Cíntia Iha

Estudo filogenômico e caracterização genômica organelar de algas gracilarioides (Gracilariaceae, Rhodophyta)

Phylogenomic study and organellar genomic characterization of gracilarioids seaweeds (Gracilariaceae, Rhodophyta)

São Paulo
2018

Cíntia Iha

Estudo filogenômico e caracterização genômica organelar de algas gracilarioides (Gracilariaceae, Rhodophyta)

Phylogenomic study and organellar genomic characterization of gracilarioids seaweeds (Gracilariaceae, Rhodophyta)

São Paulo
2018

Cíntia Iha

Estudo filogenômico e caracterização genômica organelar de algas gracilarioides (Gracilariaceae, Rhodophyta)

Phylogenomic study and organellar genomic characterization of gracilarioids seaweeds (Gracilariaceae, Rhodophyta)

Tese apresentada ao Instituto de Biociências da Universidade de São Paulo, para a obtenção do Título de Doutor em Ciências, na Área de Botânica.

Orientadora: Profª Drª. Mariana Cabral de Oliveira

Colaborador: Dr. Heroen Verbruggen
(The University of Melbourne, Australia)

São Paulo
2018

Iha, Cíntia

Estudo filogenômico e caracterização genômica organelar de algas gracilarioides (Gracilariaceae, Rhodophyta).
118 p.

Tese (Doutorado) – Instituto de Biociências da Universidade de São Paulo. Departamento de Botânica, 2018.

1. Gracilariaceae, 2. Genoma organelar, 3. Filogenômica. I. Universidade de São Paulo. Instituto de Biociências. Departamento de Botânica

Comissão Julgadora

_____ _____

_____ _____

_____

Profª. Drª. Mariana Cabral de Oliveira
Orientadora

Dedico esta tese a todas as pessoas que direta ou indiretamente me ajudaram, influenciaram ou permitiram que este trabalho fosse realizado.

Two roads diverged in a yellow wood,
And sorry I could not travel both
And be one traveler, long I stood
And looked down one as far as I could
To where it bent in the undergrowth;

Then took the other, as just as fair,
And having perhaps the better claim,
Because it was grassy and wanted wear;
Though as for that the passing there
Had worn them really about the same,

And both that morning equally lay
In leaves no step had trodden black.
Oh, I kept the first for another day!
Yet knowing how way leads on to way,
I doubted if I should ever come back.

I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I—
I took the one less traveled by,
And that has made all the difference.

The Road Not Taken – Robert Frost

Agradecimentos

Em primeiro lugar, quero agradecer ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) pela bolsa regular ao longo dos quatro anos do curso de doutorado (N° 152939/2014-8), à CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pela bolsa concedida para realizar o estágio-sanduiche na University of Melbourne, Austrália (N° 88881.134422/2016-01) e por outros auxílios vindos do Programa PROEX e, por fim, a FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo), que também auxiliou neste projeto.

Quero dar um agradecimento especial à minha orientadora Profª. Drª. Mariana C. de Oliveira por me auxiliar por todos esses anos (que também contam o tempo de iniciação científica e mestrado) por todo ensinamento, orientação, apoio, muita paciência e por ter acreditado em mim desde o começo. Me sinto muito sortuda por ter sido sua aluna e levarei comigo não apenas os ensinamentos acadêmicos e científicos, mas o também exemplo de uma ótima professora, orientadora, cientista, administradora e, mais do que tudo, uma ótima pessoa.

I would like to thank Dr. Heroen Verbruggen for my internship in his lab in the University of Melbourne, Australia. Thank you for the hospitality, teachings and all contribution to this work. They were important and have made all the difference. Thank you also to the people in Melbourne: Vanessa, Chris Jackson, Joana, Chiela, Emilia, Sonja, Pilar, Marisa, Lupita, Francesco, and Sam.

Agradeço à Drª. Goia Lyra pela parceria e colaboração que permitou aumentar bastante a quantidade de dados do projeto e resultou em um trabalho melhor e mais completo. Além disso, as trocas de informações e conversas foram essenciais para o desenvolvimento de ideias e resolução de problemas. I would like to thank Professor Dr. Charles Davis and Chris Grassa from Harvard University Herbaria for the collaboration with my project, which resulted in a good paper.

Agradeço aos professores do Laboratório de Algas Marinhas (LAM) Flávio, Estela, Valéria e Fungyi. Agradeço aos técnicos: Willian, pela amizade e abraços gostosos de todos os dias que eu aparecia; Rosário, por ter me ajudado muito no cultivo das algas; e André, por toda ajuda para fazer os reagentes dos experimentos. Agradeço também aos colegas que convivi nesses quatro anos: Allyson, Ale, Luzka, Sabrina, Mari Melo, Patrícia, Talissa, Karol, Talita, Carol Azevedo, Vanessa, Nuno, Fabi, André, Bruno, Ana e todos que passaram por lá…

Agradeço às pessoas que apareceram durante esses quatro anos e foram essenciais em alguma etapa desse projeto: André Martins, pela amizade e por sempre me ajudar na bioinformática. Ao pessoal do CEFAP, em especial Susan e Tiago, por me ajudarem a montar o plano do projeto e também no auxílio do preparo das bibliotecas de DNA. Ao Jonas por me auxiliar no servidor SEAL. Pessoal da PensaBio, em especial Danusa e Kadu que me auxiliaram no preparo das bibliotecas de DNA. Prof° Daniel Lahr e Anush por terem emprestado o kit para RT-PCR. Anush, Danusa e Felipe Braga da Roche, por terem me auxiliado muito para entender a RT-PCR.

À Universidade de São Paulo, ao Instituto de Biociências e ao Departamento de Botânica, pelo programa de Pós-Graduação. Aos funcionários do IB, em especial ao Norberto, por estar sempre

disposto a tirar dúvidas e ajudar. E ao pessoal da limpeza que dá duro pra manter o ambiente limpo. Agradeço à Erika da secretaria da pós-graduação que sempre foi muito solícita e levava docinhos nas reuniões da CPG.

Por fim, mas mais importante, agradeço aos meus amigos, em especial, ao Allyson pelo carinho e por me dar um suporte nesse final de doutorado que é sempre tenso. E agradeço principalmente à minha família, minha mãe Sueli, meu pai Julio, meu irmão Renan e minha vó Iracy, que sempre me apoiam nessa decisão estranha de seguir a carreira acadêmica. Muito obrigada com todo amor!

Summary

List of tables 3
List of figures 5
Abbreviations 7
Resumo 9
Abstract 11
Introduction 13
The Gracilariaceae family 13
Economic importance 16
High-throughput sequencing 17
Genomic studies in red algae 18
Objectives 23
General 23
Specific 23
Material and Methods 25
Choice of HTS machine 25
Taxon sampling and culturing 26
Molecular techniques and high-throughput sequencing 28
Samples added to the project 31
Organellar genome assembly and annotation 33
Plasmid genomes assembly and annotation 33
Phylogenetic analyses 34
Organellar genomic comparison 35
*leu*C and *leu*D genes 35
Results and Discussion 37
DNA library 37
High-throughput Sequencing output and genome assemblies 41
Quantity and quality of samples sequenced in HiSeq2500 41
Quantity and quality of samples from Dr. G. Lyra database 42
Genome assemblies 43
Organellar genome structure 45
Extrachromosomal plasmid from Gracilariaceae 46
Genomic comparison and plasmid-derived sequences of mitochondria 48
Genomic comparison and plasmid-derived sequences of chloroplast 52
*leu*C/*leu*D operon 55

Organellar phylogenomics and systematics of Gracilariaceae 58
Conclusion 65
Bibliography 67
Supplementary material i
Supplementary tables ii
Genome maps xii

List of tables

Table 1 - Genera of Gracilariaceae, number of currently recognized species and notes about currently taxonomical situation. Data were obtained from Guiry and Guiry (2018). 15

Table 2 – Mitochondrial and chloroplast genomes from Gracilariaceae published in GenBank. 20

Table 3 - Information of red algae plasmids available in GenBank. 21

Table 4 - Species cultured and sequenced in this work. 27

Table 5 - Parameters used in Covaris S2 DNA fragmentation tests using only *Gracilaria tenuistipitata*. *Standard parameters indicated by the Covaris S2 manual to obtain 500 bp fragments. The parameters selected for DNA library preparation are in bold. 31

Table 6 – Gracilariaceae species sequenced by Dr. G. Lyra and assembled in this work. 32

Table 7 - List of Gracilariaceae protein-coding genes used in the ML tree search. 35

Table 8 - Quantity of total fresh mass used for DNA extraction. Concentration of DNA extracted, eluted in MilliQ water, in both NanoDrop™ (total nucleic acid) and Qubit® (double-stranded DNA). NanoDrop™ 260/280 and 260/230 ratios of absorbance values. Concentration of DNA libraries measured in Qubit® and RT-PCR. 37

Table 9 - High-throughput sequencing output information. Expected coverage calculated using all read output. Number of reads normalized to 100x coverage using only reads sequenced in lane 1. Percentage of normalized reads. Reads dropped in Trimmomatic v0.036 using normalized reads. 42

Table 10 – HTS output information of samples from Dr. G. Lyra database. Part of the sequenced samples was not normalized and the trimming and cleaning processes were done with total sequencing output. 43

Table 11 - General information of the assembly and structure of mitochondrial and chloroplast genomes of samples sequenced in HiSeq 2500, *Gracilaria vermiculophylla* and *Melanthalia intermedia* from G. Lyra database. Depth of coverage was done using normalized reads. Std – standard deviation. 44

Table 12 - General information of the assembly of mitochondrial and chloroplast genomes of samples from Dr. G. Lyra database. 45

Table 13 - Information of the plasmid genomes found in *Gracilaria ferox* and *Gracilaria vermiculophylla.* 46

Table 14 – Homologous ORFs found in plasmids of Gracilariaceae species. 47

List of figures

Figure 1 – Molecular workflow to prepare the DNA libraries. 29

Figure 2 – Bioanalyzer result of first test of DNA fragmentation. Y axis, fluorescence (FU) and X axis, length of the fragments (bp). 38

Figure 3 – The best Bioanalyzer results in fragmentation tests. 40

Figure 4 – Linear schematic representation of circular Gracilariaceae plasmids. Plasmids with bold names were sequenced in this study, others were obtained from GenBank. Light grey arrows are ORFs. Black arrows are repeated regions. Regions identified as 'Repeat 1, 2 and 3' are entirely repeated sequences in different positions of the genome. Regions identified as 'Repeat region' are repetitions in tandem. Scale line in base pairs.... 48

Figure 5 – Mitochondrial genome phylogeny and genomic comparison of Gracilariaceae. A. Left, maximum likelihood phylogenetic tree based on an amino acid matrix with 23 coding genes sequences. Right, mitochondrial progressive MAUVE alignment. Each linear genome corresponds to the name in the phylogenetic tree. Horizontal axis above refers to genome length in base pairs (bp). Synteny between genomes is represented by the colour Locally Collinear Blocks (LCB). LCB above the line is the forward strand; below the line is the reverse strand. The same color line of LCB along the alignment indicate similarity region among the genomes. Gene maps are shown below the LCBs, protein coding genes in white, rRNA genes in red, tRNA genes in dark green and intron in light green. Position of boxes above (forward) and below (reverse) line also refers to gene orientation. B and C. Loci marked in black line in A in detail. Protein coding genes are represented in arrow box in yellow, plasmid-derived sequence in grey, and each *trn* is marked in different colors. Arrow boxes do not represent gene length and the scheme represents only the synteny and not alignment. 51

Figure 6 – Chloroplast genome phylogeny and genomic comparison of Gracilariaceae. A. Left, maximum likelihood phylogenetic tree based on an amino acid matrix with 187 coding genes sequences. Right, mitochondrial progressive MAUVE alignment. Each linear genome corresponds to the name in the phylogenetic tree. Horizontal axis above refers to genome length in base pair (bp). Synteny between genomes is represented by Locally Collinear Blocks (LCB). LCB above the line is the forward strand, below the line is the reverse strand. The same color line of LCB along the alignment indicate similarity region among the genomes. Gene maps are shown below the LCBs, protein coding genes in white, rRNA genes in red, tRNA genes in black, intron in light green and *leu*C/*leu*D operon in yellow (showed in *Melanthalia intermedia*., *Gracilaria chilensis*, and *G. tenuistipitata*). Position of boxes above (forward) and below (reverse) line also refers to gene orientation. Continuous orange line boxes represent plasmid-derived sequences. Red arrows indicate plasmid-insertion similar to the Gle4293 plasmid. Green arrows indicate plasmid-insertion similar to the Gro4970 plasmid. B and C. Loci with plasmid-insertions marked in A in detail. Protein coding genes are represented in arrow box in yellow, including plasmids derived ORFs (FSO)..... 53

Figure 7 – Gene synteny of the plasmid-derived sequence regions found in chloroplast genomes. These regions are identified in orange line box in Figure 6..... 54

Figure 8 – Maximum likelihood phylogeny of *leu*C. Numbers at nodes indicate bootstrap replicates. Bootstrap support values are shown only above 50%. 56

Figure 9 – Maximum likelihood phylogeny of *leu*D. Numbers at nodes indicate bootstrap replicates. Bootstrap support values are shown only above 50%. 57

Figure 10 – Mitochondrion maximum likelihood phylogenetic tree based on an amino acid matrix with 23 coding genes sequences. Bootstrap values are shown in the branches. Sequences from GenBank are indicated with accession number..... 60

Figure 11 – Chloroplast maximum likelihood phylogenetic tree based on an amino acid matrix with 187 coding genes sequences. Bootstrap values are shown in the branches. Sequences from GenBank are indicated with accession number. 61

Figure 12 – Schematic figure including only the *Gracilaria* sensu lato from chloroplast genomes (Figure 11). The characteristics about thallus (flat or cylindrical) and spermatangia structure are represented in all species and were based on several literature descriptions. These features are not discrete and some cases the same species could presents both. 63

Abbreviations

aa – amino acid

bp – base pairs

CDS – coding protein sequence

FSO – freestanding ORFs of plasmids

G. – *Gracilaria*

Gh. – *Gracilariophila*

Gp. – *Gracilariopsis*

HGT – horizontal gene transfers

HTS – High-throughput sequencing

LCB – locally collinear blocks

ML – Maximum likelihood

nt – nucleotide

ORF – Open-reading frame

PCR – polymerase chain reaction

PDS – plasmid-derived sequence

RT-PCR – Real-Time PCR

s.l. – *sensu lato*

s.s. – *sensu strictu*

Resumo

A família Gracilariaceae está globalmente distribuída e inclui várias espécies economicamente importantes. Atualmente, cinco gêneros são reconhecidos em Gracilariaceae: *Gracilariophila* (parasita), *Curdiea*, *Melanthalia*, *Gracilaria* e *Gracilariopsis*. Algumas espécies de *Gracilaria* foram taxonomicamente transferidas para *Hydropuntia*. Entretanto, esse gênero é bastante controverso. Técnicas de sequenciamento de alta performance (HTS) levaram a um aumento de estudos usando genomas organelares completos, que têm sido usados para inferir relações filogenéticas em Rhodophyta e na investigação de outros aspectos dos genomas de algas vermelhas, incluindo sintenia gênica e transferências horizontal de genes (HGT). O HTS também facilitou a busca por plasmídeos extracromossômicos e sua influência nos genomas organelares por HGT. Nós utilizamos HTS para montar e anotar genomas organelares (mitocôndrias e cloroplastos) de sete espécies de Gracilariaceae usando a plataforma Illumina HiSeq 2500 e recebemos sequências de 31 amostras Gracilariaceae da Dr. Goia Lyra que foram montadas, anotadas e incluídas em nossas análises. Utilizamos esses dados, combinados com genomas publicados, para inferir filogenias e comparar a arquitetura do genoma dessas espécies representando as principais linhagens em Gracilariaceae. Os genomas mitocondrial e plastidial são altamente conservados na sintenia gênica e a variação ocorreu principalmente em regiões onde foram encontradas inserções de sequências derivadas de plasmídeos (PDS), similares aos plasmídeos extracromossômicos conhecidos de algas vermelhas. Nos genomas mitocondriais, as inserções de PDS estavam em duas regiões onde a direção da transcrição muda: entre os genes *cob* e *trn*L e os genes *trn*A e *trn*N. As inserções de PDS no genoma do cloroplasto estavam em posições diferentes, mas geralmente encontradas entre os genes *psd*D e *rrs*. O operon bacteriano *leu*C/*leu*D foi encontrado nos genomas dos cloroplastos de *Gracilaria tenuistipitata*, *G. chilensis*, *M. intermedia* e também no plasmídeo de *G. vermiculophylla*. As árvores filogenéticas mostram duas origens diferentes de *leu*C/*leu*D: os genes encontrados no cloroplasto e no plasmídeo estavam próximos de proteobactérias, e os genes codificados no núcleo estavam próximos de Viridiplantae e cianobactérias. Gracilariaceae pode ser um bom modelo para estudar o impacto de PDS na evolução de genomas devido à presença frequente de inserções PDS em genomas organelares. Nossas análises filogenéticas demonstraram histórias evolutivas similares entre cloroplasto e mitocondria. No entanto, a filogenia de cloroplasto foi melhor resolvida com valores máximos de Bootstrap em todos os ramos. Nossa amostragem taxonômica corrobora a presença de três linhagens principais: *Melanthalia*/*Curdiea*, *Gracilariopsis* e *Gracilaria*. *Melanthalia intermedia* aparece como grupo-irmão do clado monofilético incluindo *Gracilaria* e *Gracilariopsis*, que foram resolvidos como gêneros monofiléticos. Além disso, baseado nas características da arquitetura do genoma organelar, os gêneros *Gracilariopsis* e *Gracilaria* se distinguem pela perda do gene plastidial *pet*P em *Gracilaria* e pela posição de rearranjo do gene *trn*H no genoma mitocondrial. Nós não encontramos evidências para a permanencia o gênero *Hydropuntia* como originalmente proposto.

Abstract

Gracilariaceae is a worldwide distributed family that includes numerous economically important species. Currently, five genera are recognized in the Gracilariaceae: *Gracilariophila* (parasitic), *Curdiea*, *Melanthalia*, *Gracilaria*, and *Gracilariopsis*. Some species of *Gracilaria* were taxonomically transferred to *Hydropuntia*. However, this genus is quite controversial. High-throughput sequencing (HTS) techniques has led to an increase in studies using complete organellar genomes, which have been used to infer phylogenetic relationships in Rhodophyta and the investigation of other aspects of red algal genomes, including gene synteny and horizontal gene transfers (HGT). HTS also facilitated the search for extrachromosomal plasmids and its influence in the organellar genomes by HGT. We applied HTS to assemble and annotate organellar genomes (mitochondria and chloroplast) from seven species of Gracilariaceae using Illumina HiSeq 2500 platform. We also received raw reads of 31 samples of Gracilariaceae from Dr. Goia Lyra that were analysed and included in our work. We used these data, combined with published genomes, to infer phylogenies and compare the genome architecture of these species representing the main lineages in Gracilariaceae. The mitochondrial and chloroplast genomes were highly conserved in gene synteny among the species, and variation mainly occurred in regions where insertions of plasmid-derived sequences (PDSs) were found, which were similar to known red algae extrachromosomal plasmids. In mitochondrial genomes, the PDS insertions were in two regions where the transcription direction changes: between *cob* and *trn*L genes, and *trn*A and *trn*N genes. PDS insertions in chloroplast genome were in different positions, but generally found between *psd*D and *rrs* genes. The bacterial *leu*C/*leu*D operon was found in *Gracilaria tenuistipitata*, *G. chilensis, M. intermedia* chloroplasts genomes, and also in *G. vermiculophylla* extrachromosomal plasmid. Phylogenetic trees show two different origins of *leu*C/*leu*D: genes found in chloroplasts and plasmids were close to proteobacteria, and genes encoded in the nucleus are close to Viridiplantae and cyanobacteria. Gracilariaceae may be a good model to study the impact of PDS in genome evolution due to the frequent presence of these sequences inserted in organellar genomes. Our phylogenetic analyses demonstrated similar evolutionary histories between the chloroplast and mitochondrial genomes. However, chloroplast phylogeny was better resolved with full support. Our taxonomical sampling supports the presence of three main lineages: *Melanthalia/Curdiea*, *Gracilariopsis* and *Gracilaria*. *Melanthalia intermedia* was sister to a monophyletic clade including *Gracilaria* and *Gracilariopsis*, which were resolved as monophyletic genera. Furthermore, the characteristics of organellar genome architecture, *Gracilariopsis* and *Gracilaria* genera are also supported by the loss of the plastid gene *pet*P in *Gracilaria* and the rearrangement position of the gene *trn*H in the mitochondrial genome. Beside this, we found no support for the genus *Hydropuntia* as originally proposed.

Introduction

The Gracilariaceae family

The family Gracilariaceae Nägeli (1847) used to belong to the order Gigartinales F.Schmitz in Engler (1892), but based on morphological characteristics specific to the Gracilariaceae species, the family was elevated to the level of order (Fredericq and Hommersand 1989a). Therefore, the order Gracilariales Fredericq & Hommersand (1989a) was determined to contain only one family Gracilariaceae. The fundamental characteristics of the order and family are based on vegetative structure, such as the development of a pseudoparenchymatous thallus and presence of pit-plugs with a single plug cap and a cap membrane (Pueschel and Cole 1982; Fredericq and Hommersand 1990a) and characteristics of the female reproductive structure: i) a supporting cell that carries a two-celled carpogonial branch flanked by sterile branches; ii) the fertilized carpogonium fusing with its supporting cell to form the fusion cell, which cuts off several gonimoblast initials; iii) secondary fusions between gonimoblast derivatives and cells of gametophytic tissues (Fredericq and Hommersand 1989a, 1990a; Gurgel and Fredericq 2004). Moreover, molecular studies including species of Gracilariales or with higher taxonomical level have confirmed the monophyly of Gracilariales (Lyra et al. 2015; Yang et al. 2015; Lee et al. 2016a).

The genera of Gracilariaceae are characterized mainly by the morphology and ontogenetic development of the female and male reproductive structures (Fredericq and Hommersand 1990b). These characteristics correspond to the spermatangia distribution and cystocarp structure (Fredericq and Hommersand 1989a; Lyra et al. 2015). The use of the distribution of spermatangia to species separation was initially proposed by Yamamoto (1978), who recognized three basic types: (i) *chorda*-type, considered the most primitive, it is characterized by the distribution of spermatangia in a layer on the surface of the thalli; ii) *textorii*-type, in which the spermatangia are found in shallow conceptacles in the cortex; (iii) *verrucosa*-type, in which the spermatangia are located in deep conceptacles that penetrate the subcortex (Yamamoto 1978; Bellorin 2002; Lyra et al. 2015). Two distribution patterns were added posteriorly: the s*ymmetrical*-type, in which the spermatangia located in the superficial cell layer are interrupted by sterile cortical cells; and the *henriquesiana*-type, in which the spermatangia are produced in deep conceptacles composed of several cavities that open through a single pore, causing slight elevation of the thallus surface (Bird and McLachlan 1984; Yamamoto 1984; Bellorin 2002). *Symmetrical*-type has not been used anymore and is considered a variation of *chorda*-type. The characteristics of the cystocarp structure are based on the presence or absence of the nutrient tubular cells and its location in the cystocarp cavity, and cystocarp completely filled or not by gonimoblast (Fredericq and Hommersand 1990b).

Currently, five genera are recognized in the Gracilariaceae: *Gracilariophila* Setchell & H.L. Wilson (Wilson 1910), which is parasitic; *Curdiea* Harvey (1855), *Melanthalia* Montagne (1843), *Gracilaria* Greville (1830), and *Gracilariopsis* Dawson (1949), which are benthic and free-living

(Lyra et al. 2015; Guiry and Guiry 2018) (Table 1). Two parasitic genera used to be recognized in Gracilariaceae: *Gracilariophila* (hereafter *Gh.*), with two species; and the *Congracilaria* Yamamoto (1986), with the only species *C. babae* Yamamoto (1986) (Table 1). Both genera are adelphoparasites, which evolved from the most recent common ancestor with their host (Setchell 1918; Goff et al. 1996; Hancock et al. 2010). *Gracilariophila* had cystocarps lacking nutrient tubular cells linking gonimoblast and pericarp, and superficial spermatangia. *Congracilaria* had cystocarps with nutrient tubular cells linking gonimoblast and pericarp, and spermatangia organized in conceptacles (Fredericq and Hommersand 1990b). *Gracilariophila oryzoides* Setchell & H.L.Wilson (Wilson 1910) is the most common species of *Gracilariophila* and it is mainly parasitic of *Gracilariopsis andersonii* (Grunow) E.Y.Dawson (1949) (Hancock et al. 2010). *Congracilaria babae* was mainly parasitic of *Gracilaria salicornia* (C.Agardh) E.Y.Dawson (1954) (Yamamoto 1986). However, based on molecular analysis of DNA sequences from the nuclear, mitochondrial and plastid genomes and anatomical observations, *C. babae* was transferred to *Gracilaria* (Ng et al. 2014).

The genera *Melanthalia* and *Curdiea,* with a few species each, are restricted to Oceania, especially to southern Australia, Tasmania, and New Zealand. They were also found in Antarctica (Gurgel and Fredericq 2004). *Melanthalia* is represented currently by four species, and *Curdiea* by eleven species (Guiry and Guiry 2018) (Table 1). The cystocarp structure of both genera is characterized by presenting a cavity completely filled by gonimoblasts, the absence of nutrient tubular cells, carposporangia in long lines However, *Melanthalia* cystocarp presents thick-walled sterile gonimoblast tissue, which is absent in *Curdiea*. The spermatangia are unknown and tetrasporangia are cruciately divided and are located between raised cortical filaments (nemathecia) (Fredericq and Hommersand 1990b).

The genus *Gracilaria* is the most diverse, with about 186 species (Guiry and Guiry 2018) (Table 1). Most species are concentrated in the tropical region, with diversity decreasing towards the poles (Plastino and Oliveira 1994). *Gracilaria* is characterized primarily based on the female and male reproductive structures. The cystocarp structure presents nutrient tubular cells in both pericarp and floor of the cystocarp cavity, which is not completely filled by gonimoblast. The spermatangia are always produced in conceptacles, which may be simple, shallow or deep (*textorii, verrucose* and *henriquesiana*-type, respectively) (Fredericq and Hommersand 1990b; Bellorin 2002).

Gracilariopsis is the second larger genus with 22 species (Guiry and Guiry 2018) (Table 1). The genus was separated from *Gracilaria* by Dawson (1949) based on anatomical details of the cystocarp, which lack nutrient tubular cells (Dawson 1949; Fredericq and Hommersand 1989b). However, Papenfuss (1966) observed that the nutrient tubular cells may be present or not in some *Gracilaria* species, thus *Gracilariopsis* was reduced to a taxonomic synonym under *Gracilaria*. *Gracilariopsis* has been widely accepted after the morphological and taxonomical studies based on *Gp. lemaneiformis* (Bory de Saint-Vincent) E.Y. Dowson, Acleto & Foldvik (1964) (Fredericq and Hommersand 1989b). In addition to the absence of the nutrient tubular cells in the cystocarps (Fredericq and Hommersand 1989b, Liao and Hommersand 2003), *Gracilariopsis* was recognized as a

distinct genus based on the male reproductive apparatus. The spermatangial cells developed on the surface layer (*chorda*-type) on the branches, close to the apices (Fredericq and Hommersand 1989b). This characteristic was observed by Yamamoto (1975), but he kept *Gracilaria* and *Gracilariopsis* merged and proposed three subgenera to *Gracilaria* based on the three main types of spermatangia development (Yamamoto 1978).

Some species of *Gracilaria* were taxonomically transferred to the genus *Polycavernosa* C.F. Chang et B.M. Xia (1963). However, the morphological characteristics that defined *Polycavernosa* included the species *P. urvillei* (Montagne) Xia & Abbott (1987), which was the type of the genus *Hydropuntia* Montagne (1842). Therefore, as *Hydropuntia* was the oldest name, all species of *Polycavernosa* were transferred to *Hydropuntia* (Wynne 1989). This genus is still quite controversial, dividing the opinion of researchers about its existence. The separation of *Gracilaria* was made based on characteristics of the reproductive system: the nutrient tubular cells connect to the gonimoblasts only at the floor of the cystocarp, and the spermatangia are produced in deep conceptacles composed of several cavities (Fredericq and Hommersand 1990b, Liao and Hommersand 2003), the *henriquesiana*-type (Bird and McLachlan 1984; Bellorin 2002). The genus *Hydropuntia* did not have general acceptance since some authors considered compound conceptacles as a variation of the simple *verrucosa* conceptacle (Bird and McLachlan 1984; Wynne 1989; Bellorin 2002). In addition, some species presented compound and simple conceptacles mixed in the same specimen (Bird et al. 1986; Abbott et al. 1991; Plastino and Oliveira 1997). However, some authors considered *Hydropuntia* as a distinct genus (Wynne 1989; Liao and Hommersand 2003; Gurgel and Fredericq 2004).

Table 1 - Genera of Gracilariaceae, number of currently recognized species and notes about currently taxonomical situation. Data were obtained from Guiry and Guiry (2018).

Genus	Species number	notes	Synonym font
Congracilaria Yamamoto, 1986	0	Parasite. Synonym of *Gracilaria*	Ng et al. 2014
Curdiea Harvey, 1855	11	Currently accept genus	
Gracilaria Greville, 1830	186	Currently accept genus	
Gracilariophila Setchell & H.L.Wilson, 1910	2	Parasite. Currently accept genus	
Gracilariopsis E.Y.Dawson, 1949	22	Currently accept genus	
Hydropuntia Montagne, 1842	0	Synonym of *Gracilaria*	Lyra et al. 2015
Melanthalia Montagne, 1843	4	Currently accept genus	

All these genera of Gracilariaceae presented a limitation in relation to the taxonomically informative characters used in the species identification, leading to discussions on the taxonomic condition of these genera. The high number of species, the wide geographic distribution and the presence of cryptic species contribute to the taxonomic confusion (Oliveira 1984; Plastino and Oliveira 1994). In addition, the recognition of species, or genera, is a complex task since it usually requires fertile plants and is only possible after microscopic analysis and the study of characteristics related to the ontogeny of spermatogenesis and to the development of post-fertilization carpogonia (Oliveira 1984; Fredericq & Hommersand 1989a). The addition of molecular studies with a

combination of morphological and anatomical data was important to the development of systematics of Gracilariaceae.

Molecular phylogenetic studies that included *Melanthalia* and/or *Curdiea* supported the monophyly of the genera (Bird et al. 1992, Bellorin et al. 2002, Gurgel and Fredericq 2004, Lyra et al. 2015). The recognition of *Gracilariopsis* as a distinct genus was also supported by molecular analysis (Bird et al. 1992, 1994; Bellorin et al. 2002; Gurgel et al. 2003, Lyra et al. 2015). However, molecular studies with Gracilariaceae were controversial in relation to the permanency of *Hydropuntia* (Bellorin et al. 2002). One of the first phylogenetic analyses with Gracilariaceae was done combining parsimony, distance and maximum-likelihood (ML) analysis using nuclear-encoded small subunit ribosomal gene (SSU rDNA) indicated that *Gracilaria* and *Hydropuntia* might be distinct taxa (Bird et al. 1992). However, the sampling of this research was very limited (Bird et al. 1992). This study also suggested that *Gracilaria chilensis* C.J.Bird, McLachlan & E.C.Oliveira (1986) would represent another group (Bird et al. 1992). Posteriorly, a phylogenetic analysis with SSU rDNA and internal transcribed spacer (ITS) regions combining parsimony, distance, and ML, with a larger dataset, resulted that species bearing *henriquesiana*-type spermatangia (*Hydropuntia*) and *Gracilaria* were not phylogenetically separated (Bellorin et al. 2002). A while later, another molecular study including the chloroplast ribulose 1,5-bisphosphate carboxylase large subunit gene (*rbc*L), which was considered to provide better resolution for inferring species level, and Bayesian analysis, beside the molecular markers and phylogenetic methods previous reported, determined a major clade including *Gracilaria* (most of the *Gracilaria* species – called *Gracilaria sensu stricto*), *Hydropuntia* complex and a new unnamed genus containing *G. chilensis* (Gurgel and Fredericq 2004). This major clade was called *Gracilaria sensu lato*. They considered that these three lineages should be distinct genera (Gurgel and Fredericq 2004). The largest and most well-resolved phylogeny of Gracilariaceae to date showed that *Hydropuntia* is not a monophyletic genus and some species with *henriquesiana*-type spermatangia were grouped in the *Gracilaria sensu stricto* clade (Lyra et al. 2015). Furthermore, the morphological features used to distingue both genera overlap in some cases. Therefore, they concluded that *Hydropuntia* should be synonymized to *Gracilaria* (Lyra et al. 2015).

Economic importance

There are species of Gracilariaceae that have great economic importance, being highly relevant in biotechnology, food, pharmaceutical, and textile industries, due to the wide applicability of the hydrocolloids (agar-agar) present in their cell walls (Zemke-White & Ohno 1999). In addition to the production of agarans, pharmaceutical studies of *Gracilaria* species, including with Brazilian species, show the presence of molecules with antiviral, anti-inflammatory and antihypertensive properties (Smit 2004, Vanderlei et al., 2011, Amorim et al., 2012).

The species of Gracilariaceae that are indicated as ideal for the cultivation are generally terete and wherefore most of the studies are focused on terete species (Bellorin et al. 2002). Economically species must have a high content of agar-agar, fast growth and be easy to reproduce vegetatively.

Therefore, precise taxonomic identification of species selected for cultivation is important, specially considering the occurrence of cryptic species (Santelices & Doty 1989).

Gracilariaceae terete species are cultivated in several parts of the world and currently represent the main source for the production of agar-agar (McHugh 2003). In Brazil, *Gracilaria* is the most economically important genus. In the 1970s, the exploitation of these agarophytes was made by collecting biomass directly from natural banks on the Northeast coast. The biomass was sold for the production of colloids, which resulted in a decrease of these reserves along the coast. Sea cultivation attempts were made, but they achieved little success due to the lack of basic knowledge about the species selected for cultivation (Carneiro 2011). Currently, with incentives to sea aquaculture, some *Gracilaria* species have been used in pilot experiments in the Northeast coast and better results are being obtained (Bezerra & Marinho-Soriano 2010, Carneiro 2011). Therefore, the interest in studies with Brazilian species for industrial-scale production has increased (Marinho-Soriano et al., 2007; Carneiro 2011).

High-throughput sequencing

Since the advent of the DNA sequencing, molecular studies become highly important to understand the complexity and diversity of the life. Sequencing of complete or nearly complete genomes is a key step in better understanding the biology of living organisms. The first genomes were sequenced using the traditional chain-terminated Sanger DNA sequencing method (Sanger 1977) in automatic sequencers, which were posteriorly called first generation sequencing (Dijk et al. 2014a). Only a single fragment, between 700-900 base pairs (bp), previously amplified by polymerase chain reaction (PCR) and cloned, was sequenced in each capillary of the sequencer, which had 96 capillaries. For the study of genomes, this method is highly costly and time-consuming (Kim et al. 2014a). For example, the Human Genome Project (International Human Genome Sequencing Consortium 2004) spent more than 10 years to publish the complete genome at a cost of about US$ 3 billion (Dijk et al. 2014a). It was clear that faster, higher throughput, and cheaper technologies were required for genome sequencing (Dijk et al. 2014a). In the last two decades, new DNA sequencing technologies have arisen decreasing the time and costs, making it possible to obtain larger molecular datasets. These new technologies were first called Next Generation Sequencing, but nowadays are better called High-Throughput Sequencing (HTS).

The number of genomic studies has increased rapidly since of the last decade thanks to HTS (Chan and Ragan 2013). HTS allows the generation of millions of DNA reads – the HTS output – in a single process, without the need to go through cloning steps and in some cases without the need for PCR before entering the sequencer (Dijk et al 2014a). A range of technologies is now available for use in HTS. Among these are several short-read sequencing platforms, which produce much shorter reads than traditional Sanger sequencers, as well as some long-read platforms. For short-read sequencing, Illumina is the dominant system in the HTS market (Reuter et al. 2015; Oliveira et al. 2018), which possess different types of sequencing instruments with a large range of output quantity, from low-

throughput to very large output (up to 900 Gb in Illumina HiSeq X) (Goodwin et al. 2016). Error rates are very low for all these platforms, compared with other technologies. PacBio and MinION technologies can produce long sequence reads but produce a lower yield and generally have a higher error rate than Illumina (Goodwin et al. 2016, Oliveira et al. 2018). They are typically used to complement the more accurate and higher-throughput Illumina methods for assembly and genome finishing (Oliveira et al. 2018).

Before sequencing, the HTS platforms require that a DNA library be constructed for each sample. A DNA library is a collection of DNA fragments each containing oligonucleotide adapters at their ends that are compatible with the sequencing system (Head et al. 2014). Library protocols consist of four main steps: DNA fragmentation, size selection, adapter ligation, and final DNA library product quantification for sequencing (Head et al. 2014). Illumina sequencing method is based on sequencing by synthesis, which each nucleotide is identified during the sequencing after its incorporation into an elongating strand (Goodwin et al. 2016). The sequencing process involves clonal amplification of adaptor-ligated DNA fragments on the flow cell, which is a glass slide with adaptors on the surface, where the DNA fragments are linked (Bentley et al. 2008; Reuter et al. 2015; Goodwin et al. 2016). After this process of amplification of each DNA fragment in clonal clusters, the sequencing by synthesis is performed (Goodwin et al. 2016).

The advantage of HTS is the wide range of its application to generate different types of output. For example, a whole-genome sequencing (WGS) including the total genomic DNA of the organism, which include organelle DNA, plasmid DNA, nuclear DNA, and genomes from endosymbionts. It is also possible to sequence a set of different organisms in the same run, using the multiplexing approach, in which the DNA fragments of each sample is ligated to a short sequence tag (called index or barcode) and the reads of each sample are latter identified by bioinformatic tools (Oliveira et al. 2018). The biggest challenge about HTS technologies is the high demand for bioinformatic tools capable of sorting, analyzing and managing large amounts of data (Dijk et al. 2014a; Oliveira et al. 2018). Each HTS run produce millions of sequences reads that require specific computational analysis workflows to produce interpretable results. Individual steps in a workflow can be carried out by different programs, with customizable sets of features (Oliveira et al. 2018).

Genomic studies in red algae

The HTS techniques have transformed many areas of biology, including the study of algae, allowing large-scale approaches in the evolutionary investigation, systematics, biodiversity, genetic population, gene expression, and epigenetic (Kim et al. 2014a; Bhattacharya et al. 2015; Oliveira et al. 2018). At the time of writing of this thesis, eight studies of nuclear genome have been performed so far for red algae. The extremophilic unicellular red alga *Cyanidioschyzon merolae* P.De Luca, R.Taddei & L.Varano (1978) was the first complete algal genome (Matsuzaki et al. 2004). Two other unicellular red algae had their nuclear genomes sequenced: the mesophilic *Porphyridium purpureum* (Bory) K.M.Drew & R.Ross (1965) (Bhattacharya et al. 2013) and another extremophilic red alga

Galdieria sulphuraria (Galdieri) Merola (Merola et al. 1982), which live in hot, toxic metal-rich, acidic environments (Schönknecht et al. 2013). For the seaweeds, two Bangiophyceae, *Pyropia yezoensis* M.S.Hwang & H.G.Choi in Sutherland et al. (2011) (Nakamura et al. 2013) and *Porphyra umbilicalis* Kützing (1843) (Brawley et al. 2017); and three Florideophyceae, including two Gracilariaceae, *Gracilariopsis lemaneiformis* (Zhou et al. 2013) and *Gp. chorda* (Lee et al. 2018), and, the carrageenan-producing macroalga, *Chondrus crispus* Stackhouse (1797) (Collén et al 2013).

The known nuclear genomes of Rhodophyta are relatively small compared to other eukaryotic strains (Matsuzaki et al., 2004; Collén et al., 2013; Lee et al. 2018). The smaller known genome of a red algae is *Galderia sulphuraria*, with 13 Mb (Schönknecht et al. 2013) and *Chondrus crispus* has the largest sequenced genome, so far, with around 105 Mbp (Collén et al. 2013). The draft genomes of the two *Gracilariopsis* are 97 Mb and 92 Mb length to *Gp. lemaneiformis* (Zhou et al. 2013) and *Gp. chorda* (Lee et al. 2018), respectively. The reduced genome is probably because the apparent absence of polyploidization and majority of monoexonic genes (88% in *C. crispus*) (Collén et al. 2013; Lee et al. 2018). The knowledge about the genomes of red algae is highly important because this group presents great economic importance, as exemplified above for gracilarioid algae, and, also have great evolutionary importance since they originated through secondary endosymbiosis the complex chloroplasts of other eukaryotic lineages: Heterokonta, Alveolata, Cryptophyta and Haptophyta (Kim et al. 2014). Therefore, red algae genes are spread throughout the Eukarya Tree of Life and the genetic and genomic knowledge of red algae are crucial for understanding the evolution of Eukaryotes (Collén et al. 2013; Kim et al. 2014).

The use of HTS techniques has led to an increase in studies using complete organellar genomes. Dozens of chloroplast and mitochondrial genomes have been published in GenBank in the last years. At the time of the writing of this work, there were available eight chloroplasts and eleven mitochondrial genomes of Gracilariaceae species in Genbank (Table 2). Most of these are from economically important crop species, e.g. *Gracilariopsis chorda* (Yang et al. 2014b), *Gp. lemaneiformis* (Zhang et al. 2012, Zhang et al. 2016), *Gracilaria firma* Chang & Xia (1976) (Ng et al. 2017), *G. tenuistipitata* var. liui Zhang & Xia (1988), which was the first complete chloroplast genome published of Florideophyceae (Hagopian et al. 2004). The family was also investigated in the parasitic relationship between the parasite *Gracilariophila oryzoides* and its host *Gp. andersonii* using mitochondrial genomes (Hancock et al. 2010). Organellar genomes have been used to infer phylogenetic relationships in Rhodophyta, first for the study of higher taxonomical levels (Janouškovec et al. 2013, Yang et al. 2015, Lee et al. 2016a) and more recently for order and families (Costa et al. 2016, Díaz-Tapia et al. 2017), using phylogenomics approaches, which is the study of evolutionary relationships based on the comparison of data analyzes at the genomic scale (Chan & Ragan 2013).

Table 2 – Mitochondrial and chloroplast genomes from Gracilariaceae published in GenBank.

Species	INSDC	Size (Kb)	GC%	Protein	rRNA	tRNA	Other RNA	Intron	Reference
Mitochondrion									
Gracilaria changii (B.M.Xia & I.A.Abbott) I.A.Abbott, J.Zhang & B.M.Xia	KX980031	25,729	27.7	25	3	24	-	1	Song et al. 2017
Gracilaria chilensis C.J.Bird, McLachlan & E.C.Oliveira	KP728466	26,898	27.6	25	3	26	-	1	Lee et al. 2015
Gracilaria chilensis	MF401962	26,897	27.6	25	3	25	-	1	Liu et al. unpublished
Gracilaria chouae Zhang & B.M.Xia	MF351970	25,224	28.2	25	3	20	-	1	Zhang and Liu unpublished
Gracilaria salicornia (C.Agardh) E.Y.Dawson	KF824534	25,272	28.4	25	3	21	-	1	Campbell et al. 2014
Gracilaria salicornia	KT373903	25,915	28.6	25	3	24	-	1	Song et al. 2016
Gracilaria vermiculophylla (Ohmi) Papenfuss	KJ526627	25,973	28.1	25	3	22	-	1	Chi et al. unpublished
Gracilariophila oryzoides Setchell & H.L.Wilson	KX687879	25,184	28.1	25	3	21	-	1	Salomaki and Lane 2016
Gracilariopsis andersonii (Grunow) E.Y.Dawson	KX687878	27,011	28	25	3	21	-	1	Salomaki and Lane 2016
Gracilariopsis chorda (Holmes) Ohmi	KC875851	26,534	27.6	26	3	24	-	1	Yang et al. 2014b
Gracilariopsis lemaneiformis (Bory de Saint-Vincent) E.Y.Dawson, Acleto & Foldvik	JQ071938	25,883	27.5	26	3	21	-	1	Zhang et al. 2012
Chloroplast									
Gracilaria chilensis	KT266788	185,637	29.3	205	3	30	2	1	Lee et al. 2016b
Gracilaria chilensis	MF401963	185,640	29.3	202	3	30	2	1	Liu et al. unpublished
Gracilaria firma Chang & Xia	KX601051	187,001	28.1	216	3	30	2	1	Ng et al. 2017
Gracilaria salicornia	KF861575	179,757	28.8	204	3	30	2	1	Campbell et al. 2014
Gracilaria tenuistipitata var. *liui* Zhang & Xia	AY673996	183,883	29.2	206	3	30	2	1	Hagopian et al. 2004
Gracilariopsis chorda	KX284722	182,459	27.4	203	3	30	2	1	Lee et al. 2016a
Gracilariopsis lemaneiformis	KU179794	182,505	27.4	208	3	30	2	1	Zhang et al. 2016
Gracilariopsis lemaneiformis	KP330491	183,013	27.4	203	3	31	2	1	Du et al. 2016

Organellar genomes have been also facilitated the investigation of other aspects of red algal genomes, including gene synteny and horizontal gene transfers (HGT). For example, sequences derived from previous reported Rhodophyta plasmids (e.g. Goff and Coleman 1990, Villemur 1990a,

1990b, Moon and Goff 1997) were observed in chloroplast genomes in Rhodophyta (Lee et al. 2016b). Furthermore, HGT from a bacterial plasmid, which contained the genes *leu*C and *leu*D, were also reported in the chloroplast genomes of *Gracilaria tenuistipitata* var. *liui* (Hagopian et al. 2004, Janouškovec et al. 2013) and *G. chilensis* (Lee et al. 2016b).

Plasmids are extrachromosomal DNA molecules, autonomously replicating, generally circular, AT-rich, double-stranded and may have active open reading frames (Goff and Coleman 1990, Moon and Goff 1997, Lee et al. 2016b). They were first found in some species of Rhodophyta in a study of delimitation of species and populations of Gracilariaceae, using endonuclease digests of chloroplasts genomes (Goff and Coleman 1988). Up to now, only five species contain one or more extrachromosomal plasmids published in GenBank (Table 3), most of them are Gracilariaceae species (Goff and Moon 1990, Villemur 1990a, Lee et al. 2016b) which means that this family is interesting to investigate the evolution of plasmids in red algae and their relationship with organellar and nuclear genomes. These genetic elements have no known function in these seaweeds, conferring no obvious phenotype to species possessing them (Moon and Goff 1997). They occur in apparently in high proportions in the genomic DNA and may be present in high copy number per cell. Their exact cellular location is not yet known but may be reside in the nucleus (Goff and Coleman 1990). Previous hybridization studies indicated that plasmid sequences were not present integrated in the nuclear and organelles genomes. Also, these studies indicated that these plasmids are autonomously replicated, each possessing its own origin of replication site (Goff and Coleman 1990). However, analysis of complete mitochondrial and chloroplast genomes suggests that plasmid sequences behave as parasitic elements that spread plasmid-derived DNA regions in organellar genomes (Lee et al. 2016b).

Table 3 - Information of red algae plasmids available in GenBank.

Plasmid	Accession number	Organism	Sequence length	GC %	Topology	Nº of ORFs	Reference
Gch3937	AF034721	*Gracilaria chilensis*	3,937	25	Circular	5	Goff and Moon, unpublished/Lee et al. 2016b
Gch7220	AF034719	*Gracilaria chilensis*	7,220	24.9	Circular	16	Goff and Moon, unpublished/Lee et al. 2016b
Gch_Gc2	X51317	*Gracilaria chilensis*	3,827	24.9	Circular	1	Villemur 1990b
Gro4059	AF034717	*Gracilaria robusta*	4,059	26.5	Circular	2	Goff and Moon, unpublished/Lee et al. 2016b
Gro4970	AF034718	*Gracilaria robusta*	4,970	25.8	Circular	7	Goff and Moon, unpublished/Lee et al. 2016b
Gle4293	AF034720	*Gracilariopsis lemaneiformis*	4,293	26.8	Circular	5	Goff & Coleman 1990
Gle_GL3.5	X56441	*Gracilariopsis lemaneiformis*	3,491	25.9	Linear	2	Goff & Coleman 1990
Pp6427	AF002711	*Pyropia pulchra*	6,427	35.5	Circular	5	Moon & Goff 1997
Pp6859	AF002712	*Pyropia pulchra*	6,859	32.9	Circular	7	Moon & Goff 1997
pPT4	FJ956744	*Pyropia tenera*	1,820	41.7	Circular	2	Choi et al. 2001

Objectives

General

Considering the relative ecological and economic importance of Gracilariaceae, our study aimed obtain organellar genomes of representative species of the main lineages of Gracilariaceae with HTS and use these data to investigate the evolution of the family using organellar genomic architecture and a phylogenomic approach.

Specific

- Characterize new mitochondrial and chloroplast genomes of Gracilariaceae, expanding the number of different lineages within the family;
- Evaluate the evolutionary history of the family using phylogenomics;
- Compare the genomic architecture of each of the organelles, including an investigation of plasmid derived sequences.

Material and Methods

Choice of HTS machine

The choice of the HTS platform was the first important step in this project because its output capacity would determine the number of samples to be sequenced and the sequence coverage. Our focus in this project was to obtain full organellar genomes but considering future research and the investment to obtain and process data of high quality, we decided to do the whole-genome sequencing of each sample with a high coverage in order to obtain also nuclear data. Considering the machines available for this project at University of São Paulo (USP), we chose the Illumina platform because there were two different types of Illumina sequencers at the USP, which would be more accessible. GENIAL (Genome Investigation and Analysis Laboratory) in the Core Facility for Scientific Research (CEFAP-USP) of the Institute of Biomedical Sciences had the MiSeq® sequencer, which can generate up to 15 Gb of data, with reads with 300 bases. The Human Genome Center and Stem Cell Research (HUG-CEL) of the Institute of Biosciences – USP had the HiSeq® 2500 sequencer that had two run modes: Rapid-Run, which generates up to 150 Gb of data, with reads with 250 bases, using a flow cell, and High-Output, which generates up to 500 Gb of data, with reads with 125 bases, using a flow cell as well. This sequencer could do two runs at the same time, generating double data with two flow cells.

We determined the expected coverage of the reads in relation to the expected size of the genome. The expected coverage is the average number of times that each nucleotide is expected to be sequenced in a certain number of reads of a given length in the total genome assuming that the reads are randomly distributed across the genome (Sims et al. 2014). Kapraun and Freshwater (2012) estimated the size of Gracilariaceae genome to be around 196 and 245 Mbp using DNA quantification by the densitometry method. However, the draft nuclear genome size of *Gracilariopsis lemaneiformis* is estimated to be 95.96 Mbp (Zhou et al. 2013) and the nuclear genome of *Chondrus crispus*, which is a Florideophyceae species, the same subphyla of Gracilariaceae, was around 105 Mbp (Collén et al. 2013). We considered an estimate of 200Mbp for the nuclear Gracilariaceae genomes. The organelle genomes were estimated to be around 30 Kbp of the mitochondrial genome and 200 Kbp of the plastid genome. These values were based on Gracilariaceae organelle genomes that have already been published, available in the GenBank (Table 2). Thus, we estimated 210 Mbp in total for nuclear plus organellar genomes of each species.

To determinate the expected coverage for a *de novo* assembly it is not an easy task. For a standard multicellular organism, the assembly of reads by HTS assemblers usually requires at least 30x coverage for a successful assembly without gaps (Chitsaz et al. 2011). However, simulated data study identified that the minimum depth of sequencing required for *de novo* assembly for differently sized genomes can vary between 50X and 100X (Desai et al. 2013). Therefore, we stipulate that the expected coverage for each sequenced genome should be greater than 50x.

Following this line of reasoning, we could calculate the expected coverage dividing the total output range of the sequencer per the estimated genome length. We estimated that each Gracilariaceae species genome had 210 Mbp and the MiSeq run produces up to 15 Gb, thus a MiSeq run would produce enough data to generate about 71x of coverage for one sample. In the case of the HiSeq 2500, a flow cell can generate 150 Gb of data in a single run, using the Rapid-Run mode, which would allow us sequencing 10 samples with 71x of coverage using multiplexing technique. Therefore, we chose to sequence 10 samples in the HiSeq 2500, because in a single run we could generate data with a higher coverage for each sample. Even if we estimated that a 50x coverage would be sufficient, we decided to aim for a higher coverage (71x), as in practice there is always some loss in clusterization and part of the reads may not have good quality.

Taxon sampling and culturing

We chose ten Gracilariaceae species (Table 4), which represent different lineages in Gracilariaceae (Lyra et al. 2015). *Gracilariopsis longissima* (S.G.Gmelin) M.Steentoft, L.M.Irvine & W.F.Farnham (1995), *Gp. mclachlanii* Buriyo, Bellorin & M.C.Oliveira (Bellorin et al. 2008) and *Gp. tenuifrons* (C.J.Bird & E.C.Oliveira) Frederiqc & Hommersand (1989b) were chosen because they are representants of *Gracilariopsis* genus. *Gracilariopsis tenuifrons* is a cylindrical species, common on the Brazilian coast. This species, presents high growth rates, it is easy to cultivate in the laboratory (Plastino 1991; Serra 2013) and produces agar of good quality (Brito and Silva 2005), therefore, it has been considered one of the species with potential to be produced in large scale in Brazil.

Gracilaria tenuistipitata var. *liui* is a cylindrical species, common in East Asia and has been extensively cultivated in salt lakes in Southern China and Taiwan. This alga has a high rate of growth and production of agar. Also, it presents great tolerance to environmental factors, such as salinity, temperature, and pH (Haglund and Pedersén 1992; Macchiavello et al. 1998), which makes this species quite suitable for cultivation in nature and in vitro. This species has been used in several physiological and molecular studies and had its life history completed in our laboratory (Barufi et al. 2010). The molecular studies include the draft sequencing of the mitochondrial genome (Takahashi 2010), the complete sequencing of chloroplast genome (Hagopian et al. 2004), and characterization of about 3,600 ESTs (expressed sequence tags) (Nyvall Collén et al. 2012). Due to its tolerance to different environmental factors, its ease of maintenance in culture, and previous knowledge, this species was included in this work because it represents one of the model organisms for red algae. In addition, this species represents a basal lineage within the genus *Gracilaria* and it has been proposed to belong to a distinct genus (Gurgel and Fredericq 2004; Lyra et al. 2015).

Gracilaria caudata Agardh (1852), *G. cornea* Agardh (1852), and *G. rangiferina* (Kützing) Piccone (1886) were included in the genus *Hydropuntia* (Wynne 1989; Gurgel and Fredericq 2004). More recently molecular and morphological data have kept these species in the genus *Gracilaria* (Lyra et al. 2015). *Gracilaria caudata* is a cylindrical species, occurring in along the Brazilian coast (Plastino and Oliveira 1997). It has been frequently collected in the Brazilian northeast to produce agar

for the national market, but the exploitation of natural banks is subject to environmental impacts in some regions (Miranda 2000). The species has the potential for cultivation in the sea (Assad-Ludewigs 1984) and tolerates a wide range of salinity (10 to 60 ups) (Yokoya and Oliveira 1992). *Gracilaria rangiferina* is close related to the type species of the *Hydropuntia* genus, *H. urvillei* Montagne, and it is an only species that have flat thalli, while all other species in the lineage have cylindrical thalli (Lyra et al. 2015). *Gracilaria isabellana* Gurgel, Fredericq & J.N. Norris (2004), *G. ferox* J.Agardh (1852), and *G. gracilis* (Stackhouse) M.Steentoft, L.M.Irvine & W.F.Farnham (1995) are flat species and belong to the *Gracilaria* clade *sensu stricto* (Gurgel and Fredericq 2004; Lyra et al. 2015).

All species, except *Gracilaria caudata*, were obtained from the LAM-USP Germplasm Bank (Costa et al. 2012). *Gracilaria caudata* has been maintained in laboratory culture but has no Germplasm Bank voucher (Table 4). These samples were cultivated in vitro for about one month. Apical fragments obtained from the bank were transferred to Erlenmeyer flasks 150 mL with 50 mL of modified von Stosch enriched seawater solution (Ursi and Plastino 2001). The cultures were maintained under 150 µmol · photons · m-2 · s-1 photosynthetically active radiation (PAR) provided by 40 W daylight fluorescent tubes on a 14 h light/10 h dark cycle and aerated for 30 min · h-1. We renewed the medium weekly when we rinsed the algae in the fresh water and MilliQ water to clean possible contaminants. Twenty-four hours before DNA extraction we added to the 50 mL of modified von Stosch medium with 50 mg · L-1 of penicillin and 25 mg · L-1 of streptomycin to eliminate possible bacterial contamination; plus 2 mL · L-1 of Micostatin to eliminate possible fungi contamination; and 1-5 mL · L-1 of germanium dioxide to eliminated possible diatom contamination.

Table 4 - Species cultured and sequenced in this work.

Species	Collection information	Germplasm voucher	Herbaria voucher
Gracilaria caudata J. Agardh	Rio do Fogo, RN, Brazil. 30/Sept/2011. E. Plastino and L. Ayres-Ostrock	-	SPF57390
Gracilaria cornea J. Agardh	Guajiru, Trairi, CE, Brazil. Feb/1992. E. Plastino	BG0113	SPF55662
Gracilaria ferox J.Agardh	Ubatuba, SP, Brazil. 28/Oct/1981. E.C. Oliveira	BG0005	SPF26055
Gracilaria gracilis (Stackhouse) M.Steentoft, L.M.Irvine & W.F.Farnham	Pajuçara, AL, Brazil. 03/Mar/1994. E. Plastino	BG0084	SPF55734
Gracilaria isabellana Gurgel, Fredericq & J.N.Norris	Cabo Frio, RJ, Brazil. 06/Nov/1987. E. Plastino	BG0036	SPF52189
Gracilaria rangiferina (Kützing) Piccone	Trairi, CE, Brazil. 16/Sept/1989. E.C. Oliveira	BG0092	SPF56055
Gracilaria tenuistipitata var. *liui* Zhang & Xia	Haikou, Hainan Island, China. E.C. Oliveira	BG0062	SPF56193
Gracilariopsis longissima (S.G.Gmelin) Steentoft, L.M.Irvine & Farnham	England	BG0051	-
Gracilariopsis mclachlanii Buriyo, Bellorin & M.C.Oliveira	Zanzibar, Tanzania. Jun/1992. E.C. Oliveira	BG0072	-
Gracilariopsis tenuifrons (C.J.Bird & E.C.Oliveira) Fredericq & Hommersand	Cabo Frio, RJ, Brazil. 06/Nov/1987. E. Plastino	BG0039	SPF52192

Molecular techniques and high-throughput sequencing

We extracted total genomic DNA from fresh samples using the manual method described in (Faugeron et al. 2001), with small modifications, mainly in relation to the initial amount of fresh mass. The algae were withdrawn from the liquid culture medium, washed with fresh water and MilliQ water and dried with absorbent paper. The extractions were done in duplicate or triplicate using a total of 190 mg to 520 mg of fresh mass. DNA was extracted from the tissue ground in liquid nitrogen until a fine powder was obtained. The ground powder was incubated at room temperature for 1 h in 800 µL of lysis buffer (0.1 M Tris base, pH 8.0, 0.05 M EDTA, 0.2 M NaCl, 2.5 M KAc), 100 µL of Tween 20 (10%), 8 µL of proteinase K (20 mg · mL^{-1}) and 40 µL of RNAse (10 mg · mL^{-1}). To purify the DNA solution we added 700µL of phenol:chlorophorm: isoamyl alcohol (25:24:1) directly to the lysis buffer, homogenizing by inversion during at least 3 min before centrifugation for 20 min at 14,000 RPM and 4°C. The aqueous phase was extracted a second time by one volume chlorophorm: isoamyl alcohol (24:1) and centrifuged for 3 min at 14,000 RPM and 4°C. The supernatant was recovered and the DNA was precipitated by 0.6 volume of isopropanol and pelleted by 15 min centrifugation at 14,000 RPM and 4°C. The pellet was washed with 70% ethanol, vacuum dried, and resuspended in 200 µL of MilliQ water. All DNA extraction has been done with caution to keep the DNA molecule as entire as possible. We needed 2 µg of unbroken double helix DNA. However, the amount of DNA extract desired was at least 6 µg of DNA measured on the NanoDrop™ 2000 Spectrophotomters (ThermoFisher) because of this equipment measure only the total amount of nucleotides, which includes single helix DNA, double helix DNA, RNA and free nucleotides (Figure 1).

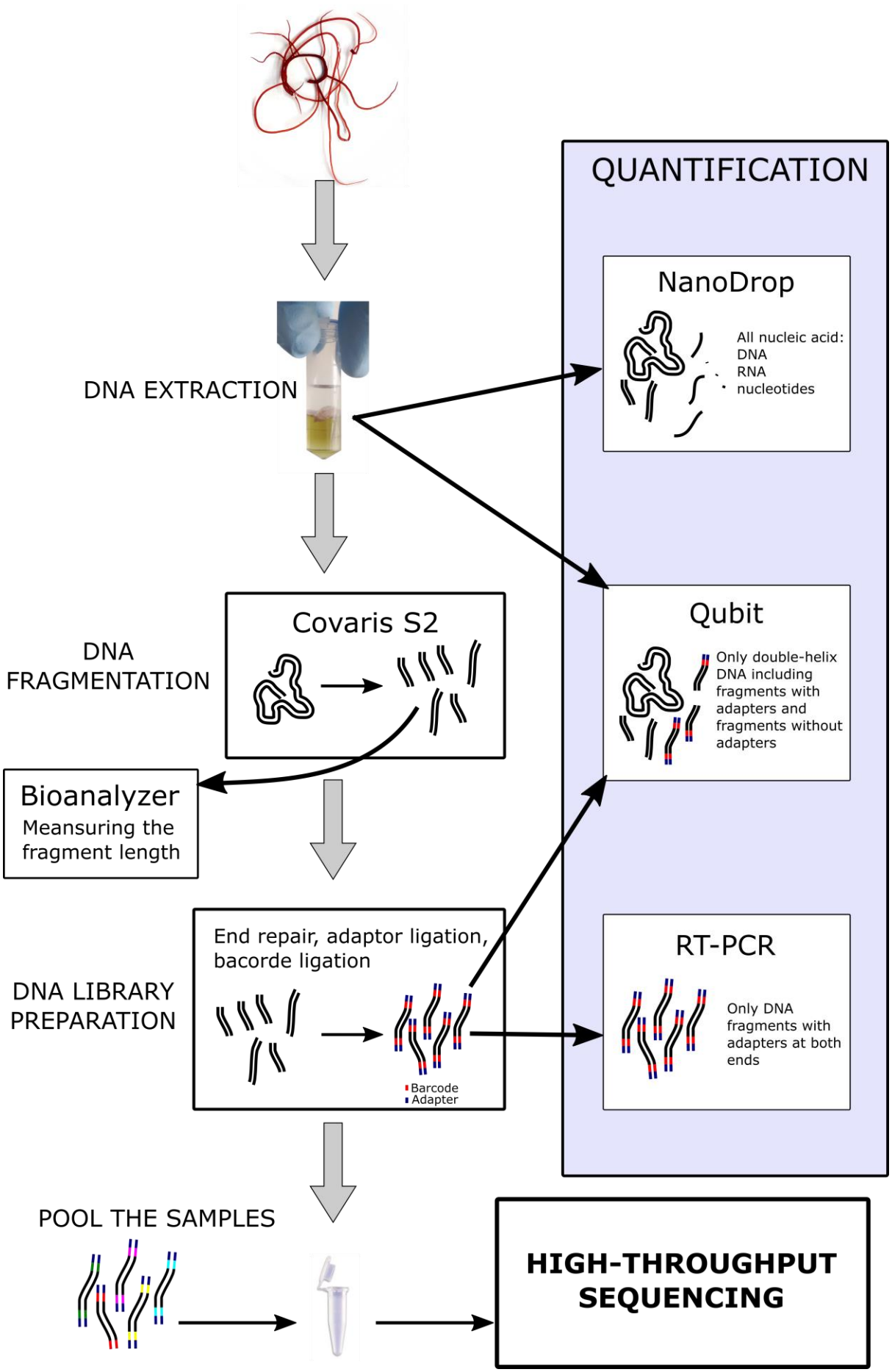

Figure 1 - Molecular workflow to prepare the DNA libraries.

The DNA purity was an important factor to make high quality DNA libraries and it was measured according to quality parameters based on the ratios of absorbance values of the NanoDrop™. The 260/280 ratio should be above 1.8 indicating relatively pure DNA in relation to proteins, and the 260/230 ratio of 1.8-2.2 indicated a relatively pure DNA in relation to carbohydrates and other extraction products. When the values indicated carbohydrate impurities, a purification step was performed using 0.1 volume of absolute ethanol, centrifugation at 14,000 RPM for 20 minutes at 4°C, the supernatant was removed and the DNA precipitation process was performed, according to Bellorin et al. (2002). When the values indicated impurities by proteins, we repeated the last cleaning step of the DNA extraction protocol with phenol: chloroform: isoamyl alcohol.

The library preparation kit chosen was the TruSeq DNA PCR-Free LT Library preparation kit (Illumina, San Diego, USA), because it does not have the PCR step. PCR may generate a bias in GC content, as there are indications that GC-neutral fragments are amplified more efficiently than fragments with high GC or high AT content (Kozarewa et al. 2009; Dijk et al. 2014b). The expected size of the library was 550 bp, since the sequencing was pair-end, with reads of 250 bases.

After extraction and measurement of the amount and quality of DNA by NanoDrop™, the quantification of double helix DNA was done on the Qubit® 3.0 Fluorometer using the Qubit® dsDNA HS Assay Kit (Thermo Fisher Scientific Inc, Waltham, USA), which allow us to measure only double helix DNA (Figure 1). Before the DNA library preparation, we diluted the genomic DNA to a final volume of 55 µL at 40 ng/µL.

The next step to prepare the DNA libraries is the fragmentation of the DNA in the size of around 550 bp. The TruSeq DNA PCR Free kit recommends the fragmentation by sonication, that it was made with Covaris S2 (Massachusetts, USA). We tested different combinations of the parameter for the DNA fragmentation with Covaris S2 (Table 5) and the size of the fragments was measured in Bioanalyzer 2100 (Agilent Technologies, USA) (Figure 1). The final parameters used in the extracted DNA were: 175 W of incident power, 5% duty factor, 200 cycles, intensity level 2, for 40 s. The DNA libraries were prepared following the TruSeq DNA PCR Free kit manufacturer's instruction. During the DNA library preparation, a barcode sequence was added in the fragments of each sample that allowed us to mix together all the samples in a single pool and make a unique multiplexing sequencing run (Figure 1). We did the final quantification of the DNA libraries with Real-Time PCR (RT-PCR) using LightCycler® 480 Instrument II (Roche Molecular Systems, USA) with KAPA Illumina Library Quantification kit (Kapa Biosystems, Wilmington, USA). The quantification by RT-PCR was necessary because it was possible to ascertain how many fragments had the adapters attached at both ends of each fragment. Qubit® measures only the total amount of double-stranded DNA, regardless of whether the fragments are bound to the adapters (Figure 1). All the prepared DNA libraries were diluted in ddH2O to a final concentration of 2 nM and were pooled together. We sequenced the pool using Illumina HiSeq 2500 at Human Genome and Stem Cell Research Center (HUG-CELL) at USP (São Paulo, Brazil). We implemented the Rapid Run mode using the HiSeq Rapid PE Cluster v2 and

HiSeq Rapid SBS (500 cycles) kits to generate reads of 250 bases length. The pool was sequenced in two different lines to avoid sequencing error.

Table 5 - Parameters used in Covaris S2 DNA fragmentation tests using only *Gracilaria tenuistipitata*. *Standard parameters indicated by the Covaris S2 manual to obtain 500 bp fragments. The parameters selected for DNA library preparation are in bold.

	Duty Factor (%)	Intensity	Duration (seconds)
Illumina	10	2	45
Test 1	10	2	35
Test 2	10	2	25
Test 3	10	2	15
Test 4	10	1	45
Test 5	10	1	35
Test 6	10	1	25
Test 7	5	2	35
Test 8	5	2	15
Test 9	5	2	15
Test 10	2	2	35
Test 11	2	2	25
Test 12	5	2	50
Test 13	5	2	5
Test 14	5	1	40
Test 15	5	1	15
Test 16*	5	3	80
Test 17	**5**	**3**	**40**
Test 18	2	2	50
Test 19	2	2	80
Test 20	2	3	45
Test 21	2	3	80
Test 22	5	3	60
Test 23	5	4	30
Test 24	5	4	55
Test 25	5	4	70
Test 26	5	5	40

Samples added to the project

We also received 31 raw reads of whole genome sequencing of Gracilariaceae species that were made by Dr. Goia Lyra during her pos-doc at Harvard University Herbaria, Cambridge, Massachusetts (Table 6). We assembled and annotated these data and used them to improve the phylogenomics analysis. Part of these samples was from her personal collection and the other was from S. Fredericq's lab collection.

Table 6 – Gracilariaceae species sequenced by Dr. G. Lyra and assembled in this work.

Species	Collection information	Voucher
Gracilaria baiana G.M. Lyra, C.F.D. Gurgel, M.C. Oliveira & J.M.C. Nunes	Penha, Vera Cruz, BA, Brazil; coll. G.M. Lyra, J.M.C. Nunes, 16 Oct. 2012	ALCB103804
Gracilaria birdiae E.M.Plastino & E.C.Oliveira	-	-
Gracilaria caudata J. Agardh	Villas do Atlântico, Lauro de Freitas, BA, Brazil; coll. G.M. Lyra, J.M.C. Nunes, J.C.G.Matos, 01 Nov. 2013	ALCB103876
Gracilaria cearensis (A.B.Joly & Pinheiro) A.B.Joly & Pinheiro in Pinheiro & A.B.Joly	Villas do Atlântico, Lauro de Freitas, BA, Brazil; coll. G.M. Lyra, J.M.C. Nunes, J.C.G.Matos, 01 Nov. 2013	ALCB103875
Gracilaria cervicornis (Turner) J. Agardh	Gravatá, Ilhéus, BA, Brazil; coll. G.M. Lyra, 17 Apr. 2011	ALCB65155
Gracilaria changii (B.M.Xia & I.A.Abbott) I.A.Abbott, J.Zhang & B.M.Xia changii	Zamboanga City, in front of Zamboanga State College, Phillipines; 28 Apr. 1998	FG54
Gracilaria chilensis C.J.Bird, McLachlan & E.C.Oliveira	Playa Charya, Coquimbo, Chile, 19 Jan. 1995	FG197
Gracilaria cornea J. Agardh	Penha, Vera Cruz, BA, Brazil; coll. G.M. Lyra, J.M.C. Nunes, 16 Oct. 2012	ALCB103802
Gracilaria crassissima (P.Crouan & H.Crouan) P.Crouan & H.Crouan	Belize	-
Gracilaria domingensis Sonder ex Kützing	Arembepe, Camaçari, BA, Brazil; coll. G.M. Lyra, J.M.C. Nunes, 23 Sept. 2010	ALCB103152
Gracilaria edulis (S.G.Gmelin) P.C.Silva	Little Santa Cruz, Phillipines, 28 Apr. 1998	FG82
Gracilaria ferox J.Agardh	Mucugê, Arraial D'Ajuda, Porto Seguro, BA, Brazil; coll. G.M. Lyra, J.M.C. Nunes, 06 Nov. 2011	ALCB65143
Gracilaria flabelliformis subsp. *simplex* Gurgel, Fredericq & J.N.Norris	Praia do Forte, Mata de São João, BA, Brazil; coll. G.M. Lyra, J.M.C. Nunes, 25 Sept. 2010	ALCB65135
Gracilaria isabellana Gurgel, Fredericq & J.N.Norris	Praia do Forte, Mata de São João, BA, Brazil; coll. G.M. Lyra, J.M.C. Nunes, 25 Sept. 2010	ALCB60808
Gracilaria multipartita (Clemente) Harvey	Carantec, Brittany, France	-
Gracilaria ornata Areschoug	Apuã, Santa Cruz de Cabrália, BA, Brazil; coll. G.M. Lyra, J.M.C. Nunes, 05 Nov. 2010	ALCB99680
Gracilaria rangiferina (Kützing) Piccone	Apuã, Santa Cruz de Cabrália, BA, Brazil; coll. G.M. Lyra, J.M.C. Nunes, 05 Nov. 2010	ALCB99703
Gracilaria salicornia (C.Agardh) E.Y.Dawson	-	-
Gracilaria tikvahiae McLachlan	-	-
Gracilaria urvillei (Montagne) I.A.Abbott	Lee Point, Darwin, Australia	FG122
Gracilaria usneoides (C.Agardh) J.Agardh	Champoton Area, Campeche Bay, Gulf of Mexico, Mexico; 14 Feb. 1999	-
Gracilaria vermiculophylla (Ohmi) Papenfuss	Virginia, Hog Island Bay, Eastern Shore, USA, 02 Out. 1999	FG107
Gracilariopsis chorda (Holmes) Ohmi	-	-
Gracilariopsis longissima (S.G.Gmelin) Steentoft, L.M.Irvine & Farnham	Venetian Lagoon, Adriatic Sea, Italy, 9 Jul. 1998	FG98
Gracilariopsis tenuifrons (C.J.Bird & E.C.Oliveira) Fredericq & Hommersand	Villas do Atlântico, Lauro de Freitas, BA, Brazil; coll. G.M. Lyra, J.M.C. Nunes, J.C.G.Matos, 01 Nov. 2013	ALCB103878
Melanthalia intermedia Harvey	Warrnambool, Australia, 13 Jul. 1995	FG46

Organellar genome assembly and annotation

Raw Illumina read quality was analyzed in the FastQC program (Andrews 2010). We used Trimmomatic v0.036 (Bolger et al. 2014) to clean and trim low-quality reads and bases. Trimmomatic parameters changed depending on the quality of the raw reads. We assembled the genomes with SPAdes v3.11.0 (Bankevich et al. 2012) setting different k-mer sizes (22, 33, 55, 77, 99, and 127). Mitochondrial and chloroplast assembled contigs were identified using local BLASTn (Altschul et al. 1990) searches against a custom-built database containing published organellar genomes from Gracilariaceae (Table 2). These blasted contigs were imported to Geneious 9.1.8 (Biomatters) and original reads were mapped using the Bowtie 2 (Langmead and Salzberg 2012) plugin to confirm the assembly. When the assembly was not good, i.e. the genome was fragmented in several small contigs, we normalized the raw reads to 100x coverage using BBnorm (Bushnell 2014) for the DNA libraries for which the coverage was too high and repeated the assembly with SPAdes. When the coverage depth is too high the quality of the assembly starts degrading (Lonardi et al. 2015). Circularity was verified by changing the start and end position and remapping the original reads.

For the genomes sequenced in this project and *G. vermiculophylla* and *Melanthalia intermedia* from Dr. G. Lyra database, we did a careful annotation for both mitochondrial and chloroplast genomes. These genomes were annotated using MFannot (http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl) to find coding sequences (CDS). To find RNA and intron sequences, we used ARAGORN (http://130.235.46.10/ARAGORN/) and RNAweasel (http://megasun.bch.umontreal.ca/cgi-bin/RNAweasel/RNAweaselInterface.pl). We used the "The Mold, Protozoan, and Coelenterate Mitochondrial Code and the Mycoplasma/Spiroplasma Code" (Genetic code table 4) (Pritchard et al. 1990; Boyen et al. 1994) to annotate the mitochondrial genomes and "The Bacterial, Archeal, and Plant Plastid Code" (Genetic code table 11) to annotate the chloroplast genomes (https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi). Large and Small ribosomal RNA were checked using the SILVA rRNA database (Quast et al. 2013). We also performed manual inspections and corrections looking for open-reading frames (ORFs) using the ORF finder plugin available in Geneious 9.1.8. Doubtful ORFs were verified using BLASTx. We reannotated the organellar genomes obtained from GenBank when genes were missing or differed from the same organellar genomes from different species. These revised annotations of published genomes are found in Supplementary material (Table S1, Table S2). For the other assembled genomes from the Dr. G. Lyra raw data, we just annotated the CDS because we used these data only for the phylogenetic analysis (see bellow).

Plasmid genomes assembly and annotation

We searched for plasmid sequences in the assembled contigs from the sequenced species in HiSeq 2500 and *Gracilaria vermiculophylla* and *Melanthalia intermedia* from Dr. G. Lyra dataset using BLASTn (-task blastn) with published Rhodophyta plasmids (Table 3) as query. The resulting contigs from BLASTn were imported to Geneious 9.1.8 and original reads were mapped using the

Bowtie 2 plugin to confirm the assembly. Circularity was verified by changing the start and end positions and remapping the original reads. The ORFs were identified using the ORF finder plugin and verified using BLASTx on the web. We also compared all plasmid homologous genes of known Gracilariaceae plasmids.

Phylogenetic analyses

We used two approaches for the phylogenetic analyses: first, we used phylogenetic trees as references for inferring the evolution of Gracilariaceae organellar genome architecture using a low number of terminals (taxa); and second, we used a larger number of terminals to obtain robust phylogenetic trees to address the main questions concerning the systematics of Gracilariaceae. For the smaller phylogenetic analysis, we selected nine newly sequenced species to a comparison of the genomic architectures. We used seven species sequenced in this project: *Gracilaria caudata*, *G. ferox*, *G. gracilis*, *G. rangiferina*, *G. tenuistipitata*, *Gracilariopsis longissima*, and *Gp. mclachlanii*, and, also *G. vermiculophylla* and *Melanthalia intermedia* from Dr. G. Lyra database. The published genomes from GenBank (Table 2), were also used in the organellar genomic comparison analysis, excluding *G. chilensis* (MF401962), *G. chouae* (MF351970), as both were published only after all the analyses were already done, and *Gp. lemaneiformis* (KP330491), which was mostly identical to the sequence for the same species. Therefore, we had 18 and 14 samples for mitochondria and chloroplast matrices, respectively. For each organelle, we constructed concatenated sequence matrix using amino acid (aa) data. For the bigger phylogenetic analysis, we included all the samples sequenced by Dr. G. Lyra (Table 6) in the previous matrices. Therefore, we had 43 and 39 samples for mitochondria and chloroplast matrices, respectively.

For all phylogenetic analysis, we used 23 CDS for the mitochondrial matrix and 187 CDS for the chloroplast matrix (Table 7). For all phylogenetic reconstructions, we used *Chondrus crispus*, *Gelidium elegans* Kützing, *G. vagum* Okamura and *Rhodymenia pseudopalmata* (J.V.Lamouroux) P.C.Silva as outgroups (Leblanc et al. 1995; Janouškovec et al. 2013; Kim et al. 2014b; Yang et al. 2014a, 2015, Lee et al. 2016a, b). We first constructed the aa matrix for each gene and, then, they were aligned using MAFFT version 7 (Katoh and Standley 2013). We concatenated the supermatrix for each organelle with FASconCAT v1.02 (Kück and Meusemann 2010), which also provides the information of each partition gene. We performed Maximum likelihood (ML) phylogenetic trees using IQ-TREE 1.5.4 (Nguyen et al. 2015) with 1,000 replicates of ultrafast bootstrap (Minh et al. 2013). For the mitochondrial matrix, we performed partitioned analyses (Chernomor et al. 2016) using PartitionFinder (Lanfear et al. 2012) available in IQ-TREE 1.5.4 (-m TESTMERGE) to select a model for each gene for the aa supermatrix. However, due the time of computer analysis, the ML analysis for chloroplast matrix was done using cpREV model (Adachi et al. 2000) plus discrete gamma model (Yang 1994) with default 4 rate categories (-m cpREV+G).

Table 7 - List of Gracilariaceae protein-coding genes used in the ML tree search.

Mitochondrion (23 genes)	*atp*4; *atp*6; *atp*8; *atp*9; *cob*; *cox*1; *cox*2; *cox*3; *nad*1; *nad*2; *nad*3; *nad*4; *nad*4L; *nad*5; *nad*6; *rpl*16; *rps*11; *rps*12; *rps*3; *sdh*B; *sdh*C; *sdh*D; *tat*C.
Chloroplast (187 genes)	*acc*A; *acc*B; *acc*D; *acp*P; *apc*A; *apc*B; *apc*D; *apc*E; *apc*F; *arg*B; *atp*A; *atp*B; *atp*D; *atp*E; *atp*F; *atp*G; *atp*H; *atp*I; *bas*1; *car*A; *cbb*X; *ccs*1; *ccs*A; *cem*A; *chl*I; *clp*C; *cpc*A; *cpc*B; *cpc*G; *cpe*A; *cpe*B; *dfr*; *dna*B; *dna*K; *dsb*D; *fab*H; *ftr*B; *fts*H; *glt*B; *gro*EL; *ilv*B; *ilv*H; *inf*B; *inf*C; *moe*B; *nbl*A; *ntc*A; *odp*A; *odp*B; *omp*R; *pbs*A; *pet*A; *pet*B; *pet*D; *pet*F; *pet*G; *pet*J; *pet*L; *pet*M; *pet*N; *pgm*A; *pre*A; *psa*A; *psa*B; *psa*C; *psa*D; *psa*E; *psa*F; *psa*I; *psa*J; *psa*K; *psa*L; *psa*M; *psb*30; *psb*A; *psb*B; *psb*C; *psb*D; *psb*E; *psb*F; *psb*H; *psb*I; *psb*J; *psb*K; *psb*L; *psb*T; *psb*V; *psb*W; *psb*X; *psb*Y; *psb*Z; *rbc*L; *rbc*R; *rbc*S; *rne*; *rnz*; *rpl*11; *rpl*12; *rpl*13; *rpl*14; *rpl*16; *rpl*18; *rpl*19; *rpl*1; *rpl*20; *rpl*21; *rpl*22; *rpl*23; *rpl*24; *rpl*27; *rpl*28; *rpl*29; *rpl*2; *rpl*31; *rpl*32; *rpl*33; *rpl*34; *rpl*35; *rpl*36; *rpl*3; *rpl*4; *rpl*5; *rpl*6; *rpl*9; *rpo*A; *rpo*B; *rpo*C1; *rpo*C2; *rps*10; *rps*11; *rps*12; *rps*13; *rps*14; *rps*16; *rps*17; *rps*18; *rps*19; *rps*1; *rps*20; *rps*2; *rps*3; *rps*4; *rps*5; *rps*6; *rps*7; *rps*8; *rps*9; *sec*A; *sec*G; *sec*Y; *suf*B; *suf*C; *syf*B; *syh*; *tat*C; *thi*G; *thi*S; *trp*A; *trp*G; *trx*A; *tsf*; *tuf*A; *ycf*17; *ycf*19; *ycf*20; *ycf*21; *ycf*22; *ycf*29; *ycf*33; *ycf*34; *ycf*35; *ycf*36; *ycf*37; *ycf*38; *ycf*39; *ycf*3; *ycf*41; *ycf*45; *ycf*46; *ycf*4; *ycf*52; *ycf*53; *ycf*54; *ycf*55; *ycf*58; *ycf*60; *ycf*63; *ycf*65

Organellar genomic comparison

We linearized the circular genomes and we standardized the start position to perform the synteny comparison. Mitochondrial genomes were linearized at the Asparagine tRNA gene (*trn*N). Chloroplast genomes were linearized at the small subunit ribosomal RNA gene (rrs). The start positions were chosen because they are the beginning of large conserved regions into the genomes. Both organellar genomes matrices were aligned using a progressive Mauve algorithm (Darling et al. 2010) in Geneious 9.1.8 applying the full alignment option, automatically calculated seed weights and automatically calculated minimum locally collinear blocks (LCB) to score and compare genetic synteny between species within genomes.

We searched for plasmid-derived sequences (PDSs) in all complete organellar genomes, including the genomes downloaded from Genbank. For this, we used 10 published red algae plasmids available in Genbank (Table 3). The PDSs were searched using tBLASTx with 52 plasmid CDS queries against the database made by Gracilariaceae organellar genomes. We also used BLASTn with complete plasmids sequences against genomes database (-task blastn) to investigated non-coding plasmid derived sequences. For both searches, we considered only hits with e-value less than 1.0e-20.

*leu*C and *leu*D genes

The chloroplast genomes of *Gracilaria tenuistipitata* and *G. chilensis* have the *leu* operon with two genes, *leu*C and *leu*D (Hagopian et al. 2004; Lee et al. 2016b) that probably were horizontally transferred from a proteobacteria to the chloroplast genome (Janouškovec et al. 2013). However,

homologous of these genes were found in other red algae nucleus genomes, as in *Gracilaria changii* and *Chondrus crispus* (Janouškovec et al. 2013; Collén et al. 2013). Therefore, we searched for *leu*C and *leu*D genes in the assembled contigs in all samples sequenced in this study. We used as query the *G. tenuistipitata* and G. *chilensis* chloroplast-encoded genes and *Chondrus crispus* nucleus-encoded genes (*leu*C = XP_005715507; *leu*D = XP_005718118). Other *leu*C and *leu*D sequences were obtained from GenBank using closest homologs from BLASTp homology searches excluding samples from environment samples and multispecies. Protein sequences were aligned using MAFFT version 7 as implemented in Geneious 9.1.8. Maximum likelihood inferences were done in IQ-TREE 1.5.4 using '-m TEST' parameter to choose the best model with ProtTest and 1,000 replicates of ultrafast bootstrap.

Results and Discussion

DNA library

Several DNA extraction tests were done to arrive at an ideal amount of fresh mass that would result in the desired amount of total DNA. The TruSeq DNA PCR-Free LT Library preparation protocol indicated that the initial amount of total DNA should be 2 μg of double-stranded DNA, diluted at a concentration of 40 ng / μL. The NanoDrop™ measures the amount of total nucleic acid, including RNA and free nucleic acid molecules. Therefore, we should use Qubit® to obtain the double-stranded DNA quantity (Figure 1). However, NanoDrop™ was much cheaper than Qubit® thus we did tests to estimate the percentage of the amount of double-stranded DNA using NanoDrop™. Besides this, NanoDrop™ measures were important to estimate the DNA quality with using the ratios of absorbance 260/280 and 260/230. With these tests, we observed that the amount of double-stranded DNA measured by NanoDrop™ was about 20-40% of the total nucleic acid concentration (Table 8). Therefore, for each extraction, it was expected to obtain at least 6 μg of total nucleic acids extracted indicated by NanoDrop™.

Table 8 - Quantity of total fresh mass used for DNA extraction. Concentration of DNA extracted, eluted in MilliQ water, in both NanoDrop™ (total nucleic acid) and Qubit® (double-stranded DNA). NanoDrop™ 260/280 and 260/230 ratios of absorbance values. Concentration of DNA libraries measured in Qubit® and RT-PCR.

	Fresh mass (mg)	DNA extraction				DNA library		
		NanoDrop (ng/μL)	NanoDrop ratios of absorbance		Qubit (ng/μL)	Qubit (ng/μL)	Qubit (nM)	RT-PCR (nM)
Species			260/280	260/230				
Gracilaria caudata	410	167.6	2.1	1.93	55.6	8.32	18.82	1.64
Gracilaria cornea	220	159.6	2.02	1.6	58.0	7.44	16.82	0.83
Gracilaria ferox	240	168.4	2.0	1.9	63.4	8.84	19.99	1.4
Gracilaria gracilis	230	193.5	2.02	1.95	70.6	7.54	17.05	1.48
Gracilaria isabellana	230	152.6	2.02	1.92	59.6	7.62	17.23	0.8
Gracilaria tenuistipitata	240	166.0	2.02	2.16	52.2	9.0	20.35	4.69
Gracilariopsis longissima	320	241.0	1.98	1.25	82.0	7.38	16.69	1.76
Gracilaria rangiferina	520	161.7	2.05	1.95	67.4	7.68	17.37	2.49
Gracilariopsis mclachlanii	350	221.7	2.12	2.16	65.6	6.34	14.34	1.66
Gracilariopsis tenuifrons	190	175.0	1.99	2.06	63.6	8.72	19.72	0.68

It was not possible to define an ideal amount of initial fresh mass for DNA extraction for all species, which would result in the desired final amount of extracted DNA. In some species, a larger quantity of fresh mass was required, as observed in *Gracilaria rangiferina* (Table 8). We also observed that the DNA was best extracted when macerated in liquid nitrogen with a mortar and pestle to form a very fine powder. We could not correlate the amount of initial biomass with the thallus morphology as *Gracilaria rangiferina*, *G. ferox* and *G. isabellana* are flat species and the others are cylindrical.

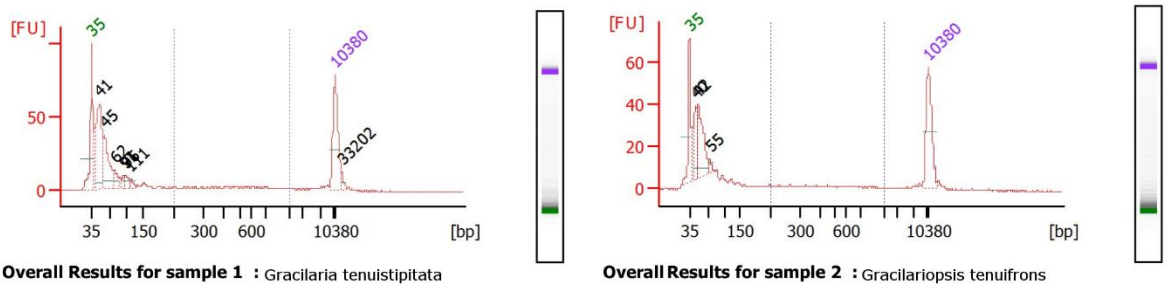

Figure 2 – Bioanalyzer result of first test of DNA fragmentation. Y axis, fluorescence (FU) and X axis, length of the fragments (bp).

Before preparing the DNA library, we had to fragment the extracted DNA with Covaris S2 in fragments around 550 bp. We did a first fragmentation test using *Gracilaria tenuistipitata* and *Gp. tenuifrons* DNA extractions with Covaris sonicator regulated according to the protocol of the library preparation kit itself (Duty Factor 10%, Intensity 2 and duration 45 seconds). After the sonication, we measured the fragmented DNA with Bioanalyzer to check the size of the DNA fragments. This step was not foreseen in the protocol of the kit, but it was suggested by the technichal support of the Illumina representative. The result of the first test in the Bioanalyzer indicated that the fragments were very short and there were no fragments around 550 bp. The peaks were around 41 bp for *G. tenuistipitata* and 42 for *Gp. tenuifrons* (Figure 2). Based on these results, several sonication tests were performed, changing the parameters of Covaris S2 (Table 5). We used only *Gracilaria tenuistipitata* to make these sonication tests because we had more biomass in the culture.

Most of the tests maintained the pattern of the first sonication (Figure 2). The best results for the fragmentation are shown in Figure 3, although none of them had the expected result, which would be a curve with the peak at 550 bp. Among the fragmentation patterns found, the parameters of test 17 Table 5 were chosen to make the first batch of DNA libraries, since it was the one that approached the average size at 550 bp.

After the DNA library preparation, we did the RT-PCR to ensure the exact quantification of the fragments with the adapters at both ends (Figure 1). This step was necessary because Qubit® measures only the total amount of double-stranded DNA, regardless of whether the fragments are bound with the adapters. The RT-PCR results showed that the adapters did not bind to most of the fragments. For example, in *Gracilariopsis tenuifrons*, only 3,44% of the total double-stranded DNA quantity were fragments with adapters (Table 8). The best result was in *G. tenuistipitata*, for which 23,04% of the total double-stranded DNA were fragments with adapters (Table 8). This result may have occurred because the adapters did not properly connect to both ends of the fragments. The lack of adapters in one or both ends of the fragment makes it impossible to sequence it.

Illumina's sequencing protocol recommends that the DNA library should be at least 2nM. However, this concentration was obtained in only two samples: *Gracilaria tenuistipitata* and *G. rangiferina* (Table 8). Therefore, for samples with a concentration of less than 1 nM (*Gp. tenuifrons*, *G. cornea*, and *G. isabellana*) were excluded from the HTS run, because low concentration could

disturb the clustering step before the sequencing run. Samples with concentrations greater than 2 nM were placed in duplicates to generate the largest possible clustering.

We chosed TruSeq DNA PCR-Free LT Library preparation kit because it does not have the PCR step during the preparation of the DNA library. PCR step may generate bias in GC content in the total amount of reads because GC-neutral fragments are amplified more efficiently than fragments with high GC or high AT content (Kozarewa et al. 2009; Dijk et al. 2014b). However, because the absence of the PCR step during the DNA library, the initial amount of extracted DNA must be very high (2 µg of double-stranded DNA) and adaptor ligation step was not fully successful, which resulted in low DNA library quantity for the sequencing. DNA library protocols with amplification step demand lower DNA quantity, because only fragments with adaptors in both ends will be amplified.

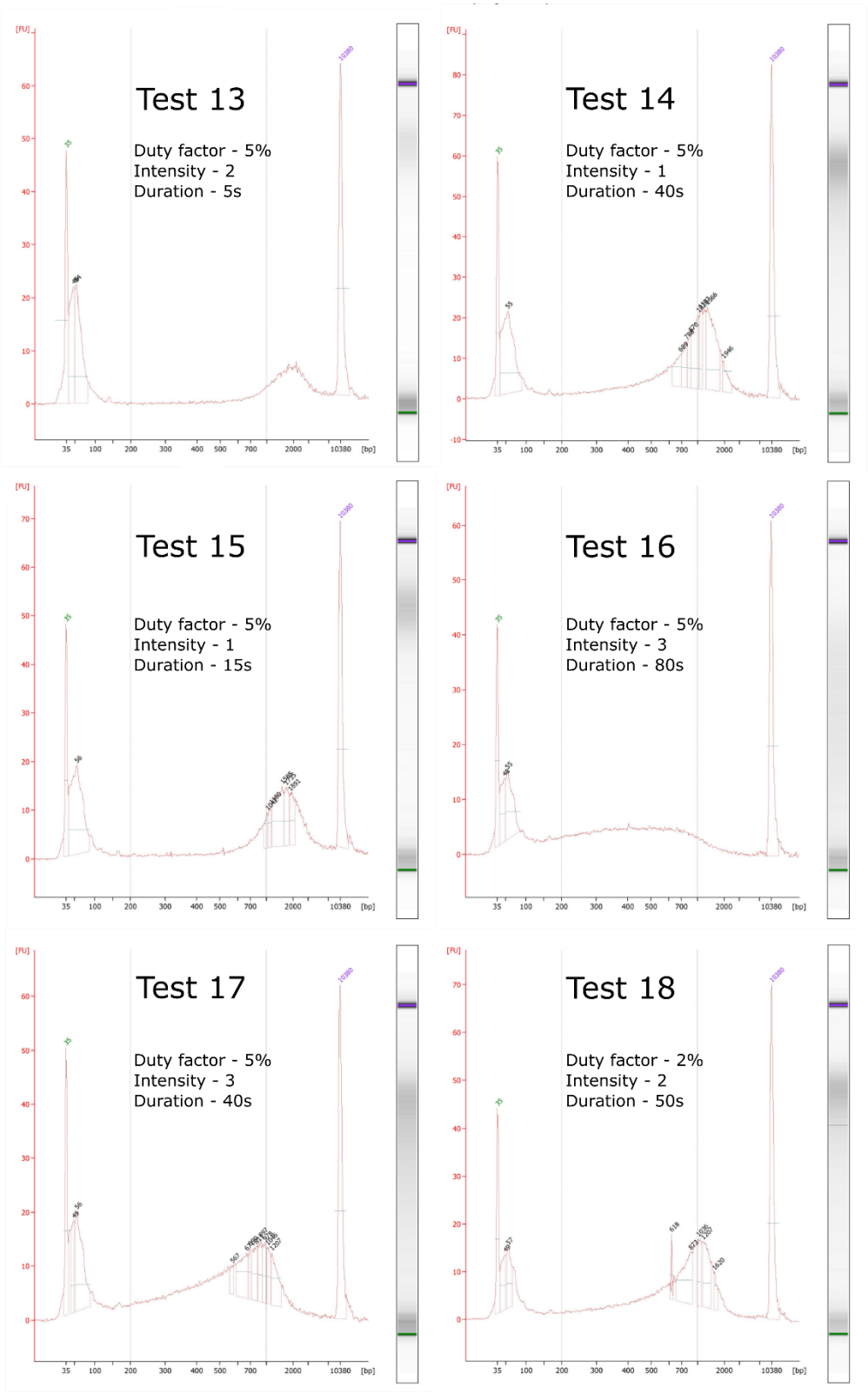

Figure 3 - The best Bioanalyzer results in fragmentation tests.

High-throughput Sequencing output and genome assemblies

Quantity and quality of samples sequenced in HiSeq2500

The HTS with Illumina HiSeq 2500 was successful despite the intial problems with fragmentation and DNA library construction mentioned above. The output capacity in the Rapid Run mode was up to 150 Gb and the output obtained was c.a. 147 Gb, which means that the run was 98,6% successful. The pool of mixed samples was sequenced in two different lanes. The number of reads sequenced in each lane is shown in Table 9. The expected coverage for each sample was higher in all sequenced species, because we excluded three species from the sequencing, especially in *Gracilaria rangiferina* and *G. tenuistipitata*, for which had a higher output because double amount of libray was included in the pool (Table 9).

We used only the read libraries sequenced in lane 1 because it was too much data to assemble the genomes using both lanes. Furthermore, our first genome assembly attempts indicated that the coverage in the organellar genomes was extraordinarily high coverage (>5,000x), which were making the organellar contigs smaller. It is known that the quality of the assembly starts degrading when the depth is too high, possibility, because of the presence of chimeric reads, sequencing error, and read duplications, or their combination thereof (Lonardi et al. 2015). Beside this, assembly of genomes with higher depth than needed is waste of computational resources (Desai et al. 2013). Therefore, we had to normalize these read libraries to 100x depth of coverage using BBnorm (Table 9). BBNorm normalizes the coverage by down-sampling reads over high-depth areas of a genome, to result in a flat coverage distribution. This process can accelerate assembly and improve assembly quality. The quantity of survivor reads after normalization of the read libraries from lane 1 was between 35.1% (*G. rangiferina*) and 67.4% (*Gp. mclachlanii*) (Table 9). We cleaned and trimmed low-quality reads and bases using Trimmomatic v0.036 and the quantity of dropped reads (excluded low quality reads) vary from 2.6% (*G. rangiferina*) to 34.3% (*Gp. mclachlanii*) (Table 9). We were expecting that c.a. 15% of the reads could be dropped during the cleaning and trimming processes. However, the larger quantity of dropped reads did not damage the assemblies.

Table 9 - High-throughput sequencing output information. Expected coverage calculated using all read output. Number of reads normalized to 100x coverage using only reads sequenced in lane 1. Percentage of normalized reads. Reads dropped in Trimmomatic v0.036 using normalized reads.

				Normalized reads (Lane 1)		
Species	Lane 1 (pair reads)	Lane 2 (pair reads)	Expected coverage	Number of total reads	Percentage of normalized read	Dropped reads (percentage)
Gracilaria caudata	16,248,938	15,960,711	77x	10,417,435	64.1%	4,562,195 (21.9%)
Gracilaria ferox	17,113,990	16,842,532	81x	9,535,124	55.7%	2,484,244 (13.0%)
Gracilaria gracilis	18,139,618	17,873,255	86x	12,196,596	67.2%	6,262,534 (25.7%)
Gracilaria rangiferina	30,873,154	30,392,254	146x	10,827,745	35.1%	567,985 (2.6%)
Gracilaria tenuistipitata	32,612,114	32,090,134	154x	14,296,493	43.8%	7,637,469 (26.7%)
Gracilariopsis longissima	18,422,330	18,130,792	87x	11,245,811	61.0%	625,310 (2.8%)
Gracilariopsis mclachlanii	15,699,262	15,443,254	74x	10,580,643	67.4%	7,262,670 (34.3%)

Quantity and quality of samples from Dr. G. Lyra database

We analyzed the 31 raw reads from Dr. G. Lyra database (Table 10). For all these species, we assembled the genomes with SPAdes v.3.11.0 without normalizing the reads. However, we could not obtain contigs larger enough to complete the organellar genomes for thirteen of the samples. Therefore, we normalized the reads for these samples and repeated the assembly (Table 10).

Most of these species were used only for phylogenomic analysis but we select *Gracilaria vermiculophylla* and *Melanthalia intermedia* in the genomic comparison analysis for full organellar genomes assembly because these species had interesting characteristics.

Table 10 – HTS output information of samples from Dr. G. Lyra database. Part of the sequenced samples was not normalized and the trimming and cleaning processes were done with total sequencing output.

Species	Read output (pair-end)	BBnorm	Normalized reads (pair-end)	Percentage of normalized reads	Total of dropped reads	Percentage of dropped reads
Gracilaria baiana	32,322,300	No	-	-	1,055,652	1.63%
Gracilaria birdiae	10,303,970	Yes	5,665,628	55.0%	2,411,602	21.28%
Gracilaria caudata	7,081,786	No	-	-	175,202	1.24%
Gracilaria cearensis	15,680,216	Yes	8,312,336	53.0%	2,264,870	13.62%
Gracilaria cervicornis	55,262,282	No	-	-	16,055,635	14.53%
Gracilaria changii	19,301,298	No	-	-	8,977,282	23.26%
Gracilaria chilensis	21,561,177	No	-	-	7,597,402	17.62%
Gracilaria cornea	5,869,324	No	-	-	3,904,021	33.26%
Gracilaria crassissima	258,815,297	Yes	52,587,154	20.3%	25,654,399	24.39%
Gracilaria domingensis	3,287,600	No	-	-	1,860,342	28.29%
Gracilaria edulis	78,038,244	No	-	-	45,920,019	29.42%
Gracilaria ferox	7,454,657	No	-	-	4,162,805	27.92%
Gracilaria flabelliformis subsp. *simplex*	61,011,636	Yes	32,173,868	52.7%	7,696,449	6.31%
Gracilaria isabellana	5,577,150	No	-	-	3,524,119	31.59%
Gracilaria multipartita	78,331,206	Yes	35,642,521	45.5%	14,542,259	20.40%
Gracilaria ornata	11,013,600	Yes	5,515,746	50.1%	2,200,772	19.95%
Gracilaria rangiferina	38,495,845	Yes	11,402,171	29.6%	5,418,413	23.76%
Gracilaria salicornia	30,809,106	Yes	18,811,220	61.1%	6,722,214	17.87%
Gracilaria tikvahiae	9,764,154	No	-	-	3,629,543	18.59%
Gracilaria urvillei	41,034,365	Yes	25,852,824	63.0%	10,057,915	19.45%
Gracilaria usneoides	14,340,190	Yes	7,445,779	51.9%	3,160,985	21.23%
Gracilaria vermiculophylla	8,874,563	Yes	4,065,742	45.8%	176,430	2.17%
Gracilariopsis chorda	6,174,430	No	-	-	3,417,462	27.67%
Gracilariopsis longissima	23,036,959	No	-	-	11,166,945	24.24%
Gracilariopsis tenuifrons	6,070,882	Yes	2,811,691	46.3%	1,160,327	20.63%
Melanthalia intermedia	34,663,104	Yes	23,320,833	67.3%	7,328,105	15.71%

Genome assemblies

For all samples sequenced in HiSeq 2500, we completed the assembly for both organellar genomes without gaps. For *G. tenuistipitata*, we completed only the mitochondrial genome because the chloroplast genome had been previously published for the same strain (Hagopian et al. 2004). We did a careful assembly, circularization and annotation of all genomes sequenced in HiSeq 2500, plus *Gracilaria vermiculophylla* and *Melanthalia intermedia* from Dr. G. Lyra database (Table 11). Each genome was completely assembled in one contig and circularized mapping the original reads against the genome contig. The mean depth of coverage of the genomes varies from 58,97x to 318.9x using normalized reads (Table 11).

Table 11 - General information of the assembly and structure of mitochondrial and chloroplast genomes of samples sequenced in HiSeq 2500, *Gracilaria vermiculophylla* and *Melanthalia intermedia* from G. Lyra database. Depth of coverage was done using normalized reads. Std – standard deviation.

Mitochondrion

Species	Accession number	Size (Kb)	GC%	Depth of coverage		Protein	rRNA	tRNA	Other RNA	Intron
				Mean	Std					
Gracilaria caudata	MH396017	26,030	28.8	111.11	18.01	25	3	24	-	1
Gracilaria ferox	MH396018	25,588	27.4	60.87	9.91	25	3	23	-	1
Gracilaria gracilis	MH396019	25,865	29.2	76.67	12.8	25	3	24	-	1
Gracilaria rangiferina	MH396020	25,908	25.4	131.35	18.97	25	3	24	-	1
Gracilaria tenuistipitata	MH396021	25,880	27.1	130.46	17.6	25	3	24	-	1
Gracilaria vermiculophylla	MH396022	26,149	28.4	154.8	13.6	25	3	24	-	1
Gracilariopsis longissima	MH396023	26,744	28.4	107.86	12.04	26	3	24	-	1
Gracilariopsis mclachlanii	MH396024	26,012	28.5	58.97	16.83	25	3	24	-	1
Melanthalia intermedia	MH396025	25,348	29.3	167.2	19.8	25	3	20	-	1

Chloroplast

Species	Accession number	Size (Kb)	GC%	Depth of coverage		Protein	rRNA	tRNA	Other RNA	Intron
				Mean	Std					
Gracilaria caudata	MH396009	182,933	28.8	114.79	11.97	207	3	30	2	1
Gracilaria ferox	MH396010	180,255	28.8	60.78	8.02	205	3	30	2	1
Gracilaria gracilis	MH396011	180,807	29.1	85.43	13.71	207	3	30	2	1
Gracilaria rangiferina	MH396012	178,841	27.7	130.54	13.3	203	3	30	2	1
Gracilaria vermiculophylla	MH396013	180,254	29.5	163.8	13.1	204	3	30	2	1
Gracilariopsis longissima	MH396014	181,896	27.3	110.61	9.92	204	3	30	2	1
Gracilariopsis mclachlanii	MH396015	184,980	27.8	60.56	13.61	205	3	30	2	1
Melanthalia intermedia	MH396016	218,416	31.8	318.9	116	224	3	30	2	1

For the genomes from Dr. G. Lyra database, we circularized all the chloroplast genomes, but we could only circularize eleven mitochondrial genomes (Table 12). The other thirteen mitochondrial genomes were complete in gene content, but we could not circularize those genomes due a repeat region. As we were just interested in the genes of these genomes, we did not see as essential to close these genomes.

Table 12 - General information of the assembly of mitochondrial and chloroplast genomes of samples from Dr. G. Lyra database.

	Mitochondrion				Chloroplast		
			Depth of coverage			Depth of coverage	
Species	Circular	Length	Mean	Std	Length	Mean	Std
Gracilaria baiana	No	25,553	2,257.2	846.5	179,484	157.9	14.8
Gracilaria birdiae	No	25,985	155.6	17.2	183,364	153.5	14.1
Gracilaria caudata	No	25,395	647	160.9	183,134	965.2	312.1
Gracilaria cearensis	Yes	26,373	149	16.2	182,079	149.4	12.6
Gracilaria cervicornis	Yes	25,714	251.2	113	180,132	943.5	127.4
Gracilaria changii	No	25,769	1,626.5	1,135.3	180,042	2,621.2	278.1
Gracilaria chilensis	Yes	25,746	1,280.2	240.7	185,684	1936	382.9
Gracilaria cornea	No	25,683	336.4	345	182,454	190.7	54.6
Gracilaria crassissima	No	25,379	950.1	118.7	181,978	948.8	111.3
Gracilaria domingensis	Yes	25,838	360.6	60.5	179,662	931.6	292.9
Gracilaria edulis	Yes	25,722	175.3	25	179,371	182.8	20.9
Gracilaria ferox	Yes	25,596	134.3	97	180,170	262.4	67.7
Gracilaria flabelliformis subsp. *simplex*	Yes	25,923	167.6	44.9	179,295	163.3	16.8
Gracilaria isabellana	No	25,674	378.5	174.1	194,277	456.8	104.6
Gracilaria multipartita	Yes	25,634	161.4	18.8	179,145	163.5	17.1
Gracilaria ornata	Yes	25,775	548.1	164.7	178,973	159.4	15.5
Gracilaria rangiferina	Yes	25844	169.6	29.8	178,841	466.7	117.9
Gracilaria salicornia	No	25,340	182.6	20.1	179,752	182.9	14.8
Gracilaria tikvahiae	Yes	25,761	265.4	68.3	190,895	560.6	81
Gracilaria urvillei	No	25,851	165.2	18.5	178,239	163	13.4
Gracilaria usneoides	No	25,239	158.3	20.9	183,376	157	14.4
Gracilariopsis chorda	No	26,804	723.4	134	182,537	1,842.5	334.6
Gracilariopsis longissima	No	26,120	788.3	108.8	181,896	1,665.5	195.7
Gracilariopsis tenuifrons	No	27,466	154.1	18	184,308	152.9	14.4

Organellar genome structure

All general characteristics of the organellar genomes of the species carefully assembled and annotated are described in (Table 11). The nine complete mitochondrial genomes sequenced in this study were similar in GC content and length. The GC content was 28.1% on average (min = 25.4% in *G. rangiferina,* max = 29.2% in *G. gracilis*). The length was 26,006 bp on average. *Gracilaria gracilis* had the shortest (25,865 bp) and *Gp. longissima* had the longest genome (26,744 bp). The complete mitochondrial genomes were also similar in gene content, including the mitochondrial genomes published in GenBank (Table 2, Table 11), comprising 25 protein-coding genes, three rRNAs (*rrl*, *rrs*, *rrn*5) and 20 to 26 tRNAs. The exceptions were *Gp. longissima*, *Gp. chorda* and *Gp. lemaneiformis* with 26 protein-coding genes (Table 2, Table 11). All mitochondrial genomes of Gracilariaceae had a Group II intron (ranging from 416 bp in *G. rangiferina* to 499 bp in *G. chilensis*) in the *trn*I tRNA gene. The genes of the nine complete mitochondrial genomes sequenced and the genomes from GenBank used in this study are listed in Table S3 in Supplementary material. Genome maps are shown in Figures S1 to Figure S9 in Supplementary material.

All *Gracilaria* and *Gracilariopsis* complete chloroplast genomes sequenced in this study were also similar in GC content and length. However, *M. intermedia* chloroplast genome is larger and had higher GC content (Table 11). The GC content was 28.8% on average (min = 27.3% in *Gp. longissima*, max = 31.8% in *M. intermedia*). The length was 186.048 bp on average. *Gracilaria rangiferina* had the shortest (178,841 bp) and *M. intermedia* the longest genome (218,416 bp). The complete chloroplast genomes of *Gracilaria* and *Gracilariopsis* species were also similar in gene content comprising of 203 to 207 protein genes, while *M. intermedia* chloroplast had 224 protein genes. All chloroplast genomes had three rRNA (*rrl*, *rrs*, *rrn*5), 30 tRNA, one tmRNA and one npDNA (*rnp*B) (Table 2, Table 11). All chloroplast genomes had a Group II intron in *trn*Me tRNA gene (ranging from 1,998 bp in *Gp. longissima* to 2,111 bp in *Gp. lemaneiformis*). We found the operon *leu*C and *leu*D genes in *M. intermedia* chloroplast. The genes of the eight chloroplast genomes sequenced and the genomes from GenBank used in this study are listed in Table S4 in Supplementary material. Genome maps are shown in Figures S10 to Figure S17 in Supplementary material.

Extrachromosomal plasmid from Gracilariaceae

From the assembled contigs data from all species sequenced in this study, we found three circular plasmid sequences in only two species: one in *G. ferox* (Gfe3115) and two in *G. vermiculophylla* (Gve4548 and Gve7464) (Figure 4; Table 13).

Table 13 - Information of the plasmid genomes found in *Gracilaria ferox* and *Gracilaria vermiculophylla*.

Plasmid	Accession number	Organism	Sequence length	GC%	Topology	Nº of ORFs
Gfe3115	MH396026	*Gracilaria ferox*	3,115	25.1	Circular	4
Gve4548	MH396027	*Gracilaria vermiculophylla*	4,548	27.5	Circular	5
Gve7464	MH396028	*Gracilaria vermiculophylla*	7,464	26.2	Circular	6

We compared all plasmid homologous genes of known Gracilariaceae plasmids. The homologous freestanding ORFs were classified in different categories called FSO (Table 14).

The Gfe3115 plasmid was 3,115 bp (82x average coverage with normalized reads to 100x) and contained four ORFs: FSO1, FSO2, FSO3 and ORF1, of which ORF1 had no similarity to other sequences in GenBank. The Gve4548 plasmid was 4,548 bp (157.7x average coverage) and had FSO1, FSO5, FSO6, FSO7and FSO8 homologous ORFs. This plasmid shares a strong similarity in ORF content and synteny with the *Gp. lemaneiformis* plasmid Gle4293 (Figure 4), which seems to be the most common plasmid type found in Gracilariaceae. The Gve7464 plasmid was 7,464 bp (154.2x average coverage) with six ORFs including *leu*C and *leu*D genes, FSO1, FSO2, FSO4 and ORF2, of which ORF2 had no similarity to other sequences in GenBank.

Table 14 – Homologous ORFs found in plasmids of Gracilariaceae species.

Freestanding ORF	Organism	Plasmid	ORF	Length (bp)
FSO1	*Gracilaria chilensis*	Gch3937	ORF1	1218
	Gracilaria chilensis	Gch7220	ORF1	813
	Gracilaria chilensis	Gch7220	ORF4	1218
	Gracilaria chilensis	Gch7220	ORF6	243
	Gracilaria chilensis	Gch7220	ORF7	153
	Gracilaria chilensis	GC2	Unique ORF	1236
	Gracilaria ferox	Gfe3115	ORF2	1254
	Gracilaria robusta	Gro4970	ORF1	1236
	Gracilaria robusta	Gro4059	ORF2347	573
	Gracilaria vermiculophylla	Gve7464	ORF1	1248
	Gracilaria vermiculophylla	Gve4548	ORF2	1242
	Gracilariopsis lemaneiformis	Gle4293	ORF1	1257
FSO2	*Gracilaria ferox*	Gfe3115	ORF3	456
	Gracilaria robusta	Gro4970	ORF5	585
	Gracilaria vermiculophylla	Gve7464	ORF4	534
FSO3	*Gracilaria chilensis*	Gch3937	ORF4	348
	Gracilaria ferox	Gfe3115	ORF4	351
	Gracilaria robusta	Gro4970	ORF6	405
FSO4	*Gracilaria chilensis*	Gch3937	ORF2	543
	Gracilaria chilensis	Gch7220	ORF5	594
	Gracilaria vermiculophylla	Gve7464	ORF2	204
FSO5	*Gracilaria vermiculophylla*	Gve4548	ORF1	519
	Gracilariopsis lemaneiformis	Gle4293	ORF2	495
FSO6	*Gracilaria vermiculophylla*	Gve4548	ORF3	600
	Gracilariopsis lemaneiformis	Gle4293	ORF5	564
FSO7	*Gracilaria vermiculophylla*	Gve4548	ORF4	687
	Gracilariopsis lemaneiformis	Gle4293	ORF4	543
FSO8	*Gracilaria vermiculophylla*	Gve4548	ORF5	804
	Gracilariopsis lemaneiformis	Gle4293	ORF3	789
FSO9	*Gracilaria robusta*	Gro4059	ORF1478	465
FSO10	*Gracilariopsis lemaneiformis*	GL3	ORF2	435
FSO11	*Gracilaria robusta*	Gro4970	ORF3	414
FSO12	*Gracilaria chilensis*	Gch7220	ORF2	213
	Gracilaria chilensis	Gch3937	ORF3	213
FSO13	*Gracilaria chilensis*	Gch7220	ORF8	348
	Gracilaria chilensis	Gch7220	ORF9	213
	Gracilaria chilensis	Gch7220	ORF10	159
	Gracilaria chilensis	Gch7220	ORF16	330
	Gracilariopsis lemaneiformis	GL3	ORF1	381

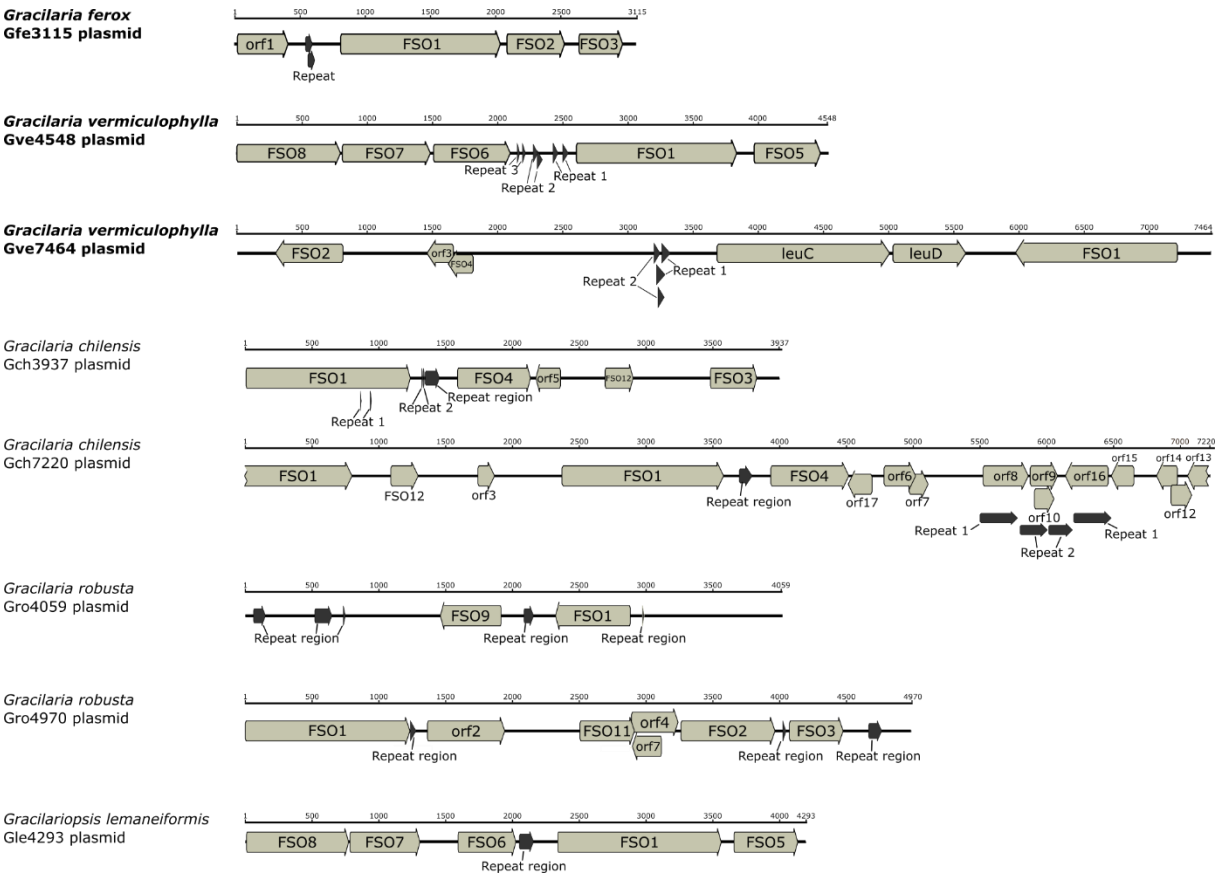

Figure 4 – Linear schematic representation of circular Gracilariaceae plasmids. Plasmids with bold names were sequenced in this study, others were obtained from GenBank. Light grey arrows are ORFs. Black arrows are repeated regions. Regions identified as 'Repeat 1, 2 and 3' are entirely repeated sequences in different positions of the genome. Regions identified as 'Repeat region' are repetitions in tandem. Scale line in base pairs.

The plasmid of *G. vermiculophylla* Gve7464, contains two genes of the proteobacterial operons related to leucine biosynthesis *leu*C and *leu*D genes (Figure 4) and this is the first record of these genes in a eukaryotic plasmid. The confirmation that this plasmid is not a contaminant bacterial plasmid is that it contains Rhodophyta exclusive plasmid ORFs: FSO1, FSO2, and FSO4 (Figure 4, Table 14). Other three species have these genes in their chloroplast genome: *Gracilaria tenuistipitata*, *G. chilensis* (Hagopian et al. 2004; Janouškovec et al. 2013; Lee et al. 2016b) and *Melanthalia intermedia*. This operon is potentially functional and it was probably originated from a protobacterial horizontal gene transfers (Janouškovec et al. 2013).

Genomic comparison and plasmid-derived sequences of mitochondria

We used the mitochondrial and chloroplast ML trees with lower number of species to support the discussion about the evolution of Gracilariaceae organellar genome architecture. The mitochondrial ML tree with lower samples was highly supported (Figure 5). Gracilariaceae is monophyletic, *M. intermedia* is sister to a monophyletic clade containing both monophyletic genera *Gracilaria* and *Gracilariopsis*. Mitochondrial tree presented lower support in some nodes, including

the position of *G. rangiferina* as a sister taxon to the other *Gracilaria* species in the tree (83%), *G. vermiculophylla* grouped with the *G. tenuistipitata/G. chilensis* clade (81%), *G. ferox* grouped with *G. salicornia* (84%), *Gp. mclachlanii* grouped with the *Gp. lemaneiformis/Gp. chorda* clade (81%), and between *Gp. andersonii* and *Gh. oryzoides* (64%).

Mitochondrial genomes are highly conserved especially in two main regions. The first one between *trn*N and *cob* genes (blue locally collinear blocks, LCB, in Figure 5A) and the other between *trn*L and *trn*A genes (purple LCB in Figure 5A). Regions that were specific for each genome were generally caused by plasmid-derived sequences (PDS) insertion. We investigated the occurrence of PDS in our newly generated and in previously published genomes. These insertions occurred mainly in two regions of the genome where the transcription direction changes (black line boxes in Figure 5A).

All PDS insertions found in the mitochondrial genomes (Figure 5B, C) were similar (e-value between 1.0e-24 to 1.0e-46) to the plasmid previously described for *Gracilaria robusta* Setchell in F.S. Collins, Holden & Setchell (1899) (Goff and Coleman 1988), identified as Gro4059 (Figure 4, Table 3). *Gracilariopsis andersonii* had the largest mitochondrial genome (Table 2) due to the ~2,000 bp PDS insertion between the genes *trn*H and *trn*L (Figure 5B). *Gracilariopsis longissima* has a PDS located between *cob* and *trn*H (Figure 5B). In *Gp. longissima*, one of the plasmid-derived CDS is homologous to FSO9 (Figure 5B, Table 14) that is also found in *G. robusta* Gro4059 plasmid (Figure 4). This gene was the only plasmid-derived CDS found in the mitochondrial genomes. The other region where PDSs were found was between *trn*A and *trn*N genes (Figure 5C). We did not find PDS insertions in the mitochondrial genomes of *G. vermiculophylla* from Asia, *G. changii*, *G. ferox, G. salicornia* from Hawaii, *Gh. oryzoides* and *M. intermedia*. All the PDS indels occurred between tRNA genes, with one exception (*Gp longissima* PDS between *cob* and *trn*H).

The gene *trn*H is in different positions in *Gracilaria* and *Gracilariopsis* and it is absent in *M. intermedia* (Figure 5B). In *Gracilariopsis*, it is found between *cob* and *trn*L. In *Gracilaria*, it is found between *trn*G and *sdh*B. This gene is in the other strand in *G. salicornia* from Hawaii when compared to the other genomes (Figure 5B). *Gracilariopsis chorda* and *Gp. lemaneiformis* had a hypothetical protein (HP) gene between *trn*G and *sdh*B (Figure 5B) that did not match with any known gene in Genbank. The region between *trn*A and *trn*N is highly variable (Figure 5A, C). For example, *Gracilaria rangiferina, G. caudata*, and *Gp. mclachlanii* had an inversion in the region including the genes *trn*Y, *trn*R, and *trn*S (Figure 5C). The *trn*Y gene is translocated in *G. vermiculophylla*, *G. chilensis* and *G. tenuistipitata* (Figure 5C). *Gracilaria chilensis* had a duplication and inversion of *trn*R and *trn*S genes between two PDS insertions. *Gracilaria salicornia* from Hawaii, *Gp. lemaneiformis, Gh. oryzoides, Gp. andersonii* and *M. intermedia* missed the *trn*Y, *trn*R, and *trn*S tRNA genes. *Gracilaria vermiculophylla* from Asia missed *trn*R and *trn*S tRNA genes while *G. ferox* missed *trn*Y gene. It is interesting that *G. salicornia* from Hawaii had only 38 bp between *trn*A and *trn*N while *G. salicornia* from Malaysia had 682 bp which contained three tRNA genes and a PDS

(Figure 5C). *Gracilaria vermiculophylla* from the USA had 806 bp between *trn*A and *trn*N, which contain two tRNA genes, while *G. vermiculophylla* from Asia had 420 bp without tRNA genes.

The new mitochondrial genomes are consistent with the others known for Gracilariaceae and other Florideophyceae, presenting a highly conserved gene content and synteny (Figure 5A) (Campbell et al. 2014; Yang et al. 2015). The 18 complete mitochondrial genomes, six for *Gracilariopsis*, 11 for *Gracilaria* and one for *Melanthalia* are highly conserved except in two regions that present variations in tRNA genes and inserted PDS (Figure 5A). The rearrangement of *trn*H between *Gracilariopsis* and *Gracilaria* (Figure 5B) was previously reported by Lee et al. (2015) and this gene is absent in *M. intermedia*. However, this rearrangement occurred several times in Florideophyceae (Yang et al. 2015). The locus with *trn*L+*nad*6+*trn*G genes (Figure 5B) is highly conserved in Florideophyceae, except in Hildenbrandiophycidae, and it is between two highly divergent loci (Yang et al. 2015). The region between *trn*A and *trn*N (Figure 5C) is also highly variable in terms of tRNA content in Florideophyceae, except in Nemaliophycidae and Hildenbrandiophycidae (Yang et al. 2015). This variation is not consistent with the phylogenetic relationship (Yang et al. 2015) and could also be observed within species, as in *G. salicornia* and *G. vermiculophylla* from two different locations (Figure 5C). Both loci in the mitochondrial genome show a propensity to variability and they could be an interesting tool for population genetics, biogeographic and phylogeographic studies (Song et al. 2016). For example, *G. salicornia* is an invasive species in Hawaii and its type locality is the Philippines. The mitochondrial genomes of *G. salicornia* specimens from Hawaii and Malaysia have 96.9% bp identity. However, if the 682 bp sequence between *trn*A and *trn*N—which include three tRNA genes and two PDS (Figure 5C)—from the Malaysian specimen is excluded the identity is 99.4%.

The hotspots of PDS insertions in mitochondrial genomes are exactly where the origin and terminus of replication may occur, where strand direction changes (Figure 5). These PDS were previously reported to *G. chilensis*, *Gp. chorda* and *Gp. lemaneiformis* (Zhang et al. 2012; Yang et al. 2014b; Lee et al. 2015). Moreover, all the PDS insertions found in mitochondria were similar to *Gracilaria robusta* plasmid Gro4059 (Figure 4), which suggest that similar plasmids could have been inserted in the mitochondria. The intron in the *trn*I gene was found in all sequenced mitochondrial genomes of Florideophyceae, suggesting that this intron was gained in the florideophycean ancestor (Yang et al. 2015).

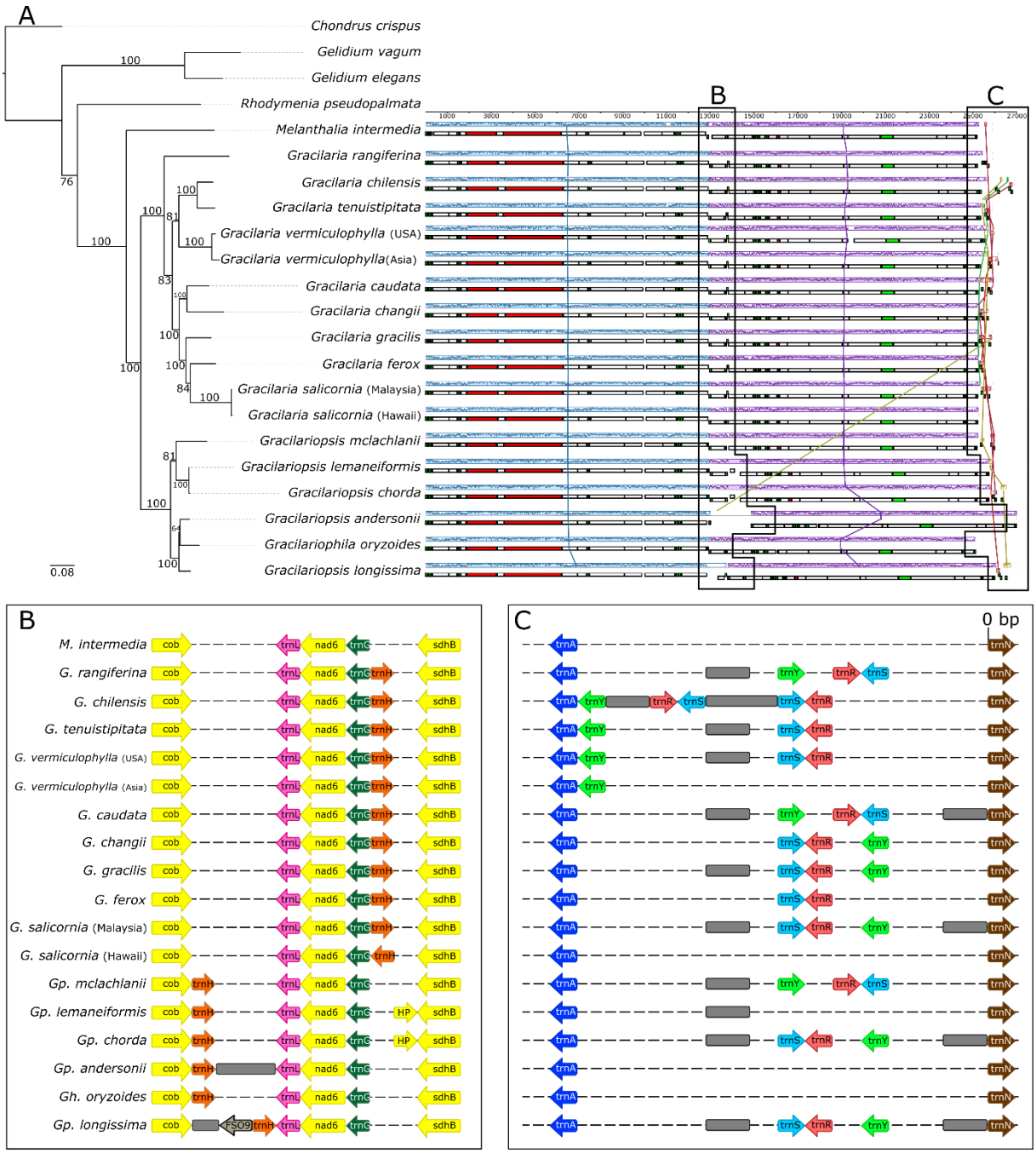

Figure 5 - Mitochondrial genome phylogeny and genomic comparison of Gracilariaceae. A. Left, maximum likelihood phylogenetic tree based on an amino acid matrix with 23 coding genes sequences. Right, mitochondrial progressive MAUVE alignment. Each linear genome corresponds to the name in the phylogenetic tree. Horizontal axis above refers to genome length in base pairs (bp). Synteny between genomes is represented by the colour Locally Collinear Blocks (LCB). LCB above the line is the forward strand; below the line is the reverse strand. The same color line of LCB along the alignment indicate similarity region among the genomes. Gene maps are shown below the LCBs, protein coding genes in white, rRNA genes in red, tRNA genes in dark green and intron in light green. Position of boxes above (forward) and below (reverse) line also refers to gene orientation. B and C. Loci marked in black line in A in detail. Protein coding genes are represented in arrow box in yellow, plasmid-derived sequence in grey, and each *trn* is marked in different colors. Arrow boxes do not represent gene length and the scheme represents only the synteny and not alignment.

Genomic comparison and plasmid-derived sequences of chloroplast

Chloroplast ML tree was congruent in topology with mitochondrial ML tree. However, chloroplast tree had full support for all nodes (Figure 6A). Like the mitochondrial genomes, chloroplast genomes are very conserved, presenting three main conserved regions. The first one was between *rrs* and *cpe*B genes (yellow LCB in Figure 6A). The second was between *nbl*A and *dna*K (green LCB in Figure 6A). The last one was between *rpl*3 and *psb*D (blue LCB in Figure 6A). Species-specific features of chloroplast genomes are generally caused by PDS insertion (Figure 6, Figure 7). They are mainly found between *psb*D and *rrs* genes. Despite the overall conservation, some differences were found. The gene *pet*P was present only in *Gracilariopsis* and *M. intermedia*. *Gracilaria salicornia, G. rangiferina, Gp. mclachlanii* and *M. intermedia* missed *ycf*23, while *G. caudata* and *G. gracilis* had a putative pseudogene. *Gracilaria chilensis* and *M. intermedia* missed *pbs*A and *G. firma* had a small CDS, probably a pseudogene. *Melanthalia intermedia* had a pseudogene of *pgm*A. In the *trn*Me-intron, all *Gracilariopsis, G. firma,* and *M. intermedia* had a CDS (*mat*) while the other *Gracilaria* species had two CDSs (Table S4).

Melanthalia intermedia had at least nine species-specific features (white spaces in LCBs in Figure 6A). Two of them were larger and had ORFs homologous to plasmids which indicated that they are PDS insertions (orange line block in Figure 6A, B; Figure 7). The first one is between ORF684 and *ycf*58 and had the *leu*C and *leu*D operon (yellow rectangle in Figure 6A; Figure 7) and homologous to plasmid ORFs as FSO1, FSO4, FSO10 (Figure 7; Table 14). The second one is between *omp*R and *rrs* and had two complete plasmid insertions (Figure 6A, B) similar to Gle4293 and Gro4970 (Figure 4). Both plasmid insertions are flanked by the homologous FSO1. *Gracilaria rangiferina* had a unique PDS insertion between *dna*K and *rpl*3, which contains two genes homologous to FSO1 and FSO6 (Figure 6; Figure 7; Table 14). *Gracilaria chilensis* and *G. tenuistipitata* had an insertion including *leu*C/*leu*D genes (yellow rectangle in Figure 6A) between *psd*D and *omp*R genes (Figure 7). *Gracilaria chilensis* PDS had two small ORFs similar to FSO1 and one small ORF similar to FSO8 between *psd*D and the *leu*C gene (Figure 7) and *Gracilaria tenuistipitata* PDS had two small ORFs similar to FSO1 between *omp*R and *rrs* genes (Figure 7). *Gracilaria caudata* had a PDS insertion with four CDSs before the *omp*R gene similar to FSO4, FSO11, FSO1, and an unknown gene (Figure 7). *Gracilaria firma* had a unique, large PDS insertion region with about 7,000 bp and 11 ORFs. In this large PDS, we found a complete plasmid insertion similar to the Gle4293 plasmid (Figure 6A, C). This species also had another PDS with two small ORFs homologous to FSO4 between *omp*R and *rrs* genes (Figure 7). *Gracilaria gracilis* had a small PDS insertion without ORFs between *rpl*11 and *moe*B genes (Figure 6A). *Gracilaria ferox* had a PDS insertion before the *omp*R gene with three degenerated ORFs, which share similarity to FSO5, FSO8, and FSO6 (Figure 7). *Gracilaria salicornia* also had a PDS insertion before ompR and had a gene homologous to FSO1 (Figure 7). *Gracilariopsis mclachlanii* had a larger PDS insertion c.a. 5,000 bp with several fragments of five ORFs similar to FSO1, FSO8, FSO7, FSO6 and FSO1 between *omp*R and *rrs* (Figure 7). *Gracilariopsis lemaneiformis* and *Gp. chorda* had a PDS insertion after *omp*R. In *Gp. lemaneiformis* the PDS had with three small

CDS homologous o FSO7, FSO6, and FSO1 (Figure 7). We did not find any PDS insertions in the *Gp. longissima* and *G. vermiculophylla* chloroplast genome.

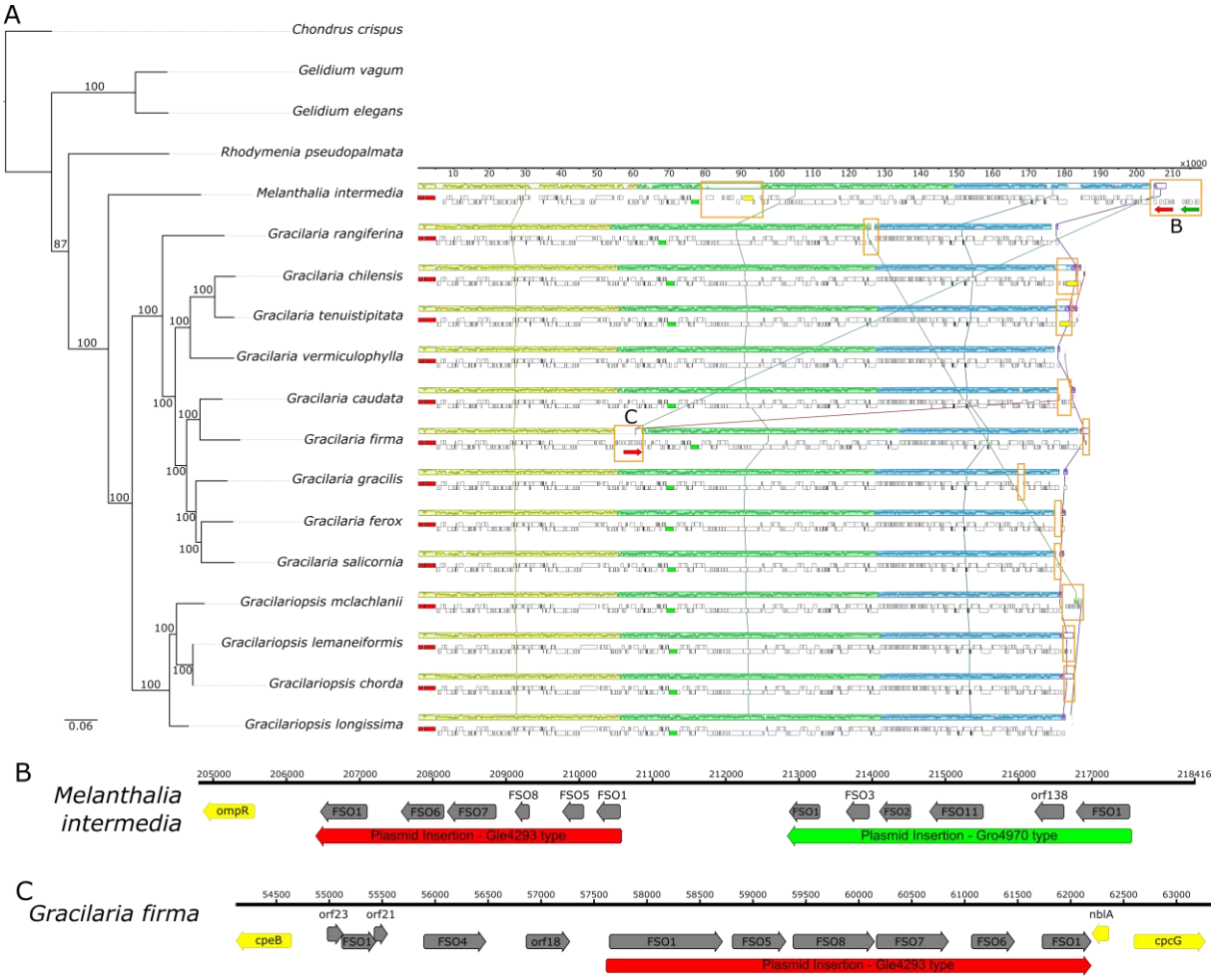

Figure 6 - Chloroplast genome phylogeny and genomic comparison of Gracilariaceae. A. Left, maximum likelihood phylogenetic tree based on an amino acid matrix with 187 coding genes sequences. Right, mitochondrial progressive MAUVE alignment. Each linear genome corresponds to the name in the phylogenetic tree. Horizontal axis above refers to genome length in base pair (bp). Synteny between genomes is represented by Locally Collinear Blocks (LCB). LCB above the line is the forward strand, below the line is the reverse strand. The same color line of LCB along the alignment indicate similarity region among the genomes. Gene maps are shown below the LCBs, protein coding genes in white, rRNA genes in red, tRNA genes in black, intron in light green and *leu*C/*leu*D operon in yellow (showed in *Melanthalia intermedia*., *Gracilaria chilensis*, and *G. tenuistipitata*). Position of boxes above (forward) and below (reverse) line also refers to gene orientation. Continuous orange line boxes represent plasmid-derived sequences. Red arrows indicate plasmid-insertion similar to the Gle4293 plasmid. Green arrows indicate plasmid-insertion similar to the Gro4970 plasmid. B and C. Loci with plasmid-insertions marked in A in detail. Protein coding genes are represented in arrow box in yellow, including plasmids derived ORFs (FSO).

To infer chloroplast genome evolution in Gracilariaceae, we compared gene synteny from 14 complete genomes available (Figure 6), four *Gracilariopsis*, nine *Gracilaria* and one *Melanthalia*. Similar to what was observed for the mitochondria, chloroplast genomes organization is highly similar except for regions with PDS insertions (Figure 6). PDS insertions are common in Rhodophyta (Lee et al. 2016b) and it was found in almost all Gracilariaceae plastids except in *G. vermiculophylla* and *Gp. longissima* (Figure 6A). This last one is the only species that a plasmid homologous ORF (FSO4) was found in the mitochondrial genome. Most PDS insertions have ORFs, which are homologous to ORFs

found in previously published plasmids (Table 14) and they are positioned in between *psd*D-*omp*R-*rrs* locus. This region is the hotspot for plasmid insertion in chloroplast genomes in Gracilariaceae (Lee et al. 2016b; Ng et al. 2017), but the correlation with the origin and terminus of replication is not as clear as it in the mitochondrial genome.

Melanthalia intermedia had two whole plasmid insertions similar in homologous ORFs and synteny with the Gle4293 and Gro4970 plasmids in the chloroplast genomes (Figure 6B) and *Gracilaria firma* had a whole plasmid insertion similar to the Gle4293 plasmid (Figure 6C) (Ng et al. 2017). The fact that all these insertions are flanked by FSO1 (Figure 6B, C), which is a protein belonging to the family DUF1368 with unknown function and specific to red algae (Lee et al. 2016b; Ng et al. 2017), could means that this homologous gene can be associated with the mobility of the complete plasmid or plasmid regions. FSO1 is the most common and widespread gene found in plasmids and chloroplast genomes of red algae (Lee et al. 2016b) and the origin of this plasmid-derived gene is unknown (Lee et al. 2016b). The presence of red algal plasmid regions in the organellar genome has been reported in Rhodophyta with or without the co-occurrence of extrachromosomal plasmid (Ng et al. 2017).

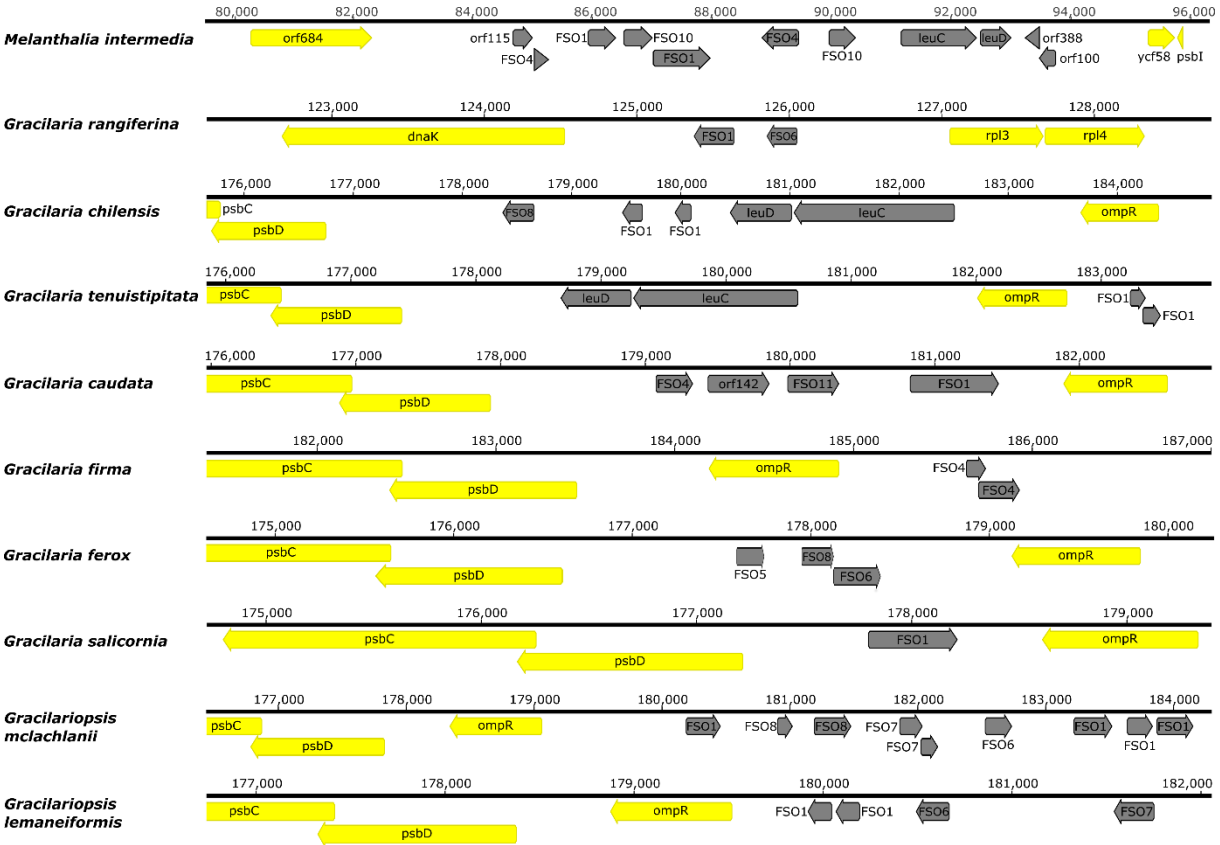

Figure 7 - Gene synteny of the plasmid-derived sequence regions found in chloroplast genomes. These regions are identified in orange line box in Figure 6.

It seems that the extrachromosomal circular plasmids found in *G. ferox* and *G. vermiculophylla* is not related with the PDS insertions in the chloroplast genomes. The PDS insertion in *G. ferox* chloroplast had three degenerated ORFs that are different from the extrachromosomal

plasmid Gfe3115 (Figure 4): FSO5, FSO8, and FSO6, which are also found in the Gle4293 plasmid. In *G. vermiculophylla*, there is no plasmid derived ORFs in organellar genomes.

*leu*C/*leu*D operon

As we found the *leu*C/*leu*D genes in the chloroplast genomes of *Melanthalia intermedia*, *Gracilaria tenuistipitata*, *G. chilensis* (Figure 6; Figure 7) and in *G. vermiculophylla* plasmid Gve7464 (Figure 4), we searched for homologous genes in all assembled contigs of the other species. We found *leu*C/*leu*D encoded in nuclear contigs in all species except *G. chilensis*, *G. tenuistipitata*, and *G. vermiculophylla*, while in *M. intermedia* we found only pseudogenes. These species are exactly the species that present these genes in the chloroplast or plasmid. The *M. intermedia leu*D nucleus encoded gene was so divergent that we could not align it in the matrix used to reconstruct the phylogeny. The ML phylogenies done for each of those two genes are not congruent (Figure 8; Figure 9) as was previous observed by Janouškovec et al. (2013). However, all genes encoded in the chloroplasts and in the plasmid (*G. vermiculophylla*) grouped together close to proteobacteria, despite proteobacteria is not monophyletic in the *leu*D phylogeny (Figure 9). All *leu*C/*leu*D nucleus-encoded grouped in an clade with cyanobacteria and Viridiplantae taxa. Gracilariaceae nucleus-encoded genes grouped with other Rhodophyta taxa, which also have these genes in the nuclear genome (Figure 8; Figure 9). These ML phylogenies clearly show that there are two origins for those genes in Rhodophyta. It is interesting that the species that have chloroplast/plasmid-encoded genes do not seem to have a functional nucleus-encoded gene, as *Melanthalia intermedia* had pseudogenes. The fact that there are more copies of chloroplast and plasmid genomes than nucleus genomes may lead to a prevalence of the chloroplast/plasmid-encoded genes in these species and the nucleus-encoded genes may be lost from the nuclear genome.

The origin and the function of most plasmid genes are unknown and difficult to determine, because, so far, the only match observed for those genes is with organellar genomes and plasmids (Lee et al. 2016b). Besides, the copy number and the position of the homologous plasmid-derived ORFs in the organellar genomes are inconsistent in the Rhodophyta phylogeny (Lee et al. 2016b, Ng et al. 2017). It is possible that the plasmids may have integrated to the organellar genomes randomly along the evolutionary history of the organisms and this incorporation into the maternally inherited organelles may have rendered their fixation in a population (Lee et al. 2016b; Ng et al. 2017). However, all plasmid-derived genes in organellar genomes seem to be pseudogenes because they use to be smaller and highly divergent comparing with the genes in the plasmids. Therefore, apparently, the plasmid genes lost their function when they are inserted in the organellar genomes. The only exception is the *leu*C and *leu*D genes, which seem to be functional. It is not possible to know whether these genes were inserted in chloroplast genome by a Rhodophyta plasmid or if the plasmid contained these genes was disintegrated from the chloroplast genome. The plasmids are analogous to transposable elements with mobility and can contribute to the gain or loss of genes in the genomes (Lee et al. 2016b, Ng et al. 2017).

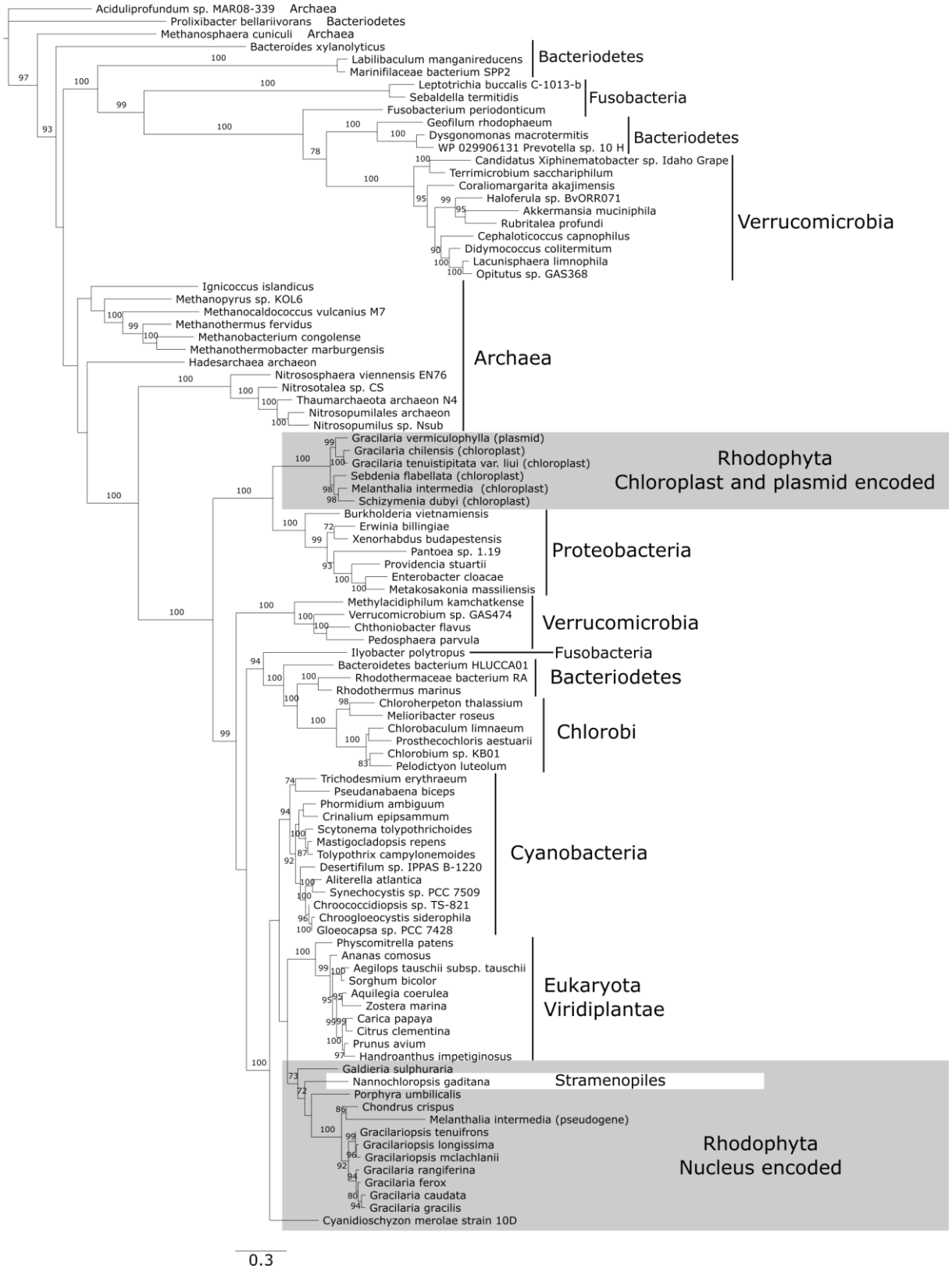

Figure 8 - Maximum likelihood phylogeny of *leu*C. Numbers at nodes indicate bootstrap replicates. Bootstrap support values are shown only above 50%.

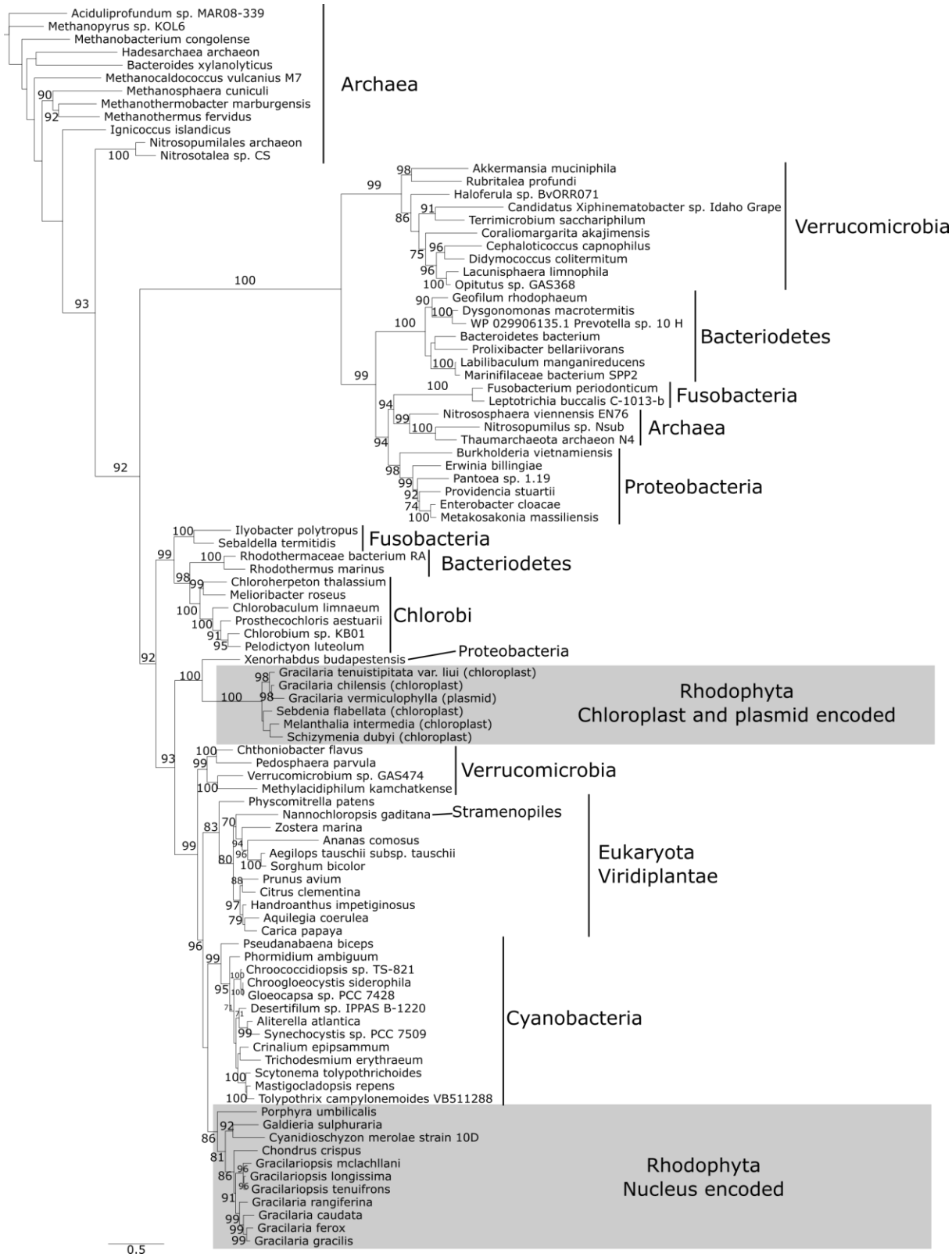

Figure 9 - Maximum likelihood phylogeny of *leu*D. Numbers at nodes indicate bootstrap replicates. Bootstrap support values are shown only above 50%.

Organellar phylogenomics and systematics of Gracilariaceae

Our ML phylogenetic analysis of mitochondria and chloroplast genes showed similar evolutionary histories (Figure 10; Figure 11). Gracilariaceae is monophyletic, *M. intermedia* is sister to a monophyletic clade containing both genera *Gracilaria* and *Gracilariopsis*, which are also monophyletic. Our phylogenies are the first using larger gene content matrices (23 genes for the mitochondrial and 187 genes for the chloroplast) for Gracilariaceae and had the highest supports. Comparing the trees of both organellar genomes, the mitochondrial tree presented lower supports in some nodes (Figure 10). On the other hand, the chloroplast tree had full support for all those nodes (Figure 11). It has been observed in other studies for red algae that chloroplast phylogeny is better supported than mitochondrial phylogeny (Yang et al. 2015; Lee et al. 2016a). Chloroplast genomes have been appointed as a promising resource to resolve red algal phylogenies, due to its highly conserved nature and adequate phylogenetic signal (Janouškovec et al. 2013; Costa et al. 2016; Lee et al. 2016a).

Our taxonomical sampling (43 species for mitochondrial and 39 species for chloroplast) represents the main lineages that have been discussed in Gracilariaceae systematics literature (Gurgel and Fredericq 2004; Lyra et al. 2015). The generic circumscription within Gracilariaceae has been controversial, particularly in relation to the status of *Hydropuntia* as a genus distinct from *Gracilaria* (Bellorin et al. 2002; Gurgel and Fredericq 2004; Lyra et al. 2015). Furthermore, a yet unnamed genus proposed for the clade represented by *G. chilensis*, *G. tenuistipitata*, and *G. vermiculophylla* has also been suggested (Gurgel and Fredericq 2004; Lyra et al. 2015). The largest and most well-resolved phylogeny of Gracilariaceae reduced *Hydropuntia* and the unnamed genus to *Gracilaria*, forming a monophyletic genus, based on three molecular markers (*rbc*L, UPA, and COI-5P) (Lyra et al. 2015).

The monophyletic clade of *Gracilariopsis* contains two monophyletic lineages (Figure 10; Figure 11): first one with *Gp. mclachlanii* and *Gp. chorda/Gp. lemaneiformis*; second one with *Gh oryzoides*/*Gp. andersonii* (in mitochondrial tree, Figure 10) and *Gp. tenuifrons* with *Gp. longissima*. In the mitochondrial ML tree, the support is lower (Figure 10), but the chloroplast tree presents full support (Figure 11). The mitochondrial and chloroplast genomes of *Gp. chorda* and *Gp. lemaneiformis* are very similar (Figure 5A; Figure 6A). In mitochondrial genome, they share 97.4% bp identity. The divergence observed between these two genomes is mainly caused by the region between *trn*A and *trn*N that includes three tRNA genes (*trn*R, *trn*S and *trn*Y in *Gp. lemaneiformis*) (Figure 5C). Without this region, the identity is 99.92% (Yang et al. 2014b). Chloroplast genomes of these species shared 99.96% identity and are identical in synteny and *rbc*L marker is identical between these specimens. However, *Gp. chorda* and *Gp. lemaneiformis* specimens were considered different species (Yang et al. 2014b), which we think is a problem of misidentification of one of the species. There are two chloroplast genomes of *Gp. lemaneiformis* published in Genbank (KU179794 and KP330491) (Du et al. 2016, Zhang et al. 2016) from specimens collected in China. However, the type locality of *Gp. lemaneiformis* is Paita, Peru, and this species may not have a worldwide distribution (Gurgel et al. 2003). Comparing the *rbc*L marker of these Chinese specimens with the *Gp. lemaneiformis* topotype

specimen (GenBank access number: AY049415) they presented 95.2% of identity. Which showed that those genomes from Chinese specimens do not belong to *Gp. lemaneiformis* and, probably, are *Gp. chorda*, for which the type locality is Enoura, Japan (Holmes 1896).

Although we considered *Gracilariopsis* is a monophyletic genus, the parasite *Gracilariophila oryzoides* belongs taxonomically to another genus. However, it is known that this parasite species is an adelphoparasite – *i.e*. the parasite has evolved from the most recent common ancestor with their host – of *Gp. andersonii* (Goff et al. 1996; Hancock et al 2010). Therefore, based on the molecular phylogenetic evidence (Figure 10; Figure 11) and morphological features (Fredericq and Hommersand, 1990b), *Gracilariophila* should be transferred to *Gracilariopsis*, even though *Gh. oryzoide*s (parasite) have a totally different lifestyle from its closely related species host (free-living). This is the same situation of the parasite *Congracilaria babae*, which was transferred to *Gracilaria babae* based on molecular and morphological data (Ng et al. 2014).

The *Gracilaria sensu lato (s.l.)* are represented in the monophyletic clade with all *Gracilaria* species in our phylogenies (Figure 10; Figure 11). To explain the previous taxonomical conflicts in this lineage we made a schematic figure including only the *Gracilaria s.l.* from chloroplast genomes (Figure 12). The characteristics about thallus (flat or cylindrical) and spermatangia structure are represented in all species and were based on several literature descriptions. However, these features are not discrete and in some cases the same species could present both. *Gracilaria* lineage is composed by four main monophyletic clades (Figure 12). The clade I used to belong to *Hydropuntia*, including the type species of the genus *G. urvillei* (ex *H. urvillei*) (Gurgel and Fredericq 2004). This clade is composed mainly of cylindrical species, except *G. rangiferina*, which is flat, and all species have *henriquesiana*-type spermatangia (Gurgel and Fredericq 2004, Lyra et al. 2015). The clade II is composed by *G. vermiculophylla*, *G. tenuistipitata*, and *G. chilensis*. This clade was proposed to be elevated to another genus, but this was not formally done (Gurgel and Fredericq 2004). All these species have cylindrical thalli (Gurgel and Fredericq 2004), but they present diverse spermatangia types. *Gracilaria chilensis* and *G. tenuistipitata* have *textorii*-type spermatangia while *G. vermiculophylla* has *verrucose*-type spermatangia (Lyra et al. 2015) (Figure 12).

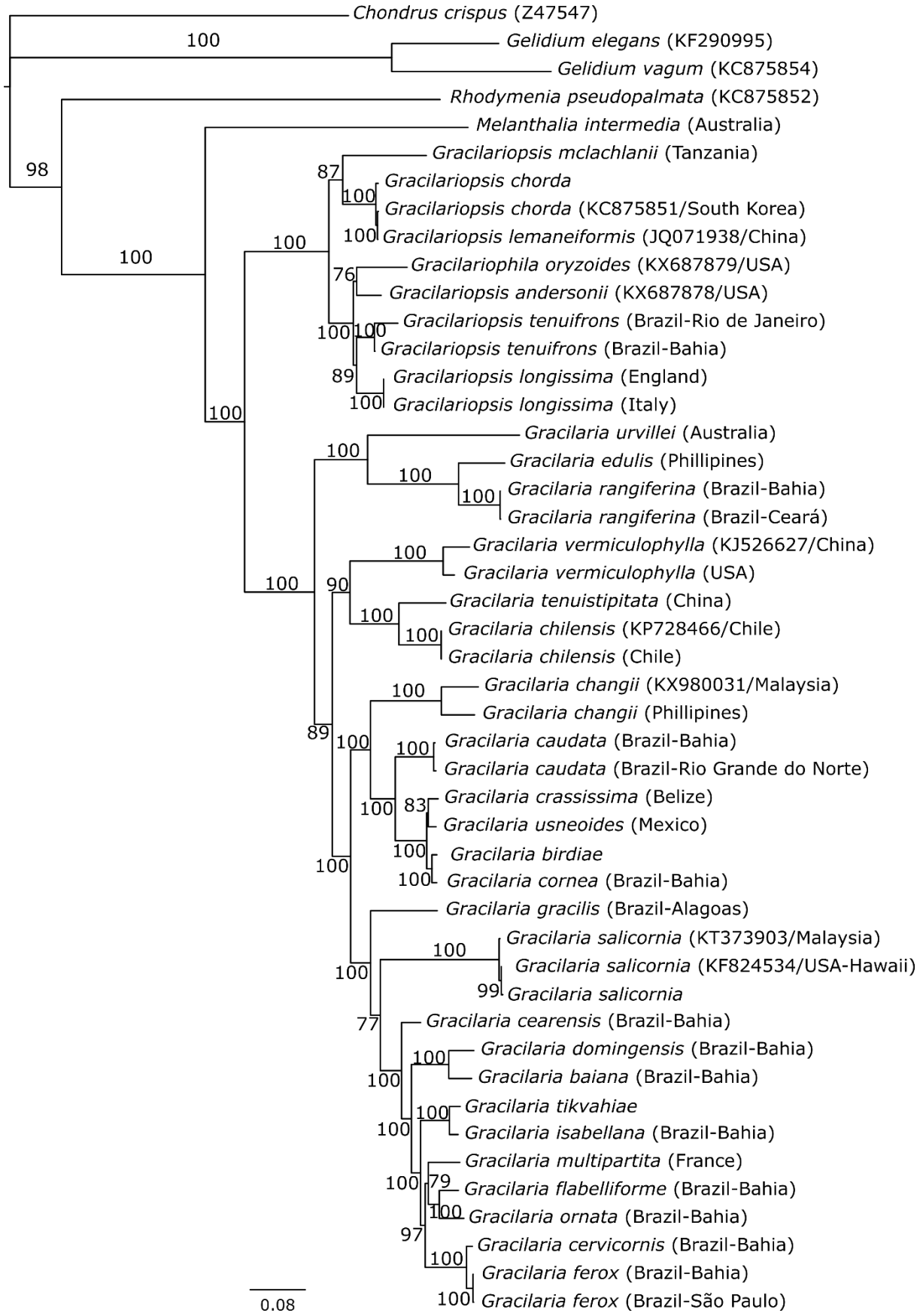

Figure 10 – Mitochondrion maximum likelihood phylogenetic tree based on an amino acid matrix with 23 coding genes sequences. Bootstrap values are shown in the branches. Sequences from GenBank are indicated with accession number.

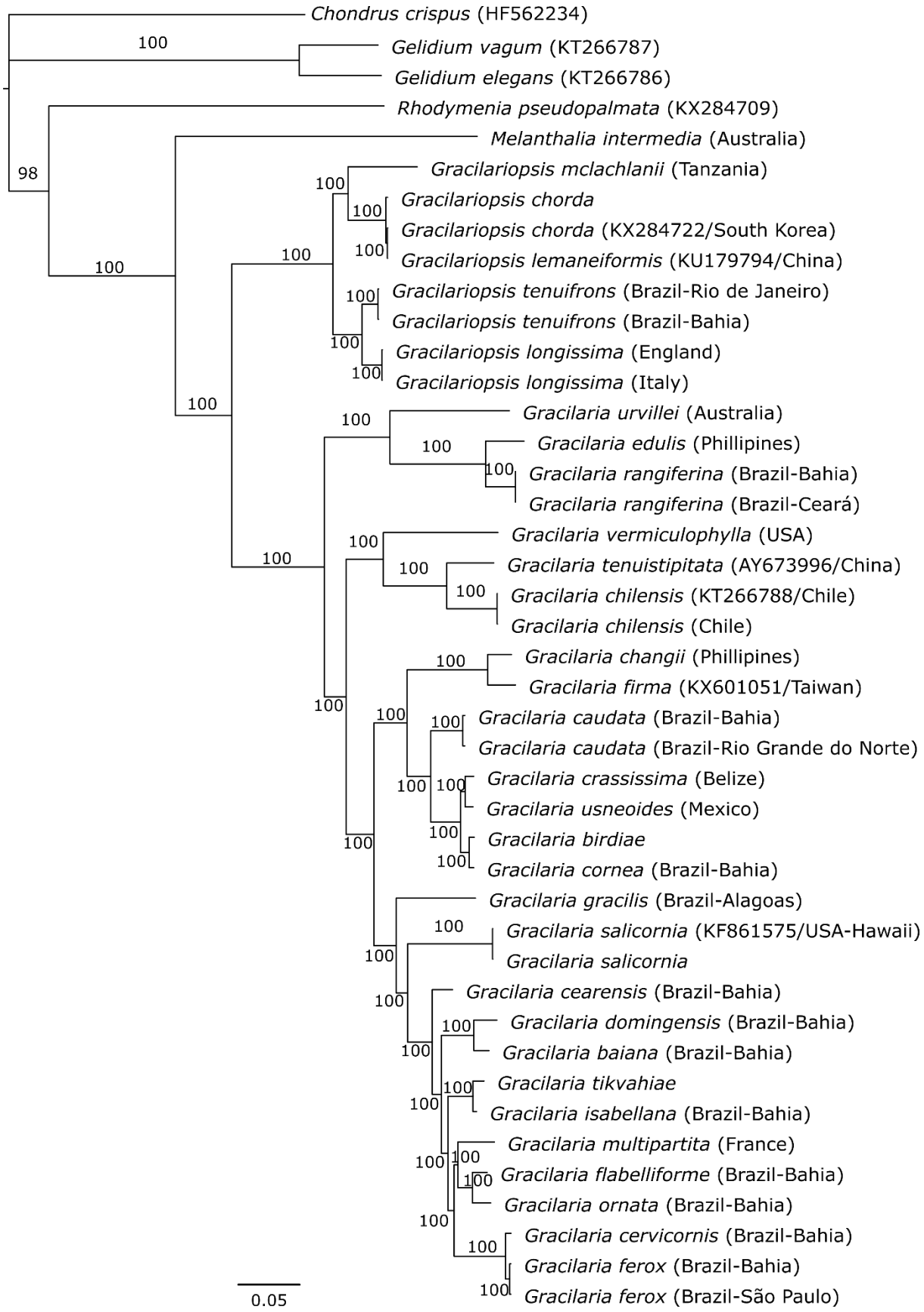

Figure 11 - Chloroplast maximum likelihood phylogenetic tree based on an amino acid matrix with 187 coding genes sequences. Bootstrap values are shown in the branches. Sequences from GenBank are indicated with accession number.

All species of clade III are cylindrical, but *G. crassissima* sometimes may display a compressed prostrate thallus (Gurgel and Fredericq 2004). These species also present spermatangia in deep conceptacles, *henriquesiana*-type, typically of *Hydropuntia*. Because this morphological characteristic of the male reproductive structure and molecular closeness with *Hydropuntia* (clade I in our ML analysis) containing *G. urvillei*, species belonging to clade III were elevated to *Hydropuntia* (Gurgel and Fredericq 2004). However, posterior analysis did not maintain these lineages (clade I + III) as closely related (Lyra et al. 2015). Furthermore, morphological analysis of some species in clade III may present a mixture of spermatangia type (Bird et al. 1986; Abbott et al. 1991; Plastino and Oliveira 1997; Lyra et al. 2015). Our results support the non-monophyly of what was once considered *Hydropuntia* (clade I + clade III) what corroborates previous findings based on three molecular markers, *rbc*L and UPA from the plastid and COI-5P from the mitochondrial genomes (Lyra et al. 2015). On the other hand, *G. chilensis*, *G. tenuistipitata*, and *G. vermiculophylla* (clade II) form a monophyletic group, but within *Gracilaria sensu lato (s.l.)*, therefore the transfer of these three species to a new genus would make *Gracilaria* non-monophyletic.

Finally, clade IV is the previous reported *Gracilaria sensu stricto* (*s.s.*) (Gurgel and Fredericq 2004; Lyra et al. 2015), which is sister group of clade III (Figure 12). This lineage contains almost all flat species of Gracilariaceae and most species present *textorii*-type spermatangia. *Gracilaria gracilis* and *G. salicornia* are cylindrical species and have *verrucosa*-type spermatangia. These two species are basal in *Gracilaria* s.s. lineage. The other species are generally flat, with some exceptions as *G. tikvahiae*, which exhibit a wide range of habit morphologies from entirely flat to terete phenotypes (Gurgel and Fredericq 2004). The flat species *G. domingensis* Sonder ex Kützing and *G. cervicornis* (Turner) J. Agardh used to be confused identified because they have similar habits and ecological niche (Lyra et al. 2016). However, *G. domingensis* possess *verrucosa*-type spermatangia while *G. cervicornis* possess *textorii*-type spermatangia (Oliveria et al. 1983; Lyra et al. 2016). A study based on molecular approaches for delimiting species and morphological analysis identified four cryptic species: *G. cervicornis, G. ferox, G. domingensis*, and *G. baiana* G.M. Lyra, C.F.D. Gurgel, M.C. Oliveira & J.M.C. Nunes (Lyra et al. 2016). *Gracilaria domingensis* and *G. baiana* are sister groups, morphologically similar and present *henriquesiana*-type spermatangia (Figure 12). *Gracilaria cervicornis* and *G. ferox* were previously synonymized by Oliveira et al. (1983), but *G. ferox* was resurrected by Lyra et al. (2016). *Gracilaria ferox* may exhibit terete to subcylindrical branches in some parts of the thalli, while *G. cervicornis* is completely flattened. Both species present *textorii*-type spermatangia.

Based on our results, two taxonomic outcomes are possible for *Gracilaria,* (i) keep only one genus *Gracilaria s.l.* based on molecular and morphological data; or (ii) split the genus into four genera based on molecular data: *Hydropuntia s.s.* (clade I) because it contains *G. urvillei* (the type species *H. urvillei* of the genus); a new unnamed genus for clade II; a new unnamed genus for clade III (ex. *Hydropuntia*); and *Gracilaria s.s.* (clade IV).

Our analyses show the genera *Gracilaria s.l.* and *Gracilariopsis* as monophyletic, with full support in both organellar trees. Furthermore, based on features of organellar genome architecture, these genera are also supported by the loss of the plastid gene *pet*P in *Gracilaria* and the position of the gene *trn*H in the mitochondrial genome (Figure 5). In *Gracilaria* this gene is between *trn*G and *sdh*B and in *Gracilariopsis* this gene is between *cob* and *trn*L genes. As previous reported by Lyra et al. 2015, we do not find any significant molecular or morphological data to split *Gracilaria* lineage into four different genera. Furthermore, in ML analysis, the length of the branches is important and it must be considered. Thus, if *Gracilaria* lineage would be split into two or more genera, the *Gracilariopsis* should also be split into two genera, because the last common ancestor of each genus is in almost same distance from the root (Figure 10; Figure 11).

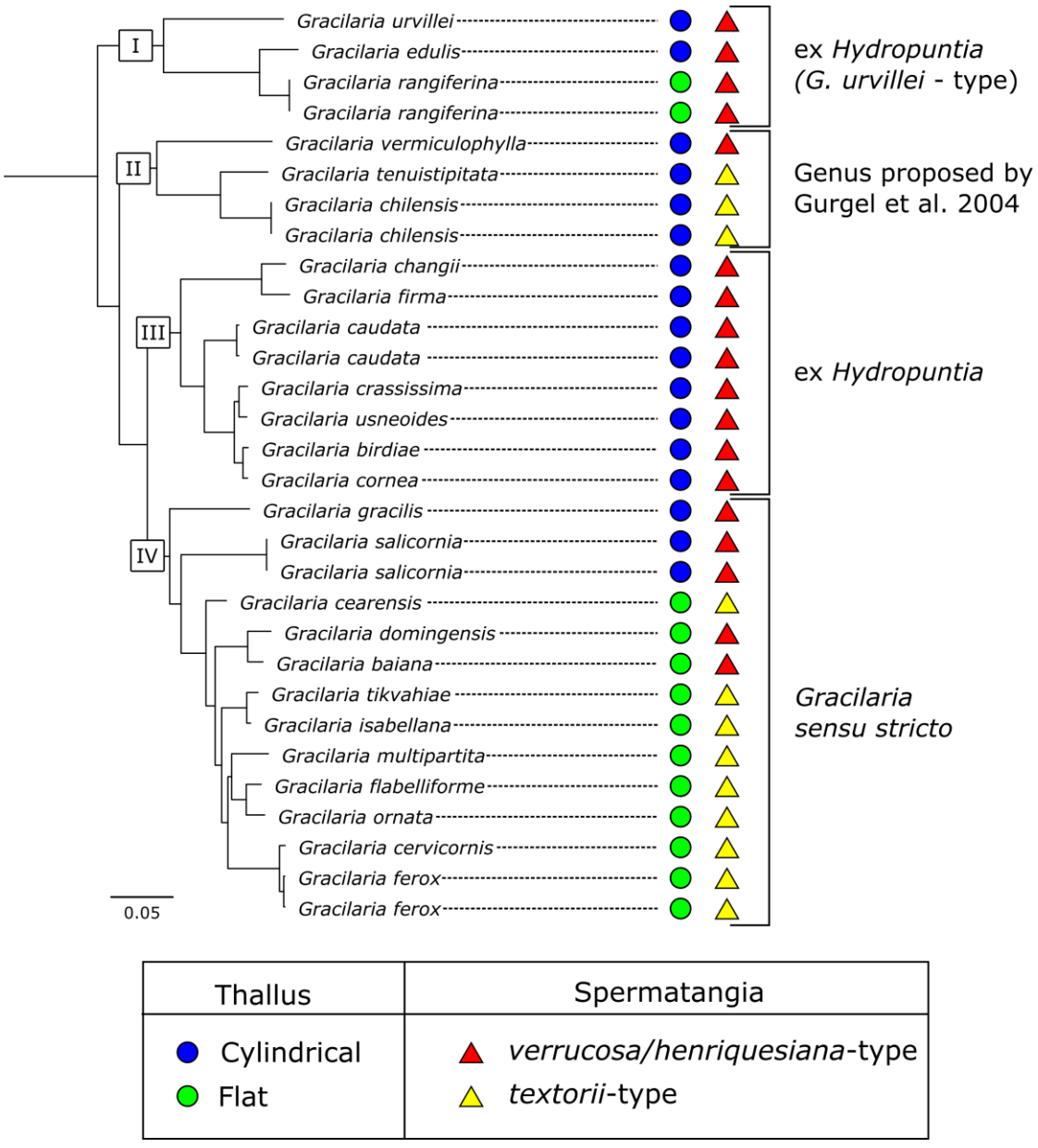

Figure 12 - Schematic figure including only the *Gracilaria* sensu lato from chloroplast genomes (Figure 11). The characteristics about thallus (flat or cylindrical) and spermatangia structure are represented in all species and were based on several literature descriptions. These features are not discrete and some cases the same species could presents both.

Conclusion

High-throughput sequencing is a very powerful technique to obtain organellar genomes from whole-genome sequencing of algae. We could obtain all the organellar genomes from the reads of each sample. Even when it was not possible to close the circular genome, we could extract most of the genes used for phylogenomics analysis. Furthermore, we could also extract the plasmid genomes from the reads of some species, which improved our understanding about the influence of the plasmid-derived sequences in the organellar genomes. Beside this, all the nuclear genomes are available to future research. Therefore, with one single run (or two, including Dr. G. Lyra database) we obtained a large and high-quality data for this research and, possible, many other future projects.

Organellar genomes are attractive to investigate the evolution of the organisms. Our phylogenies are the first using larger gene content matrices (23 genes for the mitochondrial and 187 genes for the chloroplast) for Gracilariaceae and had the highest supports and congruent results between chloroplast and mitochondrial phylogenomics. Furthermore, chloroplast tree had full support for all those nodes, which corroborates the use of chloroplast genomes as a promising resource to resolve red algal phylogenies. Our taxonomical sampling represents the main lineages that have been discussed in Gracilariaceae systematics literature and supports the presence of three main lineages: *Melanthalia/Curdiea*, *Gracilariopsis* and *Gracilaria*. Furthermore, the characteristics of organellar genome architecture, *Gracilariopsis* and *Gracilaria* are also supported by the loss of the plastid gene *pet*P in *Gracilaria* and the rearrangement position of the gene *trn*H in the mitochondrial genome. Beside this, we found no support for the genus *Hydropuntia* as originally proposed.

The genomic architecture is useful to understand the nature and source of change that occurred during the evolution of organisms. Organellar genomes are highly conserved in Rhodophyta, but also have variable regions that can be useful for comparisons at species or population levels, e.g. the plasmid insertion regions. Plasmids have been recognized as mobile elements. Their origin and the process of horizontal gene transfer into organellar genomes of red algae remain unclear (Lee et. al. 2016b). Moreover, Gracilariaceae is a good model to study the impact of PDS in genome evolution due to the significant presence of these sequences in organellar genomes.

Bibliography

Abbott I.A., Junfu Z. and Bangmei X. (1991) *Gracilaria mixta* sp. nov. and other western Pacific species of the genus (Rhodophyta: Gracilariaceae). *Pacific Sci*. 45:12–27

Adachi J., Waddell P.J., Martin W. and Hasegawa M. (2000) Plastid Genome Phylogeny and a Model of Amino Acid Substitution for Proteins Encoded by Chloroplast DNA. *J. Mol. Evol.* 50:348–358. doi: 10.1007/s002399910038

Agardh J.G. (1852) Species genera et ordines algarum, seu descriptiones succinctae specierum, generum et ordinum, quibus algarum regnum constituitur. *Volumis secundi: Algas florideas complectens.* Part 2, fasc. 2. pp. 577-700 [701-720, Addenda and Index]. Lundae [Lund]: C.W.K. Gleerup.

Altschul S.F., Gish W., Miller W., Myers E.W. and Lipman D.J. (1990) Basic local alignment search tool. *J. Mol. Biol*. 215:403–410. doi: 10.1016/S0022-2836(05)80360-2

Amorim R.N.S., Rodrigues J.A.G., Holanda M.L., Quinderé A.L.G., Paula R.C.M., Melo V.M.M and Benevides N.M.B. (2012) Antimicrobial effect of a crude sulfated polysaccharide from the red seaweed *Gracilaria ornata*. *Brazilian Archives of Biology and Technology*. 55: 171–181.

Andrews S. (2010) FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Assad-Ludewigs I.Y. (1984) *Estudos sobre a biologia de uma espécie de Gracilaria (Rhodophyta, Gigartinales) do litoral norte do estado de São Paulo, Brasil*. Universidade de São Paulo;

Bankevich A., Nurk S., Antipov D., Gurevich A.A., Dvorkin M., Kulikov A.S., Lesin V.M. et al. (2012) SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* 19:455–477. doi: 10.1089/cmb.2012.0021

Barufi J.B., Oliveira E.C., Plastino E.M. and Oliveira M.C. (2010) Life history, morphological variability and growth rates of the life phases of *Gracilaria tenuistipitata* (Rhodophyta: Gracilariales) in vitro. *Sci. Mar.* 74:297–303. doi: 10.3989/scimar.2010.74n2297

Bellorin A.M. (2002) *Sistemática e Filogenia molecular de algas Gracilarióides* (Gracilariaceae, Rhodophyta). Universidade de São Paulo.

Bellorin A.M., Buriyo A., Sohrabipour J., Oliveira M.C. and Oliveira E.C. (2008) *Gracilariopsis mclachlanii* sp. nov. and *Gracilariopsis persica* sp. nov. of the Gracilariaceae (Gracilariales, Rhodophyceae) from the Indian Ocean. *J. Phycol.* 44: 1022–1032.

Bellorin A.M., Oliveira M.C. and Oliveira E.C. (2002) Phylogeny and systematics of the marine algal family Gracilariaceae (Gracilariales, Rhodophyta) based on small subunit rDNA and ITS

sequences of Atlantic and Pacific species. *J. Phycol.* 38:551–563. doi: 10.1046/j.1529-8817.2002.01152.x

Bentley D.R., Balasubramanian S., Swerdlow H.P., Smith G.P., Milton J., Brown C.G., Hall K.P., Evers D.J., Barnes C.L., Bignell, H.R., et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456, 53–59.

Bezerra A.F. and Marinho-Soriano E. (2010) Cultivation of the red seaweed *Gracilaria birdiae* (Gracilariales, Rhodophyta) in tropical waters of northeast Brazil. *Biomass and Bioenergy* 34: 1813-1817.

Bhattacharya D., Price D.C., Chan C.X., Qiu H., Rose N., Ball S., Weber A.P.M., Arias M.C., Henrissat B., Coutinho P.M., et al. (2013) Genome of the red alga *Porphyridium purpureum*. *Nat. Commun*. 4:1941.

Bhattacharya D., Qiu H., Price D.C., Yoon H.S. and Graham M. (2015) Why we need more algal genomes. *J. Phycol.* 51: 1-5. doi:10.1111/jpy.12267

Bird C.J. and McLachlan J. (1984) Taxonomy of *Gracilaria*: Evaluation of some aspects of reproductive structure. *Hydrobiologia* 116/117:41–62.

Bird C.J., McLachlan J. and Oliveira E.C. (1986) *Gracilaria chilensis* sp. nov. (Rhodophyta, Gigartinales), from Pacific South America. *Can. J. Bot.* 64: 2928–2934.

Bird C.J., Oliveira E.C. and McLachlan J. (1986) *Gracilaria cornea*, the correct name for the western Atlantic alga hitherto known as *G . debilis* (Rhodophyta, Gigartinales). *Can. J. Bot*. 64:2045–2051. doi: 10.1139/b86–267

Bird C.J., Ragan M.A., Critchley A.T., Rice E.L. and Gutell R.R. (1994) Molecular relationships among Gracilariaceae (Rhodophyta): further observations on some undertermined species. *Eur. J. Phycol.* 29: 195–202.

Bird C.J., Rice E.L., Murphy C.A. and Ragan M.A. (1992) Phylogenetic relationships in the Gracilariales (Rhodophyta) as determined by 18S rDNA sequences. *Phycologia* 31:510–522.

Bolger A.M., Lohse M. and Usadel B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. doi: 10.1093/bioinformatics/btu170

Boyen C., Leblanc C., Bonnard G., et al (1994) Nucleotide sequence of the *cox*3 gene from *Chondrus crispus*: Evidence that UGA encodes tryptophan and evolutionary implications. *Nucleic. Acids Res.* 22:1400–1403. doi: 10.1093/nar/22.8.1400

Brawley S.H., Blouin N.A., Ficko-Blean E., Wheeler G.L., Lohr M., Goodson H. V., Jenkins J.W., et al. (2017) Insights into the red algae and eukaryotic evolution from the genome of *Porphyra umbilicalis* (Bangiophyceae, Rhodophyta). *Proc. Natl. Acad. Sci.* 114:E6361–70.

Brito L. and Silva S. (2005) Fenología y ciclo de vida del alga *Gracilariopsis tenuifrons* (Gracilariaceae) en Sucre,Venezuela. *Rev. Biol. Trop.* 53:67–73.

Bushnell B. (2014) BBTools software package. https://jgi.doe.gov/data-and-tools/bbtools

Campbell M.A., Presting G., Bennett M.S. and Sherwood A.R. (2014) Highly conserved organellar genomes in the Gracilariales as inferred using new data from the Hawaiian invasive alga *Gracilaria salicornia* (Rhodophyta). *Phycologia* 53:109–116. doi: 10.2216/13-222.1

Carneiro, M.A.A. (2011) *Fenologia, aspectos ecofisiológicos e seleção de linhagens em cultivos no mar de* Gracilaria birdiae *(Rhodophyta, Gracilariales) no Estado do Rio Grande do Norte, Brasil*. Universidade de São Paulo.

Chan X.C. and Ragan M.A. (2013) Next-generation phylogenomics. *Biology Direct* 8:3.

Chang C.F. and Xia B.M. (1963) *Polycavernosa*, a new genus of the Gracilariaceae. *Studia Marina Sinica* 3: 119–126.

Chang C.F. and Xia B.M. (1976) Studies on Chinese species of *Gracilaria*. *Studia Marina Sinica* 11: 91–166.

Chernomor O., von Haeseler A. and Minh B.Q. (2016) Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices. *Syst. Biol.* 65:997–1008. doi: 10.1093/sysbio/syw037

Chitsaz H., Yee-Greenbaum J.L., Tesler G., et al (2011) Efficient de novo assembly of single-cell bacterial genomes from short-read data sets. *Nat. Biotechnol.* 29:915–921. doi: 10.1038/nbt.1966

Choi H.S., Choi K.H., Kim T.H., Lee C.H. and Rhew T.H. (2001) Characterization of natural plasmid and construction of putative transformation vector using the plasmid in the Korean red alga, *Porphyra tenera*. *Algae* 16:287–294.

Collén J., Porcel B., Carrée W., Ball S.G., Chaparro C., Tonon T., Barbeyron T., Michel G., Noel B., Valentin K., et al. (2013) Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the Archaeplastida. *Proc. Natl. Acad. Sci. USA* 110(13): 5247–5252.

Collins F.S., Holden I. and Setchell W.A. (1899) *Phycotheca boreali-americana. A collection of dried specimens of the algae of North America*. Vol. Fasc. XII-XIII, Nos. 551-650. Malden, Massachusetts.

Costa E.S., Plastino E.M., Petti R., et al (2012) The Gracilariaceae Germplasm Bank of the University of São Paulo, Brazil-a DNA barcoding approach. *J. Appl. Phycol*. 24:1643–1653. doi: 10.1007/s10811-012-9828-y

Costa J.F., Lin S., Macaya E.C., et al (2016) Chloroplast genomes as a tool to resolve red algal phylogenies : a case study in the Nemaliales. *BMC Evol. Biol*. 16:205. doi: 10.1186/s12862-016-0772-3

Darling A.E., Mau B. and Perna N.T. (2010) progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS One* 5:e11147. doi: 10.1371/journal.pone.0011147

Dawson E.Y. (1949) Studies of the northeast Pacific Gracilariaceae. *Pap Allan Hancock Found Occas* 7:1–105

Dawson E.Y. (1954) Notes on tropical Pacific marine algae. *Bulletin of the Southern California Academy of Sciences* 53: 1-7.

Dawson E.Y., Acleto O.C. and Foldvik N. (1964) The seaweeds of Peru. *Beihefte zur Nova Hedwigia* 13: 1-111.

De Luca P., Taddei R. and Varano L. (1978) "*Cyanidioschyzon merolae*": a new alga of thermal acidic environments. *Webbia* 33: 37-44.

Desai A., Marwah V.S., Yadav A., et al (2013) Identification of Optimum Sequencing Depth Especially for De Novo Genome Assembly of Small Genomes Using Next Generation Sequencing Data. *PLoS One* 8:. doi: 10.1371/journal.pone.0060204

Díaz-Tapia P., Maggs C.A., West J.A. and Verbruggen H. (2017) Analysis of chloroplast genomes and a supermatrix inform reclassification of the Rhodomelaceae (Rhodophyta). *J. Phycol.* 53:920–937.

Dijk E.L., Auger H., Jaszczyszyn, Y. and Thermes C. (2014a) Ten years of next-generation sequencing technology. *Trends in Genetics* 30: 418–426.

Dijk E.L., Jaszczyszyn Y. and Thermes C. (2014b) Library preparation methods for next-generation sequencing: Tone down the bias. *Exp. Cell. Res.* 322:12–20. doi: 10.1016/j.yexcr.2014.01.008

Drew K.M. and Ross R. (1965) Some generic names in the Bangiophycidae. *Taxon* 14: 93-99.

Du Q., Bi G., Mao Y. and Sui Z. (2016) The complete chloroplast genome of *Gracilariopsis lemaneiformis* (Rhodophyta) gives new insight into the evolution of family Gracilariaceae. *J. Phycol.* 52:441–50.

Faugeron S., Valero M., Destombe C., et al (2001) Hierarchical spatial structure and discriminant analysis of genetic diversity in the red alga *Mazzaella laminarioides* (Gigartinales, Rhodophyta). *J. Phycol.* 37:705–716. doi: 10.1046/j.1529-8817.2001.01072.x

Fredericq S. and Hommersand M.H. (1989a) Proposal of the Gracilariales ord. nov. (Rhodophyta) based on an analysis of the reproductive development of *Gracilaria verrucosa*. *J. Phycol.* 25:213–227. doi: 10.1111/j.1529-8817.1989.tb00116.x

Fredericq S. and Hommersand M.H. (1989b) Comparative morphology and taxonomic status of *Gracilariopsis* (Gracilariales, Rhodophyta). *J. Phycol*. 25:228–241.

Fredericq S. and Hommersand M.H. (1990a) Taxonomy of *Melanthalia obtusata* var. *abscissa* and its placement in the Gracilariales (Rhodophyta). *Cryptogam. Algol.* 2:4–11.

Fredericq S. and Hommersand M.H. (1990b) Diagnoses and key to the genera of the Gracilariaceae (Gracilariales, Rhodophyta). *Hydrobiologia* 204–205:173–178. doi: 10.1007/BF00040230

Goff L.J. and Coleman A.W. (1988) The use of plastid DNA restriction endonuclease patterns in dilineating red algal species and populations. *J. Phycol.* 24:357–368.

Goff L.J. and Coleman A.W. 1990. Red algal plasmids. *Curr. Genet.* 18:557–65.

Goff L.J., Moon D.A., Nyvall P., et al (1996) The evolution of parasitism in the red algae: Molecular comparisons of adelphoparasites and their hosts. *J. Phycol.* 32:297–312. doi: 10.1111/j.0022-3646.1996.00297.x

Goodwin S., McPherson J.D. and McCombie W.R. (2016) Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics* 17: 333–351.

Greville R.K. (1830) *Algae britannicae, or descriptions of the marine and other inarticulated plants of the British islands, belonging to the order Algae; with plates illustrative of the genera.* pp. [i*-iii*], [i]-lxxxviii, [1]-218, pl. 1-19. Edinburgh and London: McLachlan and Stewart; Baldwin and Cradock.

Guiry M.D. and Guiry G.M. (2018) Algaebase. In: Natl. Univ. Ireland, Galway. http://www.algaebase.org/

Gurgel C.F.D. and Fredericq S. (2004) Systematics of the Gracilariaceae (Gracilariales, rhodophyta): A critical assessment based on *rbc*L sequence analyses. *J. Phycol.* 40:138–159. doi: 10.1046/j.1529-8817.2004.02129.x

Gurgel C.F.D., Fredericq S. and Norris J.N. (2004) *Gracilaria flabelliformis* and *G. apiculata* (Gracilariaceae, Rhodophyta): restoring old names for common tropical western Atlantic species, including the recognition of three new subspecies, and a replacement name for "*G. lacinulata*". *Cryptogamie Algologie* 25: 367-396.

Gurgel C.F.D., Liao L.M., Fredericq S. and Hommersand M.H. (2003) Systematics of *Gracilariopsis* Dawson (Gracilariales, Rhodophyta) based on *rbc*L sequence analysis and morphological evidence. *J. Phycol.* 39:154–171.

Haglund K. and Pedersén M. (1992) Growth of the red alga *Gracilaria tenuistipitata* at high pH. Influence of some environmental factors and correlation to an increased carbonic-anhydrase activity. *Bot. Mar.* 35:579–587

Hagopian J.C., Reis M., Kitajima J.P., et al (2004) Comparative Analysis of the Complete Plastid Genome Sequence of the Red Alga *Gracilaria tenuistipitata* var. *liui* Provides Insights into the

Evolution of Rhodoplasts and Their Relationship to Other Plastids. *J. Mol. Evol.* 59:464–477. doi: 10.1007/s00239-004-2638-3

Hancock L., Goff L. and Lane C. (2010) Red Algae Lose Key Mitochondrial Genes in Response to Becoming Parasitic. *Genome Biol. Evol.* 2:897–910. doi: 10.1093/gbe/evq075

Harvey W.H. (1855) Short characters of some new genera and species of algae discovered on the coast of the colony of Victoria, Australia. *Annals and Magazine of Natural History* 15(2): 332-336.

Head S.R., Komori H.K., LaMere S.A., Whisenant T., Van Nieuwerburgh F., Salomon D.R. and Ordoukhanian P. (2014) Library construction for next-generation sequencing: *Overviews and challenges. Biotechniques*. 56:61-77.

Holmes M. (1896) New marine algae from Japan. *J. Linn. Soc. London, Bot*. 31:248–260.

International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature* 431, 931–945.

Janouškovec J., Liu S.L., Martone P.T., et al (2013) Evolution of Red Algal Plastid Genomes: Ancient Architectures, Introns, Horizontal Gene Transfer, and Taxonomic Utility of Plastid Markers. *PLoS One* 8:. doi: 10.1371/journal.pone.0059001

Kapraun D.F. and Freshwater D.W. (2012) Estimates of nuclear DNA content in red algal lineages. *AoB Plants* 2012:pls005-pls005. doi: 10.1093/aobpla/pls005

Katoh K. and Standley D.M. (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* 30:772–780. doi: 10.1093/molbev/mst010

Kim K.M., Park J.H., Bhattacharya D. and Yoon H.S. (2014a) Applications of next-generation sequencing to unravelling the evolutionary history of algae. *International Journal of Systematic and Evolutionary Microbiology* 64: 333–345.

Kim K.M., Yang E.C., Yi G. and Yoon H.S. (2014b) Complete mitochondrial genome of sublittoral macroalga *Rhodymenia pseudopalmata* (Rhodymeniales, Rhodophyta). *Mitochondrial DNA* 25:273–274. doi: 10.3109/19401736.2013.800501

Kozarewa I., Ning Z., Quail M.A., et al (2009) Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nat Methods* 6:291–295. doi: 10.1038/nmeth.1311

Kück P. and Meusemann K. (2010) FASconCAT: Convenient handling of data matrices. *Mol. Phylogenet. Evol.* 56:1115–1118. doi: 10.1016/j.ympev.2010.04.024

Kützing F.T. (1843) *Phycologia generalis oder Anatomie, Physiologie und Systemkunde der Tange. Mit 80 farbig gedruckten Tafeln, gezeichnet und gravirt vom Verfasser*. pp. [part 1]: [i]-xxxii, [1]-142, [part 2:] 143-458, 1, err.], pls 1-80. Leipzig: F.A. Brockhaus.

Lanfear R., Calcott B., Ho S.Y.W. and Guindon S. (2012) PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.* 29:1695–1701. doi: 10.1093/molbev/mss020

Langmead B. and Salzberg S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. doi: 10.1038/nmeth.1923

Leblanc C., Boyen C., Richard O., Bonnard G., Grienenberger J. and Kloareg B. (1995) Complete Sequence of the Mitochondrial DNA of the Rhodophyte *Chondrus crispus* (Gigartinales). Gene Content and Genome Organization. *J. Mol. Biol.* 250:484-495.

Lee J., Boo S.M., Mansilla A. and Yoon H.S. (2015) Unique repeat and plasmid sequences in the mitochondrial genome of *Gracilaria chilensis* (Gracilariales, Rhodophyta). *Phycologia* 54:20–23. doi: 10.2216/PH14-97.1

Lee J., Cho C.H., Park S.I., et al (2016a) Parallel evolution of highly conserved plastid genome architecture in red seaweeds and seed plants. *BMC Biol* 14:75. doi: 10.1186/s12915-016-0299-5

Lee J., Kim K.M., Yang E.C., et al (2016b) Reconstructing the complex evolutionary history of mobile plasmids in red algal genomes. *Sci. Rep.* 6:23744. doi: 10.1038/srep23744

Lee J., Yang E.C., Graf L., Yang J.H., Qi H., Zion U.Z., Chan C.X. et al. (2018) Analysis of the draft genome of the red seaweed *Gracilariopsis chorda* provides insights into genome size evolution in Rhodophyta. *Mol. Biol. Evol.*

Liao L.M. and Hommersand M. (2003) A morphological study and taxonomic reassessment of the generitype species in the Gracilariaceae. *J. Phycol*. 39: 1207-1232.

Lonardi S., Mirebrahim H., Wanamaker S., et al (2015) When less is more: "slicing" sequencing data improves read decoding accuracy and de novo assembly quality. *Bioinformatics* 31:2972–2980. doi: 10.1093/bioinformatics/btv311

Lyra G.D.M., Costa E.S., Jesus P.B., et al (2015) Phylogeny of Gracilariaceae (Rhodophyta): Evidence from plastid and mitochondrial nucleotide sequences. *J. Phycol*. 51:356–366. doi: 10.1111/jpy.12281

Lyra G.D.M., Gurgel C.F.D., Costa E., Jesus P.B., Oliveira M.C., Davis C.C. and Nunes J.M.C. (2016) Delimitating cryptic species in the *Gracilaria domingensis* complex (Gracilariaceae, Rhodophyta) using molecular and morphological data. *J. Phycol*. 52:997-1017.

Macchiavello J., Paula É.J. and Oliveira E.C. (1998) Growth Rate Responses of Five Commercial Strains of *Gracilaria* (Rhodophyta, Gracilariales) to Temperature and Light. *J. World Aquac. Soc.* 29:259–266. doi: 10.1111/j.1749-7345.1998.tb00986.x

Marinho-Soriano E., Camara M.R., Cabral T.M. and Carneiro M.A.A. (2007) Preliminary evaluation of the seaweed *Gracilaria cervicornis* (Rhodophyta) as a partial substitute for the industrial feeds used in shrimp (*Litopenaeus vannamei*). *Aquaculture Research* 38: 182–187.

Matsuzaki M., Misumi O., Shin-i T., Maruyama S., Takahara M., Miyagishima S., Mori T., Nishida K., Yagisawa F., Nishida K., et al. (2004) Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* 428(6983): 653–657.

McHugh D.J. 2003. A guide to the seaweed industry, Rome: FAO Fisheries Technical Paper. Food and Agriculture Organization of the United Nations.

Merola A., Castaldo R., De Luca P., Gambardella R. Musacchio A. and Taddei R (1982) Revision of *Cyanidium caladium*. Three species of acidophilic algae. *Giornale Botanico Italiano* 15: 189-195.

Minh B.Q., Nguyen M.A.T. and Von Haeseler A. (2013) Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* 30:1188–1195. doi: 10.1093/molbev/mst024

Miranda G.E.C. (2000) *Avaliação do impacto da exploração (simulada) da alga agarófita* Gracilaria caudata *J. Agardh (Rhodophyta) no litoral do Estado da Paraíba*. Universidade de São Paulo

Montagne C. (1842) *Prodromus generum specierumque phycearum novarum, in itinere ad polum antarcticum...ab illustri Dumont d'Urville peracto collectarum, notis diagnosticis tantum huc evulgatarum, descriptionibus verò fusioribus nec non iconibus analyticis iam iamque illustrandarum*. pp. [1-]16. Parisiis [Paris]: apud Gide, editorem.

Montagne C. (1843) Quatrième centurie de plantes cellulaires exotiques nouvelles. Décades VII. *Annales des Sciences Naturelles, Botanique, Seconde Série* 20: 294-306.

Moon D.A. and Goff L.J. (1997) Molecular Characterization of Two Large DNA Plasmids in the Red Alga *Prophyra pulchra*. *Curr. Genet.* 32:132–138.

Nägeli C. (1847) Die neuern Algensysteme und Versuch zur Begründung eines eigenen Systems der Algen und Florideen. *Neue Denkschriften der Allgemeinen Schweizerischen Gesellschaft für die Gesammten* 9(2): 1-275.

Nakamura Y., Sasaki N., Kobayashi M., Ojima N., Yasuike M., Shigenobu Y., et al. (2013) The First Symbiont-Free Genome Sequence of Marine Red Alga, Susabi-nori (*Pyropia yezoensis*). *PLoS ONE* 8(3): e57122. doi.org/10.1371/journal.pone.0057122

Ng P.K., Lim P.E. and Phang S.M. (2014) Radiation of the red algal parasite *Congracilaria babae* onto a secondary host species, *Hydropuntia* sp. (Gracilariaceae, Rhodophyta). *PLoS One* 9:1–11. doi: 10.1371/journal.pone.0097450

Ng P.K., Lin S.M., Lim P.E., et al (2017) Complete chloroplast genome of *Gracilaria firma* (Gracilariaceae, Rhodophyta), with discussion on the use of chloroplast phylogenomics in the subclass Rhodymeniophycidae. *BMC Genomics* 18:40. doi: 10.1186/s12864-016-3453-0

Nguyen L.T., Schmidt H.A., von Haeseler A. and Minh B.Q. (2015) IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* 32:268–274. doi: 10.1093/molbev/msu300

Nyvall Collén P., Collén J., Silva Reis M., et al (2012) Analysis of expressed sequence tags from the agarophyte *Gracilaria tenuistipitata* (Rhodophyta). *J. Appl. Phycol.* 24:641–647. doi: 10.1007/s10811-011-9681-4

Oliveira E.C. (1984) Taxonomic criteria in the genus *Gracilaria* Greville (Rhodophyta): na exprerience with the Western Atlantic species. *Hydrobiology* 116/117: 55-58.

Oliveira E.C., Bird C. J. and Mclachlan J. (1983) The genus *Gracilaria* (Rhodophyta, Gigartinales) in the westerns Atlantic. *Gracilaria domingensis*, *G. cervicornis*, and *G. ferox. Can. J. Bot.* 61:2999–3008.

Oliveira M.C., Repetti S.I., Iha C., Jackson C.J., Díaz-Tapia P., Lubiana K.M.F., Cassano V. et al. (2018) High-throughput sequencing for algal systematics. *Eur. J. Phycol.* 00:1–17.

Papenfuss G.F. (1966) Notes on algal nomenclature. V. Various Chlorophyceae and Rhodophyceae. *Phykos* 5:95–105

Plastino E.M. (1991) *Cultivo in vitro, estudos reprodutivos e biossistemática de algas gracilarióides (Rhodophyta, Gracilariales) de talo cilíndrico*. Universidade de São Paulo

Plastino E.M. and Oliveira E.C. (1994) *Gracilariaceae*. In: Akatsuka I (ed) Biology of Economic Algae. SPB Academic Publishing bv, The Hague, The Netherlands, pp 185–226

Plastino E.M. and Oliveira E.C. (1997) *Gracilaria caudata* J. Agardh (Gracilariales, Rhodophyta) - restoring an old name for a common western Atlantic alga. *Phycologia* 36:225–232. doi: DOI 10.2216/i0031-8884-36-3-225.1

Pritchard A.E., Sable C.L., Venuti S.E. and Cummings D.J. (1990) Analysis of NADH dehydrogenase proteins, ATPase subunit 9, cytochrome b, and ribosomal protein L14 encoded in the mitochondrial DNA of Paramecium. *Nucleic. Acids Res.* 18:163–171. doi: 10.1093/nar/18.1.163

Pueschel C.M. and Cole K.M. (1982) Rhodophycean Pit Plugs: An Ultrastructural Survey with Taxonomic Implications. *Am. J. Bot.* 69:703. doi: 10.2307/2442960

Quast C., Pruesse E., Yilmaz P., et al (2013) The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic. Acids Res*. 41:590–596. doi: 10.1093/nar/gks1219

Reuter J.A., Spacek D. and Snyder M.P. (2015) High-throughput sequencing technologies. *Molecular Cell* 58: 586–597.

Salomaki E.D. and Lane C.E. (2016) Red Algal Mitochondrial Genomes are More Complete than Previously Reported. *Genome Biol. Evol.* 9:evw267.

Sanger F. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*, 74(12): 5463–5467.

Santelices B. and Doty M.S. (1989) A review of *Gracilaria* farming. *Aquaculture* 78: 95-133.

Schmitz F. (1892) [6. Klasse Rhodophyceae]. 2. Unterklasse Florideae. In: *Syllabus der Vorlesungen über specielle und medicinisch-pharmaceutische Botanik. Eine Uebersicht über das gesammte Pflanzensystem mit Berücksichtigung der Medicinal- und Nutzpflanzen. Grosse Ausgabe.* (Engler, A. Eds), pp. 16-23. Berlin: Borntraeger.

Schönknecht G., ChenW., Ternes C.M., Barbier G.G., Shrestha R.P., Stanke M., Bräutigam A., Baker B.J., Banfield J.F., Garavito R.M., et al. (2013) Gene transfer from bacteria and archaea facilitated evolution of na extremophilic eukaryote. *Science* 339(6124): 1207–1210.

Serra D.R. (2013) *Resposta de* Gracilariopsis tenuifrons *(Gracilariales - Rhodophyta) a estímulos de irradiância in vitro*. Universidade de São Paulo.

Setchell W.A. (1918) Parasitism among the Red Algæ. *Proc. Amerian. Philos. Soc.* 57:155–172.

Sims D., Sudbery I., Ilott N.E., et al (2014) Sequencing depth and coverage: key considerations in genomic analyses. *Nat. Rev. Genet.* 15:121–132. doi: 10.1038/nrg3642

Smit A. J. 2004. Medicinal and pharmaceutical uses of seaweed natural products: A review. *J. Appl. Phycol.* 16:245–62.

Song S.L., Yong H.S., Lim P.E. and Phang S.M. (2017) Complete mitochondrial genome of *Gracilaria changii* (Rhodophyta: Gracilariaceae). *J. Appl. Phycol.* 29:2129–2134.

Song S.L., Yong H.S., Lim P.E., et al (2016) Complete mitochondrial genome, genetic diversity and molecular phylogeny of *Gracilaria salicornia* (Rhodophyta: Gracilariaceae). *Phycologia* 55:371–377. doi: 10.2216/15-128.1

Stackhouse J. (1797) *Nereis britannica; continens species omnes fucorum in insulis britannicis crescentium: descriptione latine et anglico, necnon iconibus ad vivum depictis...* Fasc. 2. pp. ix-xxiv, 31-70, pls IX-XIII. Bathoniae [Bath] and Londini [London]: S. Hazard; J. White.

Steentoft M., Irvine L.M. and Farnham W.F. (1995) Two terete species of *Gracilaria* and *Gracilariopsis* (Gracilariales, Rhodophyta) in Britain. *Phycologia* 34: 113-127.

Sutherland J.E., Lindstrom S.C., Nelson W.A., Brodie J., Lynch M.D., Hwang M.S., Choi H.G., Miyata M., Kikuchi N., Oliveira M.C., Farr T., Neefus C., Mols-Mortensen A., Milstein D. and Müller K.M. (2011) A new look at an ancient order: generic revision of the Bangiales (Rhodophyta). *Journal of Phycology* 47(5): 1131-1151.

Takahashi M.M. (2010) *Sequenciamento e análise do genoma mitocondrial de Gracilaria tenuistipitata (Gracilariales , Rhodophyta).* Universidade de São Paulo.

Ursi S. and Plastino E.M. (2001) Crescimento in vitro de linhagens de coloração vermelha e verde clara de Gracilaria sp. (Gracilariales, Rhodophyta) em dois meios de cultura: análise de diferentes estádios reprodutivos. *Rev. Bras.Bot.* 24:587–594.

Vanderlei E., Araújo I.W.F., Quinderé A.L.G., Fontes B.P., Eloy Y.R.G., Rodrigues J.A.G., Silva A.A.R, Chaves H.V., et al. (2011) The involvement of the HO-1 pathway in the anti-inflammatory action of a sulfated polysaccharide isolated from the red seaweed *Gracilaria birdiae*. *Inflammation research* 60(12): 1121–1130.

Villemur R. (1990a) Circular plasmid DNAs from the red alga *Gracilaria chilensis*. *Curr. Genet.* 18:251–257. doi: 10.1007/BF00318389

Villemur R. (1990b) The DNA sequence and structural organization of the GC2 plasmid from the red alga *Gracilaria chilensis*. *Plant Mol. Biol.* 15:237–243. doi: 10.1007/BF00036910

Wilson H.L. (1910) *Gracilariophila*, a new parasite on *Gracilaria confervoides*. *University of California Publications in Botany* 4: 75-84.

Wynne M.J. (1989) The Re-Instatement of *Hydropuntia* Montagne (Gracilariaceae, Rhodophyta). *Taxon* 38:476. doi: 10.2307/1222295

Yamamoto H. (1975) The relationship between *Gracilariopsis* and *Gracilaria* from Japan. *Bull. Fac. Fish. Hokkaido Univ.* 26: 217-224.

Yamamoto H. (1978) Systematic and anatomical study of the Genus *Gracilaria* in Japan. *Mem. Far. Fish., Hokkaido Univ.* 25:97–152

Yamamoto H. (1984) An evaluation of some vegetative features and some interesting problems in Japonese populations of *Gracilaria*. *Hydrobiologia*. 116/117:51-54.

Yamamoto H. (1986) *Congracilaria babae* gen. et sp. nov. (Gracilariaceae), an Adelphoparasite Growing on *Gracilaria salicornia* of Japan. *Bull. Fac. Fish. Hokkaido Univ.* 37:281–290.

Yang E.C., Kim K.M., Boo G.H., et al (2014a) Complete mitochondrial genome of the agarophyte red alga *Gelidium vagum* (Gelidiales). *Mitochondrial DNA* 25:267–268. doi: 10.3109/19401736.2013.800499

Yang E.C., Kim K.M., Kim S.Y. and Yoon H.S. (2014b) Complete mitochondrial genome of agar-producing red alga *Gracilariopsis chorda* (Gracilariales). *Mitochondrial DNA* 25:339–341. doi: 10.3109/19401736.2013.800502

Yang EC, Kim KM, Kim SY, et al (2015) Highly Conserved Mitochondrial Genomes among Multicellular Red Algae of the Florideophyceae. Genome Biol Evol 7:2394–2406. doi: 10.1093/gbe/evv147

Yang Z. (1994) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *J. Mol. Evol.* 39:306–314. doi: 10.1007/BF00160154

Yokoya N.S. and Oliveira E.C. (1992) Effects of salinity on the growth rate, morphology and water content of some brazilian red algae of economic importance. *Ciencias Mar*. 18:49–64. doi: 10.7773/cm.v18i2.893

Zemke-White W. L. and Ohno M. 1999. World seaweed utilisation: An end-of-century summary. *J. Appl. Phycol.* 11:369–76.

Zhang C.F. and Xia B.M. (1988) On two new *Gracilaria* (Gigartinales, Rhodophyta) from South China. In: *Taxonomy of economic seaweeds with reference to some Pacific and Caribbean species.* (Abbott, I.A. Eds) Vol.2, pp. 127-129. California Sea Grant College Program.

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

Zhang Y., Guo Y.M., Li T.J., Chen C.H., Shen K.N. and Hsiao C. Der (2016) The complete chloroplast genome of *Gracilariopsis lemaneiformis*, an important economic red alga of the family Gracilariaceae. *Mitochondrial DNA Part B Resour*. 1:2–3.

Zhou W., Hu Y., Sui Z., et al (2013) Genome survey sequencing and genetic background characterization of *Gracilariopsis lemaneiformis* (Rhodophyta) based on next-generation sequencing. *PLoS One* 8:e69909. doi: 10.1371/journal.pone.0069909

Supplementary material

Supplementary tables

Table S1 – Notes on annotations done in published mitochondrion genomes from Genbank. “c” indicated reverse strand.

Mitochondrion			
Species	Accession Number	Publication	Re-annotation
Gracilaria changii	KX980031	Song et al. 2017	rrn5-core c(16,149-16,177) annotated. *trn*A-Ile with intron c(20,853-21,413) annotated.
Gracilaria chilensis	KP728466	Lee et al. 2015	rrn5-core c(16,109-16,216) annotated. *trn*W c(24,663-24,736) annotated.
Gracilaria salicornia	KF824534	Campbell et al. 2014	rrn5-core c(16,106-16,136) annotated. *trn*I with intron c(20,807-21,374) annotated.
Gracilaria salicornia	KT373903	Song et al. 2016	rrn5-core c(16,105-16,135) annotated. *trn*I with intron c(20,806-21,373) annotated.
Gracilaria vermiculophylla	KJ526627	Unpublished	rrn5-core c(16,072-16,178) annotated. *trn*I with intron c(21,053-21,619) annotated.
Gracilariopsis andersonii	KX687878	Salomaki & Lane, 2016	*trn*D (7,375-7,448) annotated. *trn*H (12,950-13,026) annotated. rrn5-core c(17,936-17,964) annotated. *trn*I with intron c(22,610-23,150) annotated.
Gracilariopsis chorda	KC875851	Yang et al. 2014	rrn5-core c(16,611-16,716) annotated. *trn*W (25,113-25,185) annotated.
Gracilariopsis lemaneiformis	JQ071938	Zhang et al. 2012	rrn5-core c(16,614-16,719) annotated. *trn*I with intron c(21,340-21,889) annotated.
Gracilariphila oryzoides	KX687879	Salomaki & Lane, 2016	rrn5-core c(16,064-16,092) annotated. *trn*I with intron c(20,745-21,286) annotated. ORF148 (24,586-25,029) annotated.

Table S2 – Notes on annotations done in published chloroplast genomes from Genbank. “c” indicated reverse strand.

Chloroplast			
Species	Accession Number	Publication	Correction
Gracilaria chilensis	KT266788	Lee et al. 2016	tmRNA (58,723-59,098) annotated. *pet*M c(113,793-113,930) annotated.
Gracilaria salicornia	KF861575	Campbell et al. 2014	rrn5 was annotated in the wrong location. Annotated in right location (4,764-4,881). tmRNA (58,949-59,341) annotated. *pet*L (68,392-68,517) annotated. *pet*M c(114,147-114,245) annotated.
Gracilaria tenuistipitata	AY673996	Hagopian et al. 2004	*pet*L (68,440-68,544) annotated. *trn*Me with intron c(69,350-71,547) annotated. *pet*M c(114,358-114,456) annotated.
Gracilariopsis chorda	KX284722	Lee et al. 2016	*pet*P (119,600-119,839) annotated. rnpB c(162,037-162,406) annotated.
Gracilariopsis lemaneiformis	KU179794	Zhang et al. 2016	*ycf*35 (10,540-10,926) annotated. *ycf*20 c(57,548-57,823) annotated. tmRNA (59,321-59,636) annotated. *pet*L (68,817-68,912) annotated. *rpl*9 (169,543-170-007) annotated.

Table S3 – Mitochondrial genes from species used in the present study. Gene present is marked with colour and "+" or name. Gene absent is marked in white. Genes as generally arranged by different colour. Line indicated same locus.

Gene	*G. caudata*	*G. changii*	*G. chilensis*	*G. ferox*	*G. gracilis*	*G. rangiferina*	*G. salicornia* (Hawaii)	*G. salicornia* (Malaysia)	*G. vermiculophylla* (Asia)	*G. vermiculophylla* (USA)	*G. tenuistipitata*	*Gh. oryzoides*	*Gp. andersonii*	*Gp. chorda*	*Gp. lemaneiformis*	*Gp. longissima*	*Gp. mclachlanii*	*Melanthalia intermedia*
Electron transport and oxidative phosphorylation protein genes																		
atp4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
atp6	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
atp8	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
atp9	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
cob	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
cox1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
cox2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
cox3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
nad1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
nad2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
nad3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
nad4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
nad4L	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
nad5	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
nad6	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
sdh2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
sdh3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
sdh4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
tatC	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Ribosomal protein genes																		
rpl16	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpl20	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rps11	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rps12	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rps3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Hypothetical protein genes																		
	orf145	orf148	orf148	orf147	orf145	orf146	orf148	orf148	orf147	orf147	orf1	orf148	orf48	orf37	orf143	orf144	orf148	orf138
														orf08	orf60			
																orf1		
Ribosomal RNA genes																		
rnl	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rns	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rrn5	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Table S3 – Continued.

Gene	*G. caudata*	*G. changii*	*G. chilensis*	*G. ferox*	*G. gracilis*	*G. rangiferina*	*G. salicornia* (Hawaii)	*G. salicornia* (Malaysia)	*G. vermiculophylla* (Asia)	*G. vermiculophylla* (USA)	*G. tenuistipitata*	*Gh. oryzoides*	*Gp. andersonii*	*Gp. chorda*	*Gp. lemaneiformis*	*Gp. longissima*	*Gp. mclachlanii*	*Melanthalia intermedia*
Transfer RNA (tRNA) genes																		
*trn*A(tgc)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*C(gca)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*D(gtc)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*E(ttc)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*F(gaa)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*G(gcc)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*G(tcc)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*H(gtg)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
*trn*I(gat)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*I-intron	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*K(ttt)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*L(taa)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*L(tag)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*M(cat)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*M(cat)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*N(gtt)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*P(tgg)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*Q(ttg)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*R(acg)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*R(tct)	+	+	+	+	+	+		+		+	+			+		+	+	
*trn*R(tct)			+															
*trn*S(gct)	+	+	+	+	+	+		+		+	+			+		+	+	
*trn*S(gct)			+															
*trn*S(tga)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*V(tac)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*W(tca)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*Y(gta)	+	+	+		+	+		+	+	+	+			+		+	+	

Table S4 – Chloroplast genes from species used in present study. Gene present is marked with colour and "+" or name. Gene absent is marked in white. Genes as generally arranged by different colour. Line indicated same locus, except to unique ORFs. Asterisk (*) indicated pseudogene.

Gene	*G. caudata*	*G. chilensis*	*G. ferox*	*G. firma*	*G. gracilis*	*G. rangiferina*	*G. salicornia*	*G. tenuistipitata*	*G. vermiculophylla*	*Gp. lemaneiformis*	*Gp. chorda*	*Gp. longissima*	*Gp. mclachlanii*	*Melanthalia intermedia*
Metabolism, photosynstesis, transport and genetic systems genes														
accA	+	+	+	+	+	+	+	+	+	+	+	+	+	+
accB	+	+	+	+	+	+	+	+	+	+	+	+	+	+
accD	+	+	+	+	+	+	+	+	+	+	+	+	+	+
acpP	+	+	+	+	+	+	+	+	+	+	+	+	+	+
acsF	+	+	+	+	+	+	+	+	+	+	+	+	+	+
apcA	+	+	+	+	+	+	+	+	+	+	+	+	+	+
apcB	+	+	+	+	+	+	+	+	+	+	+	+	+	+
apcD	+	+	+	+	+	+	+	+	+	+	+	+	+	+
apcE	+	+	+	+	+	+	+	+	+	+	+	+	+	+
apcF	+	+	+	+	+	+	+	+	+	+	+	+	+	+
argB	+	+	+	+	+	+	+	+	+	+	+	+	+	+
atpA	+	+	+	+	+	+	+	+	+	+	+	+	+	+
atpB	+	+	+	+	+	+	+	+	+	+	+	+	+	+
atpD	+	+	+	+	+	+	+	+	+	+	+	+	+	+
atpE	+	+	+	+	+	+	+	+	+	+	+	+	+	+
atpF	+	+	+	+	+	+	+	+	+	+	+	+	+	+
atpG	+	+	+	+	+	+	+	+	+	+	+	+	+	+
atpH	+	+	+	+	+	+	+	+	+	+	+	+	+	+
atpI	+	+	+	+	+	+	+	+	+	+	+	+	+	+
bas1	+	+	+	+	+	+	+	+	+	+	+	+	+	+
carA	+	+	+	+	+	+	+	+	+	+	+	+	+	+
cbbX	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ccs1	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ccsA	+	+	+	+	+	+	+	+	+	+	+	+	+	+
cemA	+	+	+	+	+	+	+	+	+	+	+	+	+	+
chlI	+	+	+	+	+	+	+	+	+	+	+	+	+	+
clpC	+	+	+	+	+	+	+	+	+	+	+	+	+	+
cpcA	+	+	+	+	+	+	+	+	+	+	+	+	+	+
cpcB	+	+	+	+	+	+	+	+	+	+	+	+	+	+
cpcG	+	+	+	+	+	+	+	+	+	+	+	+	+	+
cpeA	+	+	+	+	+	+	+	+	+	+	+	+	+	+
cpeB	+	+	+	+	+	+	+	+	+	+	+	+	+	+
dfr	+	+	+	+	+	+	+	+	+	+	+	+	+	+
dnaB	+	+	+	+	+	+	+	+	+	+	+	+	+	+
dnaK	+	+	+	+	+	+	+	+	+	+	+	+	+	+
dsbD	+	+	+	+	+	+	+	+	+	+	+	+	+	+
fabH	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ftrB	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ftsH	+	+	+	+	+	+	+	+	+	+	+	+	+	+
gltB	+	+	+	+	+	+	+	+	+	+	+	+	+	+
groEL	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Table S4 – Continued.

Gene	*G. caudata*	*G. chilensis*	*G. ferox*	*G. firma*	*G. gracilis*	*G. rangiferina*	*G. salicornia*	*G. tenuistipitata*	*G. vermiculophylla*	*Gp. lemaneiformis*	*Gp. chorda*	*Gp. longissima*	*Gp. mclachlanii*	*Melanthalia intermedia*
ilvB	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ilvH	+	+	+	+	+	+	+	+	+	+	+	+	+	+
infB	+	+	+	+	+	+	+	+	+	+	+	+	+	+
infC	+	+	+	+	+	+	+	+	+	+	+	+	+	+
moeB	+	+	+	+	+	+	+	+	+	+	+	+	+	+
nblA	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ntcA	+	+	+	+	+	+	+	+	+	+	+	+	+	+
odpA	+	+	+	+	+	+	+	+	+	+	+	+	+	+
odpB	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ompR	+	+	+	+	+	+	+	+	+	+	+	+	+	+
pbsA	+		+	*	+	+	+	+	+	+	+	+	+	
petA	+	+	+	+	+	+	+	+	+	+	+	+	+	+
petB	+	+	+	+	+	+	+	+	+	+	+	+	+	+
petD	+	+	+	+	+	+	+	+	+	+	+	+	+	+
petF	+	+	+	+	+	+	+	+	+	+	+	+	+	+
petG	+	+	+	+	+	+	+	+	+	+	+	+	+	+
petJ	+	+	+	+	+	+	+	+	+	+	+	+	+	+
petL	+	+	+	+	+	+	+	+	+	+	+	+	+	+
petM	+	+	+	+	+	+	+	+	+	+	+	+	+	+
petN	+	+	+	+	+	+	+	+	+	+	+	+	+	+
petP										+	+	+	+	+
pgmA	+	+	+	+	+	+	+	+	+	+	+	+	+	*
preA	+	+	+	+	+	+	+	+	+	+	+	+	+	+
psaA	+	+	+	+	+	+	+	+	+	+	+	+	+	+
psaB	+	+	+	+	+	+	+	+	+	+	+	+	+	+
psaC	+	+	+	+	+	+	+	+	+	+	+	+	+	+
psaD	+	+	+	+	+	+	+	+	+	+	+	+	+	+
psaE	+	+	+	+	+	+	+	+	+	+	+	+	+	+
psaF	+	+	+	+	+	+	+	+	+	+	+	+	+	+
psaI	+	+	+	+	+	+	+	+	+	+	+	+	+	+
psaJ	+	+	+	+	+	+	+	+	+	+	+	+	+	+
psaK	+	+	+	+	+	+	+	+	+	+	+	+	+	+
psaL	+	+	+	+	+	+	+	+	+	+	+	+	+	+
psaM	+	+	+	+	+	+	+	+	+	+	+	+	+	+
psb30	+	+	+	+	+	+	+	+	+	+	+	+	+	+
psbA	+	+	+	+	+	+	+	+	+	+	+	+	+	+
psbB	+	+	+	+	+	+	+	+	+	+	+	+	+	+
psbC	+	+	+	+	+	+	+	+	+	+	+	+	+	+
psbD	+	+	+	+	+	+	+	+	+	+	+	+	+	+
psbE	+	+	+	+	+	+	+	+	+	+	+	+	+	+
psbF	+	+	+	+	+	+	+	+	+	+	+	+	+	+
psbH	+	+	+	+	+	+	+	+	+	+	+	+	+	+
psbI	+	+	+	+	+	+	+	+	+	+	+	+	+	+
psbJ	+	+	+	+	+	+	+	+	+	+	+	+	+	+
psbK	+	+	+	+	+	+	+	+	+	+	+	+	+	+
psbL	+	+	+	+	+	+	+	+	+					

Table S4 – Continued.

Gene	*G. caudata*	*G. chilensis*	*G. ferox*	*G. firma*	*G. gracilis*	*G. rangiferina*	*G. salicornia*	*G. tenuistipitata*	*G. vermiculophylla*	*Gp. lemaneiformis*	*Gp. chorda*	*Gp. longissima*	*Gp. mclachlanii*	*Melanthalia intermedia*
psbN	+	+	+	+	+	+	+	+	+	+	+	+	+	+
psbT	+	+	+	+	+	+	+	+	+	+	+	+	+	+
psbV	+	+	+	+	+	+	+	+	+	+	+	+	+	+
psbW	+	+	+	+	+	+	+	+	+	+	+	+	+	+
psbX	+	+	+	+	+	+	+	+	+	+	+	+	+	+
psbY	+	+	+	+	+	+	+	+	+	+	+	+	+	+
psbZ	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rbcL	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rbcR	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rbcS	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rne	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rnz	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpoA	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpoB	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpoC1	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpoC2	+	+	+	+	+	+	+	+	+	+	+	+	+	+
secA	+	+	+	+	+	+	+	+	+	+	+	+	+	+
secG	+	+	+	+	+	+	+	+	+	+	+	+	+	+
secY	+	+	+	+	+	+	+	+	+	+	+	+	+	+
sufB	+	+	+	+	+	+	+	+	+	+	+	+	+	+
sufC	+	+	+	+	+	+	+	+	+	+	+	+	+	+
syfB	+	+	+	+	+	+	+	+	+	+	+	+	+	+
syh	+	+	+	+	+	+	+	+	+	+	+	+	+	+
tatC	+	+	+	+	+	+	+	+	+	+	+	+	+	+
thiG	+	+	+	+	+	+	+	+	+	+	+	+	+	+
thiS	+	+	+	+	+	+	+	+	+	+	+	+	+	+
tilS	+	+	+	+	+	+	+	+	+	+	+	+	+	+
trpA	+	+	+	+	+	+	+	+	+	+	+	+	+	+
trpG	+	+	+	+	+	+	+	+	+	+	+	+	+	+
trxA	+	+	+	+	+	+	+	+	+	+	+	+	+	+
tsf	+	+	+	+	+	+	+	+	+	+	+	+	+	+
tufA	+	+	+	+	+	+	+	+	+	+	+	+	+	+
upp	+	+	+	+	+	+	+	+	+	+	+	+	+	+
leuC		+						+						+
leuD		+						+						+
Ribosomal protein genes														
rpl1	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpl11	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpl12	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpl13	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpl14	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpl16	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpl18	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpl19	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpl2	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpl20	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Table S4 – Continued.

Gene	*G. caudata*	*G. chilensis*	*G. ferox*	*G. firma*	*G. gracilis*	*G. rangiferina*	*G. salicornia*	*G. tenuistipitata*	*G. vermiculophylla*	*Gp. lemaneiformis*	*Gp. chorda*	*Gp. longissima*	*Gp. mclachlanii*	*Melanthalia intermedia*
rpl21	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpl22	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpl23	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpl24	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpl27	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpl28	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpl29	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpl3	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpl31	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpl32	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpl33	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpl34	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpl35	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpl36	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpl4	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpl5	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpl6	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rpl9	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rps1	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rps10	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rps11	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rps12	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rps13	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rps14	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rps16	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rps17	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rps18	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rps19	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rps2	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rps20	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rps3	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rps4	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rps5	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rps6	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rps7	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rps8	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rps9	+	+	+	+	+	+	+	+	+	+	+	+	+	+
						Unknown protein genes								
ycf17	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ycf19	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ycf20	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ycf21	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ycf22	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ycf23	*	+	+	+	*			+	+	+	+	+		
ycf29	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ycf3	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Table S4 – Continued.

Gene	*G. caudata*	*G. chilensis*	*G. ferox*	*G. firma*	*G. gracilis*	*G. rangiferina*	*G. salicornia*	*G. tenuistipitata*	*G. vermiculophylla*	*Gp. lemaneiformis*	*Gp. chorda*	*Gp. longissima*	*Gp. mclachlanii*	*Melanthalia intermedia*
ycf33	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ycf34	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ycf35	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ycf36	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ycf37	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ycf38	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ycf39	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ycf4	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ycf41	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ycf45	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ycf46	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ycf52	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ycf53	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ycf54	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ycf55	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ycf58	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ycf60	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ycf61	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ycf63	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ycf65	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ycf80	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ycf91	+	+	+	+	+		+	+	+	+	+	+	+	+
	orf223	orf220	orf220	orf9	orf225	orf219	orf215	orf220	orf225	orf12		orf222	orf222	orf684
	orf439	orf446	orf342	orf8	orf393	orf316	orf450	orf446	orf446	orf11	orf445	orf445	orf327	
	orf241	orf245	orf241	orf6	orf241	orf239	orf240	orf245	orf246	orf9	orf245	orf241	orf241	orf2492
	orf134	orf149	orf134	orf4	orf134	orf511	orf144	orf149	orf135	orf7	orf149*	orf140	orf141	
	orf149	orf160	orf651	orf3	orf655	orf130	orf164	orf160	orf123	orf6	orf160	orf640	orf133	orf162
	orf265	orf65	orf240	orf10	orf230	orf220	ycf65	orf65	orf327	orf222	orf65	orf233	orf234	orf65
Intron genes														
ORFintron1	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ORFintron2	+	+	+		+	+	+	+	+					
Unique ORFs (not in the same locus)														
	orf142			orf23		orf798	orf139			orf4			orf99	orf100
	orf119			orf22		orf915				orf3			orf101	orf111
	orf205			orf21						orf2				orf112
				orf20										orf115
				orf18										orf1202
				orf17										orf138
				orf16										orf145
				orf15										orf147
				orf14										orf159
				orf13										orf200
				orf12										orf212
				orf1										orf218
				orf2										orf227

Table S4 – Continued.

Gene	*G. caudata*	*G. chilensis*	*G. ferox*	*G. firma*	*G. gracilis*	*G. rangiferina*	*G. salicornia*	*G. tenuistipitata*	*G. vermiculophylla*	*Gp. lemaneiformis*	*Gp. chorda*	*Gp. longissima*	*Gp. mclachlanii*	*Melanthalia intermedia*
orf247														
orf249														
orf2493														
orf328														
orf388														
Ribosomal RNA genes														
rnl	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rns	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rrn5	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Transfer RNA (tRNA) genes														
*trn*A(tgc)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*C(gca)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*D(gtc)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*E(ttc)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*F(gaa)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*G(gcc)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*G(tcc)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*H(gtg)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*I(gat)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*K(ttt)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*L(caa)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*L(taa)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*L(tag)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*M(cat)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*M(cat)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*M(cat)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*M-intron	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*N(gtt)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*P(tgg)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*Q(ttg)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*R(acg)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*R(ccg)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*R(tct)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*S(gct)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*S(tga)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*T(ggt)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*T(tgt)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*V(gac)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*V(tac)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*W(cca)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*trn*Y(gta)	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Other RNA genes														
rnpB	+	+	+	+	+	+	+	+	+	+	+	+	+	+
tmRNA	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Genome maps

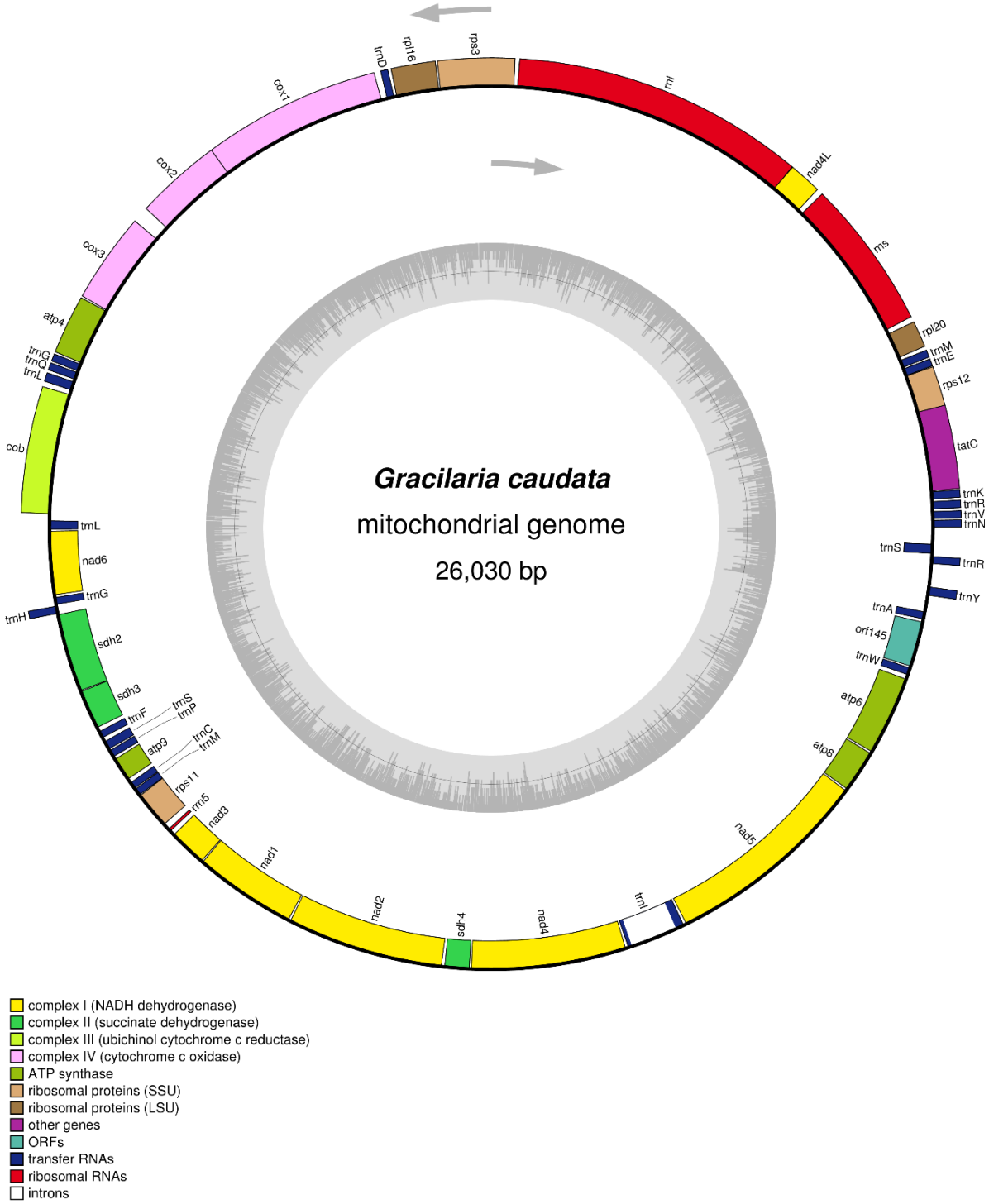

Figure S1 - Mitochondrial genome map of *Gracilaria caudata*. Genes present inside the circle are transcribed in a clockwise direction whereas those outside counter clockwise. The bar graphs on the inner circle reveal GC content in dark grey with the 50% threshold line. The annotated genes are color-coded according to the functional categories listed in the legend.

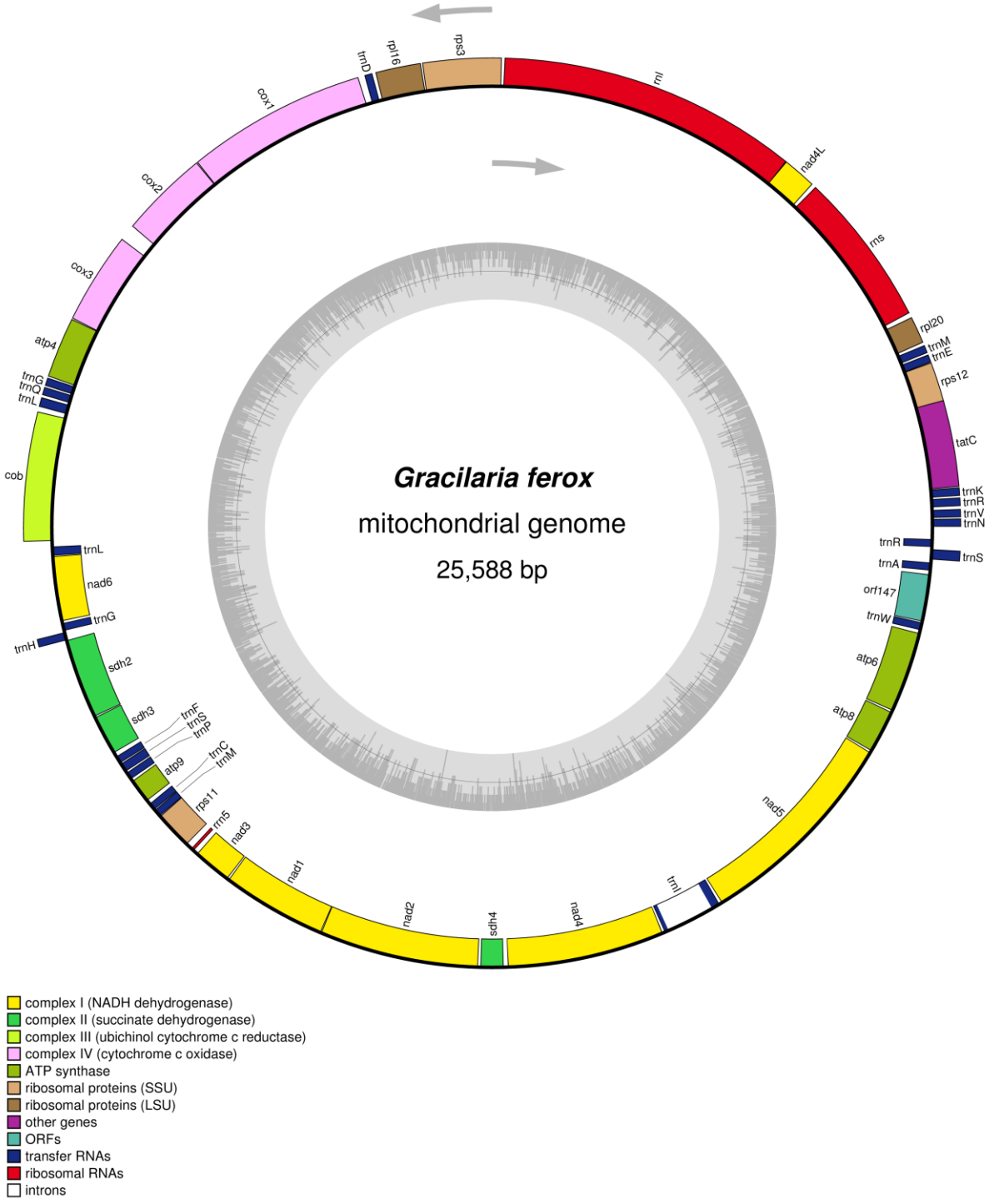

Figure S2 - Mitochondrial genome map of *Gracilaria ferox*. Genes present inside the circle are transcribed in a clockwise direction whereas those outside counter clockwise. The bar graphs on the inner circle reveal GC content in dark grey with the 50% threshold line. The annotated genes are color-coded according to the functional categories listed in the legend.

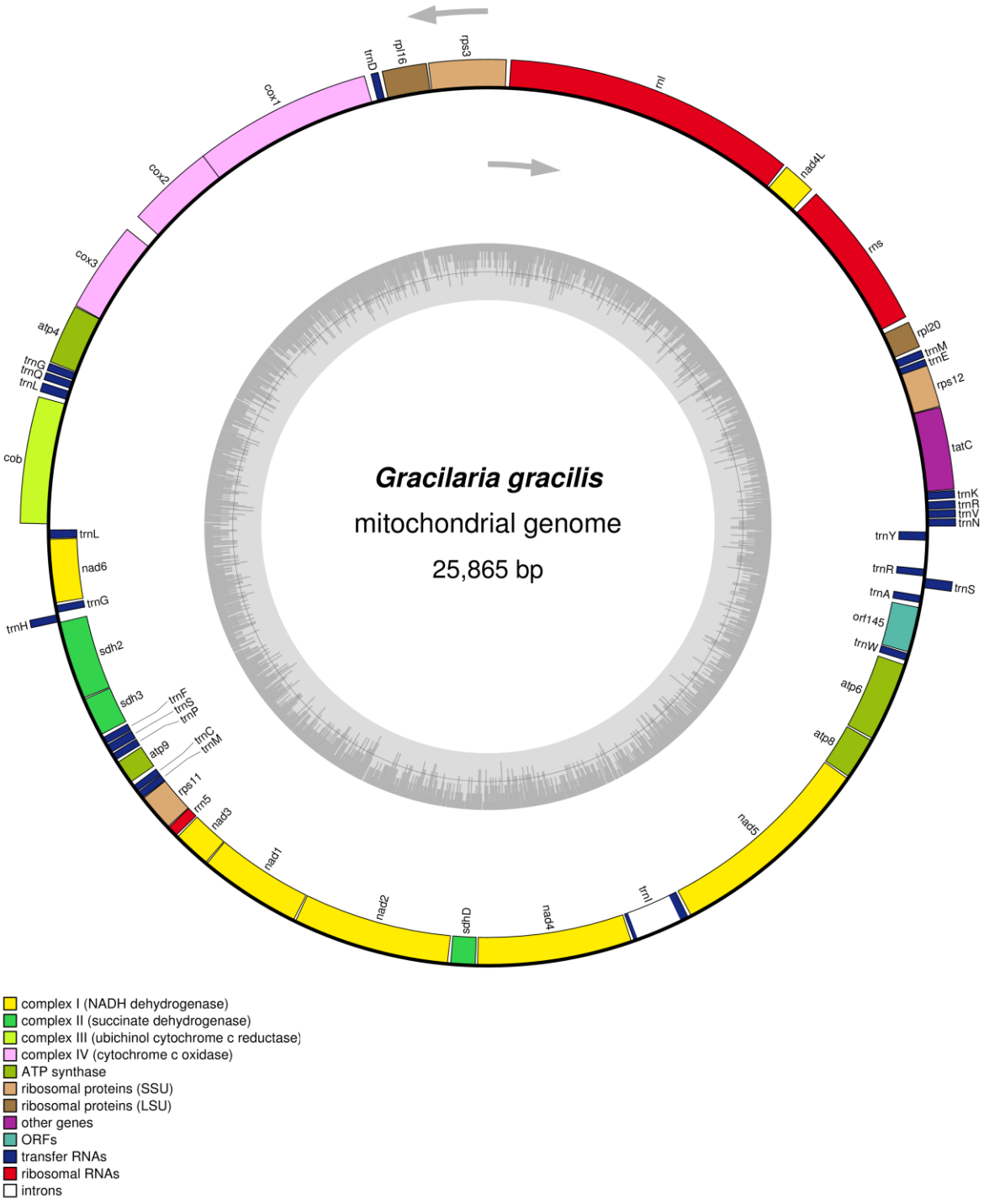

Figure S3 - Mitochondrial genome map of *Gracilaria gracilis*. Genes present inside the circle are transcribed in a clockwise direction whereas those outside counter clockwise. The bar graphs on the inner circle reveal GC content in dark grey with the 50% threshold line. The annotated genes are color-coded according to the functional categories listed in the legend.

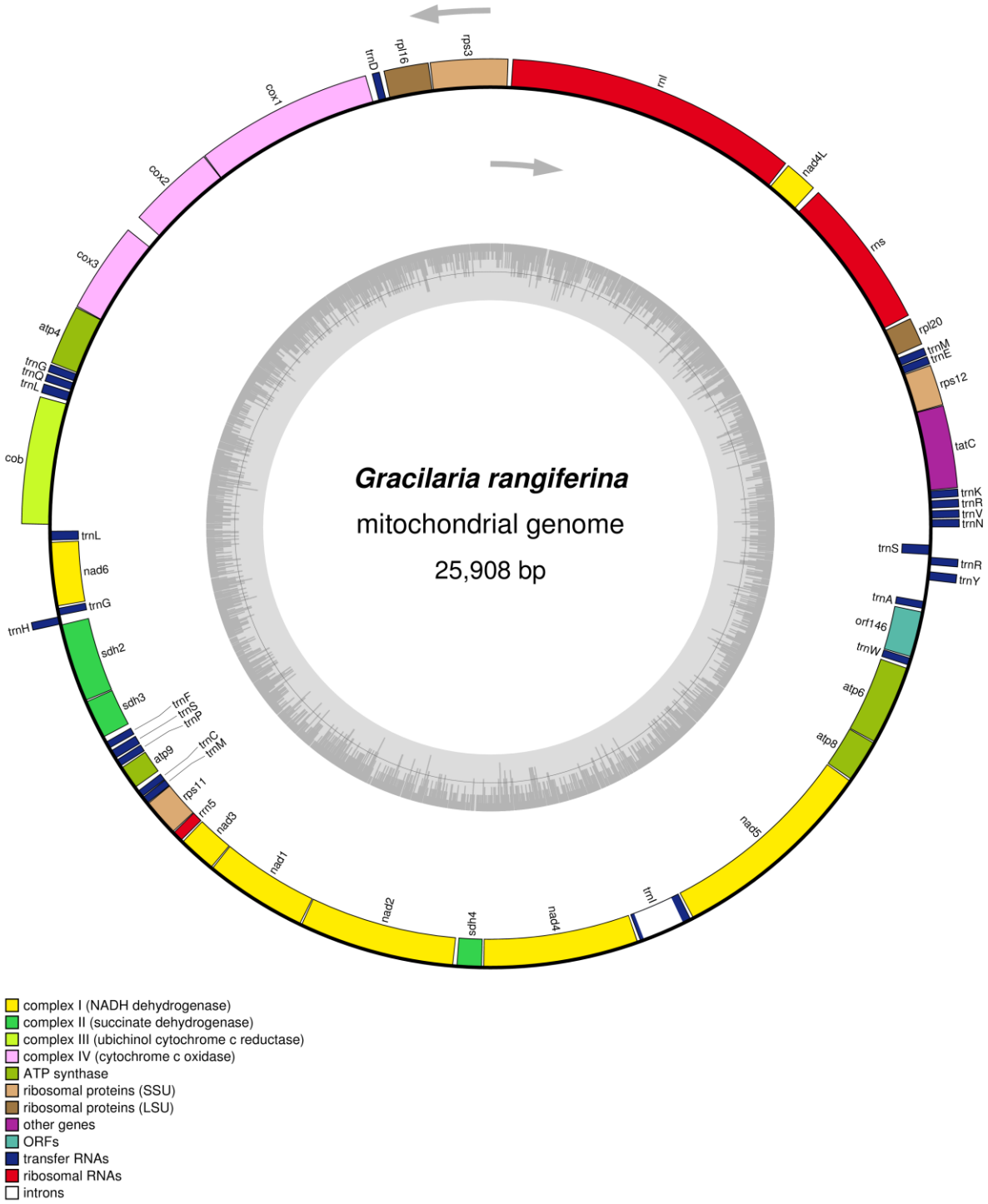

Figure S4 - Mitochondrial genome map of *Gracilaria rangiferina*. Genes present inside the circle are transcribed in a clockwise direction whereas those outside counter clockwise. The bar graphs on the inner circle reveal GC content in dark grey with the 50% threshold line. The annotated genes are color-coded according to the functional categories listed in the legend.

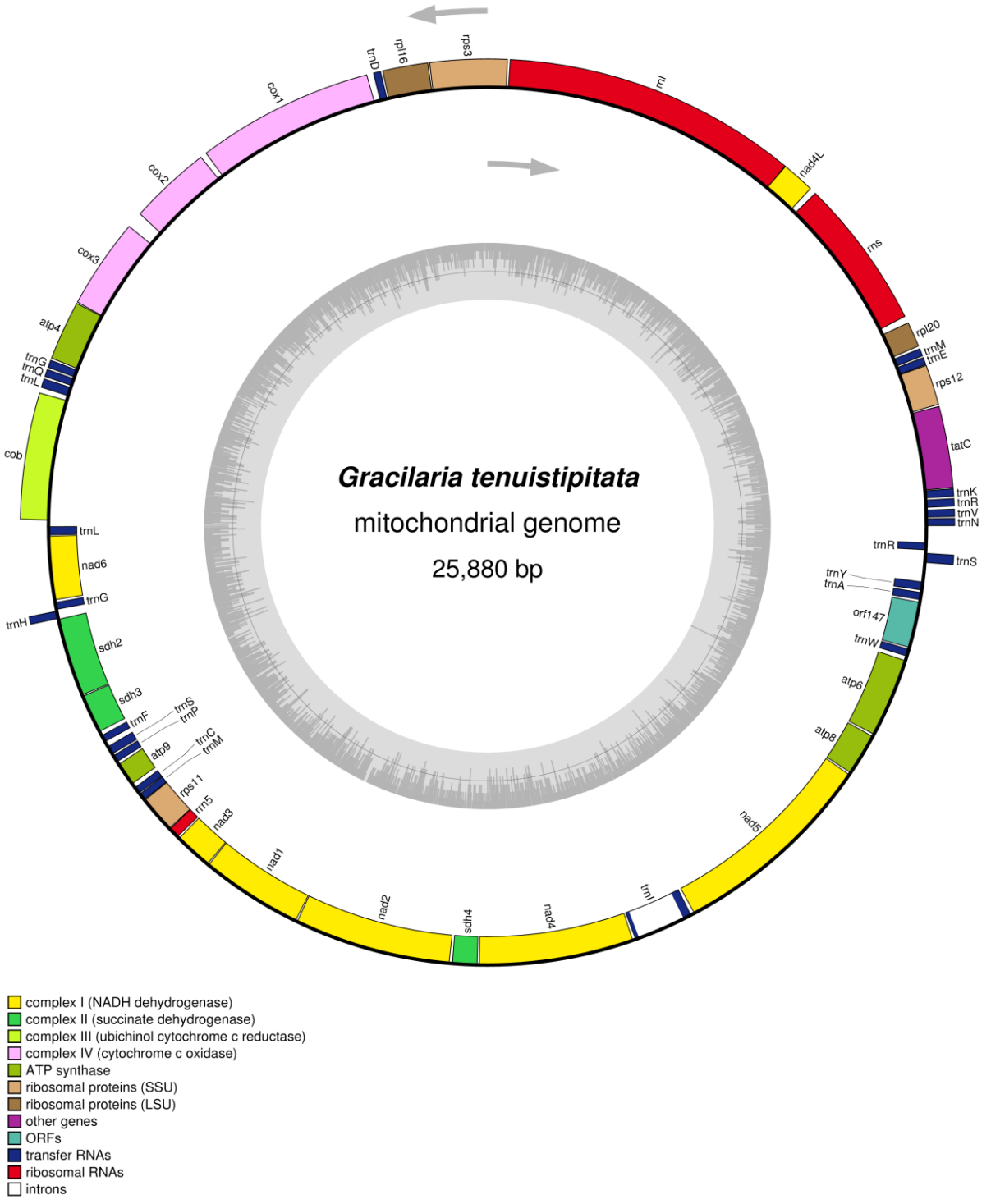

Figure S5 - Mitochondrial genome map of *Gracilaria tenuistipitata*. Genes present inside the circle are transcribed in a clockwise direction whereas those outside counter clockwise. The bar graphs on the inner circle reveal GC content in dark grey with the 50% threshold line. The annotated genes are color-coded according to the functional categories listed in the legend.

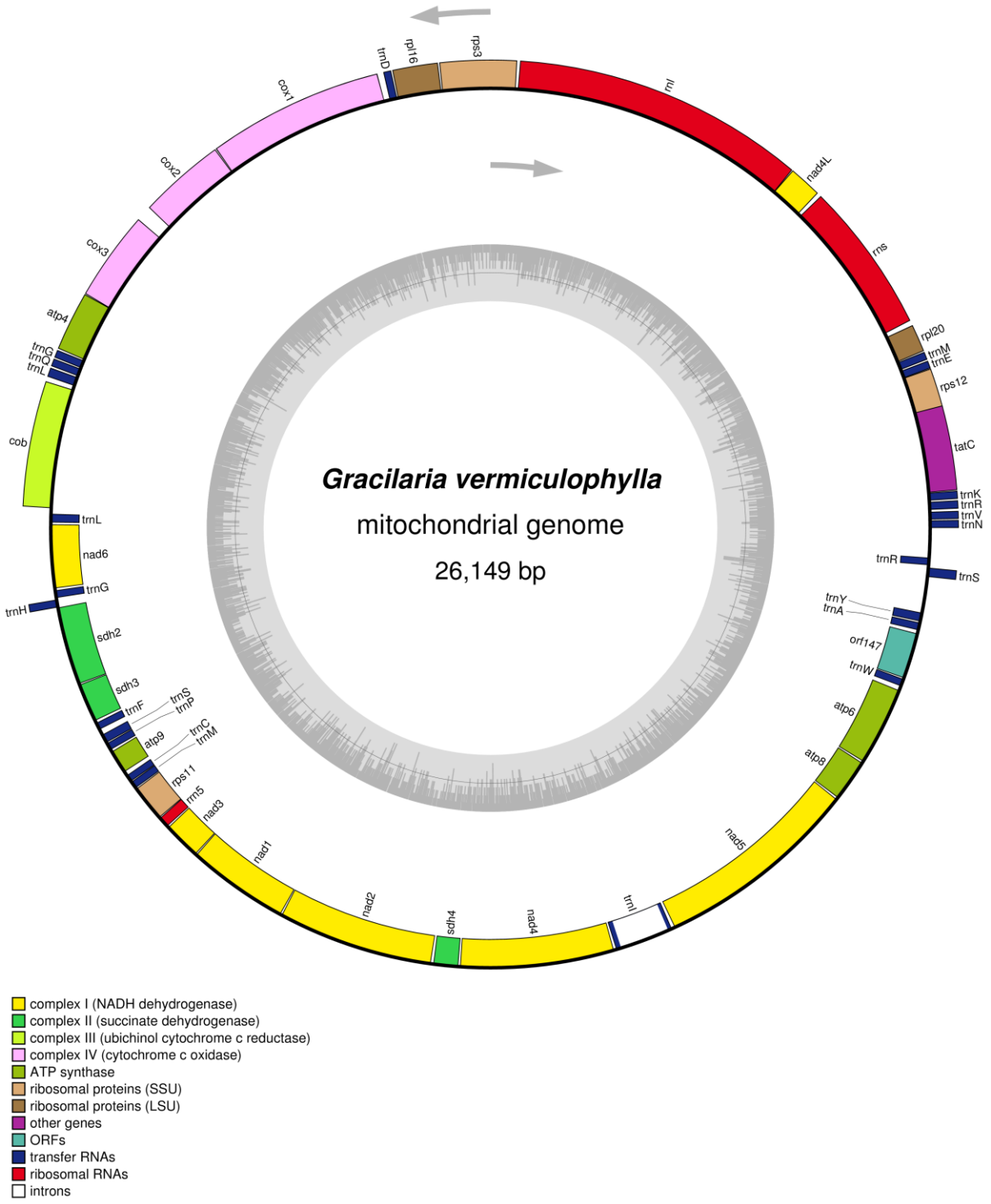

Figure S6 - Mitochondrial genome map of *Gracilaria vermiculophylla*. Genes present inside the circle are transcribed in a clockwise direction whereas those outside counter clockwise. The bar graphs on the inner circle reveal GC content in dark grey with the 50% threshold line. The annotated genes are color-coded according to the functional categories listed in the legend.

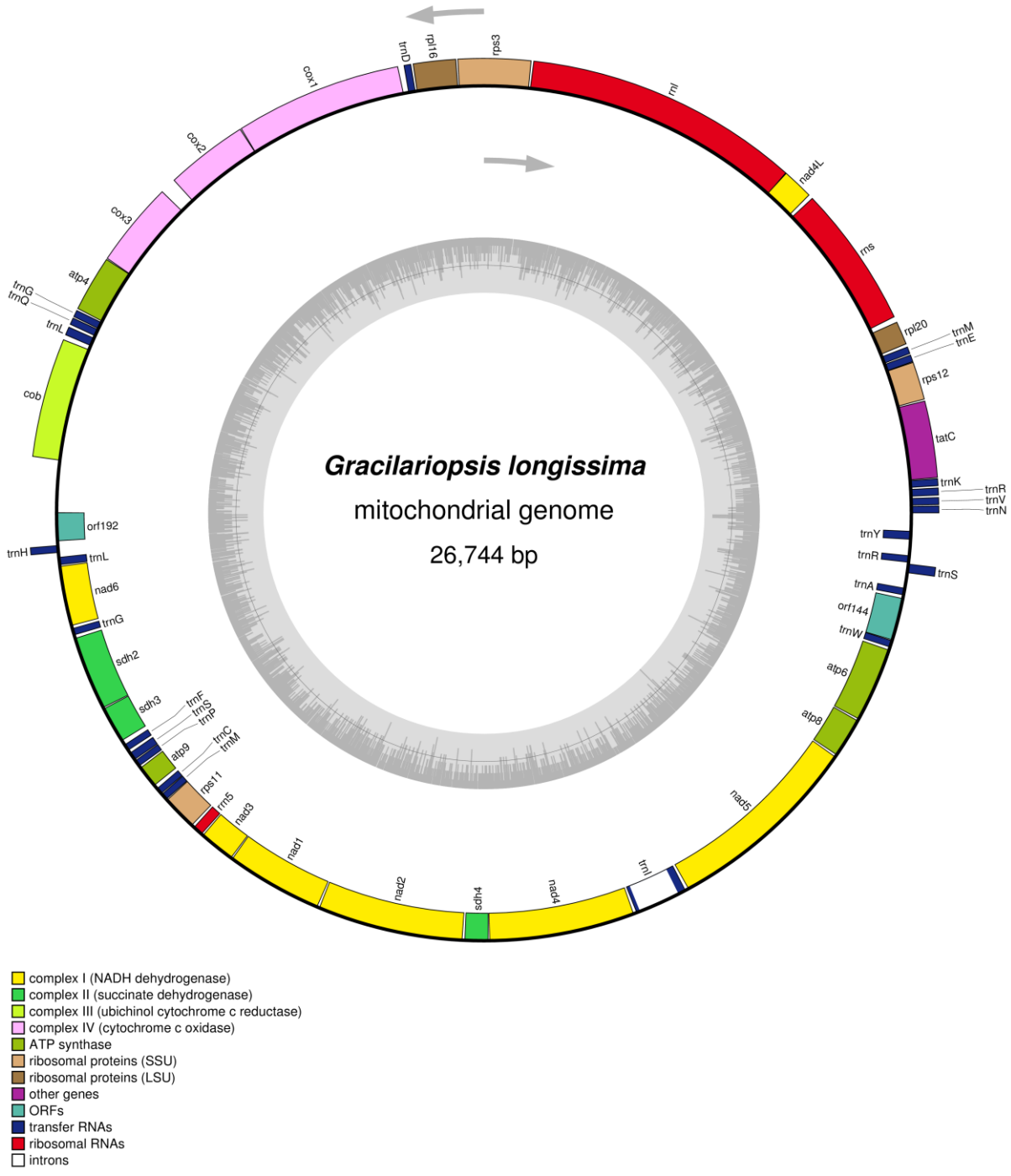

Figure S7 - Mitochondrial genome map of *Gracilariopsis longissima*. Genes present inside the circle are transcribed in a clockwise direction whereas those outside counter clockwise. The bar graphs on the inner circle reveal GC content in dark grey with the 50% threshold line. The annotated genes are color-coded according to the functional categories listed in the legend.

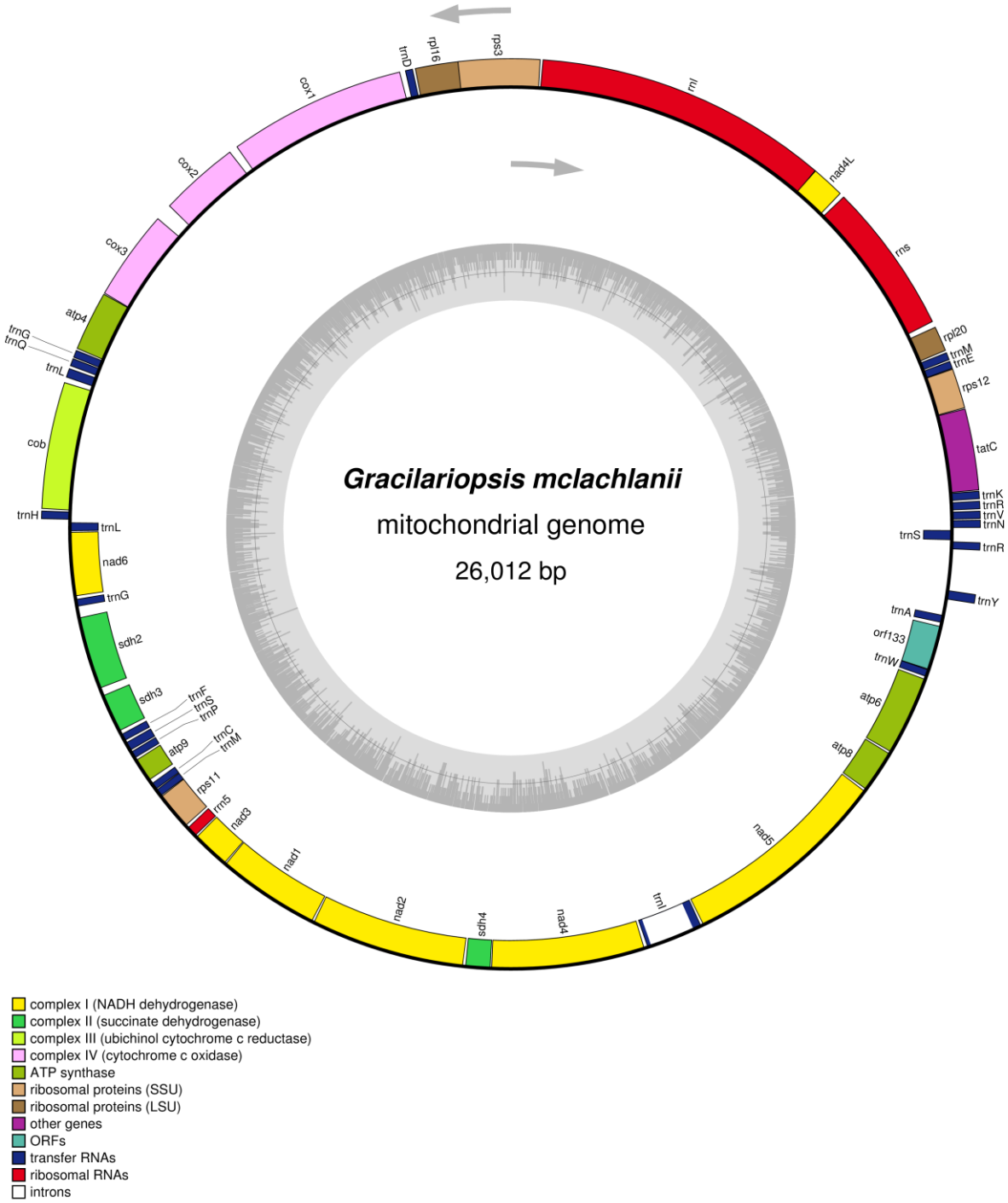

Figure S8 - Mitochondrial genome map of *Gracilariopsis mclachlanii*. Genes present inside the circle are transcribed in a clockwise direction whereas those outside counter clockwise. The bar graphs on the inner circle reveal GC content in dark grey with the 50% threshold line. The annotated genes are color-coded according to the functional categories listed in the legend.

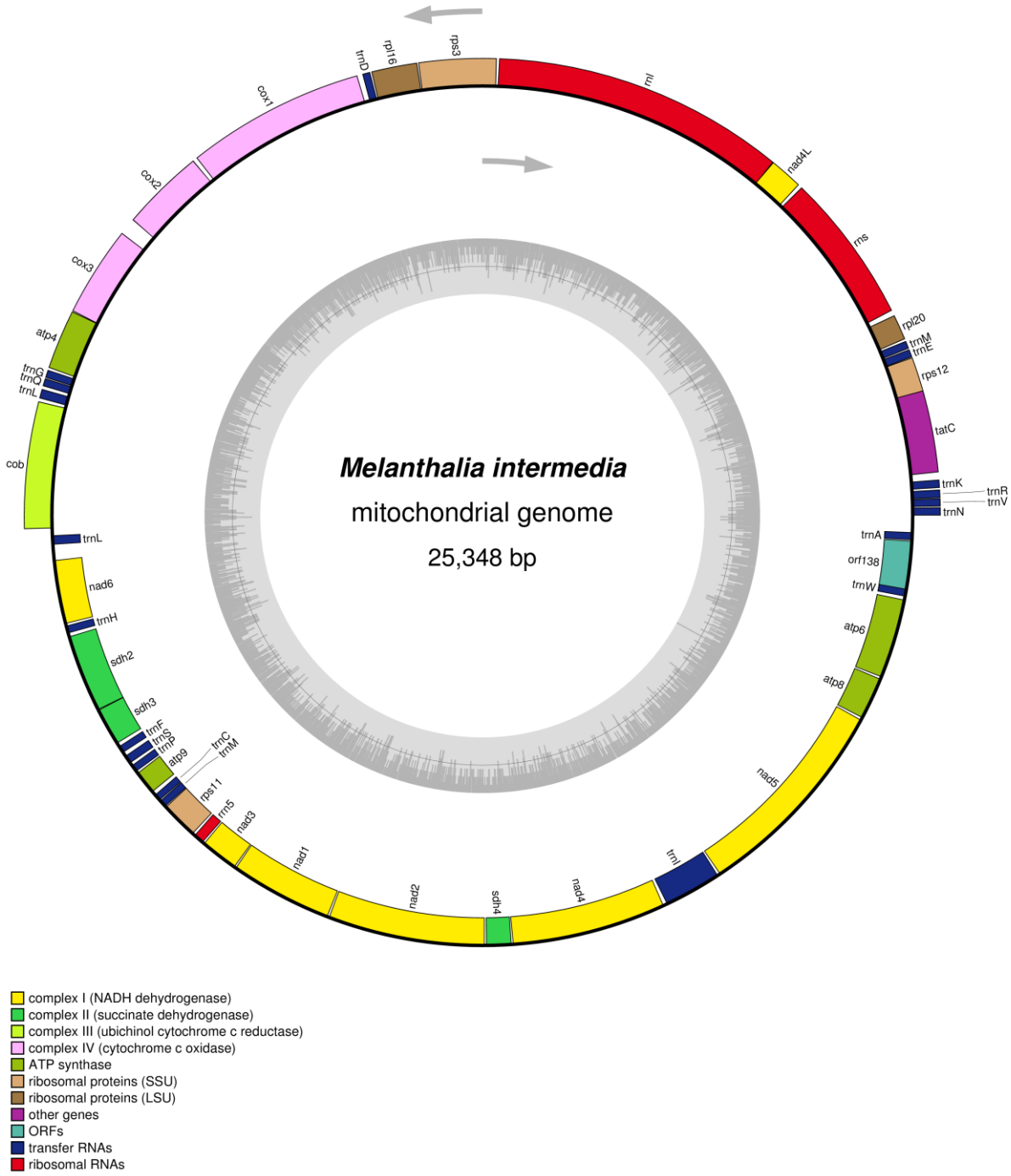

Figure S9 - Mitochondrial genome map of *Melanthalia intermedia*. Genes present inside the circle are transcribed in a clockwise direction whereas those outside counter clockwise. The bar graphs on the inner circle reveal GC content in dark grey with the 50% threshold line. The annotated genes are color-coded according to the functional categories listed in the legend.

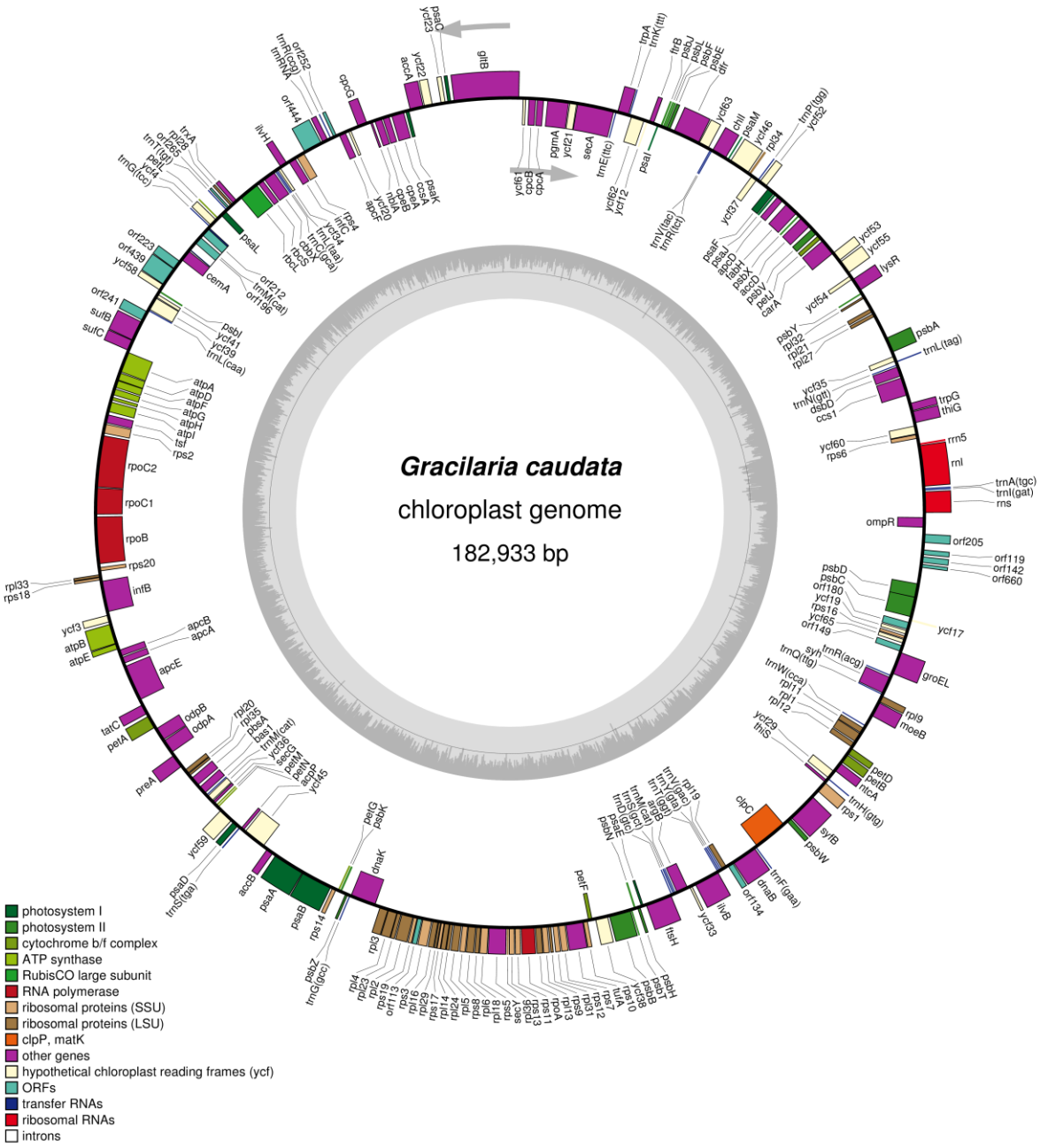

Figure S10 - Chloroplast genome map of *Gracilaria caudata*. Genes present inside the circle are transcribed in a clockwise direction whereas those outside counter clockwise. The bar graphs on the inner circle reveal GC content in dark grey with the 50% threshold line. The annotated genes are color-coded according to the functional categories listed in the legend.

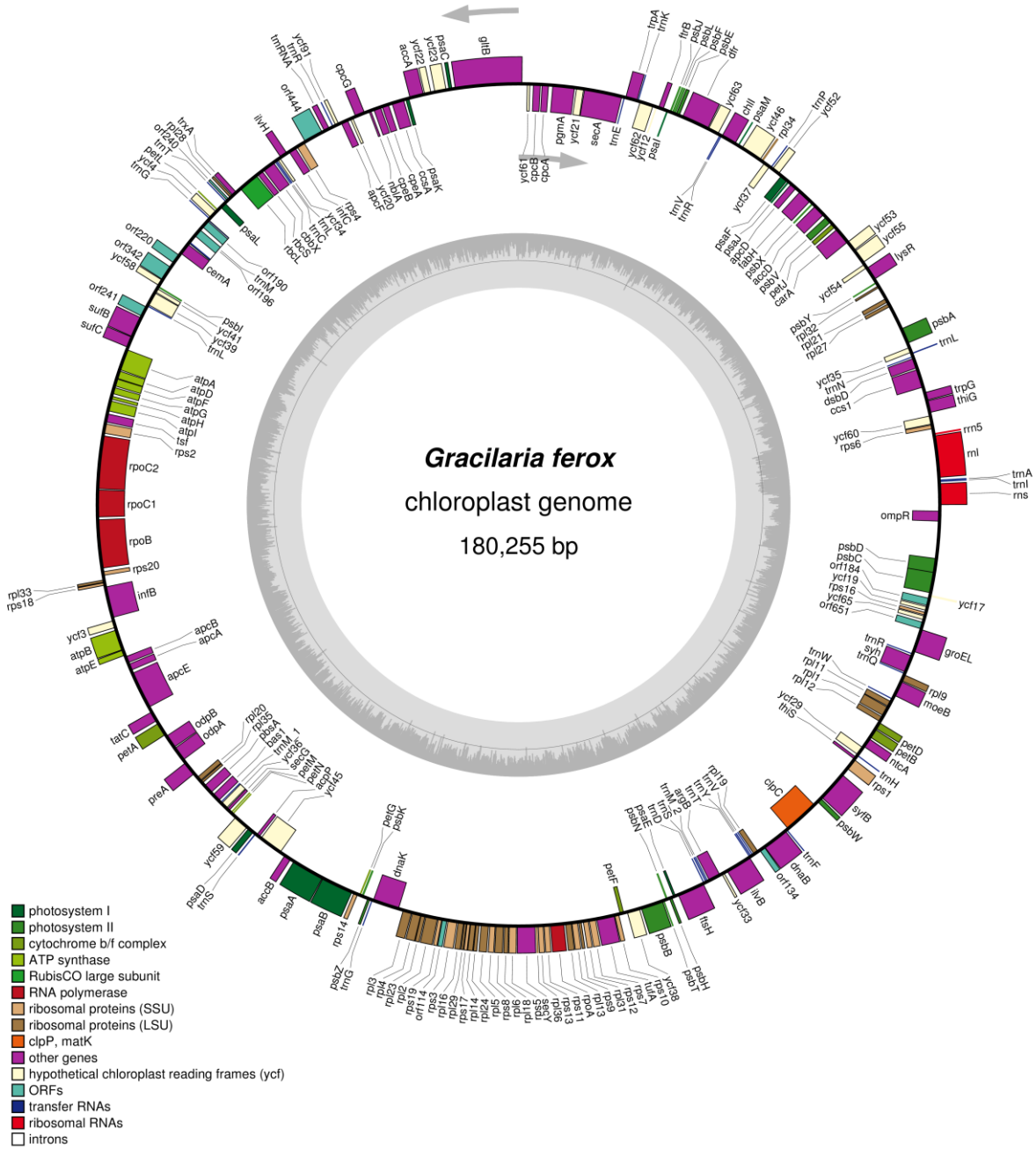

Figure S11 - Chloroplast genome map of *Gracilaria ferox*. Genes present inside the circle are transcribed in a clockwise direction whereas those outside counter clockwise. The bar graphs on the inner circle reveal GC content in dark grey with the 50% threshold line. The annotated genes are color-coded according to the functional categories listed in the legend.

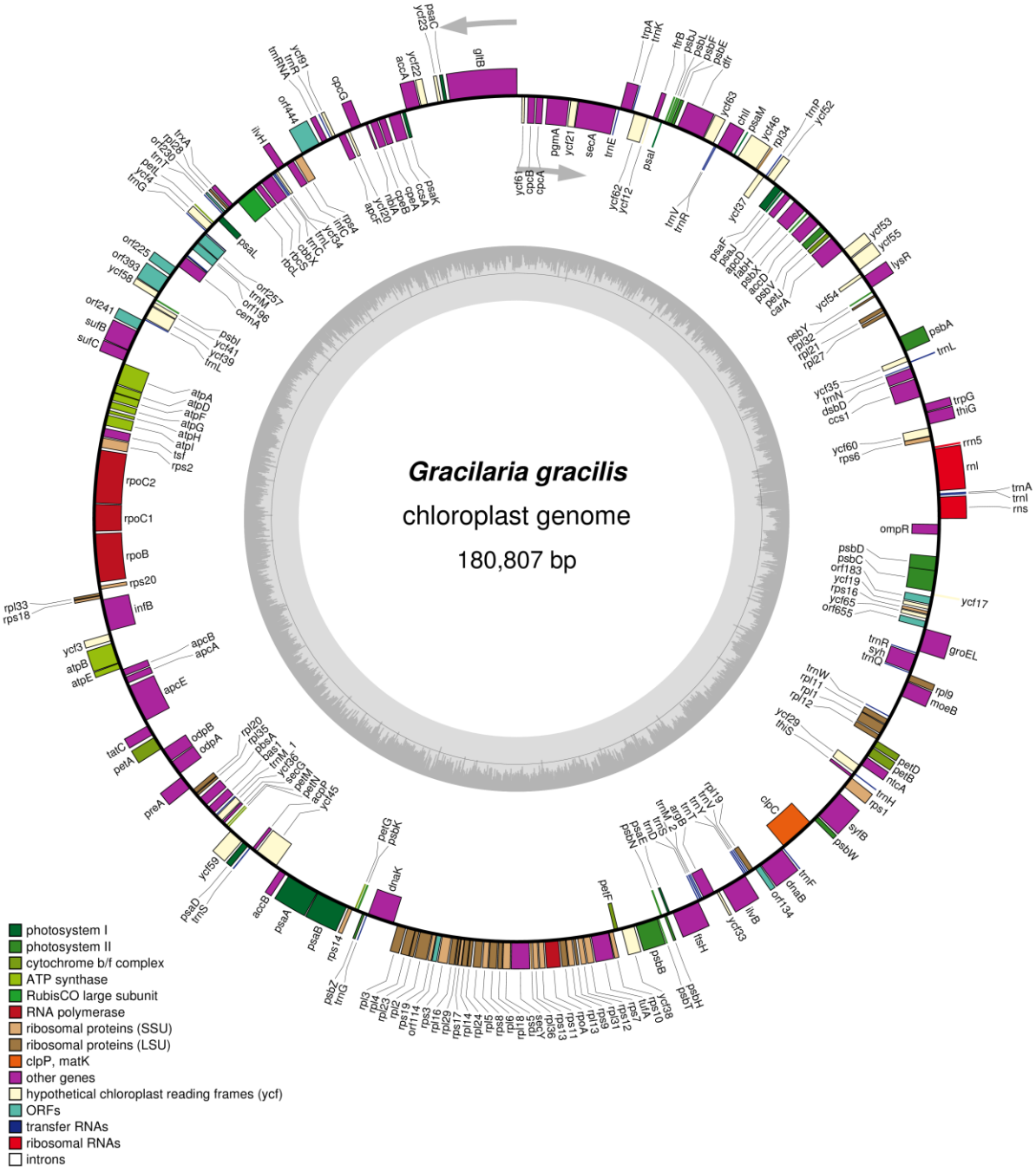

Figure S12 - Chloroplast genome map of *Gracilaria gracilis*. Genes present inside the circle are transcribed in a clockwise direction whereas those outside counter clockwise. The bar graphs on the inner circle reveal GC content in dark grey with the 50% threshold line. The annotated genes are color-coded according to the functional categories listed in the legend.

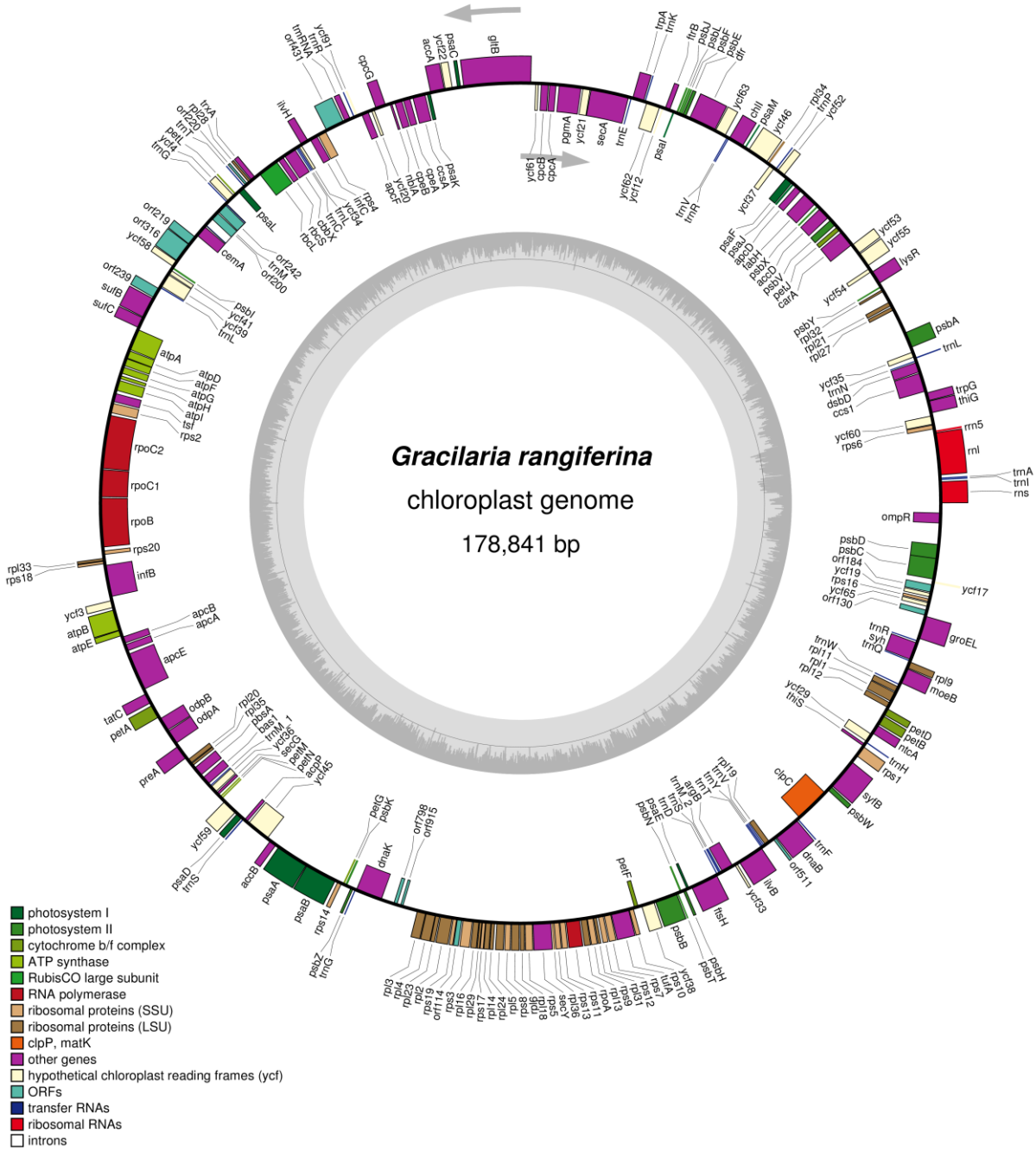

Figure S13 - Chloroplast genome map of *Gracilaria rangiferina*. Genes present inside the circle are transcribed in a clockwise direction whereas those outside counter clockwise. The bar graphs on the inner circle reveal GC content in dark grey with the 50% threshold line. The annotated genes are color-coded according to the functional categories listed in the legend.

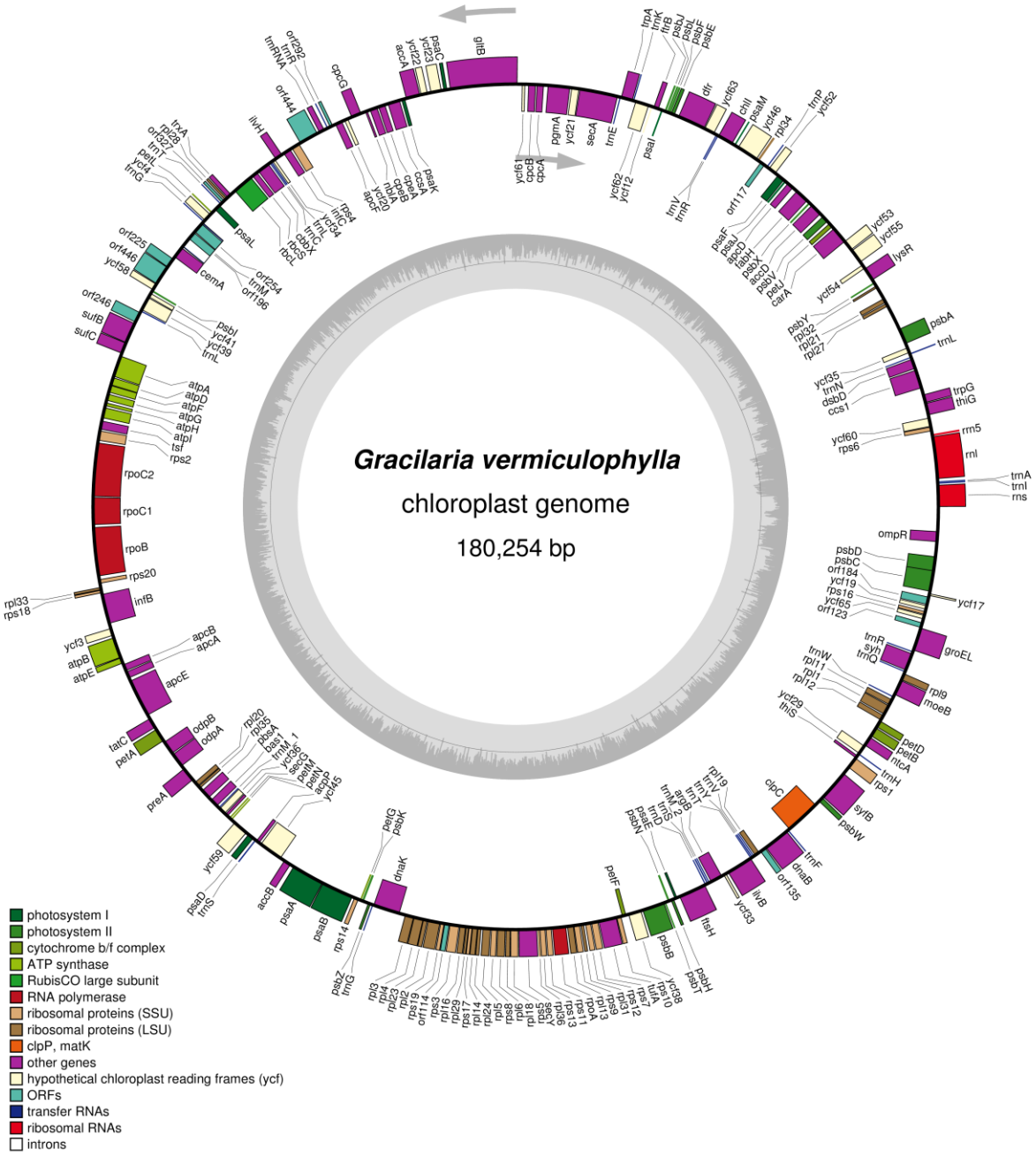

Figure S14 - Chloroplast genome map of *Gracilaria vermiculophylla*. Genes present inside the circle are transcribed in a clockwise direction whereas those outside counter clockwise. The bar graphs on the inner circle reveal GC content in dark grey with the 50% threshold line. The annotated genes are color-coded according to the functional categories listed in the legend.

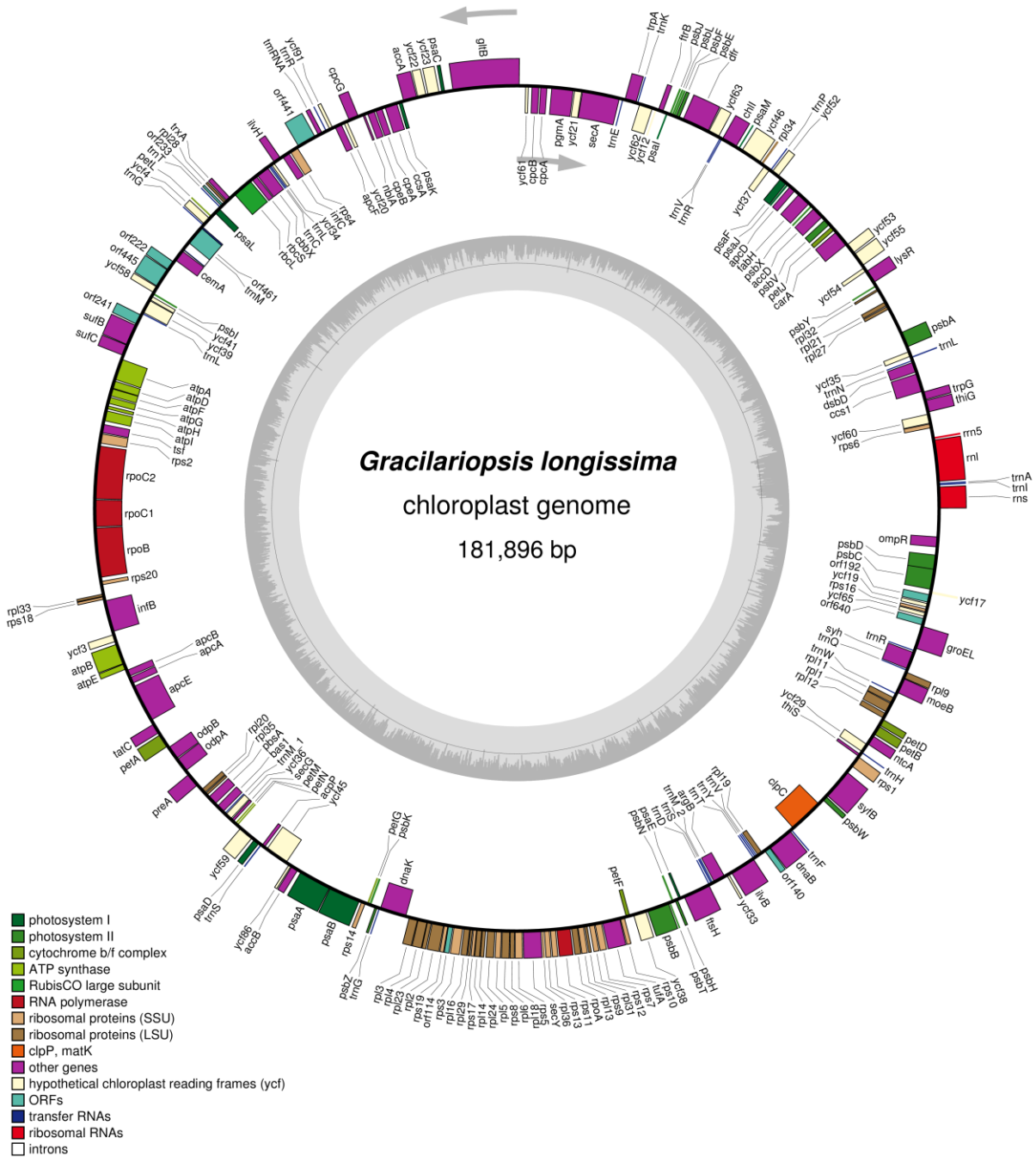

Figure S15 - Chloroplast genome map of *Gracilariopsis longissima*. Genes present inside the circle are transcribed in a clockwise direction whereas those outside counter clockwise. The bar graphs on the inner circle reveal GC content in dark grey with the 50% threshold line. The annotated genes are color-coded according to the functional categories listed in the legend.

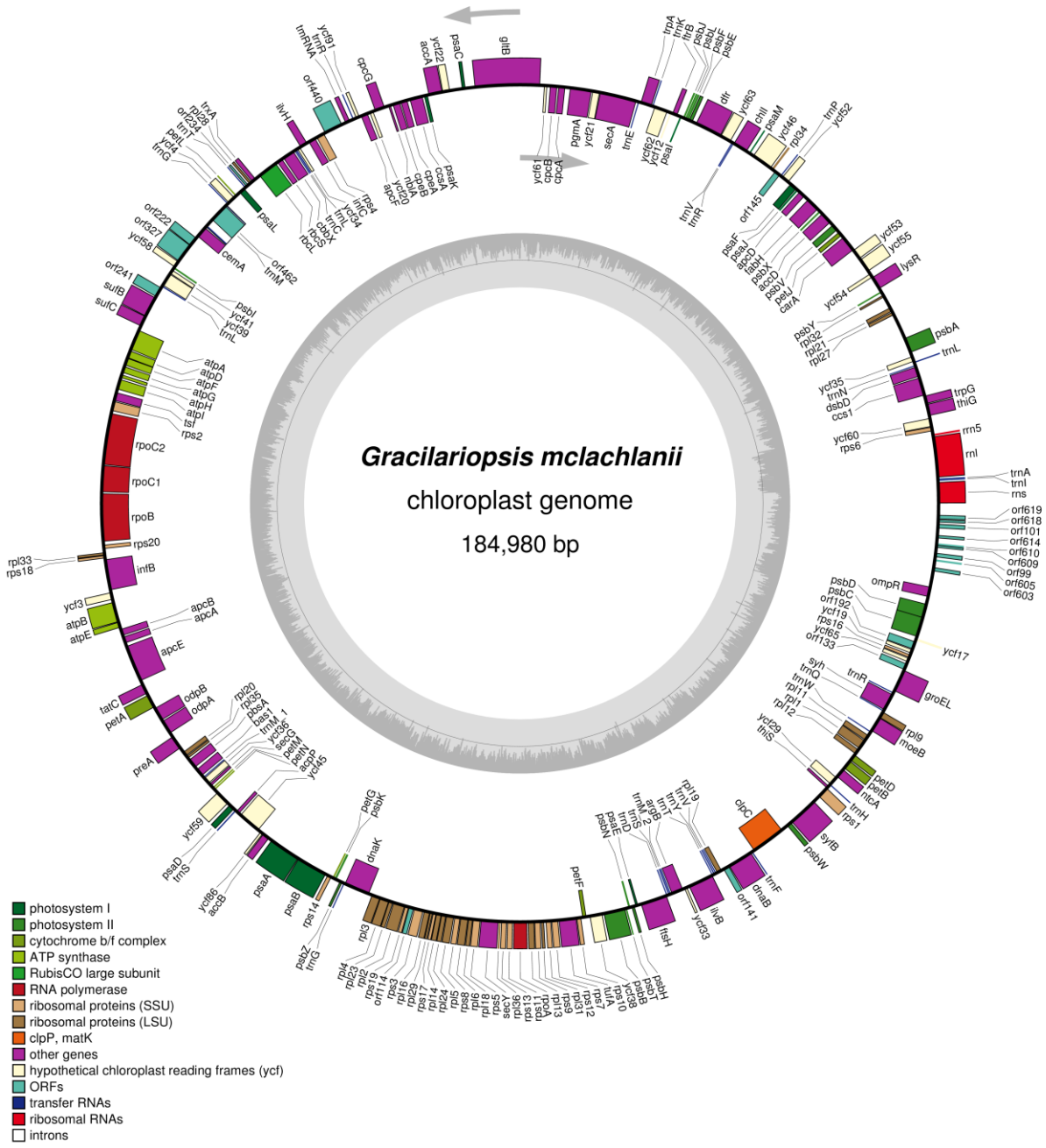

Figure S16 - Chloroplast genome map of *Gracilariopsis mclachlanii*. Genes present inside the circle are transcribed in a clockwise direction whereas those outside counter clockwise. The bar graphs on the inner circle reveal GC content in dark grey with the 50% threshold line. The annotated genes are color-coded according to the functional categories listed in the legend.

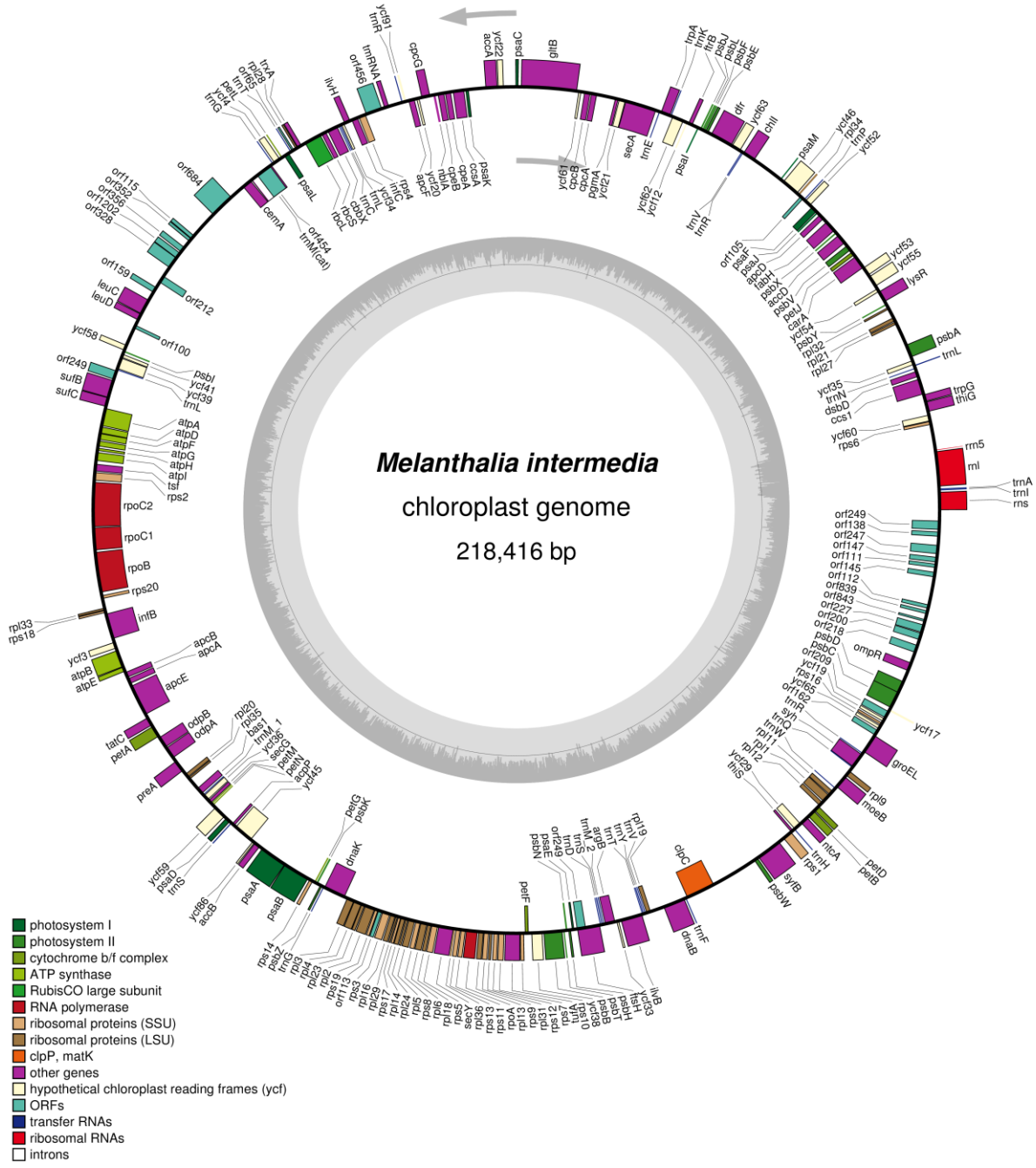

Figure S17 - Chloroplast genome map of *Melanthalia intermedia*. Genes present inside the circle are transcribed in a clockwise direction whereas those outside counter clockwise. The bar graphs on the inner circle reveal GC content in dark grey with the 50% threshold line. The annotated genes are color-coded according to the functional categories listed in the legend.