to 150 bp between adjacent amplification products. Due to the requirement of designing primers that can cover the entire organelle genome and generate overlapping regions, this method is called "primer walking"[14]. At present, this method is still the mainstream technology in small-scale organelle genome research.

With the continuous development of long PCR technology, this method has been able to successfully amplify DNA fragments exceeding 5 kb, providing strong technical support for large-scale sequencing projects such as mitochondrial genomes.

1.5.3 High throughput mitochondrial genome sequencing

High throughput sequencing technology, also known as "next-generation" sequencing technology or Massively parallel sequencing (MPS), is significantly different from traditional Sanger sequencing and is known for its ability to perform parallel sequencing on a large number of nucleic acid molecules simultaneously. This technology can efficiently generate at least 100Mb of sequencing data in one sequencing reaction, greatly improving sequencing efficiency. At present, the mainstream platforms for high-throughput sequencing technology include Roche's 454 pyrophosphate sequencing platform, Illumina's Solexa genome analysis platform, and Applied BioSystems (ABI)'s SOLiD sequencing platform. The data output of these platforms far exceeds that of traditional Sanger sequencing, hence the name "high-throughput" sequencing technology. The emergence of high-throughput sequencing technology has not only greatly promoted the research progress of genomics and functional genomics, entering a new stage of high efficiency and low

cost, but also provided new possibilities for the study of algal mitochondrial genomes, foreshadowing the arrival of the era of large-scale and efficient sequencing [15].

1.6 The purpose and significance of this research

Gracilaria seaweed has a fast growth rate and high agar content, making it a large red algae with important economic value and ecological benefits. Its industrial market prospects are broad, and it has received widespread attention from algae related researchers and markets. However, as an important economic group in red algae, its complex classification relationship has not yet been clarified.

Since the establishment of the genus Gracilaria in 1830, its early classification mainly relied on the observation of the external morphological characteristics of algae. However, due to the diversity and complexity of the growth environment, many species of algae in the genus Gracilaria exhibit significant variations in morphology, color, texture, and internal structure. In addition, some species face great challenges in classification and identification due to the lack of reproductive organs, which can easily lead to confusion between synonyms or homonyms.

With the rapid development of molecular biology, researchers have provided new molecular level evidence for the classification of the genus *Gracilaria*. Traditionally, *Gracilariopsis lemaneiformis* has been classified under the genus *Gracilaria*, but the latest molecular biology research has proposed different perspectives. Researchers have conducted a systematic evolutionary analysis of seaweed and asparagus in China using multiple sequences, including the 18S rRNA gene encoded by nuclear genes, the cox2-3 spacer in mitochondria, the Rubisco spacer in chloroplasts, and plastid genes. This research result further confirms that *Gracilariopsis lemaneiformis* does not belong to the genus *Gracilaria* and is relatively primitive in evolutionary relationships [19]. These findings not only deepen our

understanding of the genus *Gracilaria* and its related species, but also provide new perspectives and tools for future taxonomy and biodiversity research.

In the phylum Red Algae, although Graciliaceae is a clear family at all levels, the classification relationships between its genera and species are complex due to the lack of morphological and reproductive structural features. Although molecular biology methods have been widely introduced into the systematic classification of the genus Gracilaria, most studies still focus on ITS and rbcL sequences, and the results are not entirely consistent. Relying solely on a single DNA sequence for analysis may lead to vastly different or even erroneous conclusions from different studies due to differences in evolutionary rates and other factors. In view of this, this article provides a detailed comparative analysis of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* from the perspective of mitochondrial genome, aiming to provide richer genomic information for the molecular biology and genetic exploration of *Gracilaria chouae* and *Gracilariopsis lemaneiformisin* in the Gracilaceae.

CHAPTER 2
MATERIALS AND METHODS

2.1 Experimental method

2.1.1 Data sources

Using the published GenBank database *Gracilaria chouae, Gracilariopsis lemaneiformis, Gracilaria multipartita, Gracilaria ornata, Gracilaria tikvahiae, Gracilaria urvillei, Gracilaria usneoides, Gracilariopsis longissima, Gracilariopsis mclachlanii* and *Gracilariopsis tenuifrons* complete mitochondrial genomics of GenBank. Download it in .gb format and .fasta format, respectively. The structure and evolution of mitochondrial genome of Gracilaria were studied by means of bioinformatics. For details see Table 2.1. (Tab. 2.1).

Table 2.1 **Mitochondrial genome of Gracilaria for comparative genomics**

species	Family	Genus	Accession number
Gracilaria chouae	Gracilariaceae	*Gracilaria*	NC_038211.1
Gracilaria multipartita	Gracilariaceae	*Gracilaria*	NC 058681.1
Gracilaria ornata	Gracilariaceae	*Gracilaria*	NC 058682.1
Gracilaria tikvahiae	Gracilariaceae	*Gracilaria*	NC 058683.1
Gracilaria urvillei	Gracilariaceae	*Gracilaria*	NC 058684.1
Gracilaria usneoides	Gracilariaceae	*Gracilaria*	NC 058685.1
Gracilariopsis lemaneiformis	Gracilariaceae	*Gracilariopsis*	JQ071938.1
Gracilariopsis longissima	Gracilariaceae	*Gracilariopsis*	NC 039150.1
Gracilariopsis mclachlanii	Gracilariaceae	*Gracilariopsis*	NC 039151.1
Gracilariopsis tenuifrons	Gracilariaceae	*Gracilariopsis*	NC 058686.1

2.1.2 Comparative analysis of mitochondrial genome between *Gracilaria chouae* and *Gracilariopsis lemaneiformis*

For the analysis of tRNA genes in mitochondrial genomes of Gracilaria gracilaria and Gracilaria Gracilaria, tRNAscan-SE should first be used to input the sequences of mitochondrial genomes downloaded from NCBI on the tRNAscan-SE webpage and select "Run tRNAscan-SE". Thus, relevant information such as the types, lengths, corresponding amino acids, anti-codons and secondary structures of the two trnas can be obtained [23].

A series of genes such as protein-coding gene, tRNA gene and rRNA gene in mitochondrial genome of Gracilaria Gracilaria and Gracilaria longiana were analyzed by Geneious software. Download and use the Mauve plugin in Geneious to import in "Align with progressiveMauve" the mitochondrial genome sequences of Gracilaria and asparagus downloaded from NCBI in ".fasta "format. Click "Align" to carry out research work such as collinearity analysis [21].

2.1.3 Analysis of mitochondrial genome codon usage in *Gracilaria chouae* and *Gracilariopsis lemaneiformis*

Firstly, the mitochondrial genomes of the two species will be imported into Phylosuite [30], and then right-click "Extract" to extract genomic data, and check the data of codon usage in the "extract_results" folder.

Visualization of the data is performed using the Draw RSCU Figure function under the Mitogenome column of Phylosuite software [24].

2.2 Systematic Evolutionary Analysis Data Processing

2.2.1 Sequence Acquisition and Alignment

Ten mitochondrial genomes of species from the *Gracilaria* and *Gracilariopsis* were downloaded from NCBI, and *Corallina officinalis* (NC033904) was selected as an outgroup. These 11 genomes were screened for shared protein-coding genes, resulting in 20 conserved protein-coding genes: atp6, atp8, cox1, cox2, cox3, cob, nad1, nad2, nad3, nad4, nad4L, nad5, nad6, rpl16, rps3, rps11, rps12, sdh2, sdh3, and sdh4.

Firstly, the shared genes were individually named, and the sequences of these shared genes from the eleven species were saved in a single ".fasta" format file. Subsequently, using MEGA11[31], these gene sequences were translated into protein sequences and individually aligned, with input format as ".fasta" and output format as ".fas".

FASconCAT v1.0.pl script was utilized to concatenate these 20 genes into a single file[24]. Then, Gblocks online tool was employed to align the amino acid sequences of all species to remove poorly conserved regions, reducing the original 5382 translated sites to 5146 sites (approximately 95% of the original sites).

2.2.2 Construction of Phylogenetic Trees

Two main methods were employed for constructing phylogenetic trees: Bayesian inference (BI) and Maximum Likelihood (ML). Bayesian inference (BI), widely used in statistics, gradually extended to molecular phylogenetics. This method is based on posterior probability, representing the probability of the correctness of the tree, for tree construction. The core of Bayesian inference lies in using Markov chain Monte Carlo (MCMC) algorithm to estimate the posterior probability of the evolutionary tree, enabling effective handling of nucleotide and amino acid sequence data and flexible analysis of morphological data or even mixed data. Maximum Likelihood (ML) builds phylogenetic trees based on the principle of cumulative probability to analyze sequence variations. ML is particularly suitable for cases where there are many mutations between species sequences, resulting in low sequence similarity and distant relative positions in the phylogenetic tree.

Bayesian inference (BI): Sequences were converted to NEX format using MEGA11 and analyzed for tree building using Bayesian mixed models. The iteration number was set to 500,000 generations, running 4 Markov chains, with saving every 100 generations, discarding the first 1250 burn-in samples. The resulting trees were visualized using FigTree v1.4.4[32].

Maximum Likelihood (ML): The sequences were first processed in Gblocks online to output a "phy" format file. Then, the "standard-RAxML-master"[35]

software was used, selecting the automatic model and running for 1000 generations. The resulting trees were visualized using FigTree v1.4.4[32].

CHAPTER 3

RESULTS AND DISCUSSION

3.1 Results and analysis

3.1.1 Comparison of basic characteristics of Gracilaria gracilaria and *Gracilariopsis lemaneiformis*

Figure 3.1 – External morphological map of *Gracilaria chouae*

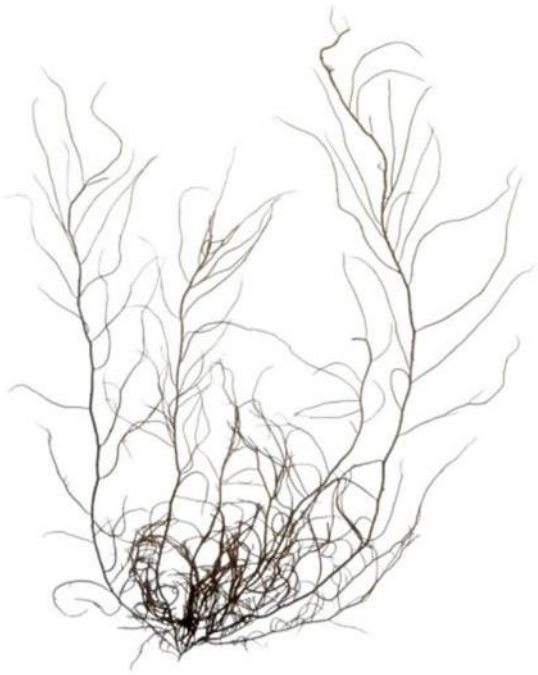


Figure 3.2 – External morphological map of *Gracilariopsis lemaneiformis*

Gracilaria chouae and *Gracilariopsis lemaneiformis* are two species of red algae in Gracilaria. The algae bodies of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* were purplish red. The algae bodies were upright, cylindrical, and the base had disc-like fixators. According to the external morphological map of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* [20] (Fig. 2.1、Fig2.2), it is difficult to distinguish between *Gracilaria chouae* and *Gracilariopsis lemaneiformis.* Therefore, it is key to conduct a comparative study on *Gracilaria chouae* and *Gracilariopsis lemaneiformis* from the perspective of molecular biology. Next, this paper will compare *Gracilaria chouae* and *Gracilariopsis lemaneiformis* from all aspects of mitochondrial genome.

3.1.2 Analysis of mitochondrial genome structure of *Gracilaria chouae* and *Gracilariopsis lemaneiformis*

The mitochondrial genome of *Gracilaria chouae* was 25,829 bp (GenBank entry number: NC_038211). The total A + T content in mitochondrial genome was 71.2%. The mitochondrial genome contains 50 genes, including 24 protein-coding

genes, 2 rRNA genes, 23 tRNA genes, and one ORF (orf148). Of the 24 protein-coding genes, 21 (87.5%) ended in TAA termination codons and 3 (12.5%) ended in TAG (*atp*8, *rps*11 and *nad*4L). All protein-coding genes of *Gracilaria chouae* use the start codon ATG. The length of the two rRNA genes is 2626 bp (LSU rRNA) and 1399 bp (SSU rRNA). The number and structure of mitochondrial genes of Gracilaria were basically similar to those published in the NCBI sequence database [17].

Species name	Length (bp)	GC content (%)	*Protein Coding genes*	tRNA genes	rRNA genes	ORF	Genbank No.
G. chouae	25,829	28.8	25	23	2	1	NC_038211
G.lemaneiformis	25,883	27.5	24	20	2	2	JQ071938

Table 3.1 – **The information of *G. chouae* and *G.lemaneiformis***

Gracilariopsis lemaneiformis mtDNA of the length of 25883 bp (GenBank login number: JQ071938), among them A + T content of 72.5%. The mtDNA of asparagus contains 48 genes. It includes 24 protein-coding genes, 2 rRNA genes, 20 tRNA genes and 2 ORFs (orf60, orf143). A long stem ring and a hairpin structure, possibly related to transcription and replication, were found in the intergenic region of the mitochondrial genome of *Gracilariopsis lemaneiformis* [18]. The secondary structure of the 20 tRNA genes contained in the mitochondrial genome of *Gracilariopsis lemaneiformis* is a standard clover shape with a length between 70 and 88 bp. All tRNA genes are scattered across two strands of the asparagus mitochondrial genome (Fig. 2.2).

3.1.3 Protein-coding genes

The coding gene products of mitochondrial genomes of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* include 1 cytochrome b (*cob*), 3 cytochrome oxidase subunits (*cox*1-3), 4 ATP synthetase subunits (*ymf*39, *atp*6, *atp*8, *atp*9), 3 succinate dehydrogenase complex subunits (*sdh*B, *sdh*C, sdhD), 7 subunits of the NADH dehydrogenase complex (*nad*1-6, *nad*4L), 5 ribosomal protein subunits (*rps*3, *rps*11, *rps*12, *rpl*16, *rpl*20), and 1 membrane protein (*sec*Y). The number of protein-coding genes and gene products of mitochondria of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* were basically consistent with those of published mitochondrial genomes of Gracilaria.

Table 3.2 **Protein-coding genes present in the mitochondrial genomes from *G. chouae and G. lemaneiformis***

gene product	*G. chouae*	*G. lemaneiformis*
Cytochrome oxidase ：3（*cox*1-3）	+	+
Apocytochrome b：1（*cob*）	+	+
ATPase subunits：4（*atp*6, *atp*8, *atp*9, *ymf*39）	+	+
NADH dehydrogenase：7（*nad*1-6, *nad*4L）	+	+
Ribosomal proteins：5（*rps*3, *rps*11, *rps*12, *rpl*16, *rpl*20）	+	+
Succinate dehydrogenase ：3（ *sdh*B, *sdh*C, *sdh*D）	+	+
Integral membrane：1（*sec*Y）	+	+

Total	24	24

Both heavy chain and light chain of mitochondrial genome of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* could encode protein genes, 13 genes were encoded by heavy chain and 11 genes （*cob*、*ymf*39、*cox*1、*cox*2、*cox*3、*rpl*16、*rps*3、*nad*4L、*rpl*20、*rps*12、*sec*Y）encoded by light chain. The total length of the protein coding sequences of *G. chouae* and *G. lemaneiformis* were 17,688 bp and 17,676 bp respectively, accounting for 68.5% and 68.3% of the whole mitochondrial genome. The gene arrangement was compact and no obvious gene overlap was found.

The length of most protein-coding genes (17) in mitochondrial genome of *Gracilaria chouae* and *Gracilariopsis lemaneiformi*s were identical. The seven genes differ in length: *atp*8、*nad*5、*nad*2、*sdh*C、*rpl*16、*rpl*20 and *sec*Y (Fig.2.4). Among them, *nad*2, *sdh*C and *sec*Y gene length had great variation.

The start codons of all protein-coding genes in mitochondrial genomes of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* were ATG, and the stop codons included TAA and TAG. The stop codon TAA was used most frequently in the mitochondrial genome of *Gracilaria chouae* and *Gracilariopsis lemaneiformis*. The start codon of all genes is ATG, and the stop codon includes TAA and TAG. The frequency of TAA as a stop codon in the mitochondrial genome of *Gracilaria chouae* was 87.5%, and that of *Gracilariopsis lemaneiformis* was 83.0%.

Only three genes of *atp*8, *rps*11 and *nad*4L in the mitochondrial genome of *G. chouae* used TAG as the stop codon. In the mitochondrial genome of *G. lemaneiformis*, four genes, *nad*4, *nad*6, *rpl*20 and *rps*12, use TAG as the stop codon. In the mitochondrial genome of algae, there are cases where the universal genetic codon is not used, In this study, the codon TGA in mitochondrial genomes of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* is not used as a stop codon, but is translated into tryptophan (Trp). All red algae mitochondrial genomes contain

different amounts of modified genetic code TGA, and it can be said that this modified genetic code appears most frequently in red algae.

Table 3.3 **Characteristics of the mitochondrial protein-coding genes of *G. chouae* and *G. lemaneiformis***

gene	length（bp）	codon			coding strand
		amino acid	initiation codon	termination codon	
*apt*6	762[a]/762[b]	253[a]/253[b]	ATG[a]/ATG[b]	TAA[a]/TAA[b]	H
*atp*8	408[a]/405[b]	135[a]/134[b]	ATG[a]/ATG[b]	TAG[a]/TAA[b]	H
*nad*5	1977[a]/1992[b]	658[a]/663[b]	ATG[a]/ATG[b]	TAA[a]/TAA[b]	H
*nad*4	1476[a]/1476[b]	491[a]/491[b]	ATG[a]/ATG[b]	TAA[a]/TAG[b]	H
*sdh*D	240[a]/240[b]	79[a]/79[b]	ATG[a]/ATG[b]	TAA[a]/TAA[b]	H
*nad*2	1497[a]/1476[b]	498[a]/491[b]	ATG[a]/ATG[b]	TAA[a]/TAA[b]	H
*nad*1	981[a]/981[b]	326[a]/326[b]	ATG[a]/ATG[b]	TAA[a]/TAA[b]/	H
*nad*3	366[a]/366[b]	121[a]/121[b]	ATG[a]/ATG[b]	TAA[a]/TAA[b]	H
*rps*11	360[a]/360[b]	118[a]/119[b]	ATG[a]/ATG[b]	TAG[a]/TAA[b]	H
*atp*9	231[a]/231[b]	76[a]/76[b]	ATG[a]/ATG[b]	TAA[a]/TAA[b]	H
*sdh*C	375[a]/372[b]	125[a]/123[b]	ATG[a]/ATG[b]	TAA[a]/TAA[b]	H
*sdh*B	753[a]/753[b]	250[a]/250[b]	ATG[a]/ATG[b]	TAA[a]/TAA[b]	H
*nad*6	609[a]/609[b]	202[a]/202[b]	ATG[a]/ATG[b]	TAA[a]/TAG[b]	H
cob	1143[a]/1143[b]	380[a]/380[b]	ATG[a]/ATG[b]	TAA[a]/TAA[b]	L
*ymf*39	543[a]/543[b]	180[a]/180[b]	ATG[a]/ATG[b]	TAA[a]/TAA[b]	L
*cox*3	819[a]/819[b]	272[a]/272[b]	ATG[a]/ATG[b]	TAA[a]/TAA[b]	L

*cox*2	792[a]/792[b]	263[a]/263[b]	ATG[a]/ATG[b]	TAA[a]/TAA[b]	L
*cox*1	1596[a]/1596[b]	531[a]/531[b]	ATG[a]/ATG[b]	TAA[a]/TAA[b]	L
*rpl*16	414[a]/411[b]	137[a]/136[b]	ATG[a]/ATG[b]	TAA[a]/TAA[b]	L
*rps*3	696 [a]/696[b]	231[a]/231[b]	ATG[a]/ATG[b]	TAA[a]/TAA[b]	L
*nad*4L	306 [a]/306[b]	101[a]/101[b]	ATG[a]/ATG[b]	TAG[a]/TAA[b]	L
*rpl*20	234[a]/243[b]	77 [a]/80[b]	ATG[a]/ATG[b]	TAA[a]/TAG[b]	L
*rps*12	369 [a]/369[b]	122[a]/122[b]	ATG[a]/ATG[b]	TAA[a]/TAG[b]	L
*sec*Y	741[a]/735[b]	244[a]/244[b]	ATG[a]/ATG[b]	TAA[a]/TAA[b]	L

Note: a: *G. chouae*; b: *G. lemaneiformis*

In the mitochondrial genome of *Gracilariopsis lemaneiformis*, the *atp*6 and *sec*Y gene spacing region has a 126-bp complete reverse repeat sequence containing poly A and T, which can form a stable stem ring structure and is located at the boundary point of the two transcription units. This reverse repeat sequence is similar to the D-ring structure of animal mitochondrial genome and can lead to transcription termination, so it is speculated that it may be the transcription-related replication initiation site of mitochondrial genome [18]. This structure is not found in the mitochondrial genomes of other species of Gracilariaceae.

3.1.4 tRNA genes

To use online software tRNAscan - SE 1.21 [23] (http://lowelab.ucsc.edu/tRNAscan-SE) for brittle gracilaria and asparagus tRNA gene prediction. Results 23 tRNAs genes were located in the mitochondrial genome of *G. chouae* and 20 tRNAs genes were located in *G. lemaneiformis* (Table 2.5). Compared with *G. Chouae* , there were less *trn*Y、 *trn*R2、*trn*S2 genes.

The 23 tRNA genes in the mitochondrial genome of *G. chouae* can encode 18 types of tRNA molecules. They are tRNA- Met、tRNA- Cys、tRNA-Pro、

tRNA-Ser、tRNA- Phe、tRNA- Gly、tRNA- Leu、tRNA- Tyr、tRNA- Arg、tRNA- Ala、tRNA- Trp、tRNA- Asn、tRNA- Val、tRNA-Lys、tRNA-Glu、tRNA-Asp、tRNA- Gln、tRNA- His. The mitochondrial genome of *Gracilariopsis lemaneiformis* contains 20 tRNA genes that encode 17 types of tRNA molecules, They are tRNA- Ala 、tRNA- Trp 、tRNA- Met 、tRNA- Cys 、tRNA-Pro 、tRNA-Ser、tRNA- Phe、tRNA- Gly、tRNA- Leu、tRNA- Asn、tRNA- Val、tRNA- Arg、tRNA-Lys、tRNA-Glu、tRNA-Asp、tRNA- Gln、tRNA- His. Thus, like other reported mitochondrial genomes of algae, the mitochondrial genomes of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* are unable to synthesize all the trnas required for their function alone.

Both chains of the mitochondrial genomes of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* can encode tRNA genes, among which 9 are encoded in the heavy chain and the rest in the light chain (Figure 2-2). The 23 tRNA genes contained in the mitochondrial genome of *Gracilaria chouae* range in length from 72 to 93 bp. The mitochondrial genome of *Gracilariopsis lemaneiformis* contains 20 tRNA genes, ranging in length from 70 to 88 bp. In this experiment, the length of the same tRNA gene was the same between *Gracilaria chouae* and *Gracilariopsis lemaneiformis.* Only four tRNA genes (*trn*C, *trn*L2, *trn*R1, *trn*V) were slightly different.

This study also predicted and analyzed the secondary structure of tRNA genes. All tRNA genes can form a typical shamrock-shaped secondary structure, which is relatively conserved, and each secondary structure is composed of 4-5 stem loops, namely, amino acid acceptance arm (binding amino acid), anti-codon ring, TΨC ring (thymidine pseudoguanosine cytidine ring, thymidine pseudoguanosine cytidine ring, etc.). Binding ribosome), DHU ring (dihydrouracil ring, binding aminoacyl tRNA synthetase) and extra ring each, some tRNA gene secondary structure does not have extra ring. The number of extra ring bases varies greatly. Different trnas have different size of extra ring, which can be composed of 3-18 nucleotides. The different length of different tRNA genes is mainly caused by the different number of extra ring

bases. In this study, most of the tRNA gene secondary structures predicted from the genomes of Gracilaria gracilaria and Gracilaria gracilaria have no extra loops, and only a few tRNA genes have extra loops, such as *trnL2*, *trnY*, *trnS1*, *trnS2*, and *trnL1* genes in *G. chauae* and the *trn*S1, *trn*L2, and *trn*L1 genes of G. lemaneiformis. Thus, tRNAs with large gene length usually have extra loops in their secondary structures.

Table 3.4 **tRNA genes of mitochondrial genomes from *G. chouae and G. lemaneiformis***

tRNA type	amino acid	length	Anticodon	Encoding chain
*trn*M1	Met	74[a]/74[b]	cat [a]/cat[b]	H
*trn*C	Cys	73[a]/72[b]	gca[a]/gca[b]	H
*trn*P	Pro	74[a]/74[b]	tgg [a]/tgg[b]	H
*trn*S1	Ser	88[a]/88[b]	tga[a]/tga[b]/	H
*trn*F	Phe	73[a]/73[b]	gaa[a]/gaa[b]	H
*trn*H	His	74[a]/74[b]	gtg[a]/gtg[b]	L
*trn*G1	Gly	73[a]/73[b]	gcc[a]/gcc[b]	H
*trn*L2	Leu	85[a]/82[b]	tag [a]/tag[b]	H
*trn*L1	Leu	86[a]/86[b]	taa[a]/taa[b]	L
*trn*Q	Gln	73[a]/73[b]	ttg[a]/ttg[b]	L
*trn*G2	Gly	74[a]/74[b]	tcc[a]/tcc[b]	L
*trn*D	Asp	72[a]/72[b]	gtc[a]/gtc[b]	L
*trn*M2	Met	74[a]/74[b]	cat[a]/cat[b]	L

*trn*E	Glu	73[a]/73[b]	ttc[a]/ttc[b]	L
*trn*K	Lys	74[a]/74[b]	ttt[a]/ttt[b]	L
*trn*R1	Arg	75[a]/76[b]	acg[a]/acg[b]	L
*trn*V	Val	73[a]/74[b]	tac[a]/tac[b]	L
*trn*N	Asn	72[a]/72[b]	gtt[a]/gtt[b]	L
*trn*Y	Tyr	85[a]/N[b]	gta[a]/N[b]	L
*trn*R2	Arg	73[a]/ N[b]	tct[a]/ N[b]	L
*trn*S2	Ser	93[a]/ N[b]	gct[a]/ N[b]	L
*trn*A	Ala	75[a]/75[b]	tgc[a]/tgc[b]	H
*trn*W	Trp	74[a]/74[b]	tca[a]/tca[b]	H

Note: a: G. chouae; b: G. lemaneiformis

3.1.5 rRNA genes

The mitochondrial genome of *Gracilaria chouae* encodes only two rRNA genes: 23S rRNA gene (LSU) and 16S rRNA gene (SSU). The mitochondrial genome of *Gracilariopsis lemaneiformis* also encodes two rRNA genes, namely 26SrRNA gene (LSU) and 16SrRNA gene (SSU). The two rRNA genes of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* were close to each other and fixed in position, with an interspaced *nad*4L gene in the middle. The LSU rRNA gene length of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* was 2,626 bp and 2606bp respectively. The SSUrRNA gene length of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* was 1,399 bp and 1390bp respectively.

3.1.6 Mitochondrial genome collinearity analysis of *Gracilaria chouae* and *Gracilariopsis lemaneiformis*

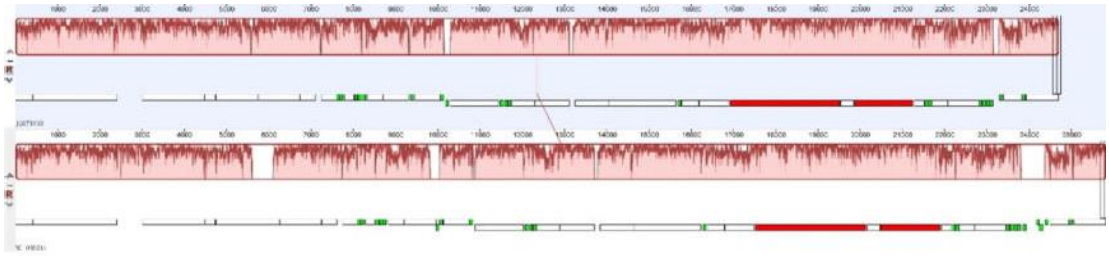


Figure 3.3 – Collinearity analysis of mitochondrial genome of *Gracilaria chouae* and *Gracilariopsis lemaneiformis*

The mitochondrial genomes of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* were analyzed collinearly by Mauve [21]. The results showed that the mitochondrial genomes of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* had high conserved sequence consistency, and the gene contents of mitochondrial gene groups were similar, but there were some differences. Their mitochondrial genomes are closed double-linked ring molecules, which are less evolved and highly conserved. The protein coding genes in mitochondrial genomes of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* were also very abundant, but the types and numbers were basically the same, with 24.

The *trn*H gene is rearranged in the mitochondrial genome of *Gracilaria* and *Gracilariopsis*. *trn*H gene is a special gene in Gracilaria. The location of the gene in *Gracilaria* and *Gracilariopsis* differs by genus, and the polarity of the cluster in which the gene is located is different in different species, which may be the splicing of the genome caused by trnH gene. However, the location of tRNA genes is generally not a basis for evolution, and the difference in this location can be used as a molecular marker between the two "genera" [19].

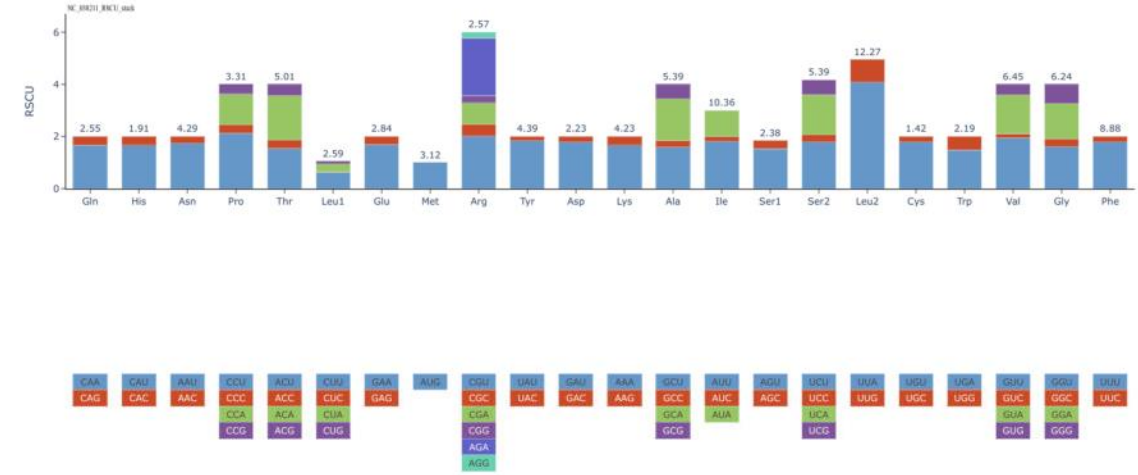
3.2 Codon analysis

Due to the degeneracy of codons, one amino acid is usually compiled by multiple codons. By analyzing its codon usage, we can study the selection of gene expression level, mRNA stability and co-translated protein folding between the two species, and further compare the differences between the two species.

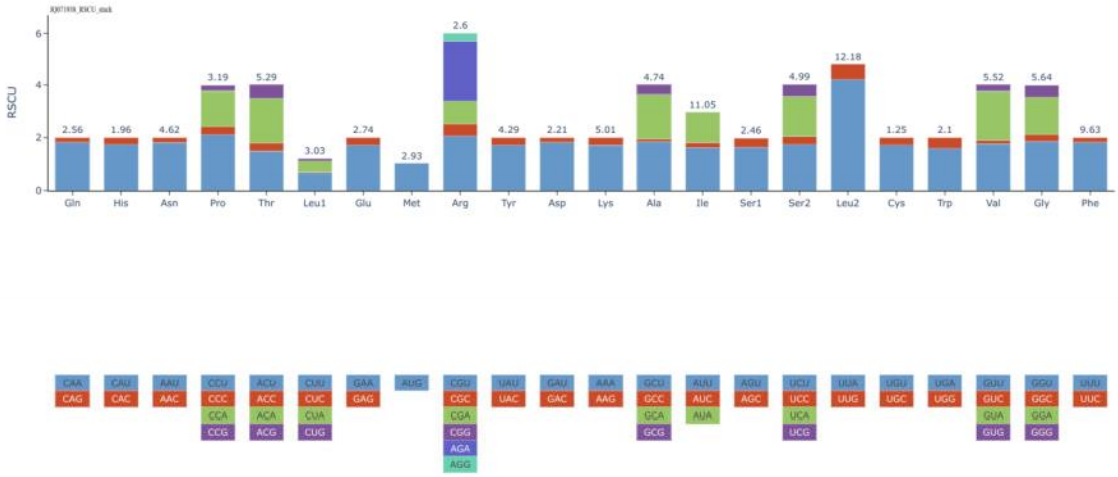
3.2.1 result and analysis

In this study, the use of synonym codons in gene expression of Gracilaria gracilaria and asparagus was studied. Synonymous codons refer to codon sequences that express the same amino acid in the genetic code. By analyzing these synonymous codons, we can understand the relationship between protein and gene expression, and how synonymous codons maintain the functional integrity and stability of the genetic code. Among many codons, if the frequency of use of a synonymous codon is significantly higher than that of other synonymous codons, it indicates that the codon has an obvious use preference in a specific individual, which is called codon preference. Due to differences in codon use preference among different species, the study of similar codon use preference in the same species or closely related species can reveal how these species strike a balance between mutation and natural selection in the process of gene expression level, mRNA stability and co-translated protein folding [29].

In this study, the CBI values of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* were 0.149 and -0.150, and ENC values were 50.07 and 48.85, respectively, using the software CodonW. Both values were greater than 35, indicating that the codon usage of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* were weak. RSCU represents the relative synonymous codon usage. When RSCU is greater than 1, it indicates that the synonymous codon has a use preference and is a high frequency codon. When RSCU is equal to 1, the codon has no use preference. When RSCU is less than 1, it indicates that the codon is a low-frequency codon and has no use preference. The results of RSCU experiment of Gracilaria gracilaria and asparagus are shown in the figure2.4.



Gracilaria chouae



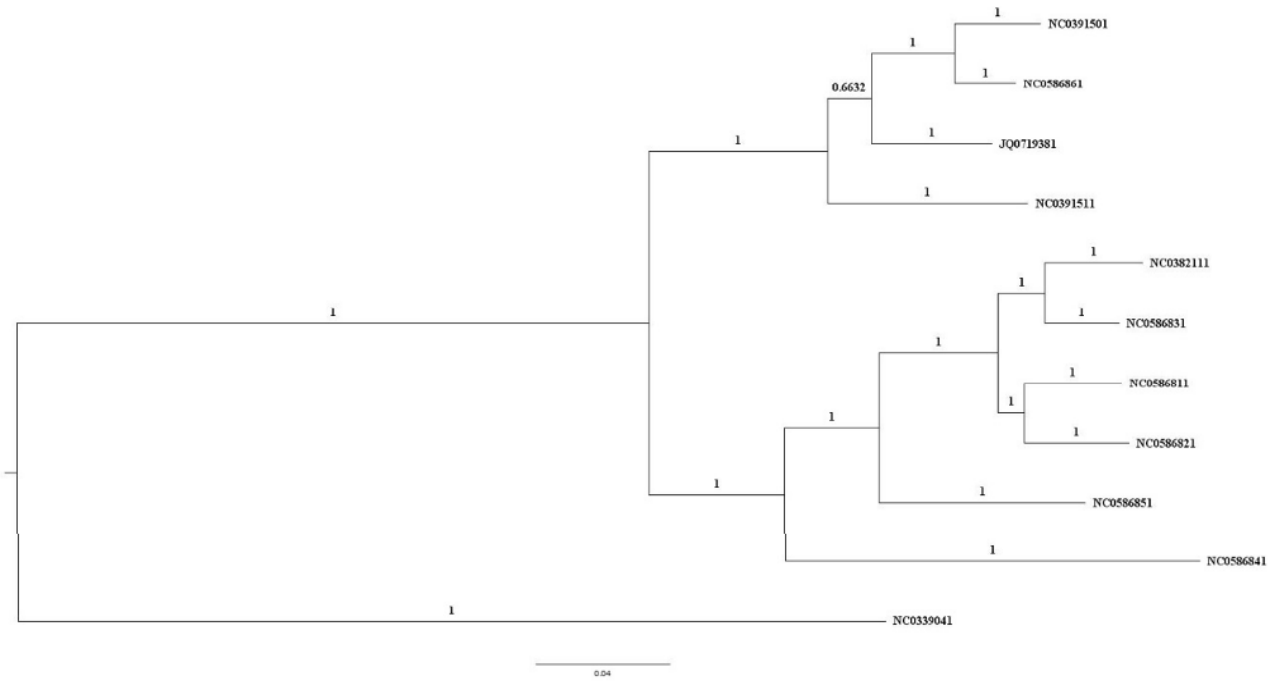
Gracilariopsis lemaneiformis

Figure 3.4 – RSCU of *Gracilaria chouae* and *Gracilariopsis lemaneiformis*

3.3 Phylogenetic Trees Results Analysis

Systematic evolutionary analysis was conducted on the 20 shared protein-coding genes of eleven species in the order Gigartinales using both ML and BI methods, as shown in Figure 3-1. (Fig.3-1)

From the figure, it can be observed that the outgroup species, *Corallina officinalis,* is distinctly separated from the other species, indicating fundamentally different evolutionary paths. The six species of the genus Gracilaria and the four species of the genus Gracilariopsis are clearly divided into two branches. Apart from the branch of *G. lemaneiformis*, *G. longissima*, and *G. tenuifrons*, which have a Bayesian support of 0.6632, the support for other branches within the genera *Gracilaria* and *Gracilariopsis* is 1, indicating close phylogenetic relationships within these genera.



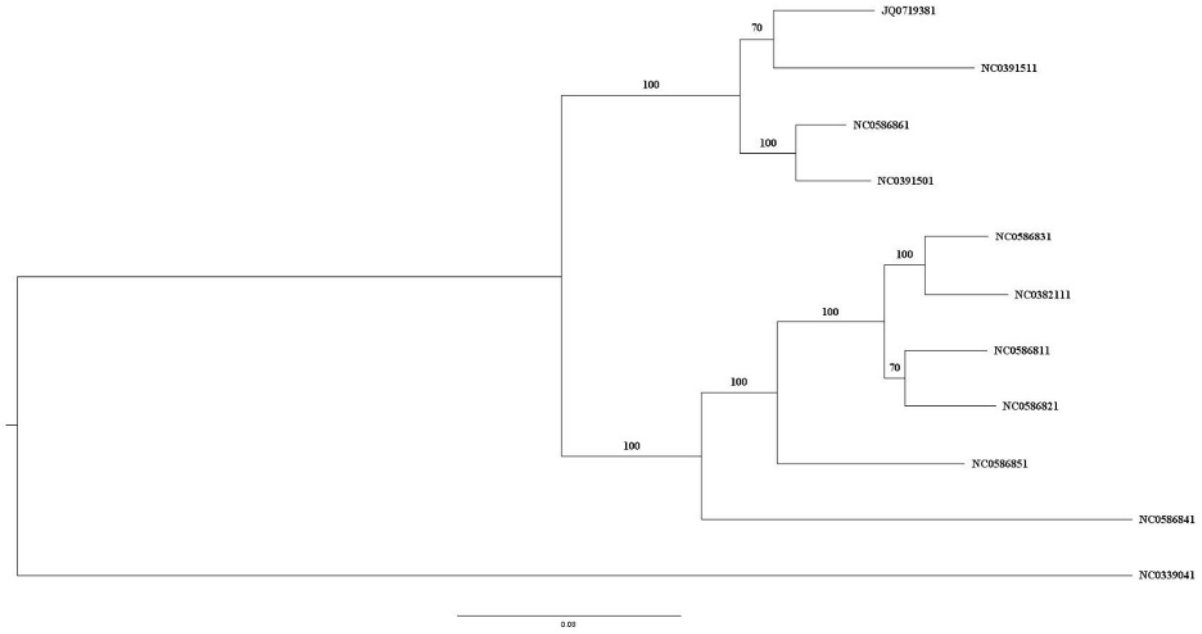


Figure 3.5 – BI&ML phylogenetic tree built using common protein-coding genes from eleven species in this work

To compare *Gracilaria* and *Gracilariopsis* from a phylogenetic perspective, this chapter utilized 20 shared protein-coding genes of eleven species including *Gracilaria, Gracilriopsis,* and the outgroup *Corallina officinalis*. After aligning and removing non-conserved sequences, systematic phylogenetic trees were constructed using both BI and ML models. The highly consistent results from both methods suggest the reliability of the evolutionary trees in this study. The results indicate a clear separation between the genus *Gracilaria* and the genus *Gracilariopsis*, demonstrating independent evolutionary paths for species within each genus.

CONCLUSIONS

1. The total length of mitochondrial genome of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* was 25,829bp and 25,883bp, respectively. *Gracilaria chouae* encoded 50 genes, including 24 protein coding genes, 2 rRNAs, 23 tRNAs and orf148. *Gracilariopsis lemaneiformis* encoded 48 genes. It included 24 protein-coding genes, 2 rRNAs, 20 tRNAs, and 2 ORFs (orf60, orf143).

2. The proportions of A, T, C and G were similar, and the content of G+C was respectively 28.8% and 27.5%. The types of tRNA in mitochondrial genomes of the two are not exactly the same, and the secondary structure of the two rRNA genes is shamrock-shaped, and most of them do not have extra rings. The lengths of the two RRNA genes are different, the types and numbers of the 24 protein-coding genes are the same, and the lengths of most of the protein-coding genes (17) are completely the same. *atp8, nad5, nad2, sdhC, rpl16, rpl20* and *secY* are different in length.

3. *trnH* gene is rearranged in mitochondrial genomes of *Gracilaria chouae* and *Gracilariopsis lemaneiformis*. It can be used as a molecular marker between the two genera. The results of collinearity analysis showed that the gene arrangement of *Gracilaria* was generally conservative.

4. The ENC values of mitochondrial genome of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* were 50.07 and 48.85, respectively, both of which were greater than 35, indicating that the codon use of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* were weak. The mean values of CBI, Fop and RSCU of *Gracilaria chouae* and *Gracilariopsis lemaneiformis* were different.

5. From a phylogenetic perspective, we compared *Gracilaria chouae* and *Gracilariopsis lemaneiformis* with 20 protein-coding genes shared by 11 species of *Gracilaria*, *Gracilriopsis* and *Corallina officinalis*. The phylogenetic tree was constructed by BI and ML models. This suggests that *Gracilaria* and *Gracilriopsis* species evolved independently in the course of evolution, and *Gracilaria* and *Gracilriopsis* are two separate genera.

LIST OF REFERENCES

1. Wang, T., et al. (2012). Large-scale seaweed in Dongtou, Zhejiang [M]. Ocean Press.
2. Chinese Seaweed Monograph. (2012). Science Press, 2(5), 41-42.
3. Jin, Y., Wu, W., & Chen, W. (2012). Influence of different temperature and salinity culture conditions on the growth and biochemical components of Gracilariopsis lemaneiformis [J]. Southern Fisheries Science, 8(02), 51-57.
4. Liu, S. J., Jie, Z. Y., & Zeng, S. F. (1997). The commercial cultivation of Gracilaria and its polyculture with prawn in China. Journal of Zhanjiang Ocean University, 17(2), 27-30.
5. Zhou, Y. (2004). Plant Biology [M]. Higher Education Press, 322.
6. Zhou, Z., He, Q., & Ma, X. (2010). Nutritional composition analysis and evaluation of Gracilaria chouae in the East China Sea. Food Science, 31(9), 284-287.
7. Zhu, L. (2015). Response characteristics and mechanism of different developmental stages of Gracilariopsis lemaneiformis to UV-B radiation [D]. Ocean University of China.
8. Fang, T. (2023). Enzymatic degradation of Gracilaria polysaccharides and study on the antioxidant activity of its enzymatic hydrolysates [D]. Ningbo University. DOI:10.27256/d.cnki.gnbou.2021.001373.
9. Liu, Y. (2017). Fermentation of Gracilaria and the effects of fermented Gracilaria on the growth, immunity, and antioxidant capacity of black porgy [D]. South China Agricultural University.
10. Li, T. (2015). Study on the complete mitochondrial genome of Undaria pinnatifida and comparative analysis of algal mitochondrial genomes [D]. Ocean University of China.

11. Zhai, Z., Wang, X., Ding, M., et al. (2014). Amplification and phylogenetic analysis of the complete mitochondrial genome of Gracilaria vermiculophylla [D]. Ocean University of China.
12. Qian, H. (2014). Amplification and systematic evolutionary analysis of the mitochondrial genome of Gracilaria vermiculophylla [Doctoral dissertation, Ocean University of China].
13. Li, T., Chen, B., Chen, S., et al. (2012). Molecular systematic analysis of 5 species of algae in the genera Gracilaria and Gracilariopsis based on ITS sequences. Journal of Sun Yat-sen University (Natural Science Edition), 51(04), 97-105.
14. Sha, M., Lin, L., Li, X., et al. (2013). Sequencing strategy and method for mitochondrial genomes. Journal of Applied Entomology, 50(01), 293-297.
15. Zhang, T. (2012). Sequencing, assembly, and comparative genomics of plant organelle genomes [D]. Zhejiang University.
16. Yu, X., Liu, P., Liu, X., et al. (2022). Complete organellar genomes and molecular phylogeny of Hypnea cervicornis (Gigartinales, Florideophyceae) from China. Journal of Applied Phycology, 34, 2705-2717.
17. Tang, X., Jia, X., Zhang, J., & Liu, T. (2019). The complete mitogenome of Gracilaria chouae and its phylogenetic analysis. Mitochondrial DNA B Resources, 4(2), 2786-2787.
18. Zhang, L., Wang, X., Qian, H., Chi, S., & Liu, T. (2012). Complete sequences of the mitochondrial DNA of the wild Gracilariopsis lemaneiformis and two mutagenic cultivated breeds (Gracilariaceae, Rhodophyta). PLoS One, 7(6), e40241.
19. Lyra, G. M., Iha, C., Grassa, C. J., et al. (2021). Phylogenomics, divergence time estimation and trait evolution provide a new look into the Gracilariales (Rhodophyta). Molecular Phylogenetics and Evolution, 165, 107294.
20. Liu, T. (2017). Common large seaweed atlas of the South China Sea [M]. Ocean Press.

21. Darlin, A. C., Mau, B., Blatter, F. R., & Perna, N. T. (2004). Mauve: Multiple alignment of conserved genomic sequence with rearrangements. Genome Research, 14(7), 1394-1403.
22. Zhao, Y., Hou, Y., Ren, J., et al. (2023). Phenotypic characteristics of taurodontism and a novel WNT10A variant in non-syndromic oligodontia family. Archives of Oral Biology, 154, 105759.
23. Lowe, T. M., & Chan, P. P. (2016). tRNAscan-SE On-line: Search and Contextual Analysis of Transfer RNA Genes. Nucleic Acids Research, 44, W54-W57.
24. Kück, P., & Meusemann, K. (2010). FASconCAT: Convenient handling of data matrices. Molecular Phylogenetics & Evolution, 56(3), 1115-1118.
25. Li, L., Li, L., Diao, S., et al. (2000). Study on the processing technology of Gracilaria flavor food. China Fisheries, (10), 50-51.
26. Yang, X., Xia, G., Qi, B., et al. (2013). Process research on Gracilaria flavor seaweed sauce: Proceedings of the Tropical Ocean Science Academic Symposium and the Eighth Guangdong Ocean and Lake Society, China, Zhanjiang, 2013 [C]
27. Chen, X. (2012). Preparation of dietary fiber from Gracilaria and its application in yogurt [D]. Ocean University of China.
28. Lu, X. (2015). Study on physiological and biochemical characteristics and tissue culture germplasm preservation of Gracilariopsis [D]. Ocean University of China.
29. Ye, Q., Song, Y., Li, M., et al. (2022). Characteristics of chloroplast genome and analysis of codon usage preference of Prunus yedoensis [J/OL]. Molecular Plant Breeding, 1-14.
30. Zhang, D., Gao, F., Jakovlić, I., et al. (2020). PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Molecular Ecology Resources, 20(1), 348–355.

31. Tamura, K., Dudley, J., Nei, M., et al. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. Molecular Biology and Evolution, 24(8), 1596.
32. Rambaut, A. (2009). FigTree v1.3.1. Retrieved from http://tree.bio.ed.ac
33. Greiner, S., Lehwark, P., & Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Research, 47, W59-W64.
34. Kurland, C. G., & Andersson, S. G. E. (2000). Origin and evolution of the mitochondrial proteome. Microbiology and Molecular Biology Reviews, 64(4), 786-820.
35. Stamatakis, A. (2006). RAxML-VI-HPC: Maximum Likelihood-Based Phylogenetic Analyses with Thousands of Taxa and Mixed Models. Bioinformatics, 22, 2688-2690.doi:10.1093/bioinformatics/btl446.