

Monica Miyuki Takahashi

Sequenciamento e análise do genoma  
mitocondrial de *Gracilaria tenuistipitata*  
(Gracilariales, Rhodophyta)

São Paulo  
2010

Monica Miyuki Takahashi

Sequenciamento e análise do genoma  
mitocondrial de *Gracilaria tenuistipitata*  
(Gracilariales, Rhodophyta)

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

Orientadora: Profa. Dra. Mariana  
Cabral de Oliveira

São Paulo

2010

Takahashi, Monica Miyuki  
T 136s

Sequenciamento e análise do genoma mitocondrial de  
*Gracilaria tenuistipitata* (Gracilariales, Rhodophyta)  
/ Monica Miyuki Takahashi. -- São Paulo : M. M. T.,  
2010. 118 p. : il.

Dissertação (Mestrado) - Instituto de Biociências da  
Universidade de São Paulo. Departamento de  
Botânica, 2010.

1. Gracilariaeae. 2. Rhodophyta. 3. Genomas. 4.  
Sequenciamento genético. I. Universidade de São  
Paulo. Instituto de Biociências. Departamento de  
Botânica. II. Título

LC QK 569.G8  
QK 569.R4

## Comissão Julgadora:

---

Prof(a). Dr(a).

---

Prof(a). Dr(a).

---

Prof(a). Dr.(a).  
Orientador(a)

À minha avó paterna, que sempre me perguntava quando eu pararia de estudar.  
Aos meus pais, que sempre me apoiaram.

“Não são os mais aptos nem os mais inteligentes os que sobrevivem, mas os que se adaptam melhor às mudanças.”

Charles Darwin

“Crescimento significa mudança, e toda mudança implica risco de passar do conhecido ao desconhecido.”

George Shinn

## AGRADECIMENTOS

À Mariana Cabral de Oliveira, pela orientação desde minha iniciação científica, e pela paciência e amizade durante esses anos. Sem suas soluções para os diversos problemas, este trabalho não teria sido concluído.

Ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) pela bolsa de iniciação científica, no início deste projeto. À Fapesp (Fundação de Amparo à Pesquisa do Estado de São Paulo) pela bolsa de mestrado, tornando possível a realização e conclusão deste trabalho.

Ao Instituto de Biociências da USP, em especial ao Departamento de Botânica, pelo programa de pós-graduação.

À minha família pelo apoio incondicional e por compreender as minhas inúmeras ausências. Principalmente aos meus pais, que sempre me incentivaram a estudar e correr atrás dos meus objetivos, além do apoio financeiro e emocional.

Aos meus dois anjos da guarda, Rosario Petti e Silvia Regina Blanco, técnicos do Laboratório de Algas Marinhas Édison José de Paula e do GaTE Lab (Genomics and Transposable Elements), respectivamente. Sem a sua ajuda e amizade, principalmente nos momentos mais difíceis, eu não teria chegado tão longe.

À minha amiga de laboratório Daniela Milstein, pelos diversos conselhos e ensinamentos. Além das muitas horas de descontração. Quase minha co-orientadora.

À professora Marie-Anne Van Sluys, por ceder seu laboratório (GaTE Lab) e pela ajuda de seus alunos. Especial agradecimento aos alunos Alessandro, Andrés, Guilherme e Jonas, que me ajudaram na montagem do genoma.

À Ana Carolina M. Junqueira, aluna de pós-doc do Laboratório de Genética Animal do Centro de Biologia Molecular e Engenharia Genética (CBMEG), na Universidade Estadual de Campinas (Unicamp). Ela me passou todo seu conhecimento sobre PCR de fragmentos longos, trazendo uma nova luz ao meu trabalho.

Aos professores Pio Colepicolo, do Departamento de Bioquímica do Instituto de Química (IQ-USP), e Carlos Frederico Martins Menck, do Departamento de Microbiologia do Instituto de Ciências Biomédicas (ICB-USP), por me permitirem usar seus equipamentos e instalações.

Aos demais professores da Fisiologia: Estela Maria Plastino, Eurico Cabral de Oliveira, Fanly Fungyi Chow Ho e Flávio Berchez, que contribuíram na minha formação

ficológica. Em especial à Fungyi, que me ensinou os primeiros passos no cultivo de algas e na biologia molecular.

Às minhas amigas Camila, Elaine, Fernanda e Natália, que muitas vezes ouviram pacientemente minhas reclamações e estiveram ao meu lado nos momentos bons e ruins.

Aos amigos “de fora” que passaram pelo laboratório e com quem tive a oportunidade de compartilhar conhecimentos: Amélia, Diógina, Lízia, Maria de Fátima, Maria Fernanda, Pedro e Valéria.

Aos meus LAMigos, que fazem minha vida mais alegre e o trabalho mais agradável: Amanda M., Amanda W., Aninha, Bia, Carol, Cintia E., Cintia S., Dani K., Fábio, Fabíola, Guilherme, Henrique, José, Lagosta, Leila, Letícia, Lígia, Luciana, Manu, Marcella, Maria Helena, Natália P., Natália R., Nelsinho, Rose e Suzana.

A todos vocês, muito obrigada.

## ÍNDICE

Listas de abreviaturas .....	09
1. Resumo .....	10
2. Abstract .....	12
3. Introdução .....	14
4. Materiais e Métodos .....	22
4.1. Condições de cultura .....	22
4.2. Extração e purificação de DNA .....	23
4.3. Gel de eletroforese .....	25
4.4. Ultracentrifugação isopícnica em gradiente de CsCl .....	26
4.5. Digestão com enzimas de restrição .....	27
4.6. <i>Primers</i> para PCR e sequenciamento .....	27
4.7. PCR (reação em cadeia da polimerase) e sequenciamento .....	28
4.8. Clonagem de produto de PCR .....	32
4.9. Análise dos <i>contigs</i> .....	33
4.10. Análise das sequências geradas .....	33
4.11. Análise do genoma completo .....	34
5. Resultados e Discussão .....	35
5.1. Obtenção de mtDNA .....	35
5.2. <i>Primer walking</i> .....	37
5.3. Sequenciamento e montagem do genoma mitocondrial .....	40
5.4. Tamanho e genes presentes no mtDNA .....	42
5.5. Genes codificadores de proteínas .....	46
5.6. Genes para RNAs estruturais (rRNA e tRNA) .....	49
5.7. Comparação dos mtDNAs entre Rhodophyta .....	55
5.8. Comparação dos mtDNAs entre organismos fotossintetizantes .....	63
6. Conclusão .....	69
7. Referências .....	70
8. Anexo 1 – Lista dos genes presentes no genoma mitocondrial de <i>Chondrus crispus</i> .....	76
9. Anexo 2 – Estruturas secundárias dos tRNAs de <i>Gracilaria tenuistipitata</i> .....	78
10. Anexo 3 – Genoma mitocondrial completo de <i>Gracilaria tenuistipitata</i> comparado ao genoma mitocondrial completo de <i>Chondrus crispus</i> .....	81

## LISTA DE ABREVIATURAS

- ATP – trifosfato de adenosina
- BLAST – “basic local alignment search tool” ou ferramenta de busca básica baseada em alinhamento local
- BSA – “bovine serum albumine” ou albumina sérica bovina
- Contigs* – “contiguous sequences” ou sequências contíguas
- DDT – ditiotreitol
- EDTA – “ethylenediamine tetraacetic acid” ou ácido etilenodiamino tetra-acético
- EST – “expressed sequence tags” ou marcadores de sequências expressas
- Kb – kilobases
- LB – “luria broth” ou meio de cultura luria
- mtDNA – DNA mitocondrial
- ORF – “open reading frame” ou quadro aberto de leitura
- pb – pares de bases
- PCR – “polymerase chain reaction” ou reação em cadeia da polimerase
- PEG – polietilenoglicol
- rpm – rotações por minuto
- rRNA – RNA ribossômico
- SDS – “sodium dodecyl sulfate” ou dodecil sulfato de sódio
- TA – temperatura de alinhamento
- TM – “temperature of melting” ou temperatura de desnaturação
- tRNA – RNA transportador
- ups – unidade prática de salinidade

## 1. RESUMO

Mitocôndrias são organelas semi-autônomas responsáveis pela respiração celular. Diversas fontes de evidências indicam a origem endossimbiótica dessas organelas, onde um organismo unicelular teria engolfado um organismo procariótico, possivelmente uma α-proteobactéria. Durante a coevolução do endossimbionte e sua célula hospedeira, ocorreu uma redução do genoma do endossimbionte, parte dos genes sendo perdida e parte transferida para o núcleo. A organização e o tamanho do genoma mitocondrial é bastante variável nas diferentes linhagens filogenéticas. O acúmulo de dados moleculares ajuda a esclarecer a origem e evolução das mitocôndrias.

Até o momento, foram sequenciados os genomas mitocondriais de apenas três espécies de Rhodophyta, *Chondrus crispus*, *Cyanidioschyzon merolae* e *Porphyra purpurea*. As algas vermelhas são economicamente importantes por serem utilizadas principalmente como alimento e para as extrações de polissacarídeos e pigmentos acessórios. *Gracilaria tenuistipitata* var. *liui* Zhang et Xia é uma macroalga vermelha que apresenta grande tolerância a fatores ambientais e alta taxa de crescimento, sendo utilizada como organismo modelo em estudos fisiológicos, bioquímicos e moleculares, incluindo o sequenciamento do genoma completo do cloroplasto feito anteriormente pelo nosso grupo de pesquisa.

Este estudo teve como objetivo o sequenciamento completo do genoma mitocondrial de *G. tenuistipitata*, a análise dos genes presentes e a comparação entre os genomas mitocondriais de Rhodophyta e de outros organismos fotossintetizantes disponíveis no GenBank. Para isso, foram realizadas extrações de DNA total para as amplificações por PCR, para posterior sequenciamento por *primer walking* e montagem do genoma mitocondrial completo de *G. tenuistipitata*.

A sequência completa do genoma mitocondrial circular da alga vermelha *G. tenuistipitata* possui 25.565 nucleotídeos e conteúdo CG de 27%. Foram identificados 50 genes, que incluem sete subunidades do complexo NADH desidrogenase (*nad1-6*, *nad4L*), três do complexo succinato desidrogenase (*sdh2-4*), o gene apocitocromo b (*cob*), três subunidades da citocromo c oxidase (*cox1-3*), três do complexo ATP sintase (*atp6*, *atp8-9*), cinco genes para proteínas ribossômicas (*rps3*, *rps11-12*, *rpl16*, *rpl20*), três genes para RNA ribossômico (*rrn5*, *rnl*, *rns*) e 22 para RNA transportador. Um intron do grupo II está inserido no tRNA para histidina. Os genes mitocondriais de *G.*

*tenuistipitata* estão organizados em dois conjuntos codificados um em cada fita do DNA, em direções de transcrição opostas. O genoma mitocondrial de *Gracilaria tenuistipitata* é bastante compacto, com poucas regiões intergênicas e utiliza um código genético modificado.

O conteúdo de genes e sua ordem no genoma mitocondrial de *G. tenuistipitata* são extremamente semelhantes ao de *C. crispus* e os genomas mitocondriais de ambas apresentam identidade de nucleotídeos superior a 70% em todas as análises realizadas. Essa semelhança pode ser explicada pelo fato de ambas as algas pertencerem à mesma classe, Florideophyceae. As quatro espécies de Rhodophyta comparadas neste trabalho possuem os componentes básicos da cadeia respiratória e síntese de ATP, além de algumas proteínas ribossômicas e de um conjunto quase completo de tRNAs e rRNAs.

Este trabalho corrobora estudos filogenéticos anteriores em relação às algas vermelhas, evidenciando que *G. tenuistipitata* e *C. crispus* são mais próximos entre si e igualmente distantes de *P. purpurea*, e que *C. merolae* se encontra em uma posição mais basal.

## 2. ABSTRACT

Mitochondria are semi-autonomous organelles responsible for cellular respiration. Many sources of evidence indicate the endosymbiotic origin of these organelles, in which an unicellular organism engulfed a prokaryotic organism, possibly an  $\alpha$ -proteobacteria. During the endosymbiont and its host coevolution, the endosymbiont genome was reduced by gene loss and gene transfer to the nucleous. The mitochondrial genome size and organization varies within different phylogenetic lineages. The accumulation of molecular data helps to elucidate the origin and evolution of mitochondria.

So far, only three species of Rhodophyta had their mitochondrial genome totally sequenced, *Chondrus crispus*, *Cyanidioschyzon merolae* and *Porphyra purpurea*. Red algae are economically important due to their use as food and for the extraction of polysacarids and accessory pigments. *Gracilaria tenuistipitata* var. *liui* Zhang et Xia is a red macroalgae very tolerant to environmental changes and shows high growth rates, being used as model organism in physiological, biochemical and molecular studies, including the sequencing of the whole chloroplast genome done by our research group.

This study aimed the sequencing of the whole mitochondrial genome of *G. tenuistipitata*, the analysis of its gene content and the comparison within mitochondrial genomes of Rhodophyta and other photosynthetic organisms available at the GenBank. For that, total DNA was extracted for PCR amplifications, which were sequenced using primer walking and assembled to obtain the complete mitochondrial genome of *G. tenuistipitata*.

The whole mitochondrial circular genome sequence of the red algae *G. tenuistipitata* was determined (25.565 nucleotides, C+G content 27%). Fifty genes were identified, including seven NADH dehydrogenase complex subunits (*nad1-6*, *nad4L*), three succinate dehydrogenase complex subunits (*sdh2-4*), the apocytochrome b gene (*cob*), three cytochrome c oxidase subunits (*cox1-3*), three ATP synthase complex subunits (*atp6*, *atp8-9*), five ribosomal proteins (*rps3*, *rps11-12*, *rpl16*, *rpl20*), three ribosomal RNA genes (*rrn5*, *rnl*, *rns*) and 22 tRNAs. One group II introns was found between the tRNA<sup>His</sup>. The mitochondrial genes of *G. tenuistipitata* are organized in two groups translated each one in each strand, in two opposite directions. The mitochondrial

genome of *Gracilaria tenuistipitata* is quite compact, with few intergenic regions and uses a modified genetic code.

The gene content and its order in the mitochondrial genome of *G. tenuistipitata* are extremely similar to the *C. crispus* mitochondrial genome, and both genomes presented nucleotide identity above 70% in all the analyses. This similarity between the mtDNA of both species can be explained by the fact that both belong to the same class, Florideophyceae. The four Rhodophyta species compared in this study have the essential components for the respiratory chain and ATP synthesis, besides some ribosomal proteins and almost all of the tRNAs and rRNAs.

This work corroborate previous phylogenetic studies about red algae, showing that *G. tenuistipitata* and *C. crispus* are more closely related between each other than to *P. purpurea*, and that *C. merolae* is in a basal position.

### 3. INTRODUÇÃO

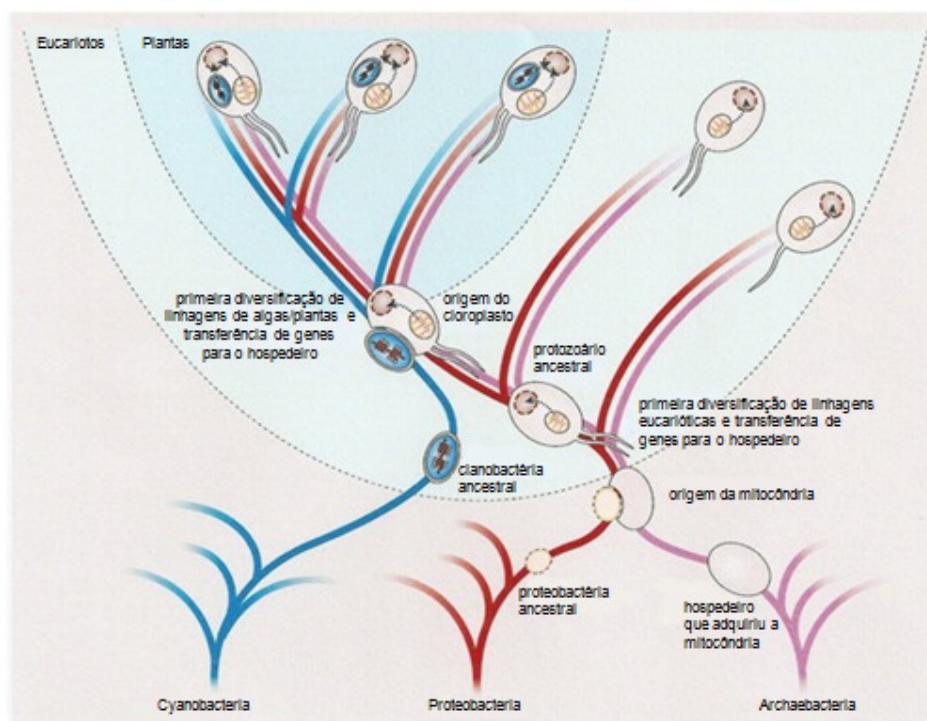
Mitocôndrias e cloroplastos são organelas citoplasmáticas cuja função primária é a conversão de energia. A mitocôndria é responsável pela respiração, enquanto o cloroplasto pela fotossíntese. Ambos os processos acoplam a transferência de elétrons à produção de ATP, usando a força próton-motora que ocorre através de suas membranas (Gillham, 1994; Alberts *et al.* 1997; Allen, 2003).

Mitocôndrias e cloroplastos possuem um sistema genético próprio contendo DNA, RNA e os componentes necessários para a replicação do DNA, para a transcrição de RNA e para a síntese de proteínas traduzidas em seu interior. Porém, essas organelas necessitam também de componentes protéicos codificados e transcritos no núcleo e traduzidos no citoplasma para operarem adequadamente, sendo portanto, descritas como organelas “semi-autônomas” (Gillham, 1994; Alberts *et al.* 1997; Allen, 2003; Adams & Palmer, 2003). A manutenção de um sistema genético na mitocôndria requer inúmeras proteínas envolvidas na replicação do DNA, reparo, recombinação, transcrição, tradução, e regulação gênica (Adams & Palmer, 2003).

As semelhanças entre a respiração bacteriana e as funções mitocondriais, assim como entre a fotossíntese de cianobactérias e as funções plastidiais nas plantas, aliadas aos dados moleculares e de ultraestrutura, apresentam evidências muito convincentes da origem bacteriana dessas organelas eucarióticas (Andersson *et al.* 2003).

Atualmente, a teoria endossimbiótica para a origem da mitocôndria e do cloroplasto é amplamente aceita. Segundo a hipótese tradicional, a mitocôndria teria se originado através de um evento endossimbiótico onde um organismo eucariótico teria engolfado um organismo procariótico, possivelmente uma  $\alpha$ -proteobactéria (Andersson *et al.* 2003). Uma hipótese alternativa sugere que a célula hospedeira seria uma arqueobactéria (Figura 1), e que a origem da mitocôndria tenha sido a força motriz para a origem da membrana nuclear e, portanto, dos eucariotos (Clark, 1999; Timmis *et al.* 2004). Ainda não está completamente estabelecido se esse evento ocorreu uma única vez (origem monofilética) ou mais de uma vez ao longo da evolução (Leblanc *et al.* 1997). Entretanto, dados moleculares, como a comparação dos genes que codificam para a subunidade pequena do RNA ribossômico (SSU rDNA), apontam para uma origem monofilética das mitocôndrias (Leblanc *et al.* 1995; Gray, 1999; Gray *et al.* 1999; Adams & Palmer, 2003). Evidências indicam ainda que eucariotos amitocondriados

podem um dia ter portado mitocôndrias. Estudos recentes detectaram a presença de genes que codificam proteínas tipicamente mitocondriais (como as chaperonas) no genoma nuclear de protistas amitocondriados (Gray *et al.* 1999). Diversas dessas proteínas mitocondriais foram encontradas também em uma organela denominada hidrogenossomo, que produz ATP anaerobicamente, sugerindo que mitocôndrias e hidrogenossomos possuem um ancestral em comum (Gray, 1999; Gray *et al.* 1999). Sendo assim, a origem da mitocôndria representa a relação simbiótica mais bem sucedida descrita até hoje (Andersson *et al.* 2003).



**Figura 1.** Diagrama esquemático representando a origem de mitocôndrias e cloroplastos, e a evolução dos eucariotos. A mitocôndria surgiu de uma  $\alpha$ -proteobactéria e o cloroplasto surgiu de uma cianobactéria. O hospedeiro seria, nesse exemplo, uma arqueobactéria unicelular. A figura também demonstra a transferência de genes que teria ocorrido entre os genomas organelares e o núcleo. Modificado de Timmis *et al.* (2004).

O sequenciamento do genoma de *Rickettsia prowazekii* (Andersson *et al.*, 1998), uma  $\alpha$ -proteobactéria parasita intracelular causadora do tifo em humanos, corrobora a teoria endossimbiótica para a mitocôndria. Tanto o parasita quanto a mitocôndria não

possuem genes para a glicólise anaeróbica, porém ambos apresentam genes para o ciclo do ácido tricarboxílico e para o complexo da cadeia respiratória. Em ambos, a produção de ATP ocorre da mesma maneira (Andersson *et al.*, 1998).

Na origem da mitocôndria, após o evento endossimbiótico e durante a coevolução do endossimbionte e sua célula hospedeira, ocorreu uma redução no genoma do endossimbionte, uma vez que grande parte do seu genoma foi perdida e centenas de genes foram transferidos para o núcleo (Figura 1). Outras centenas de proteínas eucarióticas, cujas sequências estão localizadas no genoma nuclear, foram eventualmente recrutadas para a mitocôndria (Adams & Palmer, 2003; Allen, 2003; Andersson *et al.* 2003).

Se a maioria dos genes foi perdida ou transferida para o núcleo, por que não todos? Muitas hipóteses foram levantadas acerca desta questão. Uma hipótese muito discutida é que algumas proteínas altamente hidrofóbicas seriam difíceis de importar para dentro da mitocôndria após sua tradução no citoplasma, e algumas poderiam ser desviadas para o retículo endoplasmático. Uma segunda hipótese é que alguns produtos mitocondriais seriam tóxicos quando presentes no citosol. Outra hipótese é que a expressão dos produtos para o transporte de elétrons e o acoplamento energético deveria ser rápida e diretamente regulada pelo estado redox da mitocôndria. E uma última hipótese é que o código genético usado pela mitocôndria em muitos eucariotos, incluindo animais, impediria novas transferências gênicas para o núcleo. Todas essas hipóteses ainda não foram devidamente evidenciadas experimentalmente (Adams & Palmer, 2003).

A organização e o tamanho do genoma mitocondrial é bastante variável nas diferentes linhagens filogenéticas dos eucariotos. Apesar do tamanho variar entre cerca de 16.000 pares de base (16 Kb) em humanos a até 500 Kb em algumas espécies de plantas, a maioria dos genomas mitocondriais contêm entre 12 e 20 genes codificadores de proteínas. A grande parte desses genes codifica componentes chaves da respiração aeróbica e tradução, como as citocromo oxidases e proteínas ribossômicas (Andersson *et al.* 2003; Bullerwell & Gray, 2004).

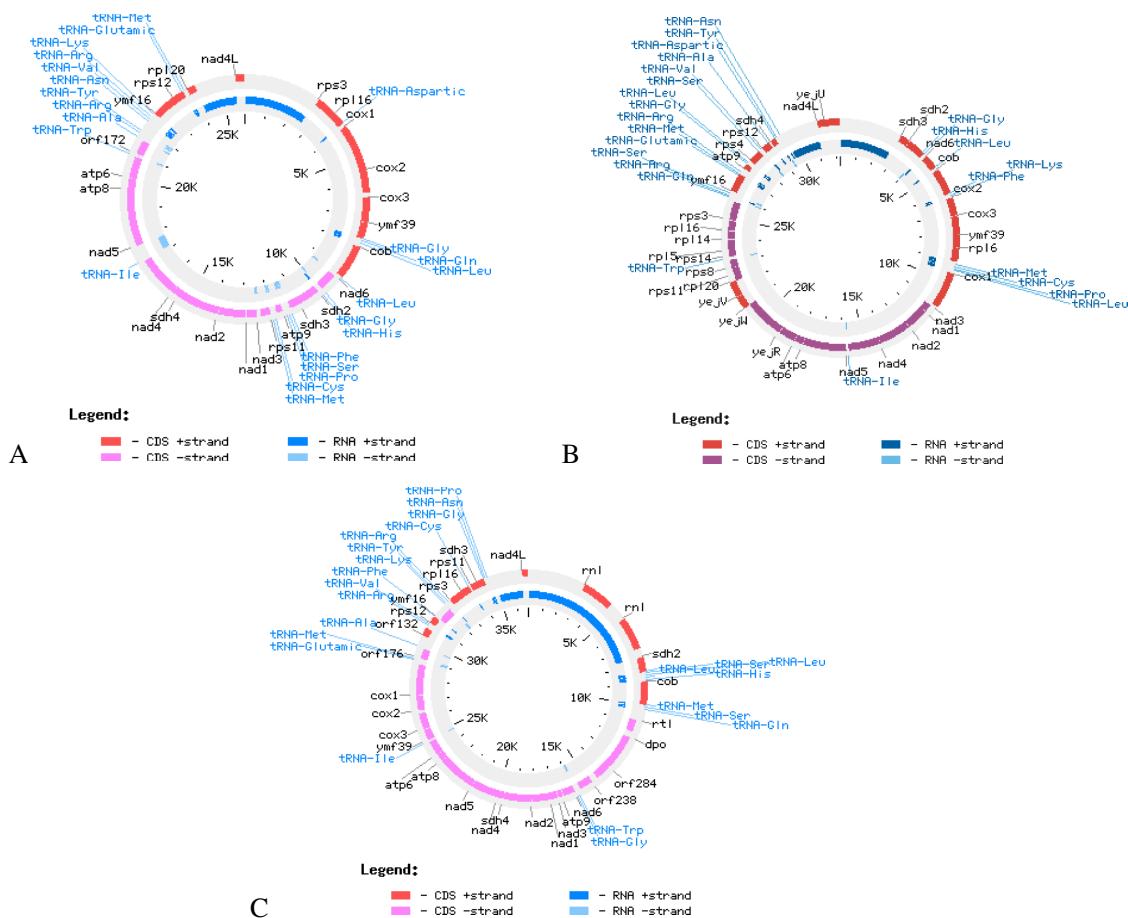
O genoma mitocondrial expandido das plantas terrestres se deve principalmente à presença de regiões intergênicas, segmentos repetidos e introns, assim como a incorporação de DNA de outras fontes como cloroplasto, núcleo e plasmídeos (Schuster & Brennicke, 1994; Bullerwell & Gray, 2004).

O acúmulo de dados moleculares, principalmente das sequências completas dos genomas organelares, ajuda a esclarecer a origem e evolução de mitocôndrias e cloroplastos nas diferentes linhagens de eucariotos (Gray, 1999). Além disso, algumas regiões do genoma mitocondrial vem sendo muito utilizadas como marcadores moleculares, como por exemplo a região 5' do gene citocromo c oxidase subunidade 1 (*cox1*). Herbert *et al.* (2003) utilizaram esse fragmento de cerca de 700 pares de bases como marcador molecular para identificação de espécies animais, técnica conhecida como DNA *barcode*. Esse método visa a identificação de organismos através do sequenciamento de pequenos fragmentos de DNA. Saunders (2005) desenvolveu *primers* para essa região em algas vermelhas (Rhodophyta), alinhando as sequências *cox1* das espécies *Cyanidium caldarium*, *Cyanidioschyzon merolae*, *Chondrus crispus* e *Porphyra purpurea*, disponíveis no GenBank, e modificando os *barcoding primers* originais publicados por Hebert *et al.* (2003). Utilizando esses *primers*, Saunders conseguiu amplificar a porção 5' do *cox1* para cerca de 250 indivíduos de 15 famílias diferentes da classe Florideophyceae. Seus resultados indicam que essa região apresenta níveis de divergência intra e interespecífica compatíveis com seu uso na delimitação de espécies, podendo ser, portanto, utilizada como DNA *barcode* em Rhodophyta.

Embora a região 5' do *cox1* utilizada como *barcode* seja curta, ela fornece informações bastante acuradas sobre a composição do genoma mitocondrial inteiro. Por exemplo, dos 648 nucleotídeos da sequência, humanos diferem de chimpanzés em aproximadamente 60 posições, e de gorila por cerca de 70 (Lane, 2009). Além disso, alguns pesquisadores defendem a hipótese de que sequências mitocondriais podem influenciar no processo de especiação, uma vez que sofrem mutações a taxas mais elevadas que o genoma nuclear (até 30 vezes mais rápido). Isso poderia causar uma incompatibilidade entre os produtos codificados entre os DNAs nuclear e mitocondrial, como por exemplo, os componentes da respiração celular, produzidos em ambos os genomas, acarretando em falha no desenvolvimento, redução de *fitness* e até divergência intraespecífica, levando a uma impossibilidade reprodutiva (Lane, 2009).

Em 1995, foi publicada a primeira sequência completa de DNA mitocondrial (mtDNA) de uma alga vermelha, *Chondrus crispus* (Figura 2A), pertencente a classe Florideophyceae (Leblanc *et al.* 1995). Em 1992, Suzuki *et al.* desenvolveram um protocolo para o isolamento dos DNAs nuclear, plastidial e mitocondrial de

*Cyanidioschyzon merolae* (Figura 2B), mas o sequenciamento completo do mtDNA só foi publicado anos depois (Ohta *et al.* 1998). O último genoma mitocondrial de uma alga vermelha a ser sequenciado foi o de *Porphyra purpurea* (Figura 2C) (Burger *et al.* 1999). A Tabela 1 apresenta informações comparativas sobre os genomas mitocondriais das três algas vermelhas.

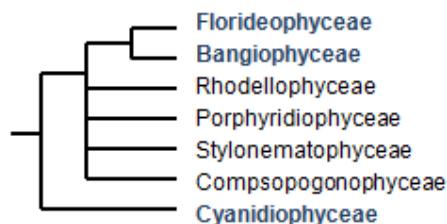


**Figura 2.** A. Mapa do genoma mitocondrial completo de *Chondrus crispus*. B. Mapa do genoma mitocondrial completo de *Cyanidioschyzon merolae*. C. Mapa do genoma mitocondrial completo de *Porphyra purpurea*. Escrito em preto estão representados os genes para proteínas e em azul, para tRNAs. Os números internos representam o tamanho do genoma. Cada fita do DNA é representado em uma cor diferente, conforme legenda junto ao mapa. Fonte: GenBank.

**Tabela 1.** Comparação entre as espécies de Rhodophyta que tiveram seus genomas mitocondriais totalmente sequenciados. Fonte: GenBank.

Espécies de Rhodophyta	<i>C. crispus</i>	<i>C. merolae</i>	<i>P. purpurea</i>
Tamanho (pb)	25.836	32.211	36.753
Conteúdo CG (%)	27	27	33
Total de genes	57	62	57
Genes para proteína	29	34	31
Genes para rRNA	4	3	2
Genes para tRNA	23	25	24
Número de acesso no GenBank	NC_001677	NC_000887	NC_002007

As relações filogenéticas dentro de Rhodophyta ainda não foram totalmente resolvidas, porém Yoon *et al.* (2006) apresentaram uma árvore filogenética baseada em dados moleculares de genes do cloroplasto (*psaA* e *rbcL*), indicando que Rhodophyta contém dois novos subfilos, Cyanidiophytina, com uma única classe, a Cyanidiophyceae, e Rhodophytina, com seis classes: Bangiophyceae, Compsopogonophyceae, Florideophyceae, Porphyridiophyceae, Rhodellophyceae e Stylopematophyceae (Figura 3). *Cyanidioschyzon merolae* pertence à classe Cyanidiophyceae, grupo mais basal dentro de Rhodophyta. *Porphyra purpurea* pertence à classe Bangiophyceae, que forma um grupo monofilético com a classe Florideophyceae, onde se encontram *G. tenuistipitata* e *C. crispus*. Essa monofilia é sustentada por dados moleculares e de ultraestrutura (por exemplo, a associação entre o Complexo de Golgi e o Retículo Endoplasmático). Além disso, há similaridades entre a fase *Conchocelis* de *Porphyra* (fase diplóide no ciclo reprodutivo) e a classe Florideophyceae (Oliveira & Bhattacharya, 2000; Saunders & Hommersand, 2004).



**Figura 3.** Árvore filogenética proposta por Yoon *et al.* (2006) para Rhodophyta. Em azul estão destacadas as classes de Rhodophyta que possuem espécies cujos genomas mitocondriais foram totalmente sequenciados.

As algas vermelhas movimentam uma indústria de cerca de 6 bilhões de US dólares por ano, sendo utilizadas como alimentos e para diversos fins industriais (Critchley & Ohno, 1998; Zemke-White & Ohno, 1999; McHugh, 2003). As algas vermelhas são também extremamente atrativas para propósitos biotecnológicos, já que possuem polissacáideos sulfatados (ficocolóides) em sua parede, as agaranas e carragenanas, e contém ficobilinas e carotenóides (pigmentos acessórios à fotossíntese), além de outros compostos de interesse (Zemke-White & Ohno, 1999).

No início do século XX, com a crescente demanda por ficocolóides, o gênero *Gracilaria* (Gracilariales, Rhodophyta) tornou-se uma importante fonte de ágar para a indústria, o que levou ao cultivo de diversas espécies em todos os continentes (Armisen, 1995). O gênero *Gracilaria* inclui mais de 100 espécies descritas (Oliveira & Plastino, 1994; Gurgel & Fredericq, 2004), sendo 168 nomes aceitos taxonomicamente até o momento ([www.algaebase.org](http://www.algaebase.org)), e possui um ciclo de vida trifásico, onde há uma alternância isomórfica entre as gerações haplóide (fase gametofítica) e diplóide (fase tetrasporofítica), e uma terceira fase diplóide (fase carposporofítica) parasita da planta haplóide feminina (Crichtley *et al.* 1993; Oliveira & Plastino, 1994).

*Gracilaria tenuistipitata* var. *liui* Zhang *et* Xia é uma macroalga vermelha que apresenta uma grande tolerância a fatores ambientais tais como salinidade (3 a 50 ppm), temperatura (13 a 33°C) e pH (5,5 a 10,0), e é de fácil crescimento *in vitro* (Macchiavello *et al.* 1998). Estas características, aliadas às altas taxas de crescimento e alta produção de ágar, tornam esta espécie extremamente adequada para cultivo (Macchiavello *et al.* 1999). De fato, *G. tenuistipitata* é extensivamente cultivada no sul da China e em Taiwan. Além disso, esta espécie tem sido frequentemente utilizada em estudos fisiológicos, bioquímicos e moleculares, tornando-se um organismo modelo dentre as algas vermelhas (Barufi *et al.* 2009). Previamente foi realizado por nosso grupo de pesquisa o sequenciamento do genoma completo do cloroplasto de *Gracilaria tenuistipitata* (Hagopian *et al.* 2002, 2004).

Comparativamente com os Metazoa, existem poucos dados sobre o genoma mitocondrial de organismos fotossintetizantes, especialmente dos diferentes grupos de algas. Atualmente existe um total de 2129 genomas mitocondriais completamente sequenciados. Destes, apenas 66 genomas mitocondriais foram sequenciados para organismos fotossintetizantes (Março/2010; <http://www.ncbi.nlm.nih.gov/genomes/>).

A classe Florideophyceae compreende cerca de 99% das espécies de Rhodophyta (Oliveira & Bhattacharya, 2000), entretanto, existe apenas uma espécie (*Chondrus crispus*) que teve seu genoma mitocondrial completamente sequenciado. A comparação de características genômicas, como a composição, a ordem (colinearidade) e a análise da sequência de diversos genes, geram dados mais completos e que permitem o estabelecimento de relações filogenéticas com maior precisão e um melhor entendimento sobre a origem e evolução dessas organelas. Até o momento, os genomas da mitocôndria e do cloroplasto (Hagopian *et al.* 2004) de Rhodophyta apresentam o maior conjunto de genes codificadores de proteínas e, portanto, trata-se de um grupo especialmente interessante para ter seus genomas organelares estudados. Entretanto, até o momento estão disponíveis apenas os genomas mitocondriais de três espécies de Rhodophyta.

### **Objetivo geral**

Este estudo visa o sequenciamento completo do genoma mitocondrial de *Gracilaria tenuistipitata* e a análise dos genes presentes.

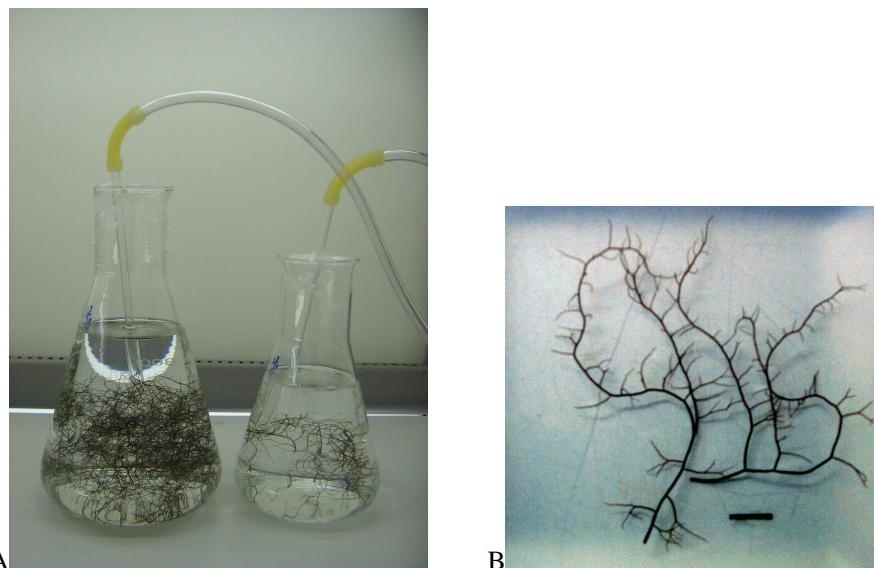
### **Objetivos específicos**

- sequenciamento completo do genoma mitocondrial de *Gracilaria tenuistipitata*;
- anotação dos quadros abertos de leitura (ORFs) e análise de regiões não codificantes;
- análise comparativa com os genomas mitocondriais de Rhodophyta e de outros organismos fotossintetizantes disponíveis no GenBank;

## 4. MATERIAIS E MÉTODOS

### 4.1. Condições de cultura

O material tetrasporofítico de *Gracilaria tenuistipitata* (Figuras 4A e 4B) foi coletado por E.C. Oliveira em 1990, em Haikou, na China, e mantido desde então no banco de germoplasma de Gracilariales do Laboratório de Algas Marinhas Édison José de Paula. O cultivo unialgal foi mantido em água do mar esterilizada, enriquecida com meio nutritivo Von Stosch (VS) 100% (Oliveira *et al.* 1995), temperatura de  $25\pm1^{\circ}\text{C}$ , fotoperíodo de 14h luz: 10h escuro, irradiância média de  $70 \mu\text{mol de fótons.m}^{-2}.\text{s}^{-1}$ , salinidade de 26 ups e aeração intermitente (30 minutos ligada: 30 minutos desligada). A troca do meio de cultura foi realizada semanalmente. Uma parte do cultivo foi congelada em nitrogênio líquido periodicamente e armazenada em ultrafreezer ( $-80^{\circ}\text{C}$ ).



**Figura 4.** A. Talos da fase tetrasporofítica de *Gracilaria tenuistipitata* em cultivo. B. A alga em detalhe. Barra corresponde a 1cm.

O cultivo algal apresentou dois tipos de contaminantes: diatomáceas e cianobactérias. A limpeza das algas foi realizada através de várias lavagens em água Milliquê, e uma última lavagem em água do mar esterilizada (salinidade: 32 ups). Os contaminantes são sensíveis à variação de salinidade, que não afeta a *G. tenuistipitata*, pois esta espécie é tolerante à variação de salinidade (Macchiavello *et al.* 1998).

Também foram realizados repiques nas bases dos talos, local de concentração dos contaminantes.

As diatomáceas possuem uma parede celular composta por sílica, e para eliminá-las foi acrescentado dióxido de germânio ( $\text{GeO}_2$ ) ao meio de cultura, pois o germânio apresenta propriedades químicas semelhantes ao silício. Desta forma, as diatomáceas incorporam o germânio em suas paredes celulares, prejudicando os processos de crescimento e reprodução (Andersen & Kawachi, 2005).

Para a descontaminação por cianobactérias, foi acrescentado o antibiótico penicilina (50mg/L) ao meio de cultura e as algas foram deixadas na solução por 48 horas. Após este período, o cultivo foi lavado como citado anteriormente, e mantido em condições ótimas de crescimento.

#### **4.2. Extração e purificação de DNA**

Foram utilizados dois métodos de extração, cada qual com rendimento e qualidade de DNA diferentes.

Todos os procedimentos foram realizados com o máximo de cautela para evitar a quebra mecânica do DNA genômico, utilizando-se ponteiras de micropipeta com as pontas cortadas e movimentos mais suaves para a homogeneização das soluções.

No primeiro método, o material algal foi macerado em nitrogênio líquido, lisado com tampão de extração e submetido à centrifugação diferencial, conforme os trabalhos de Hagopian *et al.* (2002) e de Triboush *et al.* (1998). Foi desenvolvida uma combinação desses dois protocolos para a obtenção de DNA enriquecido em A-T, típico de organelas, conforme abaixo.

Cerca de 50g de alga, lavadas com água destilada, foram maceradas em nitrogênio líquido em um almofariz por cerca de 1 hora. O pó fino resultante foi suspenso em 250mL de tampão de lise (50mM Tris pH 8,0, 25mM EDTA, 1,6M sorbitol, 10% PEG 4000, 0,1% BSA, 1mM DTT), mantendo sempre a temperatura entre 0 e 4°C. A suspensão foi filtrada em gaze 2 vezes. O filtrado foi transferido para tubos de centrífuga de 10mL e centrifugado por 10 min a 4.000 rpm, sempre a 4°C. O sobrenadante foi recuperado e o procedimento repetido por mais duas vezes. O sobrenadante foi transferido para um novo tubo e centrifugado a 12.000 rpm por 15 min

a 4°C. O precipitado foi ressuspenso em 600µL de tampão TEN (100mM Tris pH 7,2, 50mM EDTA, 100mM NaCl, 0,2% β-mercaptoetanol) a 0°C e foi adicionado 30µL de SDS 20%, misturando a solução constantemente. A solução foi transferida para tubos Eppendorf, acrescentando-se um volume igual de fenol equilibrado (650µL) e centrifugando-se por 10 min a 12.000 rpm a 4°C. O sobrenadante foi recuperado e o procedimento repetido. A fase aquosa (sobrenadante) foi transferida para um novo tubo Eppendorf e foi adicionado 1 volume de clorofórmio:álcool isoamil (24:1). A solução foi centrifugada por 10 min a 12.000 rpm a 4°C e o procedimento repetido. O DNA foi precipitado com 1/10 volume de acetato de amônio 5M (cerca de 20µL) e 1 volume de isopropanol (cerca de 220µL), e incubado a -20°C por toda a noite. A solução foi centrifugada por 15 min a 12.000 rpm, sempre a 4°C. O precipitado de DNA foi lavado com 500µL de etanol 70%, depois com etanol 95% para remover resíduos de sal. O precipitado foi dissolvido em 20µL de tampão TE (10mM Tris pH 8,0, 1mM EDTA) e acrescido de 150µL de tampão TEN. A solução de DNA foi incubada por 15 min a 65°C com 10µL de SDS 20%. Foram adicionados 50µL de acetato de potássio 5M e a solução de DNA foi incubada a -20°C por 1 hora, e em seguida a solução foi centrifugada por 15 min a 12.000 rpm. O sobrenadante foi cuidadosamente coletado em um novo tubo Eppendorf para evitar transferir partículas do precipitado. Foram acrescidos à solução 1/10 de volume de acetato de amônio 5M (20µL) e 1 volume de isopropanol (220µL), que depois foi incubada a -20°C por 1 hora. O DNA foi precipitado por 15 min a 12.000 rpm a 4°C e lavado com etanol 70% e 95%. O precipitado de cada amostra foi ressuspenso em 10µL de tampão TE e as amostras armazenadas em freezer.

No segundo método, adaptado de Rice & Bird (1990), o rendimento de DNA total é maior que o método anterior, porém com menor grau de pureza em relação aos genomas organelares. Foram selecionados 10g de alga, lavados com água destilada e imediatamente congelados em nitrogênio líquido. A amostra foi macerada em nitrogênio líquido em um almofariz por cerca de 1 hora. O pó fino resultante foi suspenso em 30mL de tampão de lise (2% Sarkosyl, 10mM EDTA, 20mM Tris-HCl – pH 7,6, 100mM NaCl) e a solução foi distribuída em tubos plásticos de 10mL. Após uma incubação a -70°C de 2 a 14 horas, as amostras foram descongeladas em temperatura ambiente. As amostras foram centrifugadas a 4.500 rpm por 5 min a 4°C. O sobrenadante foi transferido para uma proveta e acrescentou-se proteinase K (50µg/mL). A solução foi

transferida para tubos plásticos e incubadas por 30 min a 37°C. Após a incubação, foi acrescentado um volume igual de fenol:clorofórmio (1:1) e centrifugado por 10 min a 2.500 rpm. A fase aquosa foi recuperada e o procedimento repetido a 3.500 rpm. Depois acrescentou-se um volume igual de clorofórmio:álcool isoamil (24:1) e centrifugou-se por 10 min a 3.500 rpm. A fase aquosa foi novamente recuperada, acrescentou-se 1/10 do volume de acetato de sódio 3M (pH 5,2) e 1 volume de isopropanol. As amostras foram incubadas a -20°C por toda a noite. O DNA foi precipitado centrifugando-se as amostras por 30 min a 8.000 rpm a 4°C e depois lavado com etanol 70%. O precipitado de cada amostra foi ressuspenso em 300µL de tampão TE (10mM Tris – pH 8,0, 1mM EDTA) e as amostras foram armazenadas a -20°C.

#### **4.3. Gel de eletroforese**

Preparação de gel de agarose para eletroforese. A agarose foi dissolvida a 0,7% em tampão TBE (445mM Tris, 445mM Ácido Bórico, 10mM EDTA). A solução foi aquecida em micro-ondas e foi acrescida 0,5µL de solução brometo de etídio (10mg/mL), após esfriar a aproximadamente 45°C.

Recentemente, o brometo de etídio foi substituído por um produto menos tóxico, denominado “GelRed Nucleic Acid Gel Stain” (Biotium, Hayward, CA, USA). Este corante pode ser utilizado da mesma maneira que o BrEt, porém um protocolo mais econômico foi testado com sucesso. Neste protocolo, o GelRed é diluído em água na proporção 1:500. Para cada amostra de DNA, é acrescido 1µL dessa diluição e depois aplicado ao gel de agarose, ao invés de acrescentar uma quantidade maior de GelRed diretamente na solução de agarose.

Preparação das amostras para eletroforese. Misturou-se 1µL de tampão de carregamento (0,25% azul de bromofenol, 0,25% xileno cianol, 30% glicerol, 50mM EDTA) e 3µL do DNA. Usou-se 1µL e 0,5µL do marcador de peso molecular 1 Kb Ladder (Invitrogen, Carlsbad, CA, USA). A eletroforese foi corrida a 60V por aproximadamente 1 hora.

Quantificação. Para a quantificação, foi efetuada uma comparação visual com o marcador de peso molecular. Conforme indicações do fornecedor, 1µL da banda de 1,6

Kb do marcador 1 Kb Ladder (Invitrogen) corresponde a 20ng de DNA, e 0,5 $\mu$ L da mesma banda corresponde a 10ng de DNA.

Alternativamente, o DNA foi quantificado no fluorômetro (Hoefer DyNA – Quant 200). Preparou-se um tampão com 1mL de solução estoque de tampão TEN 10x (100mM Tris, 10mM EDTA, 2M NaCl), 9mL de água Milliquê e 1 $\mu$ L de solução estoque de Hoechst Dye 33258 (1mg/mL). Para calibrar o fluorômetro foram utilizados 2 $\mu$ L de “Calf thymus DNA” 1:10 (solução estoque 1mg/mL), na concentração de 100ng/mL, em 2mL do tampão preparado. Para cada quantificação foram utilizados 2 $\mu$ L de amostra.

#### **4.4. Ultracentrifugação isopícnica em gradiente de CsCl**

Esta técnica visa separar as frações de DNA pelo peso molecular (densidade), com o uso do corante Hoechst 33258 (bisbenzimidina), que acentua essa diferença. O método descrito abaixo foi modificado de Aldrich & Cattolico (1981).

Foram pesados 9,2g de CsCl e diluídos em 9,5mL de tampão TE (0,8mg CsCl/mL), em uma proveta, para cada tubo. A solução foi transferida em um tubo para ultracentrífuga de 11,5mL, já acrescida de 46 $\mu$ L de Hoechst 33258 (40 $\mu$ g/mL). A amostra de DNA foi colocada no tubo (mínimo de 40 $\mu$ g e máximo de 200 $\mu$ g), completando-se o tubo com a solução de CsCl. Os tubos foram lacrados com o auxílio de um prensador. A amostra foi centrifugada a 20°C por 46 horas a 45.000 rpm, e por 2 horas a 20.000 rpm.

Para visualizar as bandas, utilizou-se luz ultravioleta (uma vez que o corante fluoresce nessa condição). Furou-se a extremidade superior do tubo e, para se recuperar o DNA da amostra, injetou-se uma agulha logo abaixo da banda, que foi recolhida com o auxílio de uma seringa. Para remover o corante, as amostras foram transferidas para tubos de ensaio e acrescentou-se um volume igual de isopropanol saturado com NaCl 2M. As amostras foram misturadas delicadamente e a fase de cima foi descartada. O procedimento foi repetido por mais 3 ou 4 vezes, até que não restasse mais o corante (verificado pela ausência de fluorescência sob luz UV).

O DNA foi precipitado adicionando-se 2 volumes de água Milliquê autoclavada e 1 volume (do total) de etanol 95%. Incubou-se a solução a -20°C por toda a noite. Os

tubos foram centrifugados a 8.000 rpm por 30 minutos, e o sobrenadante foi descartado. O precipitado foi lavado com etanol 70% e ressuspenso em 100 $\mu$ L de tampão TE.

Alternativamente, foram testados diferentes parâmetros em relação à concentração do CsCl (0,8 a 1mg CsCl/mL) e do corante Hoechst 33258 (40 a 60 $\mu$ g/mL), além da variação na quantidade de DNA descrita acima.

#### **4.5. Digestão com enzimas de restrição**

Acrescentou-se 1 $\mu$ L de *Eco*RI e 1 $\mu$ L de tampão *Eco*RI 10x em 8 $\mu$ l de DNA (total= 10 $\mu$ L). A solução foi incubada a 37°C por 3 a 4 horas. Em seguida foi acrescentado 3 $\mu$ L de tampão de carregamento para parar a reação. As amostras de digestão foram diretamente analisadas em gel de agarose como descrito acima.

#### **4.6. Primers para PCR e sequenciamento**

Para a amplificação por PCR e sequenciamento de regiões específicas do genoma mitocondrial de *G. tenuistipitata*, foram desenhados diversos *primers*, além de terem sido utilizados os *primers* cox2F e cox3R, que amplificam a região espaçadora desses dois genes (Zuccarello *et al.*, 1999) e os *primers* gazF1 e gazR1 (Saunders, 2005), que amplificam a região 5' do gene *cox1*, todos desenhados para Rhodophyta. Todos os *primers* utilizados nesse trabalho estão listados na Tabela 2.

Inicialmente, foram alinhadas as regiões conservadas dos genomas mitocondriais das macroalgas *Chondrus crispus* e *Porphyra purpurea* (Tabela 1), utilizando o programa BioEdit Sequence Alignment Editor (Hall, 1999), versão 7.0.5.3.

Conforme a obtenção de novas sequências de *G. tenuistipitata*, foram desenhados mais *primers* localizados nas pontas dessas sequências, para o avanço no sequenciamento do genoma completo.

Embora algumas regiões do genoma mitocondrial sejam bastante conservadas, uma ou mais bases podem ser divergentes entre as espécies, dentro da sequência escolhida como *primer*. Neste caso, foram desenhados *primers* degenerados, onde há mais de uma possibilidade de base em uma determinada posição (Linhart & Shamir,

2005). Por exemplo, se a quinta base de uma determinada região de *C. crispus* for um T, e em *P. purpurea* um A, podemos desenhar um primer com w nesta posição, o que significa que metade dos *primers* sintetizados terão um T e a outra metade, um A, tudo na mesma solução. A elaboração de *primers* degenerados é muito útil em casos em que não conhecemos a sequência da espécie alvo. Porém, quanto maior a degeneração, menor a especificidade (Linhart & Shamir, 2005).

Além do tamanho e conteúdo de bases, deve-se levar em conta a sequência de nucleotídeos do *primer*. Essa sequência não pode apresentar regiões que se sobreponham, e deve-se avaliar o risco para a formação de dímeros (alinhamento de dois *primers*) e hairpins (pareamento das bases dos *primers* entre si). Os *primers* possuem entre 18 e 24 pares de base, geralmente. Esse tamanho proporciona uma boa taxa de alinhamento específico e apresenta um custo de produção razoável, uma vez que as empresas fabricantes calculam o valor conforme o número de bases de cada oligonucleotídeo. Entretanto, para as “Long PCR” (PCR de sequências longas) foram desenhados *primers* mais longos, de até 30 bases, que aumentam a temperatura de desnaturação (TM), recomendado pelo fabricante da polimerase, além de aumentar a especificidade de alinhamento e evitar a amplificação de outras regiões além da desejada. Os *primers* foram desenhados com o auxílio do programa Gene Runner, versão 3.05 (1994, Hastings Software, Inc.).

#### **4.7. PCR (reação em cadeia da polimerase) e sequenciamento**

Foram utilizados dois protocolos de PCR. No primeiro, utilizou-se “Taq DNA Polymerase” comum (Invitrogen ou Promega) no qual foram amplificados fragmentos de até 2 Kb, nas seguintes condições: 1X tampão de PCR; 1,5mM de MgCl<sub>2</sub>; 0,2mM de cada dNTP; 0,2µM de cada *primer*; 2ng de DNA e 1,25U de Taq. Foram usadas diversas temperaturas de alinhamento (TA), variando de 42°C a 47°C. O ciclo básico utilizado foi: 1. 94°C por 5 min; 2. 94°C por 1 min; 3. TA por 2 min; 4. 72°C por 4 min; e 5. 72°C por 7 min. Os passos 2 a 4 foram repetidos 35 vezes. A TA é calculada com base na temperatura de desnaturação dos pares de *primers*, sendo geralmente alguns graus menor para um melhor funcionamento dos oligonucleotídeos. Utilizou-se o termociclador Techne TC-512 (Bibby Scientific Ltd., Staffordshire, UK), que permite um gradiente de

temperatura no mesmo ciclo, ou seja, cada coluna do termociclador pode ser ajustado em temperaturas diferentes dentro de um gradiente definido.

No segundo protocolo utilizado, foram amplificados fragmentos de até 8 Kb, com o auxílio da “Elongase Enzyme Mix” (Invitrogen). As condições da reação foram: 1X tampão B de PCR (2,0mM Mg<sup>2+</sup>); 0,3mM de cada dNTP; 0,4µM de cada *primer*; 10ng de DNA e 2U de Elongase. O ciclo utilizado foi: 1. 92°C por 2 min; 2. 92°C por 20 seg; 3. TA por 1 min; 4. 60°C por 16 min; e 5. 60°C por 6 min. Os passos 2 a 4 foram repetidos 35 vezes (Barau *et al.*, 2005). Mais uma vez, foram testadas diversas temperaturas de alinhamento (TA), variando de 45°C a 55°C, conforme os pares de *primers*.

As amostras de PCR foram analisadas em gel de agarose, para se verificar o sucesso da amplificação. As que obtiveram resultado positivo foram purificadas com “Illustra GFX PCR DNA and Gel Band Purification Kit” (GE Healthcare, Buckinghamshire, UK), segundo o protocolo do fabricante.

O sequenciamento foi feito com o produto de PCR purificado e quantificado (conforme descrito acima), sendo aproximadamente 40ng para fragmentos menores de 2 Kb e cerca de 100ng para fragmentos maiores. Foi utilizado o kit de sequenciamento “BigDye™ Terminator Cycle Sequencing Kits” (Applied Biosystems, Foster City, CA, EUA). As reações de sequenciamento foram realizadas utilizando-se 40 ciclos a 96°C por 10 seg, 54°C por 20 seg e 60°C por 4 min. Os produtos foram precipitados com isopropanol 65%, lavados em etanol 60% e sequenciados no sequenciador automático ABI PRISM™ 3100 Genetic Analyzer (Applied Biosystems).

Para o sequenciamento do genoma completo, foi utilizada principalmente a técnica de *primer walking*. O sequenciamento inicial do produto de PCR foi feito com os *primers* utilizados na amplificação, obtendo-se uma sequência direta e outra reversa. A partir dessas sequências, foram desenhados novos *primers* (direto e reverso) em suas pontas, até que todo o fragmento amplificado fosse sequenciado, formando um único contig (Haas *et al.* 1998; Machida *et al.* 2002).

**Tabela 2.** Descrição dos *primers* desenhados a partir dos genes mitocondriais de *Porphyra purpurea*, *Chondrus crispus* e de sequências obtidas de *Gracilaria tenuistipitata*. Também estão descritos os *primers* cox2F e cox3R (Zuccarello *et al.*, 1999) e gazF1 e gazR1 (Saunders, 2005). Foi usado F para *primers* na direção direta e R para a direção reversa. w= A/T; r= A/G; k= G/T; d= A/G/T; y= C/T; m= A/C; s= C/G; h= A/C/T. TM= “temperature of melting”.

Primer	Gene	Sequência (5' - 3')	Localização em <i>C.crispus</i> (pb)	TM (°C)
rps12F	<i>rps12</i>	ACAAyAAGyCCTAArAAACC	1250-1269	49,2
rps12R	<i>rps12</i>	GGTTTyTTAGGGrCTTrTTGT	1250-1269	49,2
rps12R3	<i>rps12</i>	ATTTGGTTTTTAGGGCTTGTGT	1250-1273	54,1
rps12R4	<i>rps12</i>	TCACCAGGAATGTAACCAAT	1325-1344	51,4
rps12R2	<i>rps12</i>	CCTCCCACCACGAACCAAAACCATAAGAAT	1365-1394	62,0
rpl20F	<i>rpl20</i>	AACAGGAAGTATTGGT	1861-1879	46,6
rnsF2	<i>rns</i>	ATAGAAAGGCAGCAGTGAGGAAT	2233-2255	56,2
rnsR2	<i>rns</i>	ATTCCTCACTGCTGCCTTCTAT	2233-2255	56,2
rnsF3	<i>rns</i>	GAATAATACTGACACTGAGATA	2601-2622	46,8
rnsR	<i>rns</i>	GACGACAGCCATGCAAC	2864-2880	54,3
rnsR3	<i>rns</i>	AAGAGTTTGTCCGTTATTGG	2910-2931	51,4
rnsF4	<i>rns</i>	CCAATAAACGGACAAAACCTTT	2910-2931	51,4
rnsR4	<i>rns</i>	ATGAAAGTAGCCCCAATCT	3008-3025	47,9
rnsF5	<i>rns</i>	AGATTGGGCTACTTCAT	3008-3025	47,9
rnsF	<i>rns</i>	AAGTCGTAACAAGGTAGC	3269-3286	49,0
nad4LR	<i>nad4L</i>	TAAAAATAATAAGAAGTA	3388-3409	38,3
nad4LF	<i>nad4L</i>	TTTAGATGATATTACAGGTC	3520-3539	43,7
nad4LR2	<i>nad4L</i>	ATAGACGATTCCGCAGC	3572-3588	51,9
rnlF4	<i>rnl</i>	AGGAGGCATTGGCTGTGATTAT	3759-3782	57,0
rnlR4	<i>rnl</i>	ATAAATCACAGCCAATGCCCT	3759-3782	57,0
rnlF7	<i>rnl</i>	GGAGCGAAAGACTAATCAAAC	4352-4373	52,8
rnlF5	<i>rnl</i>	CAATCCAAATCAACTAAGAAT	4563-4586	46,8
rnlR5	<i>rnl</i>	TTTCACTACATTTCAACTA	4702-4721	44,3
rnlF6	<i>rnl</i>	ATGAAAGAAGAAGTAACGAC	5202-5221	47,5
rnlR7	<i>rnl</i>	AGAGTCCTGTGGTTTGT	5226-5245	51,2
rnlR	<i>rnl</i>	ACCTCCGTTACATTTAG	5669-5686	46,2
rnlF	<i>rnl</i>	TAAAATGTAACGGAGGTG	5670-5687	46,5
rnlR6	<i>rnl</i>	CACTATGACAGACTTCGT	5764-5782	49,1
rnlF3	<i>rnl</i>	CTGTAAGAAAAAGAAAAATAAAAAGCG	6010-6035	50,7
rnlR3	<i>rnl</i>	CGCTTTTATTTCTTTCTTACAG	6010-6035	50,7
rnlF2	<i>rnl</i>	CCAGTAGCATCGTCAAGTAG	6092-6111	52,4
gazF1	<i>cox1</i>	TCAACAAATCATAAGATATTGG	7584-7606	47,0
cox1-5'R	<i>cox1</i>	AATATCTTATGATTGTTGA	7584-7604	42,7
cox1F-2	<i>cox1</i>	TTTGrTTTTGGACATCC	8269-8287	47,4
gazR1	<i>cox1</i>	ACTTCTGGATGTCCAAAAAAyCA	8271-8293	53,9
cox1F3	<i>cox1</i>	TAGGTGTATTAGGATTAT	8401-8419	41,4
cox1F	<i>cox1</i>	CTTATTATGTAGTTGCGC	8671-8688	45,5
cox1R	<i>cox1</i>	GCGCAACTACATAATAAG	8671-8688	45,5
cox1R3	<i>cox1</i>	ACCGAATCAATAATAATCCAGCA	8732-8756	53,1
cox2F	<i>cox2</i>	GTACCwTCTTdrGrrkdAAATGTGATGC	9710-9738	67,0
cox3R	<i>cox3</i>	GGATCTACwAGATGrAAwGGATGTC	10114-10138	62,0
cox3F	<i>cox3</i>	CACCAACAGTAGAAATCGG	10426-10444	50,6

**Tabela 2.** Continuação.

Primer	Gene	Sequência (5'- 3')	Localização em <i>C. crispus</i> (pb)	TM (°C)
ymf39F	<i>ymf39</i>	ATAGATAAAATAGGTAGCGTCAT	11471-11493	48,2
ymf39R	<i>ymf39</i>	ATGACGCTACCTATTATCTAT	11471-11493	48,2
cobF	<i>cob</i>	ATTAGCATTGCAAGTGTAG	11924-11943	50,5
cobF2	<i>cob</i>	TTAGCATTGCAAGTGTAGAGCATA	11925-11949	56,3
cobR	<i>cob</i>	TATAATCCTCTAAAAATATG	12033-12052	38,9
cobF3	<i>cob</i>	TTAGGAGGCGTAGCAATGAT	12612-12634	58,2
nad6F2	<i>nad6</i>	CATCATAACCAACAATAAAAAT	13420-13441	45,2
nad6F	<i>nad6</i>	AAAACCTCTTACCTATTGGGG	13714-13731	49,3
rps11R	<i>rps11</i>	ATACAATAACAAAACAGAGAACAA	15891-15913	49,0
nad3F	<i>nad3</i>	TAAATCCrCACTCATAAGC	16384-16402	48,4
nad3R	<i>nad3</i>	GCTTATGAGTGyGGATTAA	16384-16402	48,4
nad3R2	<i>nad3</i>	AGTGCTTATGAGTGCGGGCT	16386-16405	59,2
nad1F	<i>nad1</i>	TTCAGGTAAATCAAAAGGAG	16901-16920	47,5
nad1R	<i>nad1</i>	CTCCTTTGATTACCTGAA	16901-16920	47,5
nad1R2	<i>nad1</i>	CGTTAGGCAGGGTATGGTGTATTCTGAT	17216-17244	61,0
nad2F3	<i>nad2</i>	GAATCAAATACATACCCTTA	17691-17711	45,1
nad2R3	<i>nad2</i>	TAAGGGTATGTATTGTGATT	17691-17711	45,1
nad2F2	<i>nad2</i>	CTCCTCCAAACACTGATAATAAAAGAAAAC	18049-18078	56,0
nad2R2	<i>nad2</i>	GTTTTCTTTATTATCAGTGTGGAGGAG	18049-18078	56,0
nad2F	<i>nad2</i>	TATGAAAwwGGTGCkGC	18325-18340	48,1
nad2R	<i>nad2</i>	GCmGCACCwTTTCATA	18325-18340	48,1
sdh4F	<i>sdh4</i>	AGTCCAAGAAAAATAAAATCCAGC	19234-19259	61,0
sdh4R	<i>sdh4</i>	GCTGGATTTTATTCTGGACT	19234-19259	61,0
nad4F3	<i>nad4</i>	GCACCTACTAATAACTAATAAT	19567-19587	44,2
nad4F2	<i>nad4</i>	GGwGTAAAAwAAATTGATGC	19945-19964	45,3
nad4R2	<i>nad4</i>	AAATGGGTGGTTATGGTCTTACG	19990-20014	55,3
nad4F	<i>nad4</i>	AAAAAATTGGTAAGCAGC	20265-20282	46,1
nad4R	<i>nad4</i>	GCTGCTTACCAATTTTTT	20265-20282	46,1
nad5F3	<i>nad5</i>	CGAACAAACCAACTTACGCTTATC	20824-20847	54,5
nad5R5	<i>nad5</i>	CTAATGAACAAGATATGCGTCG	22268-22289	51,5
nad5R3	<i>nad5</i>	AACTGCGGGGGTATTTTATTA	22561-22582	52,2
nad5R4	<i>nad5</i>	GTTCGTAGGATGAGAAGGTGTAGGTTAGC	22934-22963	58,5
nad5F	<i>nad5</i>	CCwACdCCTTCTCACsC	22940-22956	53,6
nad5R	<i>nad5</i>	GsGTGAGAAGGhGTwGG	22940-22956	53,6
nad5F2	<i>nad5</i>	GAGAGATAAAAGCCATTGAGAAAAAACACA	23239-23268	56,3
nad5R2	<i>nad5</i>	TGTGTTTTCTCAATGGCTTATCTCTC	23239-23268	56,3
atp8R2	<i>atp8</i>	ATCGTATCATGTATTCCCCA	23764-23785	49,7
atp8F	<i>atp8</i>	ACGtCTAATTGwGGCAT	23811-23798	49,7
atp8R	<i>atp8</i>	ATGCCwCAATTAGAyCGT	23811-23798	49,7
atp6F2	<i>atp6</i>	GCTTGTAAACCTGCTATTGCTAA	23870-23892	53,5
atp6R2	<i>atp6</i>	TTAGCAATAGCAGGTTACAAGC	23870-23892	53,5
atp6F	<i>atp6</i>	CATHGTyCAArCAAyCC	23968-23985	48,4
atp6R	<i>atp6</i>	GGrTTTGTyTTGrACdATG	23968-23985	48,4
atp6F3	<i>atp6</i>	GATAGAGCAAGACCAAATGTAATAA	24191-24220	53,7
atp6F4	<i>atp6</i>	CAATTATTGGwATTAkAGTATT	24386-24408	43,3
atp6F5	<i>atp6</i>	CAGGAGAAGAAGGATTGAACC	24603-24624	53,3
orf172F	<i>orf172</i>	TTTTTAGACTCCCCTTTGAA	25170-25192	50,2

#### 4.8. Clonagem de produto de PCR

Foram feitas clonagens com os produtos de PCR que apresentaram múltiplas bandas, utilizando-se o kit “pGEM-T Vector Systems” (Promega, Madison, WI, USA), conforme o protocolo do fabricante. Para isso, a banda de interesse foi cortada do gel de agarose e purificada com o kit “Illustra GFX PCR DNA and Gel Band Purification Kit” (GE Healthcare), segundo o protocolo do fabricante.

Os vetores foram inseridos em células de *E. coli* através do método de *heat-shock*, em que a permeabilidade da membrana é conferida por choque térmico, permitindo a entrada do plasmídeo na bactéria (Sambrook *et al.*, 1989).

Após a transformação, as bactérias cresceram no meio de cultura líquido SOC (Triptona 2%; Extrato de levedura 0,5%; NaCl 8,6mM; KCl 2,5mM; MgSO<sub>4</sub> 20mM; Glicose 20mM) por 45 min a 37°C com agitação. Foram então plaqueadas em meio de cultura sólido composto por LB (triptona 10g/L; extrato de levedura 5g/L; NaCl 5g/L; ágar 15g/L), IPTG (250µg/mL), X-Gal (40µg/mL) e ampicilina (100µg/mL). As placas foram colocadas em estufa por 14 horas a 37°C.

As colônias brancas foram selecionadas por repique e colocadas em meio de cultura líquido CircleGrow para se multiplicarem. Após 20 horas de incubação a 37°C no agitador, uma parte foi acrescida de uma solução de glicerol 50% (mesmo volume) e armazenada a -80°C. O restante foi centrifugado a 4.000 rpm por 10 min. O sobrenadante foi descartado e adicionou-se 240µL de tampão GTE (Glicose 50mM; Tris-HCl 25mM; EDTA 10mM, pH 8,0). Novamente foi centrifugado a 4.000 rpm por 10 min, descartado o sobrenadante e adicionados 80µL de GTE. As bactérias foram ressuspensas e foram acrescentados 5µL de RNase (10mg/mL). Foram transferidos 60µL de cultura para uma nova placa, acrescido de 60µL de NaOH/SDS (NaOH 0,2M; SDS 1%) e incubado a temperatura ambiente por 10 min. Adicionou-se 60µL de acetato de potássio 3M e incubou-se por mais 10 min a temperatura ambiente. Após a incubação a placa foi levada à estufa a 90°C por 30 min e depois imediatamente no gelo por 10 min. Centrifugou-se por 10 min a 4000 rpm e o sobrenadante foi transferido para uma placa com filtro, acoplada a uma outra placa. As placas foram centrifugadas a 4000 rpm por 6 min e a placa com filtro foi descartada. Ao filtrado foram adicionados 110µL de isopropanol 100% e centrifugado a 4.000 rpm por 45 min. O sobrenadante foi descartado e lavado

com 200 $\mu$ L de etanol 70% a 4.000 rpm por 5 min. O precipitado foi ressuspenso em 40 $\mu$ L de água ou TE.

Os plasmídeos purificados foram sequenciados com o uso de *primers* do vetor (T7 e SP6).

#### **4.9. Análise dos *contigs***

As sequências contíguas (*contigs*) que não entraram na montagem final do genoma do cloroplasto (GenBank NC 006137; Hagopian *et al.*, 2004), que foi previamente sequenciado em nosso laboratório, foram recuperados do banco de dados (<http://gracilaria.ib.usp.br/mirror/gt-chloroplast/>) e comparados (BLASTN e BLASTX) às sequências disponíveis no GenBank (<http://www.ncbi.nlm.nih.gov/>). Os resultados das buscas foram vasculhados através de palavras chave (“mitochondrial”; “*Chondrus*” e “*Porphyra*”) para mais rapidamente localizar os *contigs* que continham similaridade ao mtDNA. Foi utilizado o conjunto de programas Phred/Phrap/Consed (Gordon *et al.*, 1998) para a verificação das montagens dos *contigs* e a qualidade das sequências.

Os *contigs* mitocondriais e as sequências consenso anteriormente obtidas foram alinhadas no programa BioEdit Sequence Alignment Editor (Hall, 1999), versão 7.0.5.3, utilizando-se como referência o genoma mitocondrial completo de *C. crispus* (GenBank NC 001677).

#### **4.10. Análise das sequências geradas**

Após o sequenciamento descrito acima, as sequências obtidas nas direções direta e reversa dos produtos de PCR purificados e da clonagem, foram comparadas às sequências disponíveis no GenBank (<http://www.ncbi.nlm.nih.gov/>) por BLASTN e BLASTX (Altschul *et al.* 1997) e alinhadas no programa BioEdit (Hall, 1999), para a geração de uma sequência consenso por produto de PCR.

No caso de posições divergentes, os cromatogramas das sequências foram analisados para confirmar os resultados.

Todas as sequências geradas foram montadas também no pacote de programas Phred/Phrap/Consed (Gordon *et al.*, 1998), para a formação de novos *contigs* e

elaboração de novos *primers*. Esse conjunto de programas permite avaliar a qualidade da sequência, além de fazer um alinhamento automático levando-se em conta essa qualidade e eliminando fragmentos não desejados, como vetores, por exemplo.

#### **4.11. Análise do genoma completo**

A partir da geração dos *contigs* no pacote de programas Phred/Phrap/Consed (Gordon *et al.*, 1998), foi feita a montagem manual da sequência consenso do genoma mitocondrial completo de *G. tenuistipitata*, com o auxílio do programa BioEdit (Hall, 1999). Este programa permite selecionar a região de quebra do genoma para linearizá-lo, uma vez que o genoma mitocondrial é circular.

Para facilitar a comparação, foi escolhida como início do genoma mitocondrial de *G. tenuistipitata* a região do RNA transportador para asparagina (tRNA<sup>Asn</sup>), a mesma utilizada em *C. crispus*. Uma vez linearizado o genoma, foram feitas análises para detectar as localizações das proteínas e dos RNAs estruturais (rRNA e tRNA) .

Para a localização de proteínas foi utilizada uma ferramenta disponível no GenBank (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) denominada Open Reading Frame Finder (ORF Finder), que faz uma listagem dos possíveis quadros abertos de leitura, ou seja, dos possíveis genes codificadores de proteínas. O programa faz seis análises simultâneas do genoma, levando-se em consideração três possibilidades de início do códon em cada fita (direta ou reversa).

Para a localização dos RNAs transportadores (tRNA) foi utilizado o programa tRNAscan-SE (Lowe & Eddy, 1997 - <http://selab.janelia.org/tRNAscan-SE/>), que indica quais e onde estão os genes, além de fornecer suas estruturas secundárias.

## 5. RESULTADOS E DISCUSSÃO

Neste trabalho foram utilizadas duas estratégias para a obtenção da sequência do genoma mitocondrial de *G. tenuistipitata*. A primeira estratégia visava a obtenção de mtDNA através de centrifugações diferenciais, e a segunda, por amplificações através da técnica de PCR. Este último método foi possível de ser realizado porque já havia três sequências de mtDNA de algas vermelhas disponíveis no GenBank, que foram alinhadas com algumas sequências mitocondriais de *G. tenuistipitata* disponíveis em nosso banco de dados, resultados de um trabalho anterior no qual foi sequenciado o genoma completo do cloroplasto desta alga (Hagopian *et al.* 2004). A partir desses fragmentos, foi possível desenhar *primers* para a amplificação e sequenciamento do restante do genoma mitocondrial.

### 5.1. Obtenção de mtDNA

*Gracilaria tenuistipitata* é uma espécie de alga não nativa no Brasil, logo, toda a biomassa para os experimentos teve que ser crescida *in vitro*. Esta foi uma das limitações para a realização dos experimentos. As condições de crescimento *in vitro* foram previamente estabelecidas em nosso laboratório (Macchiavello *et al.*, 1998).

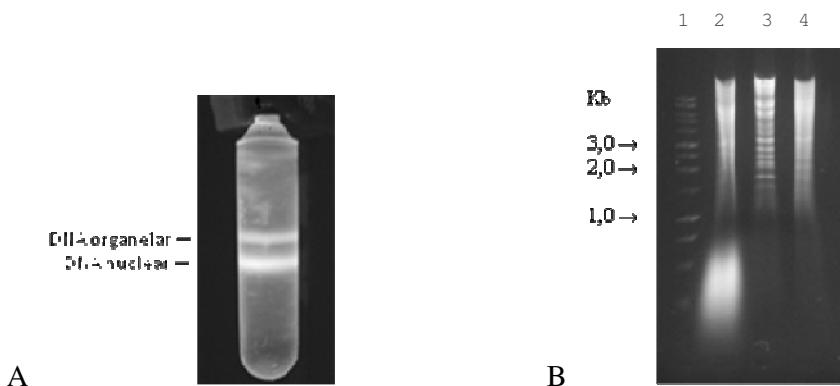
Os métodos utilizados para a obtenção de mtDNA purificado são diversos. Para a alga unicelular *Cyanidioschyzon merolae* suas células foram agitadas com contas de vidro (0,05mm) e lisadas com SDS e proteinase K. Após uma série de extrações com fenol/clorofórmio, os DNAs foram parcialmente purificados através do gradiente de densidade com cloreto de césio com o corante Hoechst 33258 e as diferentes bandas coletadas. Nesse protocolo, o DNA é separado em bandas por densidade, sendo que três bandas podem ser obtidas de acordo com o material: superior, intermediária e inferior, que correspondem aos DNAs mitocondrial, plastidial e nuclear, respectivamente (Suzuki *et al.*, 1992). Posteriormente, o mtDNA foi digerido com as endonucleases *EcoRI* e *HindIII*, e clonado (Ohta *et al.*, 1998). Já o mtDNA de *Porphyra purpurea* foi isolado após digestão da fração do DNA rico em A-T (também obtida em gradiente de CsCl) com a enzima de restrição *SacI* e recuperado por eletroeluição em gel de agarose.

A princípio, o protocolo de extração de DNA utilizado neste estudo havia sido elaborado a partir dos trabalhos de Hagopian *et al.* (2002) e de Triboush *et al.* (1998). Este protocolo visava o enriquecimento da solução com DNA organelar rico em A-T,

através de centrifugações diferenciais. Porém o rendimento era muito baixo (a cada 50g de alga fresca, era obtido cerca de 20 $\mu$ g de DNA total). Portanto, um outro protocolo foi adaptado de Rice & Bird (1990), que apresenta um rendimento maior (cerca de 800 $\mu$ g de DNA total usando-se 10g de alga fresca).

Após a mudança na extração de DNA total, sua purificação foi feita por meio de ultracentrifugação isopícnica em gradiente de CsCl (Aldrich & Cattolico, 1981), utilizando o corante Hoechst 33258 (bisbenzimidina). Este corante se liga entre as bases adjacentes da dupla fita de DNA, com maior afinidade para bases A-T (Chesnick & Cattolico, 1993), conferindo assim uma diferença de densidade e possibilitando a separação das frações de DNA em bandas. Era esperado a obtenção de três bandas: uma nuclear, uma plastidial e uma mitocondrial. Porém, foram obtidas apenas duas: uma superior, contendo os genomas organelares, e uma inferior, contendo o genoma nuclear (Figura 5A).

Foi feita uma digestão com a enzima de restrição *EcoRI*, para confirmar se a banda superior correspondia ao DNA organelar, o que foi confirmada pelo padrão de bandas apresentado no gel de eletroforese (Figura 5B). Devido à grande diferença de tamanho entre os genomas organelares e nuclear, esperava-se obter um padrão de bandas discretas para os genomas organelares e um padrão de bandas contínuas para o DNA nuclear.



**Figura 5.** **A.** Resultado da ultracentrifugação em gradiente de CsCl com o corante Hoechst 33258, com a formação de duas bandas de DNA. **B.** Gel de eletroforese corado com brometo de etídio após digestão com *EcoRI* das bandas de DNA recuperadas do gradiente de CsCl. A primeira coluna é o marcador 1Kb Ladder (Promega). A segunda coluna é o DNA total, a terceira é a banda superior e a quarta, a banda inferior. Nota-se um padrão de bandas discretas na terceira coluna. As fotos foram tiradas em luz UV.

O objetivo do gradiente de CsCl era isolar o genoma mitocondrial, para a clonagem e construção de uma biblioteca de *shotgun*. Entretanto, o total isolamento do mtDNA não foi possível, embora diferentes parâmetros tenham sido testados durante a realização deste experimento, como a utilização de diferentes concentrações de CsCl, do corante Hoescht e do DNA.

### **5.2. Primer walking**

Em paralelo aos experimentos visando a obtenção de mtDNA, foram testados vários pares de *primers* para a amplificação do genoma mitocondrial de *G. tenuistipitata* (Tabela 3) a partir do DNA total. Os *primers* variam em tamanho, temperatura de alinhamento e alguns são degenerados (apresentam bases divergentes em uma ou mais posições).

Os genomas organelares são extremamente ricos em bases A-T, cerca de 70% do total (Gillham, 1994). Devido a isso, há uma certa dificuldade em se desenhar *primers*, pois devido à baixa TM, estes podem formar dímeros e estruturas secundárias (*loops* e *hairpins*) e alinhar em diferentes regiões do genoma que possam ser parecidas. Os genomas organelares apresentam várias pequenas sequências repetitivas, que também dificultam o pareamento dos *primers* (Schuster & Brennicke, 1994; Bullerwell & Gray, 2004). Logo, para a obtenção de fragmentos maiores de PCR foram desenhados *primers* maiores (com até 30 nucleotídeos) para que se pudesse atingir TM maior, recomendado pelo fabricante da enzima elongase (Invitrogen) e aumentar a especificidade, diminuindo a possibilidade de se amplificar mais de uma região.

Alguns *primers* apresentaram resultados insatisfatórios, devido ao resultado negativo para a PCR, ou com a amplificação de diversas regiões do genoma ao mesmo tempo, formando múltiplas bandas no gel (Tabela 3). A maioria desses oligonucleotídeos foram desenhados apenas com base nos genomas mitocondriais de *C. crispus* e *P. purpurea*, e por isso o desempenho irregular.

Dada a dificuldade em se obter mtDNA purificado e ao relativo sucesso da estratégia de *primer walking*, optou-se em usar essa metodologia para finalizar o sequenciamento do genoma mitocondrial de *G. tenuistipitata*.

**Tabela 3.** Descrição de todas as combinações de *primers* testados para PCR e seus respectivos resultados. A sequência e localização de cada *primer* estão descritos na Tabela 2. Estão descritos também o tamanho esperado com base na localização dos *primers* no genoma mitocondrial de *Chondrus crispus* e o tamanho do fragmento obtido para *Gracilaria tenuistipitata*. RN= resultado negativo e MB= múltiplas bandas.

Combinação de <i>primers</i>	Tamanho esperado (Kb) <i>C. crispus</i>	Tamanho obtido (Kb) <i>G.tenuistipitata</i>	DNA Polimerase
rps12F × rnsR	1,6	MB	Taq Invitrogen
rps12F × nad4LR	1,5	1,5	Taq Invitrogen
rps12F × rnlR4	2,5	2,5	Taq Invitrogen
rps12F × rnlR4	2,5	2,0 e 0,7	Elongase Invitrogen
rpl20F × rnsR3	1,1	MB	Taq Invitrogen
rpl20F × rnsR4	1,2	MB	Taq Invitrogen
rpl20F × nad4LR	1,5	1,5	Taq Invitrogen
rpl20F × rnlR4	1,9	RN	Taq Invitrogen
rnsF3 × nad4LR	0,8	0,8	Taq Invitrogen
rnsF3 × nad4LR2	1,0	1,0	Taq Invitrogen
rnsF4 × nad4LR	0,5	0,5	Taq Invitrogen
rnsF4 × nad4LR2	0,7	0,7	Taq Invitrogen
rnsF5 × nad4LR	0,4	0,4	Taq Invitrogen
rnsF5 × nad4LR2	0,6	0,6	Taq Invitrogen
rnsF × rnlR	2,4	2,5	GoTaq Promega
rnsF × GazR1	5,0	RN	GoTaq Promega
nad4LF × rnlR4	0,3	0,3	Taq Invitrogen
nad4LF × rnlR7	1,7	1,7	Taq Invitrogen
nad4LF × rnlR	2,2	2,3	GoTaq Promega
nad4LF × rnlR3	2,5	2,0	Elongase Invitrogen
nad4LF × cox1-5'R	4,0	3,3	Elongase Invitrogen
nad4LF × GazR1	4,8	3,5	Elongase Invitrogen
nad4LF × cox3R	6,6	RN	Elongase Invitrogen
rnlF4 × rnlR3	2,3	2,3	Elongase Invitrogen
rnlF4 × cox1-5'R	3,8	3,8	Elongase Invitrogen
rnlF4 × GazR1	4,5	4,5	Elongase Invitrogen
rnlF7 × cox1-5'R	3,3	3,3	Elongase Invitrogen
rnlF × cox1R	3,0	RN	GoTaq Promega
rnlF × GazR1	2,6	2,6	GoTaq Promega
rnlF3 × GazR1	2,3	2,0	Taq Invitrogen
rnlF3 × cox3R	4,1	4,1	Elongase Invitrogen
rnlF2 × cox3R	4,0	4,0	Elongase Invitrogen
GazF1 × GazR1	0,7	0,7	GoTaq Promega
GazF1 × cox1R	1,1	1,1	GoTaq Promega
GazF1 × cox3R	2,6	2,6	GoTaq Promega
cox1F2 × cox1R	0,4	RN	GoTaq Promega
cox1F2 × cox3R	1,5	RN	GoTaq Promega
cox1F3 × cox1R3	0,4	0,4	Taq Invitrogen
cox1F × cox3R	1,5	0,6	GoTaq Promega
cox2F × cox3R	0,4	0,4	GoTaq Promega
cox2F × ymf39R	2,0	1,6	Taq Invitrogen

**Tabela 3.** Continuação.

Combinação de primers	Tamanho esperado (Kb) <i>C. crispus</i>	Tamanho obtido (Kb) <i>G.tenuistipitata</i>	DNA Polimerase
cox2F x cobR	2,3	2,3	GoTaq Promega
cox2F x nad3R	6,7	RN	Elongase Invitrogen
ymf39F x nad2R2	6,4	6,4	Elongase Invitrogen
ymf39F x nad2R3	6,0	6,0	Elongase Invitrogen
cobF x nad3R	3,5	MB	Elongase Invitrogen
cobF x nad1R	5,0	MB	Elongase Invitrogen
cobF x nad2R	6,5	6,5 e 2,5	Elongase Invitrogen
cobF x nad4R	8,5	RN	Elongase Invitrogen
cobF2 x nad2R3	5,6	5,6	Elongase Invitrogen
cobF2 x nad2R2	5,9	5,9	Elongase Invitrogen
cobF3 x nad1R2	4,6	4,6	Elongase Invitrogen
nad6F2 x nad3R2	3,0	MB	Elongase Invitrogen
nad6F2 x nad1R2	3,8	3,8	Elongase Invitrogen
nad6F x rps11R	2,2	RN	Elongase Invitrogen
nad6F x nad1R2	3,5	3,5	Elongase Invitrogen
nad3F x nad2R	2,0	MB	Taq Invitrogen
nad2F3 x nad5R2	5,9	5,6	Elongase Invitrogen
nad2F2 x nad5R3	4,5	4,5	Elongase Invitrogen
nad2F2 x nad5R4	4,9	4,8	Elongase Invitrogen
nad2F2 x nad5R2	5,5	5,5	Elongase Invitrogen
nad2F x nad4R	2,0	MB	Taq Invitrogen
nad2F x nad5R	4,7	MB	Elongase Invitrogen
nad2F x atp8R	5,4	MB	Elongase Invitrogen
nad2F x atp6R	5,6	MB	Elongase Invitrogen
nad4F3 x nad5R5	2,7	2,7	Elongase Invitrogen
nad4F3 x nad5R2	3,7	3,7	Elongase Invitrogen
nad4F2 x nad5R5	2,3	2,3	Elongase Invitrogen
nad4F x nad5R	2,7	MB	Taq Invitrogen
nad4F x atp8R	3,5	3,5	Elongase Invitrogen
nad4F x atp6R	3,7	MB	Elongase Invitrogen
nad5F3 x atp8R2	3,0	3,0	Elongase Invitrogen
nad5F3 x nad5R2	2,5	2,5	Elongase Invitrogen
nad5F3 x atp6R2	3,0	3,0	Elongase Invitrogen
nad5F x atp6R	1,0	1,5 e 0,5	Taq Invitrogen
nad5F2 x atp6R2	0,7	0,7	Taq Invitrogen
nad5F2 x rps12R	3,9	RN	Elongase Invitrogen
nad5F2 x rps12R2	4,0	MB	Elongase Invitrogen
nad5F2 x rnsR	5,5	MB	Elongase Invitrogen
atp8F x rps12R	3,3	3,8 e 0,7	Elongase Invitrogen
atp8F x rps12R2	3,5	MB	Elongase Invitrogen
atp8F x rnsR	4,9	MB	Elongase Invitrogen
atp6F2 x rps12R	3,2	4,6	Elongase Invitrogen
atp6F2 x rps12R4	3,3	MB	Elongase Invitrogen
atp6F2 x rps12R2	3,4	MB	Elongase Invitrogen

**Tabela 3.** Continuação.

Combinação de <i>primers</i>	Tamanho esperado (Kb) <i>C. crispus</i>	Tamanho obtido (Kb) <i>G.tenuistipitata</i>	DNA Polimerase
atp6F2 × rnsR	4,8	MB	Elongase Invitrogen
atp6F3 × rps12R	2,9	4,5	Elongase Invitrogen
atp6F3 × rps12R4	3,0	MB	Elongase Invitrogen
atp6F3 × rnsR2	3,9	MB	Elongase Invitrogen
atp6F3 × nad4LR	5,0	MB	Elongase Invitrogen
atp6F × rps12R2	3,3	MB	Elongase Invitrogen
atp6F × rnsR	4,7	MB	Elongase Invitrogen
atp6F5 × rps12R	2,5	MB	Elongase Invitrogen
atp6F5 × rps12R4	3,3	MB	Elongase Invitrogen
atp6F5 × nad4LR	4,6	MB	Elongase Invitrogen
orf172F × rps12R3	1,9	RN	Taq Invitrogen
orf172F × nad4LR	4,0	RN	Elongase Invitrogen

### 5.3. Sequenciamento e montagem do genoma mitocondrial

O genoma do cloroplasto de *G. tenuistipitata* foi completamente sequenciado pelo nosso grupo (GenBank NC 006137; Hagopian *et al.* 2004) a partir de uma biblioteca de *shotgun*. Essa biblioteca foi construída com DNA isolado de cloroplastos, entretanto, apesar da baixa contaminação (89% dos clones sequenciados eram DNA plastidial), foram sequenciados clones contendo DNA de origem nuclear e especialmente DNA de origem mitocondrial. Os *contigs* que não entraram na montagem final do genoma do cloroplasto foram recuperados do banco de dados e comparados às sequências disponíveis no GenBank (BLASTN e BLASTX). Os resultados dessa busca foram analisados e foram identificados sete *contigs* que continham sequências de origem mitocondrial, formando um total de 10.428 pares de base (aproximadamente 1/3 do esperado para o genoma mitocondrial completo; Tabela 4). Foram identificados nesses *contigs* os genes mitocondriais: *cox1*, *cox2*, *cox3*, *cob*, *atp6*, *nad2*, *nad4*, *nad4L* e *rns*.

As sequências desses *contigs* foram alinhadas com as sequências dos genomas mitocondriais completos de *C. crispus* e *P. purpurea*, e a partir desses alinhamentos iniciou-se a construção do *primers* para a amplificação de outros segmentos do genoma mitocondrial de *G. tenuistipitata*.

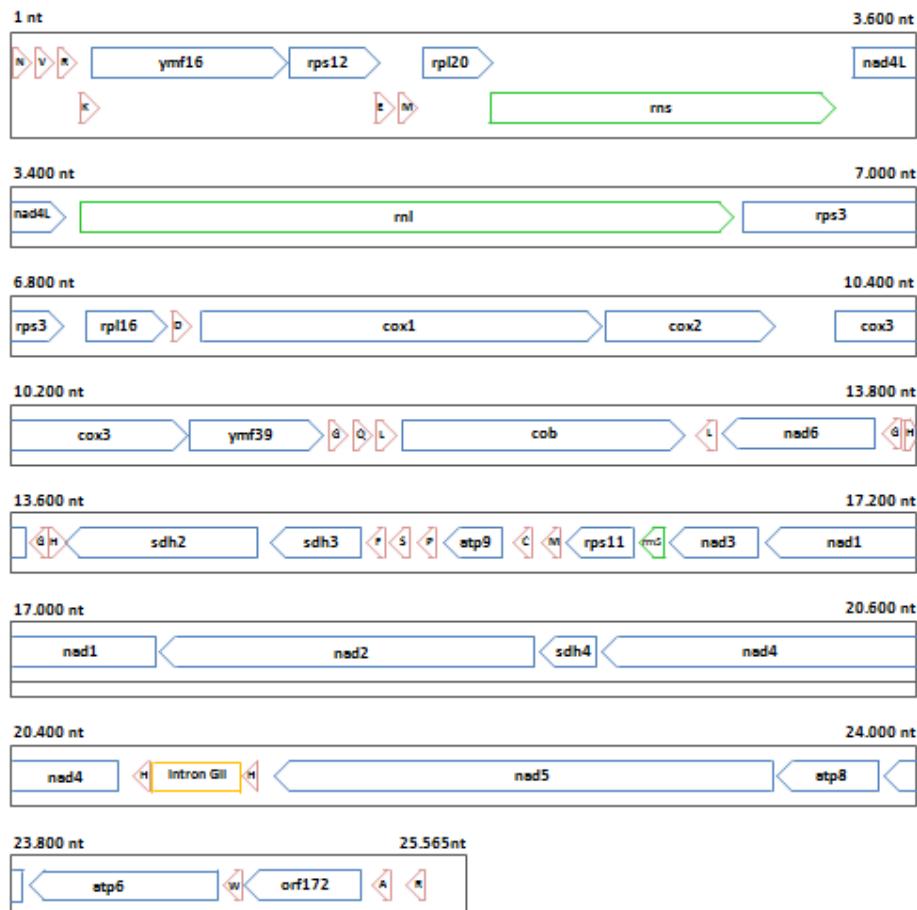
O sequenciamento do genoma mitocondrial de *G. tenuistipitata* foi realizado por *primer walking*, ou seja, a partir do sequenciamento das pontas dos produtos

amplificados por PCR, foram desenhados *primers* mais internos para a obtenção de todo o fragmento.

**Tabela 4.** Contigs recuperados da montagem do genoma plastidial de *Gracilaria tenuistipitata* que contém sequências de origem mitocondrial.

Contig #	pb	Organismo/e-value (BLASTX/BLASTN)	Produto (gene)
1	910	<i>Gracilariopsis lemaneiformis</i> 5e-51/2e-109	Citocromo c oxidase subunidade 1 ( <i>cox1</i> )
2	606	<i>Gracilariopsis lemaneiformis</i> <i>Gracilaria tenuistipitata</i> 3e-45/3e-67	Citocromo c oxidase subunidade 3 ( <i>cox3</i> ) Citochromo c oxidase subunidade 2 ( <i>cox2</i> )
3	791	<i>Chondrus crispus</i> 1e-54/1e-14	NADH desidrogenase subunidade 4 ( <i>nad4</i> )
4	751	<i>Chondrus crispus</i> 2e-104/8e-87	Apocitocromo b ( <i>cob</i> )
5	1826	<i>Chondrus crispus</i> 4e-94/9e-30	ATP sintase F0 subunidade 6 ( <i>atp6</i> )
6	2398	<i>Chondrus crispus</i> 2e-104/1e-07	NADH desidrogenase subunidade 2 ( <i>nad2</i> )
7	3146	<i>Chondrus crispus</i> 8e-36/7e-180	NADH desidrogenase subunidade 4L ( <i>nad4L</i> ) Subunidade pequena do rRNA ( <i>rns</i> )
Total	10.428		

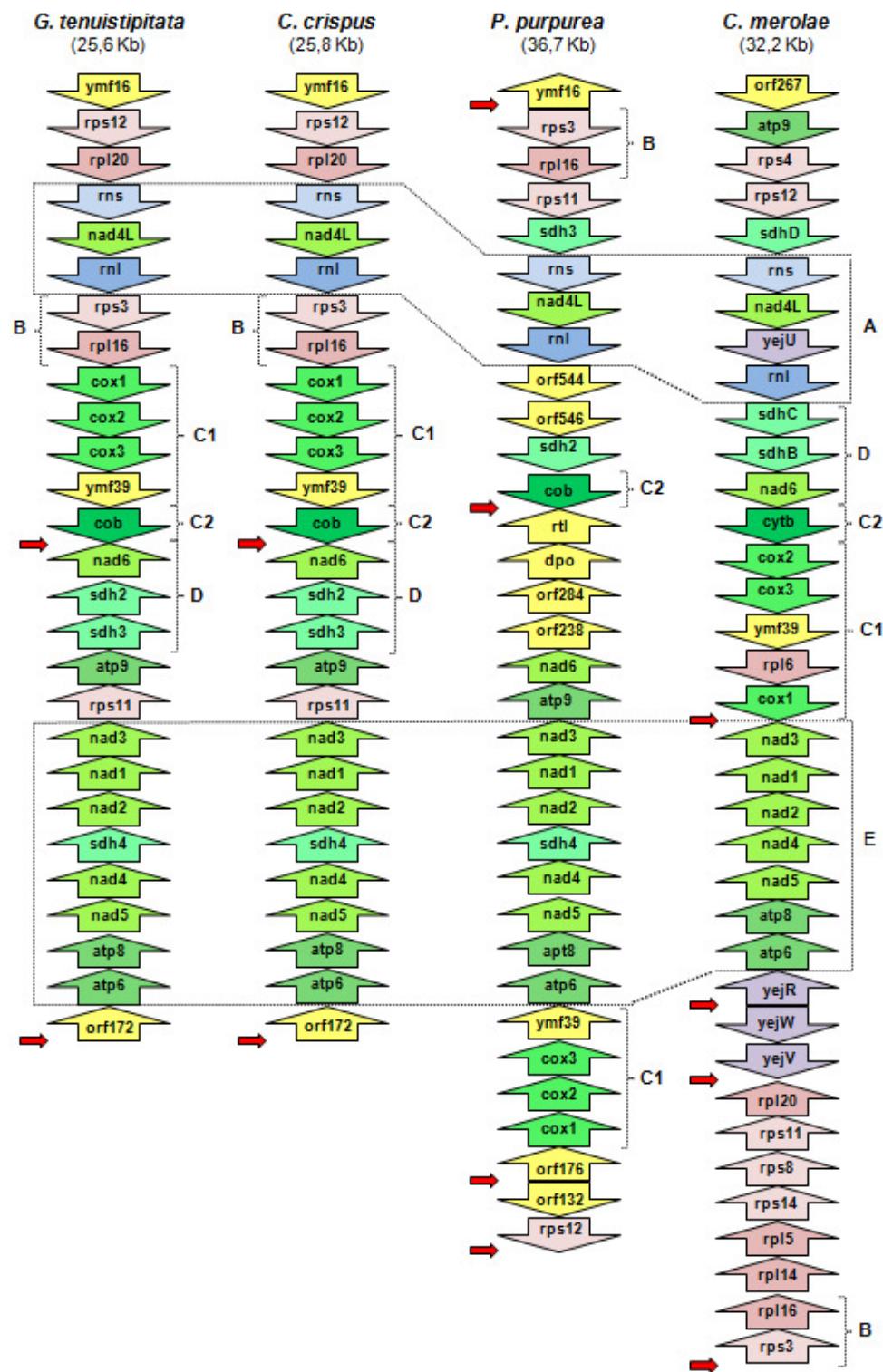
Todas as sequências obtidas foram comparadas ao banco de dados do GenBank por BLASTN e BLASTX (Altschul *et al.* 1997) e alinhadas no pacote de programas Phred/Phrap/Consed (Gordon *et al.*, 1998). A montagem final da sequência consenso do genoma mitocondrial completo de *G. tenuistipitata* (Figura 6) foi feita manualmente no programa BioEdit (Hall, 1999), onde foi delimitado o local de quebra do genoma circular para linearizá-lo, facilitando as análises posteriores. A linearização do genoma mitocondrial de *G. tenuistipitata* foi baseada no genoma mitocondrial linearizado de *C. crispus*, que apresenta duas direções de transcrição opostas, localizado entre os genes *orf172* e *ymf16*, e *cob* e *nad6*. A origem de replicação do mtDNA estaria em uma destas regiões (Leblanc *et al.* 1995).



**Figura 6.** Esquema representando o genoma mitocondrial completo e linearizado de *Gracilaria tenuistipitata*. Em azul estão os genes que codificam proteínas. Em verde, os genes para rRNA e em rosa, os tRNAs. Os RNAs transportadores estão identificados pelas letras correspondentes ao seu aminoácido. Em amarelo está representado o Intron do grupo II presente no tRNA<sup>His</sup>. Entre os genes *cox2* e *cox3* há uma região espaçadora utilizada como marcador molecular (Zuccarello *et al.* 1999). Os números acima dos quadros representam a extensão da sequência em nucleotídeos.

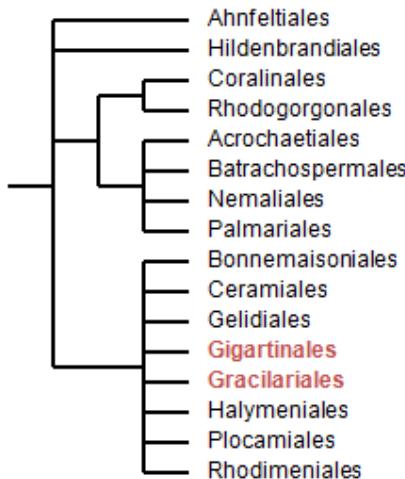
#### **5.4. Tamanho e genes presentes no mtDNA**

O genoma mitocondrial de *G. tenuistipitata* é circular, contém 25.565 pb e o total de nucleotídeos A+T representa cerca de 73%, números muito similares aos observados para o genoma mitocondrial de *C. crispus* (Tabela 1). Além disso, a composição de genes e sua distribuição no genoma mitocondrial de *G. tenuistipitata* é praticamente a mesma que em *C. crispus* (Figura 7). Ambas as espécies pertencem a ordens diferentes, mas dentro da mesma classe (Florideophyceae).



**Figura 7.** Comparação entre os quatro genomas mitocondriais completos de Rhodophyta. Em tons de azul estão representados genes para rRNA. Em tons de rosa, genes para proteínas ribossômicas. Em tons de verde, genes para a cadeia respiratória de elétrons e síntese de ATP. Em lilás estão representados genes envolvidos na biogênese do citocromo tipo c. Em amarelo, os genes para outras proteínas ou cuja função é desconhecida. A orientação das setas verticais indica o sentido dos genes. As setas horizontais (vermelhas) indicam mudança de direção dos genes. As letras (A-E) ao lado indicam regiões conservadas dos genomas mitocondriais das quatro algas vermelhas. O bloco A é composto pelos genes *rns*, *nad4L* e *rnl*, com exceção de *C. merolae*, que apresenta o gene *yεjU* entre eles. Todos os genes da região A estão orientados na mesma direção. O bloco B é composto pelos genes *rps3* e *rpl16*, também com sentido invertido apenas em *C. merolae*. O bloco C1 é composto pelos genes *cox1-3* e *ymf39*. Em *P. purpurea* o bloco está invertido, e em *C. merolae* há um rearranjo entre os genes, com *rpl6* no meio e *cox1* afastado dos demais citocromos. O bloco C2 é composto pelo gene apocitocromo b, que está próximo ao bloco C1 em três genomas mitocondriais, com exceção *P. purpurea*. O bloco D é composto pelos genes *sdh2-3* e *nad6*, cuja direção está invertida em *C. merolae* em comparação aos genomas mitocondriais de *G. tenuistipitata* e *C. crispus*, e em *P. purpurea* os genes estão afastados um do outro. A região E é a mais conservada em todos os genomas mitocondriais e é composta pelos genes *atp6*, *atp8*, *nad5*, *nad4*, *sdh4*, *nad2*, *nad1* e *nad3*, com exceção de *C. merolae*, que apresenta o gene *sdh4* afastado do bloco. Não foram representadas orfs sobrepostas a genes conhecidos. Os genes *sdhB-D* são homólogos aos genes *sdh2-4*, respectivamente, assim como os genes *cob* e *cytb*, e *ymf16* e *orf267*.

Assim como a filogenia de Rhodophyta ainda não foi completamente esclarecida (Yoon *et al.* 2006), a árvore filogenética da classe Florideophyceae proposta pelo projeto Tree of Life (<http://tolweb.org/Florideophyceae>) também não está totalmente resolvida, apresentando politomias (Figura 8). Porém, pode-se notar que as ordens de *G. tenuistipitata* (Gracilariales) e *C. crispus* (Gigartinales) estão próximas, formando um clado junto com outras seis ordens.



**Figura 8.** Árvore filogenética para a classe Florideophyceae apresentado pelo projeto Tree of Life. Em vermelho estão destacadas as ordens Gigartinales, no qual se encontra *Chondrus crispus*, e Gracilariales, onde está *Gracilaria tenuistipitata*. Nota-se que as duas ordens estão no mesmo clado.

Portanto, a grande similaridade entre os genomas mitocondriais de *G. tenuistipitata* e *C. crispus* pode indicar uma proximidade entre as ordens a que pertencem.

No genoma mitocondrial de *G. tenuistipitata* foram identificados 50 genes, sendo 25 genes codificadores de proteínas, 3 genes para rRNA e 22 para tRNA ( Figura 6 e Tabelas 5, 6 e 7)

A ordem e o sentido dos genes para proteínas são iguais em *G. tenuistipitata* e *C. crispus*, assim como para os rRNAs (Figura 7). Porém, alguns tRNAs são diferentes entre essas duas espécies, como será discutido mais adiante.

**Tabela 5:** Resumo de todos os genes identificados no genoma mitocondrial de *Gracilaria tenuistipitata*, de acordo com suas funções.

---

**A. Genes para RNA ribossômico (3)**

Subunidade pequena: *rns*

Subunidade grande: *rnl*

5S rRNA: *rrn5*

**B. Genes para RNA transportador (22 no total, Tabela 7 e Anexo 2)**

**C. Genes para proteína ribossômica (5)**

Subunidade pequena: *rps3*, *rps11*, *rps12*

Subunidade grande: *rpl16*, *rpl20*

**D. Genes para o transporte de elétrons e fosforilação oxidativa (17)**

Complexo I da cadeia respiratória: *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*

Complexo II da cadeia respiratória: *sdh2*, *sdh3*, *sdh4*

Complexo III da cadeia respiratória: *cob*

Complexo IV da cadeia respiratória: *cox1*, *cox2*, *cox3*

Complexo ATP sintase: *atp6*, *atp8*, *atp9*

**E. Genes similares a ORFs com função desconhecida (3)**

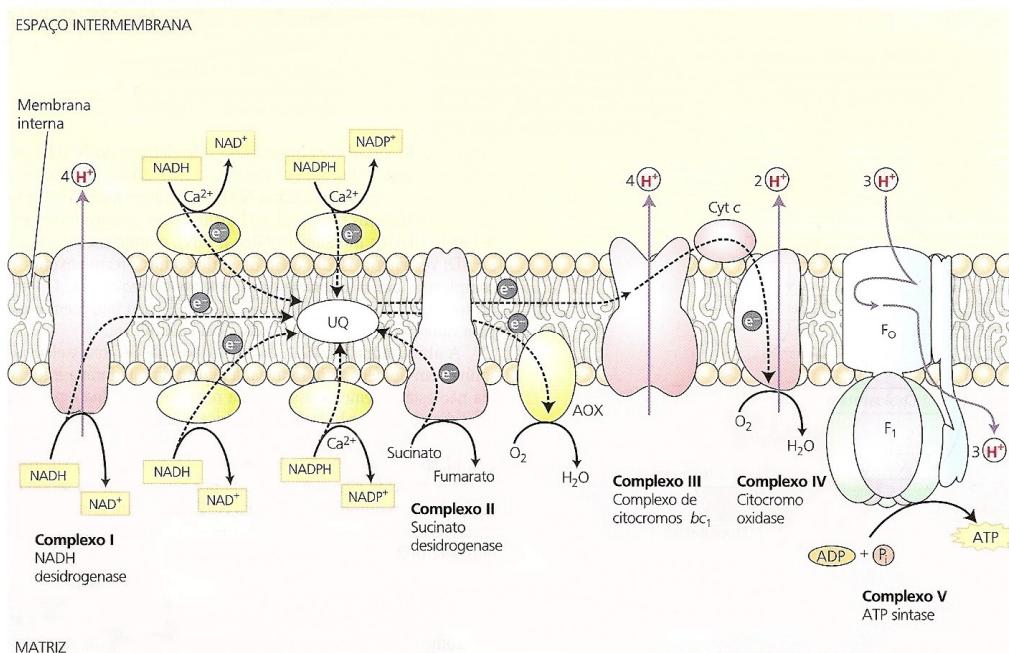
*ymf16*, *ymf39*, *orf172*

---

## 5.5. Genes codificadores de proteínas

As proteínas da cadeia de transporte de elétrons estão organizadas em quatro complexos multiprotéicos, localizados na membrana interna da mitocôndria (Figura 9). O Complexo I (NADH desidrogenase) oxida os elétrons do NADH gerados na matriz mitocondrial durante o ciclo do ácido cítrico e transfere estes elétrons à ubiquinona (UQ), que é um pequeno carregador de prótons e elétrons. O Complexo II (succinato desidrogenase) catalisa a oxidação do succinato no ciclo do ácido cítrico, transferindo elétrons também à ubiquinona. O Complexo III (complexo de citocromos bc<sub>1</sub>) recebe os elétrons da ubiquinona e transfere para o Complexo IV (citocromo c oxidase), que realiza a redução de moléculas de O<sub>2</sub> a H<sub>2</sub>O. Todo este processo de transferência de

elétrons para o oxigênio está acoplada à síntese de ATP via a ATP sintase (Taiz & Zeiger, 2004).



**Figura 9.** Organização da cadeia de transporte de elétrons e síntese de ATP na membrana interna das mitocôndrias vegetais. Nesta figura estão representados os quatro complexos da cadeia respiratória (em rosa) e o complexo ATP sintase (em azul). Modificado de Taiz & Zeiger (2004).

O genoma mitocondrial de *G. tenuistipitata* apresenta genes para os quatro complexos da cadeia respiratória e para a síntese de ATP (Tabela 5). Os sete genes do complexo I encontram-se quase todos agrupados (*nad1* ao *nad5*, interrompidos pelo gene *sdh4*), com exceção dos genes *nad6* e *nad4L* (Figura 7). Somente o gene *nad4L* é codificado no sentido direto, o restante está localizado na fita complementar do DNA. Os três genes do complexo II (*sdh2*, *sdh3* e *sdh4*) estão orientados na mesma direção, localizados entre os genes do complexo NADH desidrogenase (Figura 7). O gene apocitocromo b (*cob*) é codificado no sentido direto e os genes do complexo IV (*cox1*, *cox2* e *cox3*) formam um único bloco no mesmo sentido, próximos ao gene do complexo III (Figura 7).

Os genes *atp6* e *atp8* estão localizados ao lado do complexo NADH desidrogenase e o gene *atp9* está um pouco mais distante. Os três genes da ATP sintase

(denominado de complexo V, na Figura 9) encontram-se na mesma fita de DNA (Figura 7).

Nota-se que os genes responsáveis pela cadeia de transporte de elétrons e síntese de ATP estão agrupados, no qual os genes dos complexos I, II e V encontram-se em uma fita (direção reversa) e os genes dos complexos III e IV na outra (direção direta). A exceção é o gene *nad4L*, que se encontra entre os genes ribossômicos *rns* e *rnl* (Figura 7). Observa-se ainda, pela Figura 7 e Tabelas 6 e 7, que o genoma mitocondrial de *G. tenuistipitata* apresenta dois blocos principais: um de sentido direto (entre os genes *ymf16* e *cob*) e um de sentido reverso (entre os genes *nad6* e *orf172*).

Os cinco genes que codificam as subunidades pequenas (*rps3*, *rps11*, *rps12*) e grandes (*rpl16*, *rpl20*) do ribossomo encontram-se espalhados pelo genoma (Figura 7).

Além das proteínas componentes da cadeia respiratória e síntese de ATP, e dos genes ribossômicos, foram identificados genes cujas funções ainda não foram definidas (*ymf16*, *ymf39*, *orf172*), mas que estão presentes também em *C. crispus* (Figura 7 e Anexo 1). Os genes *ymf16* e *ymf39* também são encontrados em *P. purpurea* e *C. merolae* (o gene *orf267* seria homólogo ao gene *ymf16* das demais algas vermelhas), indicando que, por serem conservados, possivelmente são genes ativos (Figura 7).

Entre os genes *cox2* e *cox3* há uma região espaçadora (173 nucleotídeos) muito utilizada como marcador molecular, por ser uma sequência relativamente conservada, mas ao mesmo tempo informativa (Zuccarello *et al.* 1999 e 2002; Milstein *et al.* 2008). Zuccarello *et al.* (1999) desenvolveram *primers* para essa região específicos para Rhodophyta (*cox2F* e *cox3R*).

Neste estudo, não foram descritos *open reading frames* (*orf*) que se sobreponham em genes com função conhecida.

**Tabela 6.** Descrição dos genes codificadores de proteínas no genoma mitocondrial de *Gracilaria tenuistipitata*.

Gene	Descrição	Localização	Sentido
<i>ymf16</i>	Proteína transportadora independente sec-Y	314-1084	Direto
<i>rps12</i>	Proteína ribossômica S12	1084-1467	Direto
<i>rpl20</i>	Proteína ribossômica L20	1628-1912	Direto
<i>nad4L</i>	NADH desidrogenase subunidade 4L	3323-3628	Direto
<i>rps3</i>	Proteína ribossômica S3	6292-7023	Direto
<i>rpl16</i>	Proteína ribossômica L16	6995-7408	Direto
<i>cox1</i>	Citocromo c oxidase subunidade 1	7544-9139	Direto
<i>cox2</i>	Citocromo c oxidase subunidade 2	9141-9896	Direto
<i>cox3</i>	Citocromo c oxidase subunidade 3	10069-10887	Direto
<i>ymf39</i>	Proteína não identificada	10891-11433	Direto
<i>cob</i>	Apocitocromo b	11748-12887	Direto
<i>nad6</i>	NADH desidrogenase subunidade 6	13034-13642	Reverso
<i>sdh2</i>	Citocromo c oxidorredutase subunidade 2	13820-14572	Reverso
<i>sdh3</i>	Citocromo c oxidorredutase subunidade 3	14616-14978	Reverso
<i>atp9</i>	ATP sintase F0 subunidade 9	15302-15532	Reverso
<i>rps11</i>	Proteína ribossômica S11	15719-16078	Reverso
<i>nad3</i>	NADH desidrogenase subunidade 3	16201-16566	Reverso
<i>nad1</i>	NADH desidrogenase subunidade 1	16574-17554	Reverso
<i>nad2</i>	NADH desidrogenase subunidade 2	17563-19065	Reverso
<i>sdh4</i>	Citocromo c oxidorredutase subunidade 4	19087-19326	Reverso
<i>nad4</i>	NADH desidrogenase subunidade 4	19340-20815	Reverso
<i>nad5</i>	NADH desidrogenase subunidade 5	21428-23413	Reverso
<i>atp8</i>	ATP sintase F0 subunidade 8	23427-23834	Reverso
<i>atp6</i>	ATP sintase F0 subunidade 6	23850-24611	Reverso
<i>orf172</i>	Proteína não identificada	24712-25176	Reverso

## 5.6. Genes para RNAs estruturais (rRNA e tRNA)

Foram identificados 3 genes para rRNA e 22 genes para tRNA, listados na Tabela 7. Mais uma vez, a composição e a ordem dos genes são muito similares ao genoma mitocondrial de *C. crispus* (Anexo 1). As estruturas secundárias dos tRNAs encontram-se no Anexo 2.

**Tabela 7.** Descrição do genes codificadores de RNAs estruturais (rRNA e tRNA) no genoma mitocondrial de *Gracilaria tenuistipitata*. (\*) tRNA que contém um intron do grupo II.

<b>RNA Ribossômico</b>				
<b>Gene</b>	<b>Proteína</b>	<b>Localização</b>	<b>Sentido</b>	
<i>rns</i>	RNA ribossômico – subunidade pequena	1904-3290	Direto	
<i>rnl</i>	RNA ribossômico – subunidade grande	3650-6270	Direto	
<i>rrn5</i>	RNA ribossômico 5S	16078-16193	Reverso	

<b>RNA Transportador</b>				
<b>Gene</b>	<b>Aminoácido</b>	<b>Localização</b>	<b>Anti-códon</b>	<b>Sentido</b>
Asn tRNA	Asparagina	1-72	GTT	Direto
Val tRNA	Valina	81-153	TAC	Direto
Arg tRNA	Arginina	173-247	ACG	Direto
Lys tRNA	Lisina	261-334	TTT	Direto
Glu tRNA	Ácido glutâmico	1453-1525	TTC	Direto
Met tRNA	Metionina	1535-1609	CAT	Direto
Asp tRNA	Ácido aspártico	7420-7491	GTC	Direto
Gly tRNA	Glicina	11445-11518	TCC	Direto
Gln tRNA	Glutamina	11531-11603	TTG	Direto
Leu tRNA	Leucina	11619-11704	TAA	Direto
Leu tRNA	Leucina	12945-13028	TAG	Reverso
Gly tRNA	Glicina	13661-13733	GCC	Reverso
His tRNA	Histidina	13745-13819	GTG	Direto
Phe tRNA	Fenilalanina	14989-15061	GAA	Reverso
Ser tRNA	Serina	15108-15195	TGA	Reverso
Pro tRNA	Prolina	15208-15281	TGG	Reverso
Cys tRNA	Cisteína	15569-15641	GCA	Reverso
Met tRNA	Metionina	15645-15718	CAT	Reverso
His tRNA	Histidina (*)	20833-21412	GTG	Reverso
Trp tRNA	Triptofano	24648-24721	TCA	Reverso
Ala tRNA	Alanina	25194-25268	TGC	Reverso
Arg tRNA	Arginina	25340-25411	TCT	Reverso

Leblanc *et al.* (1995) descreveram a presença de dois genes para o RNA ribossômico 5S (*rrn5*) em *C. crispus*, um entre os genes *cox2* e *cox3*, e outro entre os genes *rps11* e *nad3*. Porém, quatro anos mais tarde, Burger *et al.* (1999) questionaram a presença do gene *rrn5* entre os genes *cox2-3* de *C. crispus*, tendo sido encontrado apenas

uma estrutura secundária para *rrn5* entre os genes *rps11* e *nad3*. A ocorrência do gene mitocondrial para o RNA ribossômico 5S é esporádica entre os eucariotos, sendo encontrado em plantas e em algumas algas verdes (Burger *et al.* 1999). O gene *rrn5* não foi encontrado em *P. purpurea* (Burger *et al.* 1999) e em *C. merolae* está entre os genes *rpl6* e tRNA<sup>Met</sup> (Ohta *et al.* 1998). Em *G. tenuistipitata*, o gene *rrn5* encontra-se entre os genes *rps11* e *nad3*, como em *C. crispus* (Figura 6).

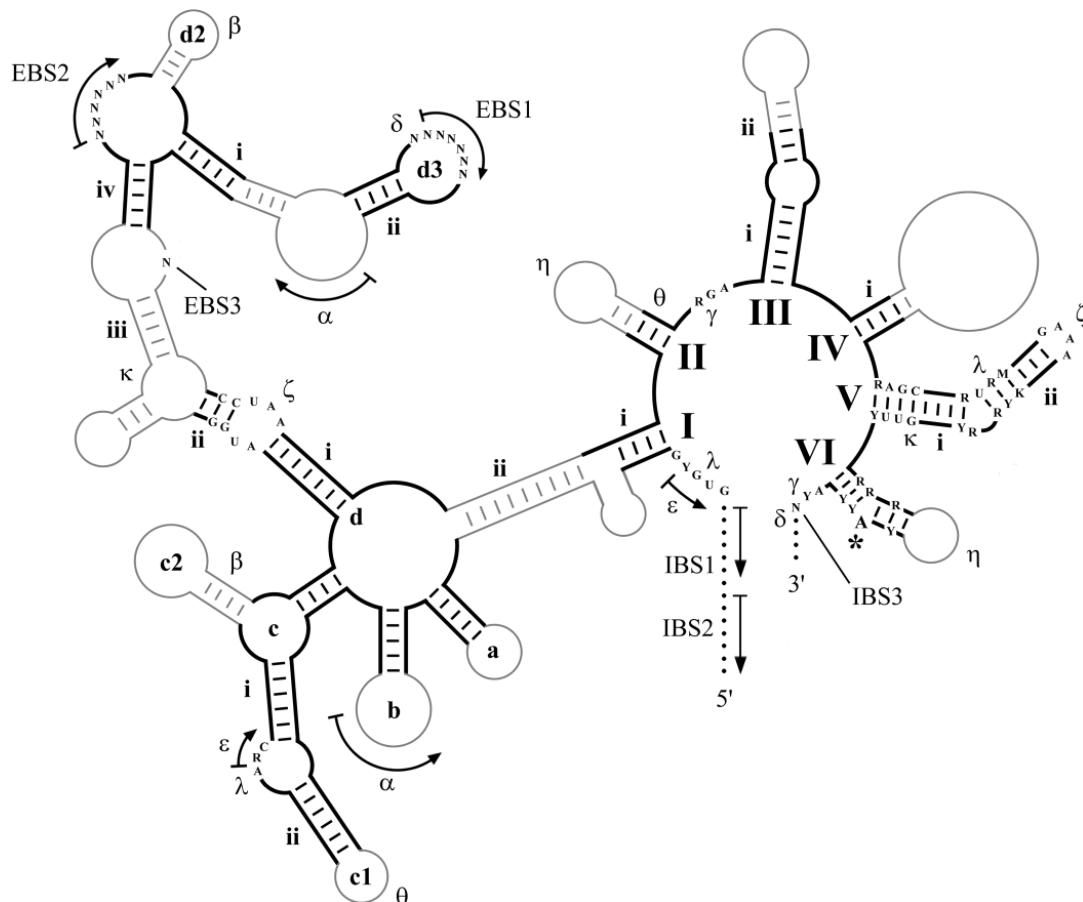
Dos 22 tRNAs encontrados no genoma mitocondrial de *G. tenuistipitata*, 20 são iguais aos de *C. crispus*, existindo apenas duas diferenças entre esses genomas. A primeira diferença é a ausência, em *G. tenuistipitata*, do RNA transportador de tirosina (tRNA<sup>Tyr</sup>), encontrado entre os genes para tRNA de arginina e asparagina do genoma de *C. crispus* (Anexo 1). Observa-se que essa é justamente a região ausente em *G. tenuistipitata*, cujo genoma mitocondrial é ligeiramente menor que o de *C. crispus* (271 nucleotídeos). E a segunda diferença é a ausência do RNA transportador de isoleucina, encontrado em *C. crispus*, e sua substituição por um tRNA de histidina em *G. tenuistipitata*. Entretanto, para uma melhor comparação, o genoma mitocondrial de *C. crispus* foi analisado no programa tRNAscan-SE (Lowe & Eddy, 1997 - <http://selab.janelia.org/tRNAscan-SE/>) e a última diferença deixou de existir. De acordo com este programa, *C. crispus* apresenta tRNA para histidina, igual ao genoma mitocondrial de *G. tenuistipitata*. O sequenciamento completo do genoma mitocondrial de *C. crispus* foi realizado por Leblanc *et al.* em 1995, e talvez as informações contidas no trabalho estejam desatualizadas, uma vez que as análises foram feitas manualmente.

Porém, os tRNAs para histidina (denominado tRNA<sup>Ile</sup> em *C. crispus*) apresentam uma característica em comum: a presença de um intron do grupo II. Em *C. crispus*, o tRNA<sup>Ile</sup> apresenta um intron de 474 nucleotídeos entre as posições 31 e 32 do gene (Leblanc *et al.* 1995). Já em *G. tenuistipitata*, o intron de 503 nucleotídeos encontra-se entre as posições 37 e 38 do gene tRNA<sup>His</sup>. Introns do grupo II são encontrados principalmente em genomas organelares de plantas e fungos, e em algumas linhagens de cianobactérias e proteobactérias, reforçando a teoria endossimbiótica das organelas (Kelchner, 2002; Saldanha *et al.* 1993). A maioria estão em genes codificadores de proteínas, alguns poucos em genes para tRNAs e rRNAs (Saldanha *et al.* 1993).

Cerca de um terço dos introns do grupo II de organelas apresentam orfs homólogos a transcriptase reversa, sugerindo que podem ser elementos móveis

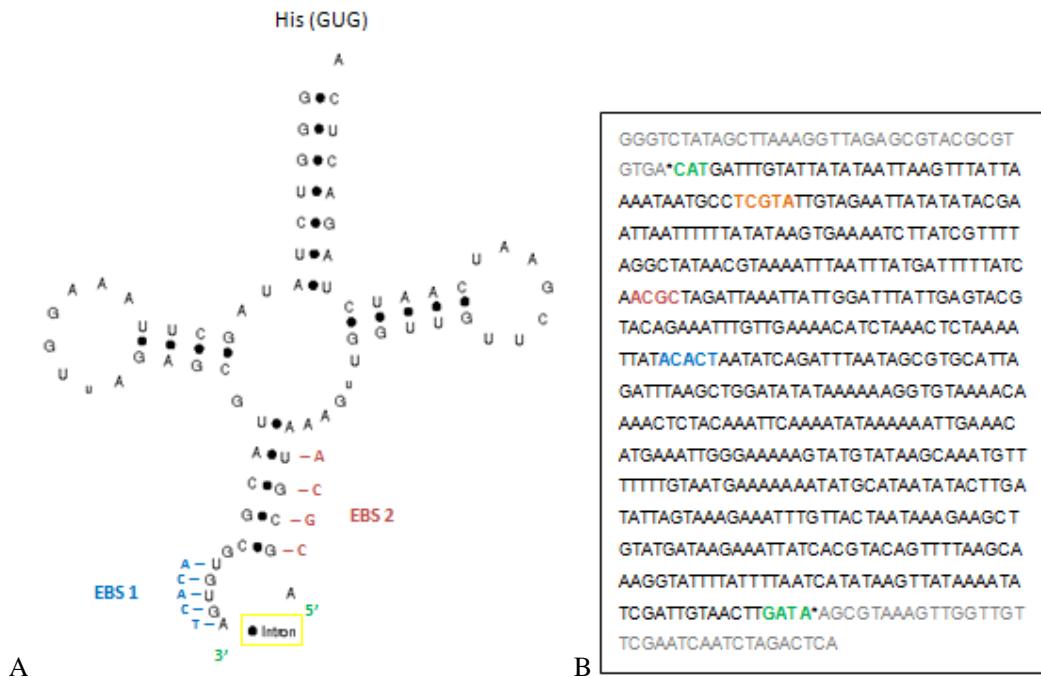
(Lambowitz & Zimmerly, 2004). Não foram encontradas orfs com qualquer semelhança a orfs de outros introns do grupo II tanto em *G. tenuistipitata* quanto em *C. crispus*.

Introns do grupo II possuem pouca similaridade de sequências, mas apresentam estrutura secundária conservada, necessária para sua atividade catalítica (Saldanha *et al.* 1993). Essa estrutura geralmente possui seis domínios (I-VI) em forma de hélices (Figura 10) que emanam de um círculo central. Os introns do grupo II possuem as extremidades 5' e 3' com sequências conservadas (GUGYG e YAY, respectivamente, sendo Y=C/T) e alguns domínios apresentam pequenas sequências conservadas essenciais para a interação intron-exon (Kelchner, 2002; Lambowitz & Zimmerly, 2004; Michel *et al.* 1989; Saldanha *et al.* 1993). As sequências EBS1, EBS2 e EBS3 (*exon binding sites* 1, 2 e 3) pareiam com as sequências IBS1, IBS2 e IBS3 (*intron binding sites* 1, 2 e 3), respectivamente, formando uma estrutura terciária. Os introns do grupo II apresentam ainda outras sequências conservadas, como no subdomínio c1 (CRA, sendo R=A/G), e o A não pareado (Figura 10, marcado com asterisco), que seria o sinal inicial para o *splicing* (Kelchner, 2002; Lambowitz & Zimmerly, 2004; Michel *et al.* 1989; Saldanha *et al.* 1993).



**Figura 10.** Estrutura secundária e interação terciária de um intron do grupo II. A numeração romana maiúscula (I-VI) representa os principais domínios dos introns do grupo II. As letras minúsculas (a-d) representam os subdomínios do domínio I. A numeração romana minúscula (i-iv) representa os segmentos de cada subdomínio. IBS1-3 interagem com EBS1-3, respectivamente, formando uma estrutura terciária. Outras interações terciárias conhecidas são representadas pelas letras gregas. O asterisco destaca o A não pareado, que seria o sinal inicial para o *splicing*. Retirado de Kelchner (2002).

Na sequência do intron de *G. tenuistipitata* foram encontradas algumas características básicas de um intron do grupo II, como as regiões EBS 1 e EBS 2, e a sequência c1 (Figura 11).



**Figura 11.** A. Estrutura secundária do tRNA para histidina em *Gracilaria tenuistipitata*. Em amarelo está destacado a localização do intron do grupo II. Em verde, estão representadas as extremidades 5' e 3' da sequência. Em azul, está destacado a região EBS1 (exon binding site 1), e em vermelho, a região EBS2 (exon binding site 2). Esses são os locais de ligação do exon. B. Sequência do tRNA para histidina com intron do grupo II em *Gracilaria tenuistipitata*. Em cinza está a sequência do his tRNA, separado do intron pelos asteriscos. Em verde estão as extremidades 5' (GATA) e 3' (CAT) da sequência. O EBS1 está em azul e o EBS2 em vermelho. Em laranja, está a sequência c1. A sequência é complementar.

*Gracilaria tenuistipitata*, assim como as demais algas vermelhas, não possui tRNAs para todos os aminoácidos (Tabela 8). O RNA transportador para treonina está ausente em todos os mtDNAs de Rhodophyta e as quatro algas codificam mais de um tRNA para um mesmo aminoácido, como por exemplo, glicina, leucina e metionina.

O código genético em *C. crispus* e *P. purpurea* é codificado de modo diferente do padrão. O códon TGA, tradicionalmente um *stop* códon, é traduzido como triptofano, e nem todas as orfs iniciam com o códon ATG (metionina), podendo começar com uma isoleucina, leucina ou valina (Leblanc *et al.* 1995; Burger *et al.* 1999). *Gracilaria tenuistipitata* também apresenta um código genético modificado, em que apenas TAA e TAG são considerados *stop* códons, e um gene (*ymf16*) se inicia com uma valina (GTT).

**Tabela 8.** Comparação de tRNAs presentes nos genomas mitocondriais de *Gracilaria tenuistipitata* (Gt), *Chondrus crispus* (Cc), *Porphyra purpurea* (Pp) e *Cyanidioschyzon merolae* (Cm).

<b>Aminoácido</b>		<b>Gt</b>	<b>Cc</b>	<b>Pp</b>	<b>Cm</b>
A	Ala Alanina	1	1	1	1
C	Cys Cisteína	1	1	1	1
D	Asp Ácido aspártico	1	1	1	1
E	Glu Ácido glutâmico	1	1	1	1
F	Phe Fenilalanina	1	1	1	1
G	Gly Glicina	2	2	2	2
H	His Histidina	2	2	1	1
I	Ile Isoleucina	0	1	1	1
K	Lys Lisina	1	1	1	1
L	Leu Leucina	2	2	2	3
M	Met Metionina	2	2	2	2
N	Asn Asparagina	1	1	1	1
P	Pro Prolina	1	1	1	1
Q	Gln Glutamina	1	1	1	1
R	Arg Arginina	2	1	2	2
S	Ser Serina	1	1	2	2
T	Thr Treonina	0	0	0	0
V	Val Valina	1	1	1	1
W	Trp Triptofano	1	1	1	1
Y	Tyr Tirosina	0	1	1	1
		total	22	23	24
					25

### 5.7. Comparação dos mtDNAs entre Rhodophyta

Os genomas mitocondriais das quatro espécies de Rhodophyta são semelhantes em relação ao tamanho, conteúdo e ordem gênica (Figura 6). *Cyanidioschyzon merolae* contém um maior número de genes, principalmente para proteínas ribossômicas (os genes *rpl5*, *rpl6*, *rpl14*, *rps4*, *rps8* e *rps14* não estão presentes nos outros genomas mitocondriais de Rhodophyta), além de ter quatro genes envolvidos na biogênese da citocromo tipo c, também não encontrados nas outras algas vermelhas (*yejR*, *yejU*, *yejV* e *yejW*). Já *P. purpurea* apresenta mais genes com função desconhecida (*orf132*, *orf176*, *orf238* e *orf284* são encontrados somente nesta alga) e uma região com dois genes descontínuos codificando para uma transcriptase reversa (*rtl*) e uma DNA polimerase (*dpo*), perfazendo um total de quase 4 Kb do genoma mitocondrial. O gene *rtl* aparenta ser um vestígio de uma orf de intron do grupo II, enquanto que o gene *dpo*

provavelmente foi introduzido por plasmídeo em algum ponto da história evolutiva da mitocôndria (Burger *et al.* 1999).

O mtDNA de *G. tenuistipitata* apresenta poucas regiões não-codificantes, cerca de 5,7% do total, proporção muito parecida com *C. crispus* (4,8% - Leblanc *et al.* 1995), *C. merolae* (4,5% - Ohta *et al.* 1998) e *P. purpurea* (9% - Burger *et al.* 1999), que apresenta o maior genoma mitocondrial entre as algas vermelhas (36 Kb).

Os genomas de *G. tenuistipitata* e de *C. crispus* podem ser divididos em dois blocos, conforme a orientação dos genes: *ymf16-cob* e *nad6-orf172*. *Porphyra purpurea* pode ser dividido em quatro: *ymf16*, *rps3-cob*, *rtl-orf176* e *orf132-rps12*, assim como *C. merolae*: *orf267(ymf16)-cox1*, *nad3-yejR*, *yejW-yejV* e *rpl20-rps3*, indicando a ocorrência de rearranjos e inversões durante a evolução de Rhodophyta (Figura 7).

Três agrupamentos de genes são conservados entre os quatro genomas mitocondriais de Rhodophyta (Figura 7). O maior é composto pelos genes *atp6-atp8-nad5-nad4-nad2-nad1-nad3*. Em *G. tenuistipitata*, *C. crispus* e *P. purpurea*, *sdh4* está inserido entre *nad4* e *nad2*. Em *C. merolae*, *sdh4* está entre *rps12* e *rns*. Nos quatro genomas mitocondriais, este agrupamento encontra-se na mesma direção (Figura 7, E). O segundo agrupamento apresenta os genes *rns-nad4L-rnl*. Mais uma vez *C. merolae* é a exceção, com o gene *yejU* inserido entre *nad4L* e *rnl*. Este agrupamento também se encontra na mesma direção nas quatro algas (Figura 7, A). E finalmente, o grupo *cox2-cox3-ymf39* é encontrado em todos os quatro mtDNAs. Em *G. tenuistipitata*, *C. crispus* e *P. purpurea*, o gene *cox1* se encontra antes do gene *cox2*. Em *C. merolae*, *cox1* se encontra à frente do agrupamento (Figura 7, C1 e C2). Em *P. purpurea*, o agrupamento está invertido em relação aos demais genomas mitocondriais das algas vermelhas.

Além desses três agrupamentos, nota-se ainda que os genes *rps3* e *rpl16* encontram-se juntos nos quatro mtDNAs (Figura 7, B), com direção oposta em *C. merolae*. Os genes *sdh3-sdh2-nad6* encontram-se agrupados em *G. tenuistipitata*, *C. crispus* e *C. merolae*, mas estão afastados um do outro no genoma mitocondrial de *P. purpurea* (Figura 7, D). O genoma mitocondrial de *P. purpurea* não possui o gene *rpl20*, encontrado nos mtDNAs das demais algas vermelhas.

Parte da região 5' do gene *cox1*, com cerca de 700pb, vem sendo muito utilizada como marcador molecular, técnica conhecida como DNA *barcode*. Este fragmento foi proposto por Herbert *et al.* (2003) como uma forma de identificar espécies animais.

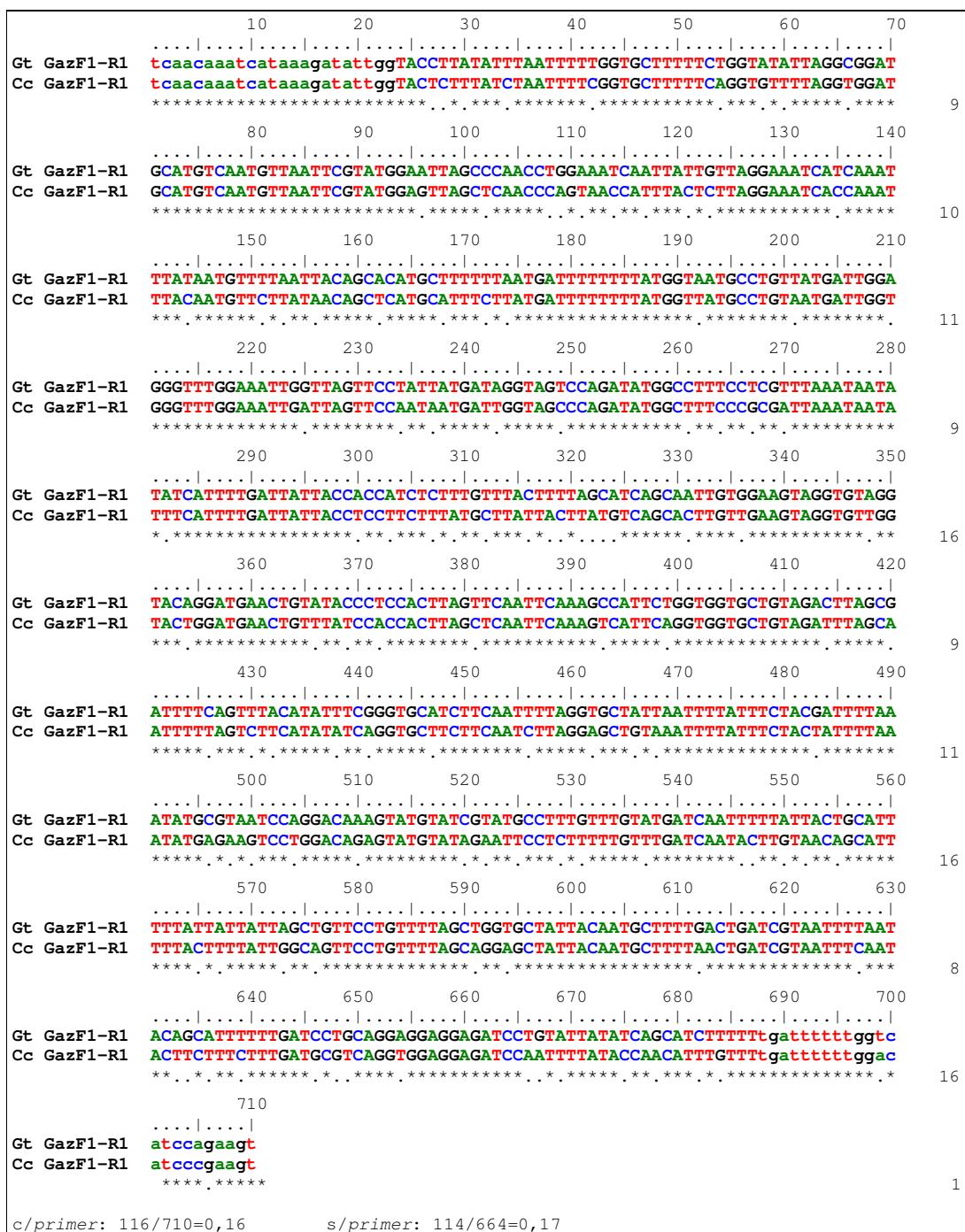
Saunders (2005) adaptou com sucesso os *primers* desenvolvidos por Herbert *et al.* (2003) para serem usados em algas vermelhas.

Para comparar o grau de divergência entre *G. tenuistipitata* e *C. crispus*, foi realizado um alinhamento da região 5' do gene *cox1*, entre os *primers* GazF1 e GazR1, utilizados para a identificação de espécies (Saunders, 2005). A Figura 12 apresenta as duas sequências alinhadas, num total de 710 nucleotídeos, incluindo as 46 bases dos *primers* (em letra minúscula). Na coluna da direita está o número de nucleotídeos diferentes entre as duas espécies, totalizando 114 bases num total de 664 (os *primers* foram excluídos), representando uma divergência de 17%.

A comparação do DNA *barcode* também foi realizada entre as demais algas vermelhas (Tabela 9). Conforme o esperado, os valores de divergência obtidos para o *barcode* refletem o grau de discrepância entre os genomas mitocondriais das quatro espécies de Rhodophyta e corroboram as análises filogenéticas baseadas em outros marcadores moleculares como o SSU rDNA (nuclear) e o *rbcL* (cloroplastidial) (Freshwater *et al.* 1994; Ragan *et al.* 1994). Segundo a árvore filogenética proposta por Yoon *et al.* (2006), *G. tenuistipitata* é mais próxima a *C. crispus* (pertencem à mesma classe) e mais distante de *C. merolae*, grupo mais basal dentre as Rhodophyta (Figura 3). *Cyanidioschyzon merolae* possui as menores porcentagens de similaridade, evidenciando ser realmente a espécie mais distante filogeneticamente.

Segundo Lane (2009), esta pequena sequência do gene *cox1* utilizada como DNA *barcode* pode conter informações bastante acuradas em relação ao genoma mitocondrial inteiro. Alguns pesquisadores defendem a hipótese de que sequências mitocondriais podem atuar no processo de especiação, uma vez que sofrem mudanças a cada geração mais rapidamente que sequências nucleares (geralmente de 10 a 30 vezes mais rápido).

A sobrevivência das células depende do processo respiratório, que ocorre nas mitocôndrias através de grandes complexos protéicos (Figura 9). Essas proteínas são produzidas tanto na mitocôndria quanto no núcleo e, portanto, eles precisam interagir perfeitamente para a respiração funcionar. Porém, se um genoma sofre mudanças mais rapidamente que outro, isso pode resultar em incompatibilidade protéica, levando a uma falha no desenvolvimento, redução do *fitness* e até a especiação, pois os organismos poderiam se tornar reprodutivamente incompatíveis (Lane, 2009).



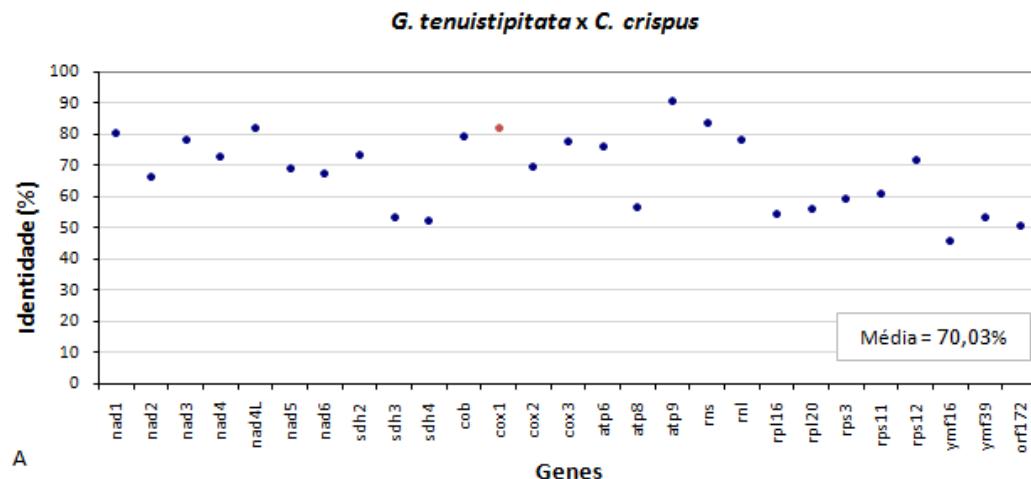
**Figura 12.** Alinhamento da região 5' do gene cox1 de *Gracilaria tenuistipitata* (Gt) e *Chondrus crispus* (Cc). Os asteriscos representam os nucleotídeos iguais nas duas sequências e na coluna da direita está o número de bases divergentes. Os primers estão em letra minúscula.

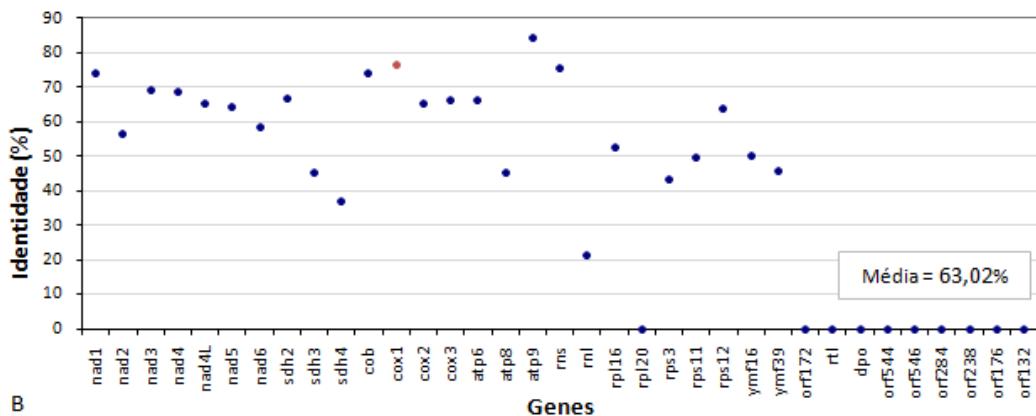
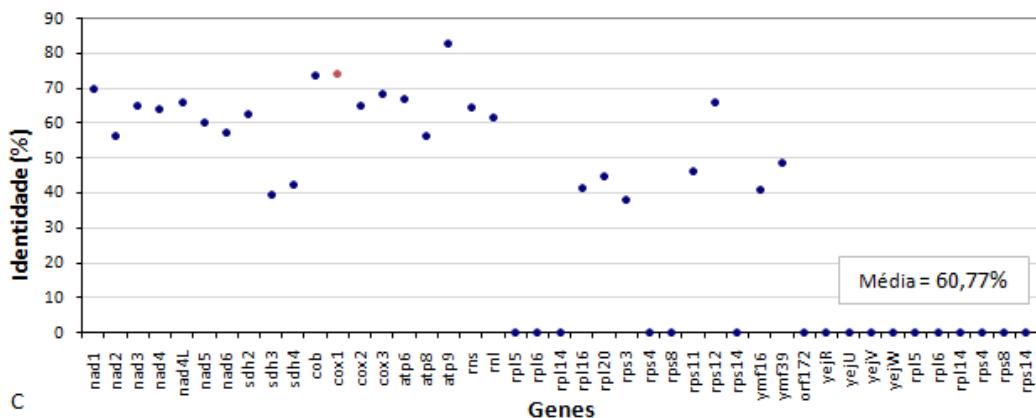
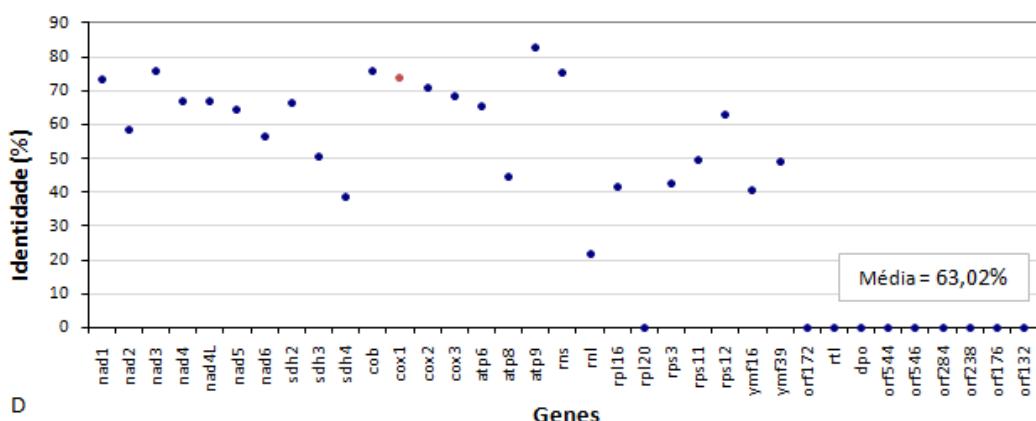
**Tabela 9.** Comparação entre as sequências da região 5' do gene *cox1* das quatro espécies de Rhodophyta que tiveram seus genomas mitocondriais completamente sequenciados. O triângulo abixo à esquerda indica a porcentagem de divergência e o triângulo acima à direita apresenta o número de nucleotídeos divergentes em um total de 664 bases na sequência.

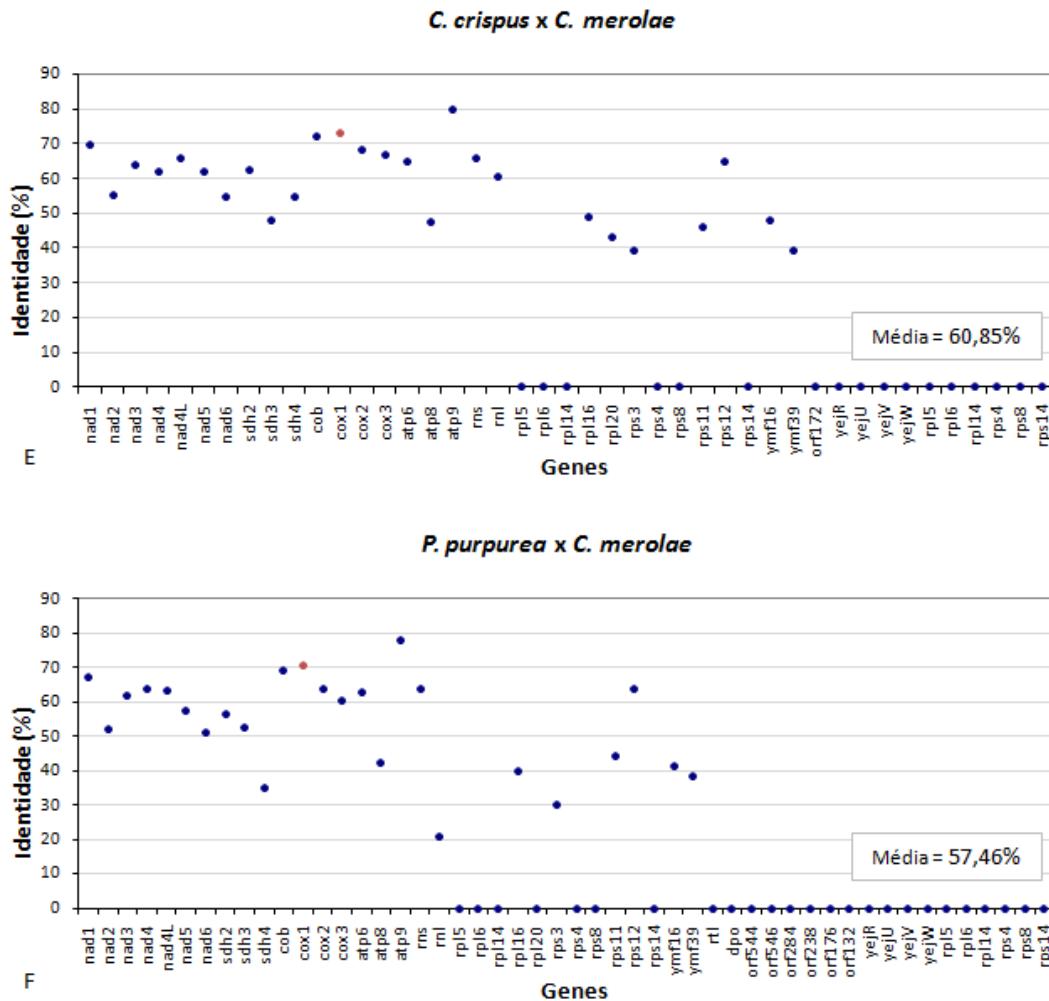
	Gt	Cc	Pp	Cm
<i>Gracilaria tenuistipitata</i> (Gt)	—	114	140	163
<i>Chondrus crispus</i> (Cc)	17,17	—	170	174
<i>Porphyra purpurea</i> (Pp)	21,09	25,61	—	171
<i>Cyanidioschyzon merolae</i> (Cm)	24,55	26,06	25,76	—

O genoma mitocondrial completo de *G. tenuistipitata* foi alinhado ao genoma mitocondrial completo de *C. crispus* (Anexo 3), apresentando similaridade de 71,72%, levando-se em conta tanto as regiões codificantes como as não codificantes. Esse alinhamento foi possível de ser realizado devido ao fato de que a ordem e a composição genômica é basicamente a mesma entre os dois organismos.

A Figura 13 apresenta os valores de identidade para cada gene que codifica proteínas, além dos genes *rns* e *rnl*. A média de identidade (nucleotídeos) foi calculada somente utilizando os genes para proteínas em comum para cada par de algas analisado, excluindo-se os genes *rns* e *rnl*. Os genes cujo valor de identidade é zero indicam que somente um dos organismos os possui. Em vermelho está destacado o gene *cox1*, um dos mais conservados nos genomas mitocondriais analisados.



*G. tenuistipitata x P. purpurea**G. tenuistipitata x C. merolae**C. crispus x P. purpurea*



**Figura 13.** Gráficos comparativos de identidade (duas a duas) entre os genes *rns*, *rnl* e codificadores de proteínas dos quatro genomas mitocondriais de algas vermelhas. Na abcissa estão representados os genes mitocondriais e na ordenada está a porcentagem de identidade. O gene *cox1* está destacado em vermelho. Os pontos na abcissa significam que o gene está presente em apenas um dos genomas mitocondriais. **A.** Comparação entre os genes mitocondriais de *Gracilaria tenuistipitata* e *Chondrus crispus*. Todos os genes estão presentes em ambas as algas, com média de identidade de 70,03%. **B.** Comparação entre os genes mitocondriais de *Gracilaria tenuistipitata* e *Porphyra purpurea*, com média de identidade de 63,02%. **C.** Comparação entre os genes mitocondriais de *Gracilaria tenuistipitata* e *Cyanidioschyzon merolae*, com média de identidade de 60,77%. **D.** Comparação entre os genes mitocondriais de *Chondrus crispus* e *Porphyra purpurea*, com média de identidade de 63,02%. **E.** Comparação entre os genes mitocondriais de *Chondrus crispus* e *Cyanidioschyzon merolae*, com média de identidade de 60,85%. **F.** Comparação entre os genes mitocondriais de *Porphyra purpurea* e *Cyanidioschyzon merolae*, com média de identidade de 57,46%.

O gene *atp9* é o mais conservado de todos, com grau de identidade acima de 75% para todos os genomas mitocondriais (entre *G. tenuistipitata* e *C. crispus* a semelhança é de 90,9%). Os genes *nad1* e *cob* também são bastante conservados em todas as algas vermelhas, além do gene *cox1*, utilizado como DNA *barcode*. O *cox1* tem sido usado para a delimitação de espécies, apresentando divergência de 0–0,3% entre espécies e de 4,5 a 13,6% entre espécies do mesmo gênero (Saunders, 2005). Por outro lado, os genes para proteínas ribossômicas, com exceção de *rps12*, são os menos conservados (Figura 13). Esse tipo de abordagem de comparação em maior escala de genes, no caso de todo o genoma mitocondrial, permite identificar outros marcadores moleculares em potencial, tanto mais conservados que o *cox1* quanto os mais variáveis, que podem ser eventualmente utilizados em outras abordagens ou em casos de uma resolução limítrofe do *cox1*.

A média de identidade dos genes para proteínas entre *G. tenuistipitata* e *C. crispus* é de 70,03% (Figura 13A), número muito próximo à porcentagem de similaridade entre os dois genomas inteiros (71,72%). Além dos genes ribossômicos, os genes *sdh3*, *sdh4*, *atp8*, *ymf16*, *ymf39* e *orf172* são os menos conservados.

Comparando as médias de identidade dos outros pares de Rhodophyta, pode-se constatar que *G. tenuistipitata* é mais próximo de *C. crispus* e mais distante de *C. merolae*, como havia sido demonstrado com a comparação da região 5' do gene *cox1* (Tabela 9 e Figura 13). *Porphyra purpurea* apresenta médias de identidade iguais para as duas algas da classe Florideophyceae (63,02%), e as médias de identidade de *C. merolae* com *G. tenuistipitata* e *C. crispus* também são próximas (60,77% e 60,85%, respectivamente), evidenciando que as ordens Gracilariales e Gigartinales são relacionadas filogeneticamente, como proposto pelo projeto Tree of Life (Figura 8).

A análise comparativa entre os genes mitocondriais codificadores de proteínas nas quatro algas vermelhas corrobora a ideia de que o gene *cox1* reflete o genoma mitocondrial inteiro pois, embora os valores de identidade das proteínas sejam menores que os do DNA *barcode*, a proporção continua a mesma.

## 5.8. Comparação dos mtDNAs entre organismos fotossintetizantes

Ao longo da evolução dos eucariotos, a mitocôndria teria se originado uma única vez através de um evento de endossimbiose de uma célula ancestral com uma  $\alpha$ -proteobactéria (Figura 1). A partir dessa interação, o genoma bacteriano do endossimbionte sofreu uma série de alterações, sendo extremamente reduzido com a perda e a migração de genes para o genoma nuclear. A comparação de genomas mitocondriais completos das diferentes linhagens eucarióticas permite uma melhor compreensão do processo de origem e evolução das mitocôndrias e dessas linhagens.

Foram selecionados 12 genomas mitocondriais de organismos fotossintetizantes, de quatro grandes grupos (Tabelas 10 e 11): Rhodophyta (algas vermelhas), Phaeophyta (algas pardas), Chlorophyta (algas verdes) e Streptophyta (plantas terrestres). Todos os genomas são circulares, com exceção de *Chlamydomonas reinhardtii*, que possui mtDNA linear (Gray & Boer, 1988).

Na Tabela 11 foram representados somente os genes que compõem a cadeia respiratória, a síntese de ATP e as proteínas ribossômicas.

As algas pardas formam um grupo extremamente homogêneo, e as três espécies selecionadas (*Dictyota dichotoma*, *Fucus vesiculosus* e *Laminaria digitata*), embora sejam de ordens diferentes (Dictyotales, Fucales e Laminariales, respectivamente), possuem exatamente os mesmos genes com funções descritas acima. Neste grupo, estão ausentes os genes do Complexo II (succinato desidrogenase) da cadeia de transporte de elétrons, porém dois genes estão presentes somente em Phaeophyta: *nad11* e *rpl31*. Nota-se ainda, o grande número de genes do Complexo I (NADH desidrogenase) e de proteínas ribossômicas, evidenciando genomas bastante conservados (Oudot-Le Secq *et al.* 2002 e 2006).

As algas verdes (*Chlamydomonas reinhardtii* – Chlamynomonadales, *Pedinomonas minor* – Pedinomonadales e *Prototheca wickerhamii* – Chlorellales), por outro lado, formam o grupo mais heterogêneo. *Chlamydomonas reinhardtii* apresenta o menor genoma mitocondrial de todos (~16 Kb), com apenas oito genes para proteínas (*cox1*, *cob*, *nad1*, *nad2*, *nad4*, *nad5*, *nad6* e *rtl*) e ausência total de genes do Complexo II da cadeia respiratória, ATP sintase e proteínas ribossômicas (Gray & Boer, 1988). *Pedinomonas minor* é intermediário em todos os quesitos, assemelhando-se mais a *C. reinhardtii* em relação ao conteúdo gênico e apresentando somente dois genes a mais,

*atp6* e *atp8*. *Prototheca wickerhamii* apresenta o maior genoma dentre as três clorófitas, com maior número de genes. Possui genes para proteínas ribossômicas, porém os genes do Complexo II da cadeia respiratória estão ausentes, como nas outras duas espécies.

As plantas terrestres foram representadas pelas espécies *Arabidopsis thaliana* e *Marchantia polymorpha*. Os genomas mitocondriais de plantas são variáveis em tamanho e complexidade (Oda *et al.* 1992). *Arabidopsis thaliana*, assim como as Chlorophyta analisadas, não apresenta os genes do Complexo II da cadeia de transporte de elétrons, e possui menor número de genes identificados quando comparado à hepática *M. polymorpha*. Embora o genoma mitocondrial de *A. thaliana* seja maior, os 57 genes codificantes representam apenas 10% do total, sendo o restante dividido entre introns, transposons, orfs com função desconhecida ou duplicações (Unseld *et al.* 1997).

As algas vermelhas foram o único grupo, com exceção de *M. polymorpha*, que apresentaram os genes do Complexo II (succinato desidrogenase) da cadeia de transporte de elétrons. Como demonstrado na Figura 13, esses genes são pouco conservados em Rhodophyta. Somente as plantas terrestres e *P. wickerhamii* possuem o gene *atp1*, e o gene *atp4* aparece apenas em *M. polymorpha*. O gene *nad7* está presente apenas nas algas pardas e nas plantas terrestres, que ainda apresentam o gene *nad9*, também compartilhado com *P. wickerhamii*.

Os únicos sete genes mitocondriais presentes em todos os organismos fotossintetizantes analisados foram *cox1*, *cob*, *nad1-2*, *nad4* e *nad5-6*, todos componentes da cadeia respiratória. Excluindo-se *C. reinhardtii* da análise, o único com genoma mitocondrial linear (Gray & Boer, 1988), há um aumento no número de genes para 13, com a inclusão de *atp6*, *atp9*, *cox2-3*, *nad3* e *nad4L*. A composição dos demais genes ribossômicos é bastante variável entre os organismos.

**Tabela 10.** Tabela comparativa entre genomas mitocondriais de diversos organismos fotossintetizantes. Rhodophyta: (Gt) *Gracilaria tenuistipitata*, (Cc) *Chondrus crispus*, (Cm) *Cyanidioschyzon merolae* e (Pp) *Porphyra purpurea*; Stramenopiles (Phaeophyceae): (Dd) *Dictyota dichotoma*, (Fv) *Fucus vesiculosus* e (Ld) *Laminaria digitata*; Chlorophyta: (Cr) *Chlamydomonas reinhardtii*, (Pm) *Pedinomonas minor* e (Pw) *Prototheca wickerhamii*; Streptophyta: (At) *Arabidopsis thaliana* e (Mp) *Marchantia polymorpha*.

Espécie	Gt	Cc	Cm	Pp	Dd	Fv
Tamanho (pb)	25.565	25.836	32.211	36.753	31.617	36.392
Conteúdo CG (%)	27	27	27	33	36	34
Total de genes	50	57	62	57	66	67
Genes para proteína	25	29	34	31	38	38
Genes para rRNA	3	4	3	2	3	3
Genes para tRNA	22	23	25	24	24	25
Número de acesso no GenBank	—	NC 001677	NC 000887	NC 002007	NC 007685	NC 007683
Referência	Este trabalho	Leblanc <i>et al.</i> 1995	Ohta <i>et al.</i> 1998	Burger <i>et al.</i> 1999	Oudot-Le Secq <i>et al.</i> 2006	Oudot-Le Secq <i>et al.</i> 2006

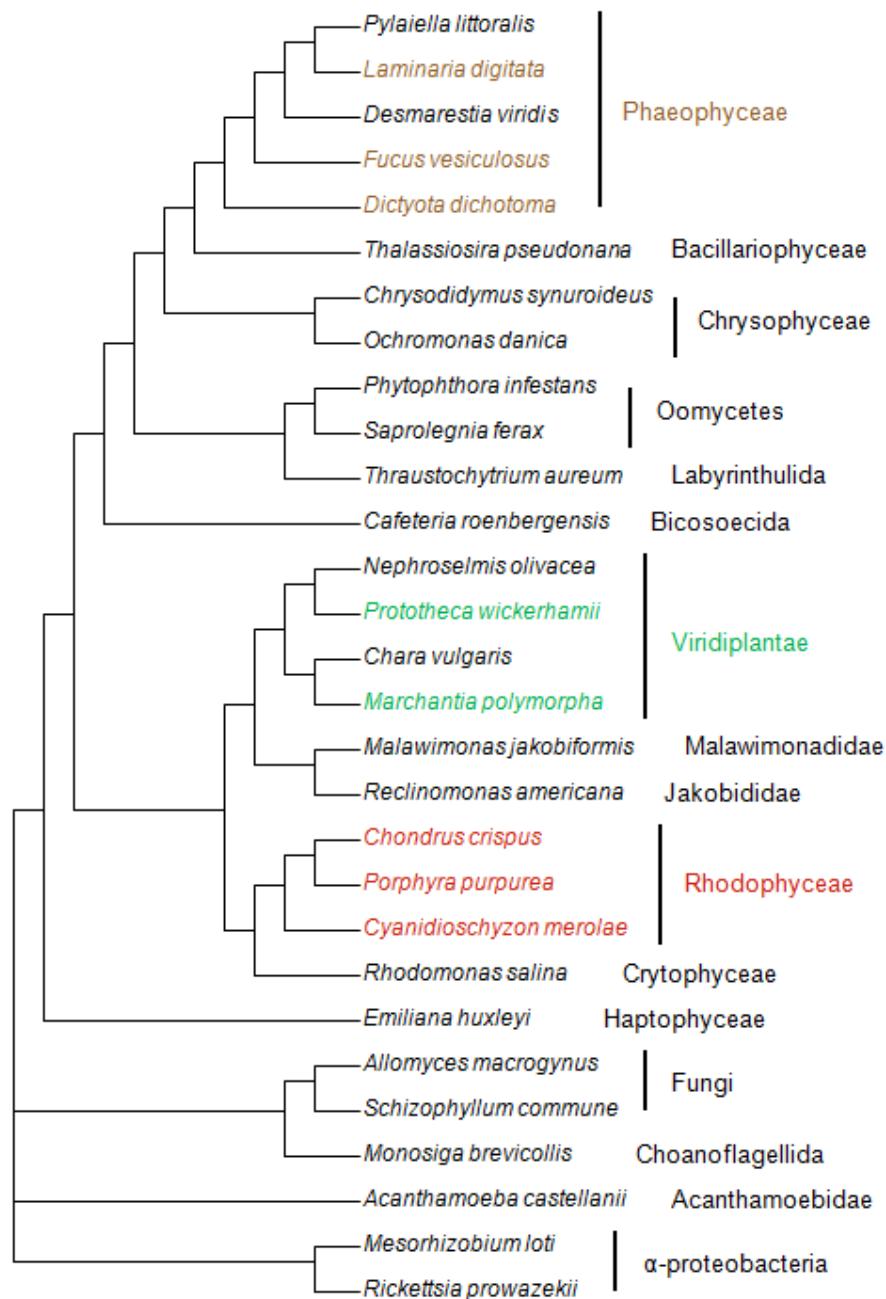
Espécie	Ld	Cr	Pm	Pw	At	MP
Tamanho (pb)	38.007	15.758	25.137	55.328	366.924	186.609
Conteúdo CG (%)	35	45	22	25	44	42
Total de genes	67	25	23	63	131	110
Genes para proteína	39	8	11	36	117	76
Genes para rRNA	3	14	3	3	3	3
Genes para tRNA	24	3	9	26	21	29
Número de acesso no GenBank	NC 004024	NC 001638	NC 000892	NC 001613	NC 001284	NC 001660
Referência	Oudot-Le Secq <i>et al.</i> 2002	Gray & Boer, 1988	Turmel <i>et al.</i> 1999	Wolff <i>et al.</i> 1994	Unseald <i>et al.</i> 1997	Oda <i>et al.</i> 1992

**Tabela 11.** Tabela comparativa entre os genes mitocondriais componentes da cadeia respiratória, síntese de ATP e codificadores de proteínas ribossômicas, presentes em diversos organismos fotossintetizantes. Rhodophyta: (Gt) *Gracilaria tenuistipitata*, (Cc) *Chondrus crispus*, (Cm) *Cyanidioschyzon merolae* e (Pp) *Porphyra purpurea*; Stramenopiles (Phaeophyceae): (Dd) *Dictyota dichotoma*, (Fv) *Fucus vesiculosus* e (Ld) *Laminaria digitata*; Chlorophyta: (Cr) *Chlamydomonas reinhardtii*, (Pm) *Pedinomonas minor* e (Pw) *Prototheca wickerhamii*; Streptophyta: (At) *Arabidopsis thaliana* e (Mp) *Marchantia polymorpha*.

Gene	G.t.	C.c.	C.m.	P.p.	D.d.	F.v.	L.d.	C.r.	P.m.	P.w.	A.t.	M.p.
Sucinato desidrogenase												
<i>sdh2</i>	+	+	+	+								
<i>sdh3</i>	+	+	+	+								+
<i>sdh4</i>	+	+	+	+								+
ATP sintase												
<i>atp1</i>										+	+	+
<i>atp4</i>												+
<i>atp6</i>	+	+	+	+	+	+	+	+	+	+	+	+
<i>atp8</i>	+	+	+	+	+	+	+	+	+	+	+	+
<i>atp9</i>	+	+	+	+	+	+	+	+	+	+	+	+
Citocromo c oxidase												
<i>cox1</i>	+	+	+	+	+	+	+	+	+	+	+	+
<i>cox2</i>	+	+	+	+	+	+	+	+	+	+	+	+
<i>cox3</i>	+	+	+	+	+	+	+	+	+	+	+	+
Apocitocromo b												
<i>cob</i>	+	+	+	+	+	+	+	+	+	+	+	+
NADH desidrogenase												
<i>nad1</i>	+	+	+	+	+	+	+	+	+	+	+	+
<i>nad2</i>	+	+	+	+	+	+	+	+	+	+	+	+
<i>nad3</i>	+	+	+	+	+	+	+	+	+	+	+	+
<i>nad4</i>	+	+	+	+	+	+	+	+	+	+	+	+
<i>nad4L</i>	+	+	+	+	+	+	+	+	+	+	+	+
<i>nad5</i>	+	+	+	+	+	+	+	+	+	+	+	+
<i>nad6</i>	+	+	+	+	+	+	+	+	+	+	+	+
<i>nad7</i>					+	+	+				+	+
<i>nad9</i>					+	+	+			+	+	+
<i>nad11</i>					+	+	+			+	+	+
Proteínas ribossômicas												
<i>rpl2</i>					+	+	+				+	+
<i>rpl5</i>					+	+	+				+	+
<i>rpl6</i>					+	+	+				+	+
<i>rpl14</i>					+	+	+					
<i>rpl16</i>	+	+	+	+	+	+	+			+	+	+
<i>rpl20</i>	+	+	+									
<i>rpl31</i>					+	+	+					
<i>rps1</i>												+
<i>rps2</i>					+	+	+			+		+
<i>rps3</i>	+	+	+	+	+	+	+			+	+	+
<i>rps4</i>					+	+	+			+	+	+
<i>rps7</i>					+	+	+			+	+	+
<i>rps8</i>					+	+	+					+
<i>rps10</i>						+	+	+				+
<i>rps11</i>	+	+	+	+	+	+	+			+		+
<i>rps12</i>	+	+	+	+	+	+	+			+	+	+
<i>rps13</i>						+	+	+		+		+
<i>rps14</i>						+	+	+	+	+		+
<i>rps19</i>						+	+	+	+	+		+

Os genes *rps3* e *rpl16* aparecem juntos nos genomas mitocondriais das quatro algas vermelhas (Figura 7), das três algas pardas (Oudot-Le Secq *et al.* 2002 e 2006), de *M. polymorpha* (Oda *et al.* 1992) e de *A. thaliana* (Unseld *et al.* 1997). *Pedinomonas minor* e *C. reinhardtii* não possuem esses genes, e em *P. wickerhamii* eles estão distantes entre si (Wolff *et al.* 1994). Outro agrupamento conservado é composto pelos genes *nad4* e *nad5*, encontrado de novo nas quatro algas vermelhas (Figura 7), nas três algas pardas (Oudot-Le Secq *et al.* 2002 e 2006), em *M. polymorpha* (Oda *et al.* 1992), em *C. reinhardtii* (Gray & Boer, 1988) e em *P. wickerhamii* (Wolff *et al.* 1994), e distantes nos demais organismos fotossintetizantes analisados. O gene *ymf16* também pode ser encontrado em outros organismos além das algas vermelhas (Figura 7). Nas algas pardas recebe o nome de *tatC* (Oudot-Le Secq *et al.* 2002 e 2006), em *M. polymorpha* é *orf244* (Oda *et al.* 1992) e em *P. wickerhamii*, *orf234.1* (Wolff *et al.* 1994).

Oudot-Le Secq *et al.* (2006) compararam sequências de dez genes mitocondriais (*atp6*, *atp9*, *cob*, *cox1-3*, *nad1*, *nad3-4* e *nad4L*) e construiram uma árvore filogenética com 29 espécies de diversos grupos (Figura 14). Os genes mitocondriais utilizados nessa análise estão entre os mais conservados, como mostra a Figura 13. As  $\alpha$ -proteobactérias foram usadas como grupo externo na análise. Os organismos analisados neste trabalho estão destacados em cores na Figura 14. Observa-se que algas verdes e plantas terrestres formam um grupo monofilético, e estão associadas à Rhodophyta. Porém, plantas verdes (embriófitas e clorófitas) e algas vermelhas estão associados às classes Cryptophyceae, Malawimonadidae e Jakobididae, não formando o grupo monofilético apresentado como Plantae em trabalhos anteriores (Burger *et al.* 1999; Baldauf *et al.* 2003; Keeling, 2007). Entretanto, Oudot-Le Secq *et al.* (2006) utilizaram apenas marcadores mitocondriais na análise, o que pode levar a conclusões equivocadas, uma vez que os genomas nuclear, cloroplastidial e mitocondrial evoluem em ritmos diferentes e possuem diferentes histórias evolutivas.



**Figura 14.** Árvore filogenética construída por análise Bayesiana a partir de sequências mitocondriais de dez genes (*atp6*, *atp9*, *cob*, *cox1-3*, *nad1*, *nad3-4* e *nad4L*). Em cores estão destacados os organismos fotossintetizantes comparados neste trabalho. Modificado de Oudot-Le Secq *et al.* 2006.

## 6. CONCLUSÃO

Genomas mitocondriais apresentam uma variedade de tamanho e organização, que apresentam principalmente duas tendências. Mitocôndrias de animais, ciliados e fungos possuem um genoma relativamente pequeno e compacto, que geralmente utiliza um código genético modificado. Já os mtDNAs de plantas terrestres são grandes, com longas regiões intergênicas, e utilizam o código genético universal (Leblanc *et al.* 1995). O genoma mitocondrial de *Gracilaria tenuistipitata* possui características de ambos, pois é extremamente compacto, com poucas regiões intergênicas (Figura 14) e utiliza um código genético modificado. Porém, assim como em *Chondrus crispus*, essa alga exibe SSU rRNA e genes para proteínas estruturalmente e filogeneticamente muito semelhantes aos de plantas terrestres (Leblanc *et al.* 1995).

O mtDNA de *G. tenuistipitata*, embora bastante condensado, apresenta poucas regiões de sobreposição de genes, esta ocorrendo apenas entre os genes tRNA<sup>Lys</sup> e *ymf16*, *rps12* e tRNA<sup>Glu</sup> e *rpl20* e *rns* (Figura 14). Apenas um intron do grupo II foi encontrado, localizado no gene tRNA<sup>His</sup>, sem ocorrência de orf. O código genético de *G. tenuistipitata*, assim como em *C. crispus* e *P. purpurea*, é modificado. O códon TGA codifica um tRNA para triptofano, e não um *stop* códon como é tradicionalmente, e um gene (*ymf16*) se inicia com o aminoácido valina, e não metionina como nos demais genes para proteínas.

A ordem e a composição do genoma mitocondrial de *G. tenuistipitata* é praticamente igual ao de *C. crispus*, com índice de identidade de nucleotídeos superior a 70% em todas as análises realizadas. Essa semelhança pode ser explicada pelo fato de ambas as algas pertencerem à mesma classe, Florideophyceae, e ainda se encontrarem no mesmo clado dentro da classe.

Este trabalho corrobora estudos filogenéticos de outros autores em relação às algas vermelhas, evidenciando que *G. tenuistipitata* e *C. crispus* são mais próximos entre si e igualmente distantes de *P. purpurea*, e que *C. merolae* é mais o divergente. O acúmulo de dados moleculares sobre genomas mitocondriais ajuda a esclarecer a origem e evolução das organelas, e contribui para o estudo das relações filogenéticas entre os diversos organismos.

## 7. REFERÊNCIAS

- Adams, K.L. & Palmer, J.D. 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Molecular Phylogenetics and Evolution* 29: 380-395.
- Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. & Watson, J.D. 1997. *Biologia Molecular da Célula*. 3ed. Porto Alegre: Artes Médicas. 1294p.
- Aldrich, J. & Cattolico, R.A. 1981. Isolation and characterization of chloroplast DNA from the marine chromophyte *Olisthodiscus luteus*: electron microscopic visualization of isomeric molecular forms. *Plant Physiology* 68: 641-647.
- Allen, J.F. 2003. The function of genomes in bioenergetic organelles. *Philosophical Transactions of the Royal Society B* 358: 19-38.
- Altschul, S.F., Madden T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research* 25: 3389-3402.
- Andersen, R.A. & Kawachi, M. 2005. Traditional microalgae isolation techniques. In: Algas culturing techniques. Andersen, R.A. (Ed.). Elsevier Academic Press, London, UK: 83-100.
- Andersson, S.G.E., Zomorodipour, A., Andersson, J.O., Sicheritz-Pontén, T., Alsmark, U.C.M., Podowski, R.M., Näslund, A.K., Eriksson, A.S., Winkler, H.H. & Kurland, C.G. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* 396: 133-143.
- Andersson, S.G.E., Karlberg, O., Canbäck, B. & Kurland, C.G. 2003. On the origin of mitochondria: a genomics perspective. *Philosophical Transactions of the Royal Society B* 358: 165-179.
- Armisen, R. 1995. World-wide use and importance of *Gracilaria*. *Journal of Applied Phycology* 7: 231-243.
- Baldauf, L.S. 2003. The deep roots of eukaryotes. *Science* 300: 1703-1706.
- Barau, J.G., Azeredo-Espin, A.M.L. & Lessinger, A.C. 2005. Conservation and versatility of a new set of primers for long-PCR amplification of complete insect mitochondrial genomes based on *Haematobia irritans* mtDNA sequences. *Molecular Ecology Notes* 5: 885-887.

- Barufi, J.B., Oliveira, E.C., Plastino, E.M. & Oliveira, M.C. 2009. Life history, morphological variability and growth rates of life phases of *Gracilaria tenuistipitata* (Rhodophyta: Gracilariales) *in vitro*. *Scientia Marina* 74: 297-303.
- Bullerwell, C.E. & Gray, M.W. 2004. Evolution of the mitochondrial genome: protist connections to animals, fungi and plants. *Current Opinion in Microbiology* 7: 528-534.
- Burger, G., Saint-Louis, D., Gray, M.W. & Lang, B.F. 1999. Complete sequence of the mitochondrial DNA of the red alga *Porphyra purpurea*. Cyanobacterial introns and shared ancestry of red and green algae. *Plant Cell* 11: 1675-1694.
- Chesnick, J.M. & Catollico, R.A. 1993. Isolation of DNA from eukaryotic algae. *Methods in Enzymology* 224: 168-176.
- Clark, C.G. 1999. The effect of secondary loss on our views of eukaryotic evolution. *Biological Bulletin* 196: 385-388.
- Critchley, A.T. 1993. *Gracilaria* (Rhodophyta, Gracilariales): an economically important agarophyte. In: Seaweed cultivation and marine ranching. Ohno, M., Critchley, A.T. (Eds.). JICA, Yolosuka, Japão: 98-112.
- Critchley, A.T. & Ohno, M. 1998. Seaweed Resources of the World. JICA, Yokosuka, Japão. 431 p.
- Freshwater, D.W., Fredericq, S., Butler, B.S., Hommersand, M.H. & Chase, M.W. 1994. A gene phylogeny of the red algae (Rhodophyta) based on plastid *rbcL*. *Proceedings of the National Academy of Sciences of the USA* 91: 7281-7285
- Gillham, N.W. 1994. *Organelle Genes and Genomes*. New York: Oxford University Press. 424p.
- Gordon, D., Abajian, C. & Green, P. 1998. Consed: A graphical tool for sequence finishing. *Genome Research* 8: 195-202.
- Gray, M.W. 1999. Evolution of organellar genomes. *Current Opinion in Genetics & Development* 9: 678-687.
- Gray, M.W. & Boer, P.H. 1988. Organization and expression of algal (*Chlamydomonas reinhardtii*) mitochondrial DNA. *Philosophical Transactions of the Royal Society B* 319: 135-147.
- Gray, M.W., Burger, G. & Lang, B.F. 1999. Mitochondrial evolution. *Science* 283: 1476-1481.

- Gurgel, C.F.D. & Fredericq, S. 2004. Systematics of the Gracilariaeae (Gracilariales, Rhodophyta): a critical assessment based on rbcL sequence analysis. *Journal of Phycology* 40: 138-159.
- Haas, S., Vingron, M., Poustka, A. & Wiemann, S. 1998. Primer design for large scale sequencing. *Nucleic Acids Research* 26 (12): 3006-3012.
- Hagopian, J.C., Nyvall, P. & Oliveira, M.C. 2002. Purification of plastid DNA from an enriched rhodoplast fraction of the red alga *Gracilaria tenuistipitata*. *Plant Molecular Biology Reporter* 20: 399-406.
- Hagopian, J.C., Reis, M., Kitajima, J.P., Bhattacharya D. & Oliveira, M.C. 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. *Journal of Molecular Evolution* 59: 464-477.
- Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Ser.* 41: 95-98.
- Hebert, P.D.N., Cywinska, A., Ball, S.L. & deWaard, J.R., 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B* 270: 313-322.
- Keeling, P.J. 2007. Deep questions in the Tree of Life. *Science* 317: 1875-1876.
- Kelchner, S.A. 2002. Group II introns as phylogenetic tools: structure, function, and evolutionary constraints. *American Journal of Botany* 89: 1651-1669.
- Lambowitz, A.M. & Zimmerly, S. 2004. Mobile group II introns. *Annual Reviews of Genetic* 38: 1-35.
- Lane, N. 2009. On the origin of bar codes. *Nature* 462: 272-274.
- Leblanc, C., Boyen, C., Richard, O., Bonnard, G., Grienenberger, J. M. & Kloareg, B. 1995. Complete sequence of the mitochondrial DNA of the rhodophyte *Chondrus crispus* (Gigartinales). Gene content and genome organization. *Journal of Molecular Biology* 250: 484-495.
- Leblanc, C., Richard, O., Kloareg, B., Viehmann, S., Zetsche, K. & Boyen, C. 1997. Origin and evolution of mitochondria: what have we learnt from red algae? *Current Genetics* 31: 193-207.
- Linhart, C. & Shamir, R. 2005. The degenerate primer design problem: theory and applications. *Journal of Computational Biology* 12 (4): 431-456.

- Lowe, T.M. & Eddy, S.R. 1997. tRANscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* 25: 955-964.
- Macchiavello, J., De Paula, E. & Oliveira, E.C. 1998. Growth rate responses of five commercial strains of *Gracilaria* (Rhodophyta, Gracilariales) to temperature and light. *Journal of the World Aquaculture Society* 29: 259-266.
- Macchiavello, J., Saito, R., Garofalo, G. & Oliveira, E.C. 1999. A comparative analysis of agarans from commercial species of *Gracilaria* (Gracilariales, Rhodophyta) grown *in vitro*. *Hydrobiologia* 398/399: 397-400.
- Machida, R.J., Miya, M.U., Nishida, M. & Nishida, S. 2002. Complete mitochondrial DNA sequence of *Tigriopus japonicus* (Crustacea: Copepoda). *Marine Biotechnology* 4: 406-417.
- McHugh, D.J. 2003. A guide to the Seaweed Industry. *FAO Fisheries Technical Paper* 441. Rome: FAO. 105p.
- Michel, F., Umesono, K. & Ozeki, H. 1989. Comparative and functional anatomy of group II catalytic introns – a review. *Gene* 82: 5-30.
- Milstein, D., Oliveira, M.C., Martins, F.M. & Matioli, S.R. 2008. Group I introns and associated homing endonuclease genes reveals a clinal structure for *Porphyra spiralis* var. *amplifolia* (Bangiales, Rhodophyta) along the Eastern coast of South America. *BMC Evolutionary Biology* 8: 308 (doi:10.1186/1471-2148-8-308).
- Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, N., Kanegae, T., Ogura, Y., Kohchi, T. & Ohyama, K. 1992. Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA: a primitive form of plant mitochondrial genome. *Journal of Molecular Biology* 223: 1-7.
- Ohta, N., Sato, N. & Kuroiwa, T. 1998. Structure and organization of the mitochondrial genome of the unicellular red alga *Cyanidioschyzon merolae* deduced from the complete nucleotide sequence. *Nucleic Acids Research* 26: 5190-5298.
- Oliveira, E.C. & Plastino, E.M. 1994. Gracilariaeae. In: Akatsuka (ed), *Biology of Economic Algae*, pp 185-226.
- Oliveira, E.C., Paula, E.J., Plastino, E.M. & Petti, R. 1995. Metodologias para el cultivo no axénico de macroalgas marinas *in vitro*. In: Alveal K., Ferrario, M.E., Oliveira,

- E.C. and Sar, E. (eds), *Manual de Métodos Ficológicos*, pp 429-447. Universidad de Concepción, Concepción, Chile.
- Oliveira, M.C. & Bhattacharya, D. 2000. Phylogeny of the Bangiophycidae (Rhodophyta) and the secondary endosymbiotic origin of algal plastids. *American Journal of Botany* 87: 482-492.
- Oudot-Le Secq, M.P., Kloareg, B. & Loiseaux-De Goér, S. 2002. The mitochondrial genome of the brown alga *Laminaria digitata*: a comparative analysis. *European Journal of Phycology* 37: 163-172.
- Oudot-Le Secq M.P., Loiseaux-De Goér, S., Stam, W.T. & Olsen, J.L. 2006. Complete mitochondrial genomes of the three brown algae (Heterokonta: Phaeophyceae) *Dictyota dichotoma*, *Fucus vesiculosus* and *Desmarestia viridis*. *Current Genetics* 49: 47-58.
- Ragan, M.A., Bird, C.J., Rice, E.L., Gutell, R.R., Murphy, C.E. & Singh, R.K. 1994. A molecular phylogeny of the marine red algae (Rhodophyta) based on the nuclear small-subunit rRNA gene. *Proceedings of the National Academy of Sciences of the USA* 91: 7276-7280.
- Rice, E.L. & Bird, C.J. 1990. Relationship among geographically distant populations of *Gracilaria verrucosa* (Gracilariales, Rhodophyta) and related species. *Phycologia* 29: 501-510.
- Sambrook, J., Fritsch, E.F. & Maniatis, T. 1989. *Molecular Cloning: A Laboratory Manual*. 2ed. 3 vol. New York: Cold Spring Harbour Laboratories Press.
- Saldanha, R., Mohr, G., Belfort, M. & Lambowitz, A.M. 1993. Group I and group II introns. *The Faseb Journal* 7: 15-24.
- Saunders, G.W. 2005. Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. *Philosophical Transactions of the Royal Society B* 360: 1879-1888.
- Saunders, G.W. & Hommersand, M.H. 2004. Assessing red algal supraordinal diversity and taxonomy in the context of contemporary systematic data. *American Journal of Botany* 91: 1494-1507.
- Schuster, W. & Brennicke, A. 1994. The plant mitochondrial genome: physical structure, information content, RNA editing, and gene migration to the nucleus. *Annual Review of Plant Physiology* 45: 61-78.

- Suzuki, K., Ohta, N. & Kuroiwa, T. 1992. Isolation of the cell-nuclear, mitochondrial, and chloroplast DNA from the ultra-small eukaryote *Cyanidioschyzon merolae*. *Protoplasma* 171: 80-84.
- Taiz, L. & Zeiger, E. 2004. *Fisiologia Vegetal*. 3ed. Porto Alegre: Artmed. 719p.
- Timmis, J.N., Ayliffe, M.A., Huang, C.Y. & Martin, W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Reviews Genetics* 5: 123-135.
- Triboush, S.O., Danilenko, N.G. & Davydenko, O.G. 1998. A method for isolation of chloroplast DNA and mitochondrial DNA from sunflower. *Plant Molecular Biology Reporter* 16: 183-189.
- Turmel, M., Lemieux, C., Burger, G., Lang, B.F., Otis, C., Plante, I. & Gray, M.W. 1999. The complete mitochondrial DNA sequences of *Nephroselmis olivacea* and *Pedinomonas minor*: two radically different evolutionary patterns within green algae. *The Plant Cell* 11: 1717-1729.
- Unseald, M., Marienfeld, J.R., Brandt, P. & Brennicke, A. 1997. The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nature Genetics* 15: 57-61.
- Wolff, G., Plante, I., Lang, B.F., Kück, U. & Burger, G. 1994. Complete sequence of the mitochondrial DNA of the chlorophyte alga *Prototheca wickerhamii*: gene content and genome organization. *Journal of Molecular Biology* 237: 75-86.
- Yoon, H.S., Müller, K.M., Sheath, R.G., Ott, F.D. & Bhattacharya, D. 2006. Defining the major lineages of red algae (Rhodophyta). *Journal of Phycology* 42: 482-492.
- Zhemke-White, W.L. & Ohno, M. 1999. World seaweed utilization: an end-of-century summary. *Journal of Applied Phycology* 11: 369-376.
- Zuccarello, G.C., Burger, G., West, J.A. & King, R.J. 1999. A mitochondrial marker for red algal intraspecific relationships. *Molecular Ecology* 8: 1443-1447.
- Zuccarello, G.C., Sandercock, B. & West, J.A. 2002. Diversity within red algal species: variation in world-wide samples of *Spyridia filamentosa* (Ceramiaceae) and *Murrayella periclados* (Rhodomelaceae) using DNA markers and breeding studies. *European Journal of Phycology* 37: 403-417.

**8. ANEXO 1.** Lista de todos os genes presentes no genoma mitocondrial de *Chondrus crispus* (fonte: GenBank). Nota-se que alguns genes sem função definida sobrepõem-se a genes conhecidos.

Gene	Descrição	Localização	Sentido
Asn tRNA	RNA transportador para asparagina	1-72	Direto
Val tRNA	RNA transportador para valina	82-154	Direto
Arg tRNA	RNA transportador para arginina	177-251	Direto
Lys tRNA	RNA transportador para lisina	295-368	Direto
<i>ymf16</i>	Proteína transportadora independente sec-Y	348-1136	Direto
<i>rps12</i>	Proteína ribossômica S12	1136-1519	Direto
Glu tRNA	RNA transportador para ácido glutâmico	1505-1576	Direto
Met tRNA	RNA transportador para metionina	1584-1657	Direto
<i>rpl20</i>	Proteína ribossômica L20	1670-1954	Direto
<i>rns</i>	RNA ribossômico subunidade pequena	1949-3324	Direto
<i>orf73</i>	Proteína não identificada	3337-3558	Reverso
<i>nad4L</i>	NADH desidrogenase subunidade 4L	3362-3667	Direto
<i>rnl</i>	RNA ribossômico subunidade grande	3689-6271	Direto
<i>rps3</i>	Proteína ribossômica S3	6303-7025	Direto
<i>rpl16</i>	Proteína ribossômica L16	7000-7416	Direto
Asp tRNA	RNA transportador para ácido aspártico	7429-7500	Direto
<i>cox1</i>	Citocromo c oxidase subunidade 1	7542-9140	Direto
<i>cox2</i>	Citocromo c oxidase subunidade 2	9143-9907	Direto
<i>rrn5</i>	RNA ribossômico 5S	9866-9988	Direto
<i>cox3</i>	Citocromo c oxidase subunidade 3	10077-10895	Direto
<i>ymf39</i>	Proteína não identificada	10908-11459	Direto
Gly tRNA	RNA transportador para glicina	11465-11539	Direto
Gln tRNA	RNA transportador para glutamina	11541-11612	Direto
Leu tRNA	RNA transportador para leucina	11625-11708	Direto
<i>cob</i>	Apocitocromo b	11751-12896	Direto
Leu tRNA	RNA transportador para leucina	12952-13032	Reverso
<i>nad6</i>	NADH desidrogenase subunidade 6	13033-13647	Reverso
Gly tRNA	RNA transportador para glicina	13688-13740	Reverso
His tRNA	RNA transportador para histidina	13746-13819	Direto
<i>sdh2</i>	Citocromo c oxidorredutase subunidade 2	13821-14573	Reverso
<i>sdh3</i>	Citocromo c oxidorredutase subunidade 3	14575-14958	Reverso
Phe tRNA	RNA transportador para fenilalanina	14973-15044	Reverso
Ser tRNA	RNA transportador para serina	15060-15143	Reverso
Pro tRNA	RNA transportador para prolina	15151-15224	Reverso
<i>apt9</i>	ATP sintase F0 subunidade 9	15236-15466	Reverso
ChcroMp46	Proteína hipotética ChcroMp46	15345-15569	Reverso
Cys tRNA	RNA transportador para cisteína	15518-15589	Reverso
Met tRNA	RNA transportador para metionina	15598-15671	Reverso
<i>rps11</i>	Proteína ribossômica S11	15674-16024	Reverso
<i>rrn5</i>	RNA ribossômico 5S	16041-16154	Reverso
<i>nad3</i>	NADH desidrogenase subunidade 3	16166-16531	Reverso

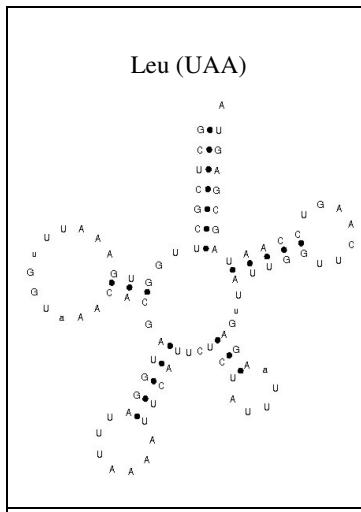
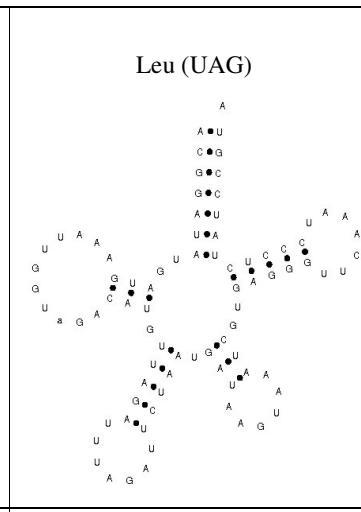
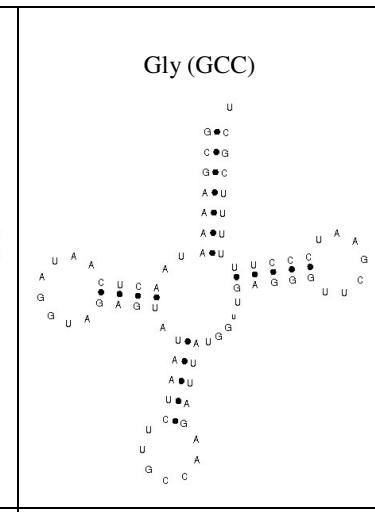
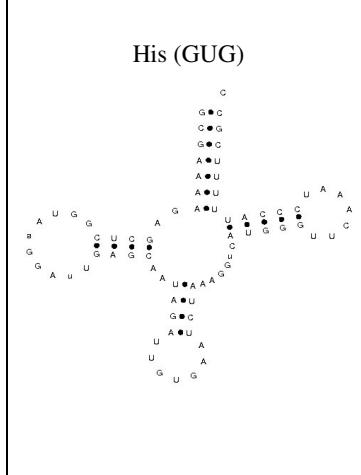
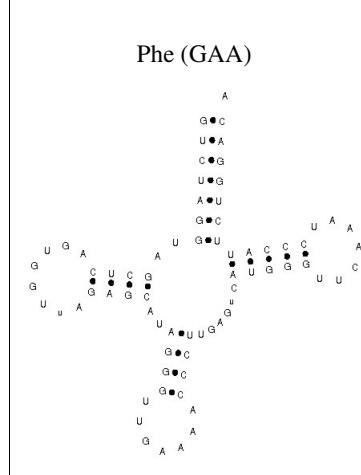
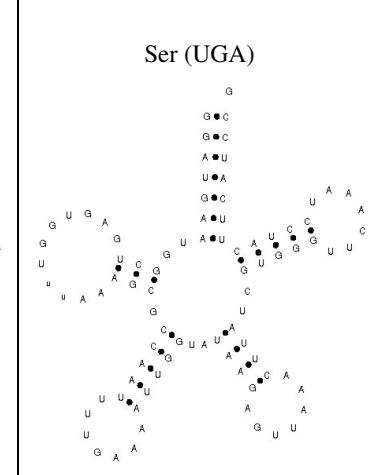
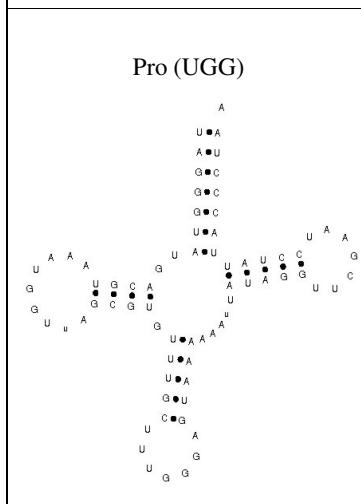
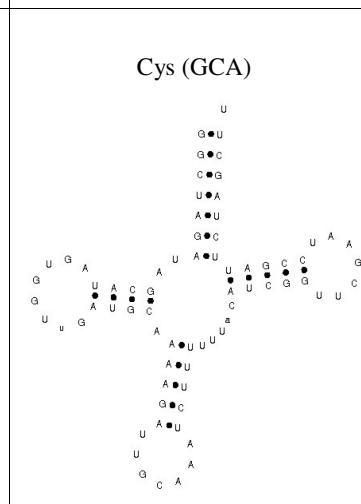
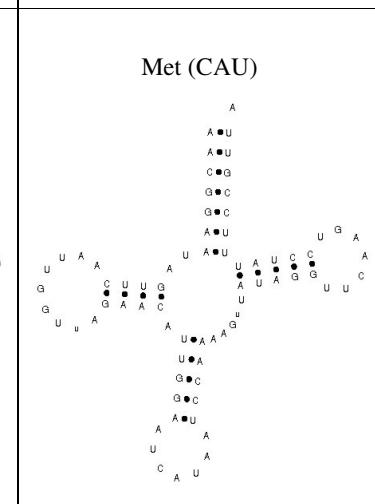
## 8. ANEXO 1. Continuação

<b>Gene</b>	<b>Descrição</b>	<b>Localização</b>	<b>Sentido</b>
<i>nad1</i>	NADH desidrogenase subunidade 1	16544-17524	Reverso
<i>nad2</i>	NADH desidrogenase subunidade 2	17543-19036	Reverso
ChcroMp51	Proteína hipotética ChcroMp51	18024-18308	Direto
<i>sdh4</i>	Citocromo c oxidoreductase subunidade 4	19038-19292	Reverso
<i>nad4</i>	NADH desidrogenase subunidade 4	19299-20777	Reverso
Ile tRNA	RNA transportador para isoleucina	20808-20849	Reverso
Ile tRNA	RNA transportador para isoleucina	21324-21354	Reverso
<i>nad5</i>	NADH desidrogenase subunidade 5	21370-23370	Reverso
<i>atp8</i>	ATP sintase F0 subunidade 8	23385-23798	Reverso
ChcroMp52	Proteína hipotética ChcroMp52	23509-23826	Direto
<i>apt6</i>	ATP sintase F0 subunidade 6	23812-24573	Reverso
Trp tRNA	RNA transportador para triptofano	24603-24674	Reverso
orf172	Proteína não identificada	24688-25206	Reverso
Ala tRNA	RNA transportador para alanina	25221-25294	Reverso
Arg tRNA	RNA transportador para arginina	25372-25447	Reverso
Tyr tRNA	RNA transportador para tirosina	25465-25554	Reverso

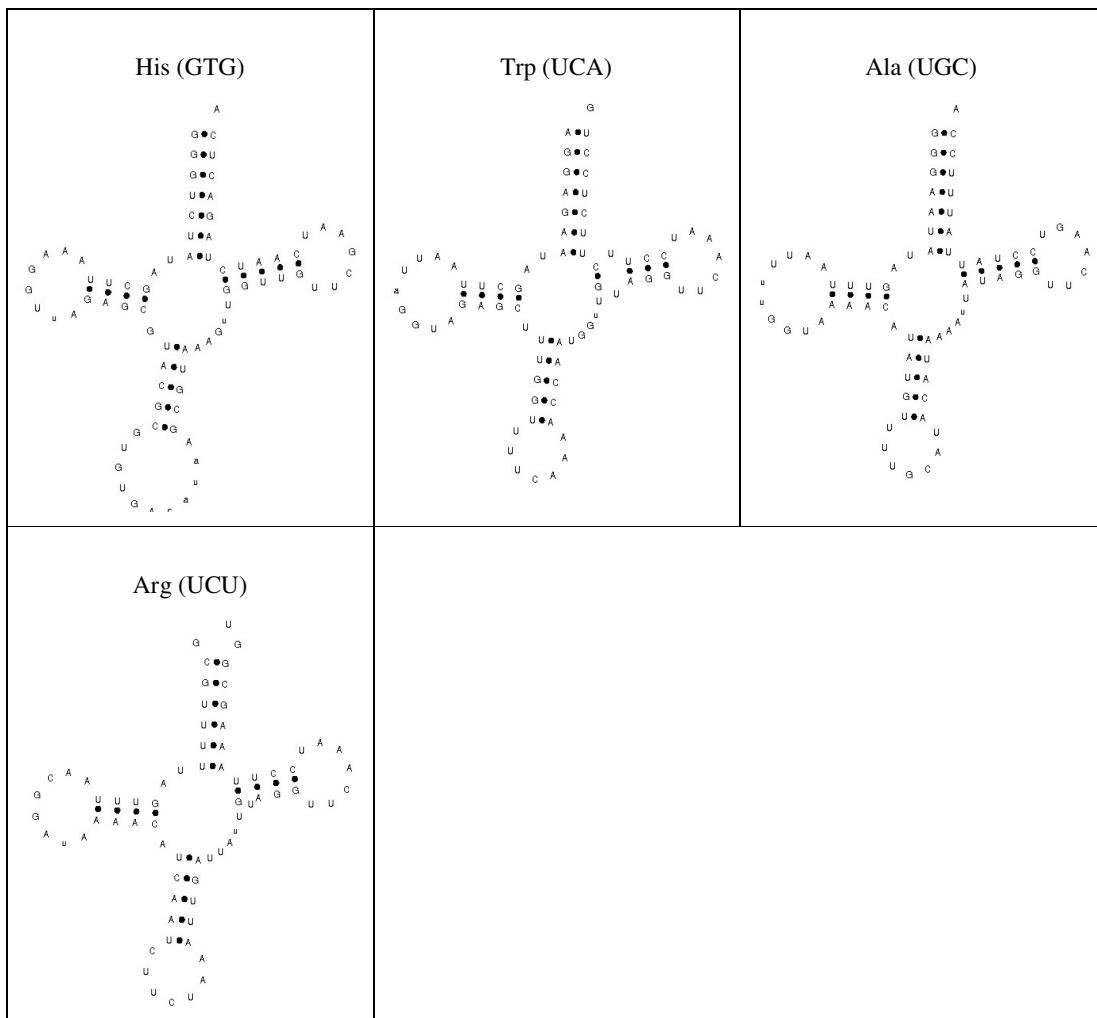
**9. ANEXO 2.** Estruturas secundárias dos tRNAs de *G. tenuistipitata* (em ordem de localização, conforme Tabela 7).

Asn (GUU)	Val (UAC)	Arg (ACG)
Lys (UUU)	Glu (UUC)	Met (CAU)
Asp (GUC)	Gly (UCC)	Gln (UUG)

## 9. ANEXO 2. Continuação

<b>Leu (UAA)</b> 	<b>Leu (UAG)</b> 	<b>Gly (GCC)</b> 
<b>His (GUG)</b> 	<b>Phe (GAA)</b> 	<b>Ser (UGA)</b> 
<b>Pro (UGG)</b> 	<b>Cys (GCA)</b> 	<b>Met (CAU)</b> 

## 9. ANEXO 2. Continuação



**10. ANEXO 3.** Genoma mitocondrial completo de *Gracilaria tenuistipitata* comparado ao genoma mitochondrial completo de *Chondrus crispus*. Os asteriscos representam os nucleotídeos iguais nos dois genomas.

10            20            30            40            50            60  
**Gt** TTCTGATAGCTAAAGGTAGAGCGTATGGCTGTTAACCAATTAGATTGAGGTCGAATCCTACTCAGGA  
**Cc** TTCTGATAGCTCAATGGTAGAGCGTATGGCTGTTAACCAATTAAAGTTAGGTTCGAATCCATTACAGGA  
\*\*\*\*\*  
  
 80            90            100          110          120          130          140  
**Gt** AGTTTTCA-AGAGCAATTAACTCAAAATGGTAGAGTATTCGTTACACTGAAAAAGTTAATGGTTCAA  
**Cc** ACTGATAAATAGGGTAATTAACTCAACTGGTAGAGTATTCGTTACATTGAAAAAGTTAATGGTTCGAA  
\*\*\*\*\*  
  
 150          160          170          180          190          200          210  
**Gt** TCCATTATTGCTCA--ATCTAAATAGTTCATAAAGTGTTGTAACCTTAATTAGGATAAAAGCGTTAAAC  
**Cc** TCCATTATTACTCATTAATTACAAAAATTAGTTATGTTGTTGTAACCTTAATTAGGATAAAAGCATTAAAC  
\*\*\*\*\*  
  
 220          230          240          250          260          270          280  
**Gt** TACGAATTAAAGATTGAAAGTTCGATTCTTCCAAACACATTA--ATAAA  
**Cc** TACGAATTAAAGGATTGAAAGTTCGATTCTTCCAAACATTTACTTATAAAAGTTAACGCAGAAAGTT  
\*\*\*\*\*  
  
 290          300          310          320          330          340          350  
**Gt** ---AACTC---GAGTGTGTAGCTTAGTTGGTAAGCAGTAAACCTTTAATTATAGATCTTAGGTT  
**Cc** AACTAATTCTAGCAAGAGTGTATAGCTTAGTCGGTAAAAGCAATAAACCTTTAATTATGGATCTTAGGTT  
\*\*\*\*\*  
  
 360          370          380          390          400          410          420  
**Gt** CAAATCCTAATACACTCA---ATTATG-TCTAAAAAAC---TTATT-TACTTGTATTCTTTAG  
**Cc** CGAACCTCTAATACACTCATTAATACACTAAACCTTTAATAAACCAATTCTTTGTTCTATTATTATTCATTG  
\*\*\*\*\*  
  
 430          440          450          460          470          480          490  
**Gt** AAATCTTTTTAGATTATTTTATGTATGTTAACGTTTTCTGTATGATTATAGCAAGTTAAATA  
**Cc** AAAATAGCTTATCG-TTTTTCTATATTATTAGTTTGTCTAATTATTGTTTTGGTAGC  
\*\*\*\*\*  
  
 500          510          520          530          540          550          560  
**Gt** TATATTATTAAATATTTTGAAGTATATCCCTTGTATATTGAAATTAAAAAAATTATGTTACAAA  
**Cc** TTCAAAATTGACTTTGATAGAAACTTACCAATTTCACAATTCTAATTAAAGTTATGCAACGCA  
\*\*\*\*\*  
  
 570          580          590          600          610          620          630  
**Gt** TGTAATGGATTATTAACATAGTATGATTCTAATTATATCCAAATCTTTTTTTATTTCCTTAT  
**Cc** ACAACAGATTAAATTGACGCCGTATGAATTACACTTTTCAAAATTCTTGTATCCATT  
\*\*\*\*\*

640            650            660            670            680            690            700  
**Gt** TGAGTTTTCATATCTATAAATTGGTAGTACTAGTTGATATAAAATTCAGCTTAAGTTCTAAAAAAAT  
**Cc** TTCGTTTATCAAATTATTTGTTTTCTAAGAATAGTTGATATGAGATCAGATGTATTTATAATTAAAT  
  
 710            720            730            740            750            760            770  
**Gt** CTTTTAACATTGCTTTTTATTTCTTTGTTCAAAATTTTTTATTACCCCTCTAT  
**Cc** TATTGTTATTCTTCGATACTTTATGTTATTCTTATTATTTGTTATTATCTTCTACCAATTAA  
  
 780            790            800            810            820            830            840  
**Gt** TTTATTTATTAAACAAAATGAGAAAATC AAAACACTAATTTA---TTTGGTATTTTATTGAATTTCGCG  
**Cc** ACTGGCTTTTATCGCAATGGAAAATAAAACAAAACAGAATCAATCTTAAATATTGAGATAGAATTAGA  
  
 850            860            870            880            890            900            910  
**Gt** ATCAATTAGTTAGTAAATGAGTTAACCTTTCGTTATTAATAGGTTCTTTAGTTTTTATATGTT  
**Cc** ATTTAAATTACGTTCAATGAATTAACTTTTCGTTATGTTTTAGTTCTATTTATATTGTAACCT  
  
 920            930            940            950            960            970            980  
**Gt** TAATACTTACATATATGAACTTTGTTA-CAATA-GATCATATTATTTTTATTAAGCATTATCGT  
**Cc** TTTTCAATTGTTTTTTCTTTTATTAAATTCTATGATAAAAATAATCTTTA-AAATTCATCGT  
  
 990            1000          1010          1020          1030          1040          1050  
**Gt** AAGGTTCTT-TTATTTATTACTATGTTTTTACTAATACCTCCGTGGATTTTACAAGTT  
**Cc** AAGATTTTATTTCTTAATACCTTAT-TTTCTTTAATTACACCACGTTACAATTCTTCAACTT  
  
 1060          1070          1080          1090          1100          1110          1120  
**Gt** TTTTGATAGGTTTCGTATTTTTTTATTGAAATAGTTTTTATTTATTTGTATAAAATTGTTAGTA  
**Cc** TTGGTTTTATTTATTCATTATTTTTATGAAATTATTTTTGTGTTACAAATGTAATTATA  
  
 1130          1140          1150          1160          1170          1180          1190  
**Gt** TAAAAT-AAATTAAATAATGCCACACTTAATCAACTATTAAAAGTATCAAGACAAAAAAAGTT  
**Cc** CATTTAGAAATTAACTTAATGCCTACTTTCAACAAATTACTAAAAAAACCAAGAAAAATAGTTCAACC  
  
 1200          1210          1220          1230          1240          1250          1260  
**Gt** GCTGAAAATCTATAGCGTTAGCAAAATGTCGGCAAAAAAGGGTGTGTTGTTAAAAATTTCACAATGA  
**Cc** AAAAGCAAAATCGATTGCTTAACTCAATGTCCTCAAAAAAGGGATTGTTAGAAAGTTTACAACAA  
  
 1270          1280          1290          1300          1310          1320          1330  
**Gt** ATCTAAAAAAACCAAATTCCGCAGAACGAAAGTAGCTAAAATCACACCTAGCAGC-GGAAAAATCAATTAA  
**Cc** GTCTAAAAAAACCTAATTCTGCAGAACGTTAAAGTTGCAAAAGT-TAAATTAAACAAATGGTAAATCCATAA







	3440	3450	3460	3470	3480	3490	3500
Gt	<b>ACTTCTATTTATTATTTTAATTAGTGTCTAGGAATATTTTGAAATCAAAAAATACTCTTGTATGT</b>						
Cc	<b>ATTCATCTTATTATTTTAGTCAGTTACTAGGAATATTTAAATCAAAAAATACTCTTGTATGT</b>						
	* . *** . . ***** . * . *** . . ***** . . ***** . . ***** . . ***** . .						
	3510	3520	3530	3540	3550	3560	3570
Gt	<b>TAATGTCCTTAGAAATGATGTTTGTCAATAAGTTAATTGATTTGCATCAATTATTTAGATGA</b>						
Cc	<b>TAATGTCACCTGAAATGATGTTTATCAATTAGTTAATTTCATCAATACGTTAGATGA</b>						
	***** . . ***** . . ***** . . ***** . . ***** . . ***** . . ***** . .						
	3580	3590	3600	3610	3620	3630	3640
Gt	<b>TATTACAGGTCAAAATTTTTCTTATTAAATTTGACAGTAGCAGCTGCCGAATCGTCATTGGATTAGCA</b>						
Cc	<b>CATTATAGGTCAAAATTTTTCTTACTCATTTAACAGTTGCAGCTGCCCAATTGGATTAGCA</b>						
	***** . . ***** . . ***** . . ***** . . ***** . . ***** . . ***** . .						
	3650	3660	3670	3680	3690	3700	3710
Gt	<b>ATTTTAGTTATTATTCATCGTACGTAATGTTATTACTGTAGAATTATGCACCTAATGAAAGGCTAAA</b>						
Cc	<b>ATTTTAGTTATATATTATCGTACGAAAGCACGATAACAGTTGAATTGATGAACCTTATGAAAGGTTAAA</b>						
	***** . . ***** . . ***** . . ***** . . ***** . . ***** . . ***** . .						
	3720	3730	3740	3750	3760	3770	3780
Gt	<b>TGTTTGGTTAAGGGATATCAAATTTTGTAACGAGCACTTAATAAAACCTTGGTAAAAAAATTGGAC</b>						
Cc	<b>TTTTTGGTTAAGAAATTAGAATCTTTAAAACAGCATTAATGAAAACCTTGTATAAAAATAAGGAC</b>						
	***** . . ***** . . ***** . . ***** . . ***** . . ***** . . ***** . .						
	3790	3800	3810	3820	3830	3840	3850
Gt	<b>GTTTGCCTACGAAAAGTTATAAGGAGGCATTGGCTTGTGATTTATAAAATCCATGTTAAAGTAAACTT</b>						
Cc	<b>GTCACGCTACGAAAGTTATGAGTAACCATAGGCTTGTATCCGTTAAATTCCTAAGAGTAAACTC</b>						
	***** . . ***** . . ***** . . ***** . . ***** . . ***** . . ***** . .						
	3860	3870	3880	3890	3900	3910	3920
Gt	<b>A-CTAAAATTTTTATAAAACAATGTAACACTGAATCATCTAAGTAAACATTGAAAATAATCAATTGAGAA</b>						
Cc	<b>AGCTTAATTTTACATATAGTAACGAGAATTGAATCATCTAGTATCGTTAATAAATCAAAACGAGAA</b>						
	***** . . ***** . . ***** . . ***** . . ***** . . ***** . . ***** . .						
	3930	3940	3950	3960	3970	3980	3990
Gt	<b>ATCGTTAGTAATGGTGAATGAACCGCAT---CTAAGAGCTTAGTAAAAGTTATAT--TTATAATTAC</b>						
Cc	<b>ATTGTAAGTAATGGTGAATGAAACCAAAAGGCTATTATTTAAATTCAGTTAAAATTATTTATTGC</b>						
	***** . . ***** . . ***** . . ***** . . ***** . . ***** . . ***** . .						
	4000	4010	4020	4030	4040	4050	4060
Gt	<b>AAAAAAAGGTATAAGTCCTGTAGTAATGTAACCTTTAATTAAAGTAATATGAATT-TAAAT</b>						
Cc	<b>TTTGGAAAGGTATAAGCCCTGTATTAAACTGAAATAAT--ATAACAAAACTAAGTAGTACGAAAGTTAC</b>						
	***** . . ***** . . ***** . . ***** . . ***** . . ***** . . ***** . .						
	4070	4080	4090	4100	4110	4120	4130
Gt	<b>TTGTACGAAAAAGGAGAACCACCTCTAACGCTTAAAGCTTAAAGGTT-AACCTATAGTGTACGAGTACCGT</b>						
Cc	<b>TTGTATGAACAAAAGGAGATCCACCTCTAACGCCAACCAACCAATTGTTAATCAATAGTAAACGAGTACTGT</b>						
	***** . . ***** . . ***** . . ***** . . ***** . . ***** . . ***** . .						







6240 6250 6260 6270 6280 6290 6300  
**Gt** TGAAATAACACTGAATGAATTAAAGTGTAAAGTATTCTTTATATTTTCATGAATAGCCTGAAAGATT  
**Cc** CGCTAAATACTGAAAGAATCAAAAGTATCAAAGCTATTTTTTATTTTCAGGAATAATCTAAAAGACT  
 \*.\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*..\*\*..\*\*..\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*..\*\*..\*\*\*\*\*.  
  
 6310 6320 6330 6340 6350 6360 6370  
**Gt** ATCAGATAGATCGGTTGAAAATACTCCTAGTAATAAATACTTTAACAAATCCGAA- TTATTCAAAA  
**Cc** ATTAGATTGATCGATTATAATTAAATTAGTAATAAAT- TTAGTAATAAACACGAAATTATTCAAAA  
 \*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*..\*\*..\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
  
 6380 6390 6400 6410 6420 6430 6440  
**Gt** AA-ATTTGATTAACTTAACCAAAA-AATTAT-GACAAAAAAATAAATCCAATAAGCCTT  
**Cc** AAGATTTAATCTAACCTAACCAAAATTCAAAATTACAATGGCACAAAAATAAATCCAATAAGTTT  
 \*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
  
 6450 6460 6470 6480 6490 6500 6510  
**Gt** AGGCTTGGTCTAACTCAAGTGTGAGCTATTACAATAACAAATTATAGTAACATAATGATTCTTATGTC  
**Cc** AGACTTGGAACAACTCAAGTTGAAATTCTACATTACAAATTACCGTAACATTCAAAATTATTCT  
 \*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
  
 6520 6530 6540 6550 6560 6570 6580  
**Gt** TATCTTTTTAAACATTGCAAGTTGATAATAATAGTAACATAAAATTAAAGATCAAATAGGTTTTAAAT  
**Cc** CAATTGTAAACAAATCAACTAACATTGCAAAATTATAATTAGATACGTTAAATCAAATGGCTTCAAAAT  
 .\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.  
  
 6590 6600 6610 6620 6630 6640 6650  
**Gt** TAACGATAAAGAATTGATAATTAAATAAACTTTTTAAACATTATTATGACGAACTTATTGAA  
**Cc** AAATTACTATGAATGGAAAATCCATAAAAATAAAACTTTTTAAGTATTATTTCTCCTTCTTTAAC  
 .\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.  
  
 6660 6670 6680 6690 6700 6710 6720  
**Gt** TATAAGTTAGATAACTATTATTGTTAACAGAAAATATCTAAAGTAATTCAAATTGATTTCCTTAA  
**Cc** AATGGAAAACCTAACT--TAACATT-TAAACAAATTCCAATTATTATCTAATTGATGTTCA-AT  
 .\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.  
  
 6730 6740 6750 6760 6770 6780 6790  
**Gt** AAATTAT-TTAAGAATGTTAAAGGATAGTTAGTAACAAACATCTAATATCGAAATTATGCC  
**Cc** AAATTATATTCACTTTATTAAATTCAAAATGATCTTGACATCTAATTAACTCTTTATACAA  
 \*\*\*\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*.\*\*.\*\*.\*\*\*\*\*.\*\*\*\*\*.  
  
 6800 6810 6820 6830 6840 6850 6860  
**Gt** TATATTATTGAAACAAAATAAGAATCCAAAATAAGTTGACAGATATCTCAATTAAACAAAGATA  
**Cc** AACTTAGAGAACAGAAACTAGGACAAAAAAATAATAGTAATTCTATGTAACCTTTACAAGATCA  
 .\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.\*\*.  
  
 6870 6880 6890 6900 6910 6920 6930  
**Gt** TCTTAATTACGGTAAAGTAGTGTATTCTACAAAAGGTATACGTTCAAGTATATTGAAAGGAATTAAATT  
**Cc** ATTAAATTCTTCAGAAAGTAATGCATTAAATTGGTTATCGAAACAAAATTAGTTGGGTCAAATT  
 \*.\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.



7640      7650      7660      7670      7680      7690      7700

Gt ATTAAACAAAGTAAAATCCTAGTTAAAATTATAAGAATAATGCAAATTTCAATTGCAATGAATTCTCG  
 Cc ATTAAATGAAGTTAAAT-----AAAGTATAAAACA-TGCAATCAATTTCCTACTCAATGAATTCTCG  
 \*\*\*\*\*..\*\*\*\*\*.\*\*\*\*\*.....\*.\*.\*\*\*\*\*.\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*

7710      7720      7730      7740      7750      7760      7770

Gt TTGAATTTCACACAAATCATAAAGATATTGGTACCTTATTTAATTGGTGCCTTCTGGTATA  
 Cc TTGAATTTCACACAAATCATAAAGATATTGGTACTCTTATCTAATTTCGGTGCCTTCTCAGGTGTT  
 \*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*

7780      7790      7800      7810      7820      7830      7840

Gt TTAGGCGGATGCATGTCAATGTTAACATTGTATGGAAATTAGCCCAACCTGGAAATCAATTATTGTTAGGAA  
 Cc TTAGGTGGATGCATGTCAATGTTAACATTGTATGGAGTTAGCTCAACCCAGTAACCATTACTCTTAGGAA  
 \*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*

7850      7860      7870      7880      7890      7900      7910

Gt ATCATCAAATTATAATGTTTAATTACAGCACATGCTTTTAATGATTTTTTTATGGTAATGCCGTGT  
 Cc ATCACCAAATTACAATGTTCTTAAACAGCTATGCAATTCTTATGATTTTTTTATGGTTATGCCGTGT  
 \*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*

7920      7930      7940      7950      7960      7970      7980

Gt TATGATTGGAGGGTTGGAAATTGGTTAGTTCCATTATGATAGGTAGTCCAGATATGGCCTTCCTCGT  
 Cc AATGATTGGTGGGTTGGAAATTGATTAGTTCCAATAATGATTGGTAGCCAGATAGGCTTCCCGCGA  
 \*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*

7990      8000      8010      8020      8030      8040      8050

Gt TTAAATAATATATCATTGTTGATTACCCACATCTCTTGTTACTTTAGCATCAGCAATTGTTGGAAG  
 Cc TTAAATAATATTCATTGTTGATTACCCCTCTTATGCTTATTACTTATGTCAGCAGTTGTTGAAG  
 \*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*

8060      8070      8080      8090      8100      8110      8120

Gt TAGGTGTTAGGTACAGGAATGAACTGTATACCCCTCACTTAGTTCAATTCAAACCCATTCTGGTGGTGCTGT  
 Cc TAGGTGTTGGTACTGGATGAACTGTTATCCACCAACTTAGCTCAATTCAAAGTCATTCAAGGTGGTGCTGT  
 \*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*

8130      8140      8150      8160      8170      8180      8190

Gt AGACTTAGCGATTTTCAGTTACATATTTCGGGTGCATCTCAATTAGTCATACTTCTAATTAAATTTCCT  
 Cc AGATTTAGCAATTTTACTGCTTCATATATCAGGTGCTCTCAATTAGGAGCTGTAATTTATTTC  
 \*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*

8200      8210      8220      8230      8240      8250      8260

Gt ACGATTTAAATATGCGTAATCCAGGACAAAGTATGTATCGTATGCCCTGGTTGATGATCAATTTC  
 Cc ACTATTTAAATATGAGAAGTCCCTGGACAGAGTATGTATAGAAATTCCCTTTGTTGATCAATACTTG  
 \*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*

8270      8280      8290      8300      8310      8320      8330

Gt TTACTGCATTTTATTATTAGCTGTTCTGTTAGCTGGTGCTATTACAATGCTTTGACTGATCG  
 Cc TAACAGCATTTCACCTTATTGGCAGTTCCCTGTTAGCAGGAGCTTACAATGCTTTAACTGATCG  
 \*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*







	10440	10450	10460	10470	10480	10490	10500
Gt	TGAAGAGATGTAATTAGGGAAAGCGACACTTGAAGGGCATCATAACAGGGATAGTTCAACGAGGGTTACGTT						
Cc	TGACGTGATGTAATAAGAGAACTCTACTTTGAAGGTCATCATAACAGGAATTGTACAGCAAGGACTAAGAT	***.**.*****.**.**.**.*****.*****.*****.*****.**.**.**.**.**.**.**.**.**.					
	10510	10520	10530	10540	10550	10560	10570
Gt	ATGGAGTAATTCTTTTATTGTATCTGAAATTCTATTTTTTCGCTTTTTCTGGCTTTTTTCATAG						
Cc	TCGGAGTCATTTTATTATTATCTGAAATTCTATTTCTTGGCTTTTTGGAGCTTTTTTCATAG	.....*****.*****.*****.*****.*****.*****.*****.*****.*****.*****.*****.					
	10580	10590	10600	10610	10620	10630	10640
Gt	TAGTTTAGCACCAACAGTAGAAAATCGGTTAACGTGACCCCTAAAGGAATAAGTGTTTAGATCCTGG						
Cc	TAGTTTAGCTCCTACAATTGAAAATAGTTCTATTCAGCTCCAAAAGGAATAATGTTAAATCCTGA	*****.*****.*****.*****.*****.*****.*****.*****.*****.*****.*****.					
	10650	10660	10670	10680	10690	10700	10710
Gt	GAAATTCCCTTTTGAATACCTTAATTATTGTTGTCAGGATGTACAGTTACTTGATGTACATCATGCAC						
Cc	GAAATTCCCTTCTAAATACCTTAATTCTATTACTTCAGGTTGACTGTAACTTGATGTACACCATTCT	*****.*****.*****.*****.*****.*****.*****.*****.*****.*****.*****.					
	10720	10730	10740	10750	10760	10770	10780
Gt	TTGTGGCTAATTAAAGAAATCAGGCTATTATTAGTTAACACTTACGGTAACCTTACCAATTATTTCAC						
Cc	TAGTTTCAAATTACGTAATCAATCAGTTAACAGTTAACATTGTTAACAAATTGTTAGCAATTGTC	*****.*****.*****.*****.*****.*****.*****.*****.*****.*****.*****.					
	10790	10800	10810	10820	10830	10840	10850
Gt	CGTATTACAAGCATAACGAATACAACATGGCTGATTTAGATTATCTGACGGTATTTAGGATCTACATT						
Cc	TACATTCAAGCATAACGAATATACTGGCTGATTCAGATTATCTGATGGGATTATGGTTCAACATT	*****.*****.*****.*****.*****.*****.*****.*****.*****.*****.*****.					
	10860	10870	10880	10890	10900	10910	10920
Gt	TATATGGCAACC GGATTTCATGGATTTCATGTAATTAAATAGGAACAAATTCTTTATTTATCTGCTTAATAC						
Cc	TACATGGCTACAGGATTTCATGGATTTCATGTTAGTTGGAACAAATTCTTTAGCAGTTGTTAAATTC	*****.*****.*****.*****.*****.*****.*****.*****.*****.*****.*****.					
	10930	10940	10950	10960	10970	10980	10990
Gt	GTTTAATACAATACCAAGTTAACGAGGGACATCATTGGATTTGAATCTTCAGCCTGGTATTGACATT						
Cc	GATTACTGCAATATCAATTAAACCAACACATCATTGGATTTGAATCTGCTGCTTGTATTGACATT	*****.*****.*****.*****.*****.*****.*****.*****.*****.*****.*****.					
	11000	11010	11020	11030	11040	11050	11060
Gt	TGTAGATGTTGGTTAATTCTTATTGTCATTTATGATGAGGGAGTAT-----GTA	*****.*****.*****.*****.*****.*****.*****.*****.*****.*****.					
Cc	TGTGGATGTTGGTTAATTCTTATTGTCATTTATGATGAGGGAGTCTAAAGATAAAT	*****.*****.*****.*****.*****.*****.*****.*****.*****.*****.					
	11070	11080	11090	11100	11110	11120	11130
Gt	ATGTTAAATTAAAGTATTGTTATTGGAGTGTCAATTAAATTCTAAAAATTATATTAAATG						
Cc	ATGCTTAATCTTAGCTTATTAGCTTAACCTCATTATTTAATTATCAGAACATTATTCTTAATG	*****.*****.*****.*****.*****.*****.*****.*****.*****.*****.					









	13940	13950	13960	13970	13980	13990	14000
Gt	<b>AAAGAGCTCGGTAAAGGATTGAGCAATAGATTGTGAATCTAAAGGTCA</b> <b>GGTTCAAAATCCCATTTCGC</b>						
Cc	<b>AGCGAGCGCGTA-GGTTT GAGCATCAGATTGTGAATCTGAAGGACG</b> <b>TGGGTTCGAATCCCATTTCGC</b>						
	***** .*****.*****.*****.*****.*****.*****.*****.*****						
	14010	14020	14030	14040	14050	14060	14070
Gt	<b>CTTATAACATTGAGATCTGATTTTGATGGATGAGATAGCTTTCCGGGATT</b> <b>CAAAAATTAGGGCAAGT</b>						
Cc	<b>CTTAATTATTATGATTCTATATTGATCGAACATAGCTTACCCG</b> <b>GATTTAAAGATTAGGACATGT</b>						
	***** .*****.*****.*****.*****.*****.*****.*****.*****.*****						
	14080	14090	14100	14110	14120	14130	14140
Gt	<b>TTTGC TACAATTCTATAATTGTATGACATCTAAATAGTCGC</b> <b>ATTTTATGATCCAAAATTAACTCGGTT</b>						
Cc	<b>CTTACTACAGTTCATATTGTATGACATCTAAATAATCGC</b> <b>ATTTTATTAGAAATTCATATTAGGTTGTAAATAATAGCAGGGC</b> <b>CTAAGTACT</b>						
	***** .*****.*****.*****.*****.*****.*****.*****.*****						
	14150	14160	14170	14180	14190	14200	14210
Gt	<b>CGGGTATT CGAGTCCCTAGAACTCTAAGATT</b> <b>CATCTATAAAGCTTGCAAAATTGCG</b>						
Cc	<b>TCAGTAGAATTATCTCGAGAACATCAACGATT</b> <b>CATCTATAAGCTTGCAAAATTGCG</b>						
	***** .*****.*****.*****.*****.*****.*****.*****.*****						
	14220	14230	14240	14250	14260	14270	14280
Gt	<b>TATCATGGTTCCACCAATAACTAGGGCAGCTAGCTG</b> <b>AAACAAGCGCACAAATACATTCA</b>						
Cc	<b>TGCTTGTATTCCACCAATACTGGGCAGCTTG</b> <b>CAGAACACAAGCACATAAAATACATT</b>						
	***** .*****.*****.*****.*****.*****.*****.*****						
	14290	14300	14310	14320	14330	14340	14350
Gt	<b>ATCCAATTCAAGATCGATCTCTTTGATTGTAAAGT</b> <b>ATTCCCTTTAGGGATTAAATTATTGTTATTAAT</b>						
Cc	<b>GTCTAATCAAGTCTATCTATTAGATTGTAAAT</b> <b>TTTCCCTTAATGAAAAACCAATTATTGATTAAT</b>						
	***** .*****.*****.*****.*****.*****.*****.*****						
	14360	14370	14380	14390	14400	14410	14420
Gt	<b>CAAGGTTAATCATTCTATATTGTGCA</b> <b>ATAAAAGTGGGTTAGGT</b>						
Cc	<b>CAAGGTTGATCAATTATATTGACCA</b> <b>ATAAAAGTTCGATAAAATCTGGAA</b>						
	***** .*****.*****.*****.*****.*****.*****.*****						
	14430	14440	14450	14460	14470	14480	14490
Gt	<b>TATGTGGTAAAGGATATAGTTAAATATT</b> <b>CTCTGTGTTAAAGATTAAAGCAGCTAAAGTGTT</b>						
Cc	<b>TATGTGGTAAATGGATAAAAGTAGTAA</b> <b>AAAATTACTTTAGTATTAAATGATCGAAGCACAGCTAATGAATT</b>						
	***** .*****.*****.*****.*****.*****.*****.*****						
	14500	14510	14520	14530	14540	14550	14560
Gt	<b>TACTCCATTATATTCAATTGAAACAACTTCCACAAATT</b> <b>CCCTCTCTACATGACCGGGAAACGCTAAAGAG</b>						
Cc	<b>TGTGCCATTCAATTTCATAGAACAGCTTCCACAT</b> <b>ATACCTCTACAGGAACGACGAAAGTTAAAGAT</b>						
	***** .*****.*****.*****.*****.*****.*****.*****						
	14570	14580	14590	14600	14610	14620	14630
Gt	<b>GAATCTTGGATATCTTTATTGAA</b> <b>TTAAAGCATCCAAAATCATTGGGCCACAA</b>						
Cc	<b>TTAAAGCATCCAAAATCATTGGGCCACAA</b> <b>ATTATTAAGTGAACAG</b>						
	***** .*****.*****.*****.*****.*****.*****.*****						

14640      14650      14660      14670      14680      14690      14700  
 Gt **GATAAAATATTAAATCAAGGTTTTTATATAAAGGGTTCATCTATATATTGTAATAATTTGGGA**  
 Cc **GGAAAATGAAACCAAGGA**TTAAGTTAAATTGGGTTCATCTATATACGAATAAACCTTTGG—  
 \* ..\*\*\*\*\*...\*\*\*.\*\*\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*..

14710      14720      14730      14740      14750      14760      14770  
 Gt **ATTTACTTAAGGTTGGA**TAAATTATTTATTGAACTTTTTGTAAAAACATTCAAATTAAAGATT  
 Cc ————**TAAGGC**GAAGAA**TTGTC**TATTGTAATAATTGTTGGTTAGGTTTTATAATCATA—CT  
 .....\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.

14780      14790      14800      14810      14820      14830      14840  
 Gt **AATAAAATAAATATAAAATAAAAC**TAATGGTTAAGAAA**TACAATATAAAATTGAT**—T  
 Cc **ATTGAAAAATTAAAGATTAAAGA**TAGCGAAAC**TAAGGAA**TAGCGAAACTTAGATAAGTATTGTC  
 \* ..\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.

14850      14860      14870      14880      14890      14900      14910  
 Gt **TAACACAAAGATCCCATAAAAGTTGTTAAACCATTAAATG**CATGATAACCTAAAAATAAGATA  
 Cc **TAGTAAGAACCAAATCTCATTTATGTC**TTGTACCATTAATAATGATAGAATAACAAAT  
 \* ..\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.

14920      14930      14940      14950      14960      14970      14980  
 Gt **ATTAATACATATAAAACCTCATAAAAAAACACTATATTATTAAACATATAATTAAATAAAATT**  
 Cc **ATAAAATAAAT-TGAAATA**—ATAATTATTTCATTGGGAAATT**CAATGAAATAAGAGTTAAAGATT**  
 \* ..\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.

14990      15000      15010      15020      15030      15040      15050  
 Gt **GTT-CATAATTAAATATGTAAGGTAAC**TGAAATAATAGATGAAAAACATTAAAGAGGATAATAAA  
 Cc **ACAAGCAAATT**CGA**ACTC**AAAATTCTAAATAAAGAAAG**TGTAAGAAAAAGCTAATCCA**  
 .....\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.

15060      15070      15080      15090      15100      15110      15120  
 Gt **ATACCTGAAATGCGATGTC**AAATAAGAGACAAAGAC**GACTCTTGAACTG**CATAATTGTAATATGAGGAG  
 Cc **ACTCCTGAAATT**CGATGTCAGATAAGAAATAAAGAGA**TAGTTGTGGAGTATACAACTGTTG**  
 \* ..\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.

15130      15140      15150      15160      15170      15180      15190  
 Gt **ATAAAGGACGATTATACATCGAATATTTATTAAACATAGTCATAATATG**CCAGAATGGGATT  
 Cc **CTATGGGGCATTAGAAATCTTAAATTGATAAACATATT-TGTAA-**ATTAATTGTCAGAATGGGATT  
 .\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.

15200      15210      15220      15230      15240      15250      15260  
 Gt **TGAACCCATGACTCAAGGGTTTCAACCC**TATGCTCTAACCACTGAGCTACCTAGACTTTAATCTTTTT  
 Cc **CGAACCCATGACTCAAGAGTTTCAGCCCTATGCTCTA-CCACTGAGCTATCTAGACAAAAAGCTATT-**  
 .\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.

15270      15280      15290      15300      15310      15320      15330  
 Gt **AACACCATAAATCCTAACTATCTATAATAATCGGATGAAGTAGGATTG**AACCCACGATAAGTTAAC  
 Cc ————**ATTTA**—GATGAAGTAGGATT**CGAACCCACGATAA**—AAA  
 .....\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.

15340      15350      15360      15370      15380      15390      15400  
 Gt TCTTATACCAATTTCAAAATTGGCGTTAAACCACTCAGCCATTCACTCCTTCAATAAATCTTAGGGT  
 Cc TATTATGTCAATTTCAAAATTGATGCTTAAACCACTCAGCCATTCACTAATTT-----TATTAGGGT  
 \* .\*\*\*\*\* ..\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*  
  
 15410      15420      15430      15440      15450      15460      15470  
 Gt AATAGGATTCGAACCTATAAATTTTTACTCCCCAAGCAAACACGCTAACATTACGTCAACCCATAAGA  
 Cc AGTAGGATTCGAACCTACAATTTTTGCTCCCCAAGCAAACACGCTAACATTACGTATAACCCTATG  
 \* .\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
  
 15480      15490      15500      15510      15520      15530      15540  
 Gt TTTATTATTAAATTAAAGTAAATAAAAATTAAGAAAGCCATCATCACGCAAATAGTGCCACAGCTT  
 Cc -----TAATTAAAT--ATTAAGTAAACAGAATTAAAAAGCCATCATTAAAGCAAATAATGCCACAGCTT  
 .\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*  
  
 15550      15560      15570      15580      15590      15600      15610  
 Gt CTGTTAATGCAAATCCTAAAATAGTATAACCAAATAATTGTTGTTTTATGATGGATTACGCGCATTAAGC  
 Cc CTGTTAGTGCAGAACCTAAAATAGTATAACCAAACATTGTTGTTTTATGATGGATTACGTGCATTAAGC  
 \*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*  
  
 15620      15630      15640      15650      15660      15670      15680  
 Gt CATTACTAATGATCCAAACACAATACCTACACCAAGCTCCTACACCAGTTAACCGATAGTTGCTAAACCT  
 Cc CATAACTAAAGATCCAAATACAATTCCCTACACCAAGCACCCTACACCAGTTAACCGATAGTTGCTAATCCT  
 .\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*  
  
 15690      15700      15710      15720      15730      15740      15750  
 Gt GCACCTATCATTTCGCACTTTGTAAGTTACGTTCAATT-TTATCAAATTATTTAT-----TAATC  
 Cc GCACCTATCATTTCGCACTTTGCAAGAGTAACGTTCAAGTTTTCAAATAATTAAATATAATTGT  
 \*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*  
  
 15760      15770      15780      15790      15800      15810      15820  
 Gt -AGCCTCTCGTC-----AAAGCTAGAAATCGGATTCGAACCGATGTAAGAAAGATTGCAATCTTTGCA  
 Cc AAGCCTCTCGCTTTATAAAAGCTAGAAATCGGATTCGAACCGATCTTTAAAGATTGCAATCTTATGCA  
 .\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*  
  
 15830      15840      15850      15860      15870      15880      15890  
 Gt TCAACCACTATGCTATCTAGCCTCT-----TAACGGAAAAGACTTGAACCTATAACTTTGGATTATG  
 Cc T-AACCACTCTGCTATCTAGCCAAATAATTGACAAAAAGACTTGAACCTATAAGCTTTGGATTATG  
 .\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*  
  
 15900      15910      15920      15930      15940      15950      15960  
 Gt ATTCCAATGTTCTAACCAATTGAACATTCCGTT--TTATATGCGTTTTAGGATTTTACATCCG  
 Cc ATTCCAATGTTCTAACCAATTGAACATTTCGTCAATTATTCACGAATTTCGTTCTGCATCCA  
 \*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*  
  
 15970      15980      15990      16000      16010      16020      16030  
 Gt TTATGTGATAGCAATGAAATTATTCATAATTGATAAAAATAGTAAAAAGATTGTTAAACACTTAATA  
 Cc TTATGTGGTCACGATATTATCAGAAATTGATAAAAATTTAAAGTTGATGTTTAATTGTTTATAA  
 .\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*

16040 16050 16060 16070 16080 16090 16100  
 Gt CTGCTTTTATGGTTATCAAATCCTTTAAATTAGATGTATAATTAACTTTAATTGAATCAAATA  
 Cc CTAATTTTGTGTTTACTAAATCCTTTGTTCTATGTGAAT-TCCTTGTAATTGACAGACCTATACT  
 \*\*\*.\*\*\*\*\*.\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
 16110 16120 16130 16140 16150 16160 16170  
 Gt TTCTATAATTGAATCTAACAGATTGAAACCAATGTTGTTAATGTTACTTTTAGCCCCACGTGTTTTTTA  
 Cc TAATATAATTGA-TCACAG-TCTAATTGATA---TTAATGTTACTTTTAGTATTTTTTGTTGTGG  
 \*.\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
 16180 16190 16200 16210 16220 16230 16240  
 Gt GATCCGGCCGATGTTCAAAATAATACCTCCCTTAAGATTTGTAATAGTACATAAAAATTATTAGATG  
 Cc ACACCTGATGATGTTCAAAATAATATTACCGTTAATTGTTCCATAATATAATATTAGCTG  
 ...\*.\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
 16250 16260 16270 16280 16290 16300 16310  
 Gt TAAATAAAATATTGAGTCTAATAGATTTAAATTATTTC-----GAATCATTT-TTAACTGT  
 Cc TAAACAGAATTGATAACACTATTAAATTGCTTTATTCCATATAAAATAGTTAAAATTTAACCGA  
 \*\*\*\*.\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
 16320 16330 16340 16350 16360 16370 16380  
 Gt CTAAGAATTCCAT-ATAAAAATAGTAGTTAGCAAAGATAGATTTATTTTGAGTTCGATATGGATCA  
 Cc TTGATATTCCCATTATTACATAGTAGTCATCAAAGAAAAATTATATGTGAGTTCGGAATGGATCA  
 .\*.\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
 16390 16400 16410 16420 16430 16440 16450  
 Gt TATATTTAATATCTACTT-TTCTAGTTAAAAAAATTATATAACTTATTCTCATCTAAATGCACCTTAT  
 Cc AATAATTAGCAATTTCGTTTAGTTAAATTGATTATTGTTATTCTCATCTCAAGTGCCTTTAT  
 .\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
 16460 16470 16480 16490 16500 16510 16520  
 Gt ATCATT CGTAAATAAA CCTATTGTTAATATAATTAAATATACTCATCTCAAAATCCAATAAAGA  
 Cc ATCATT CGTAAATAAA CCTATTGTTAATATTACTAAAAAACATCATCTCAAAACCAATTATGGA  
 \*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
 16530 16540 16550 16560 16570 16580 16590  
 Gt AATACTATTCAAAGCTAATGATCATGGAAAAGGAAACTAAATTCAAGATCAAATTAGAAACAATATA  
 Cc TATTCTCCGAGTACTAATGATCATGGAAAATAAAACATTAATTCTAAATCAAATTAAAAAGAATA  
 .\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
 16600 16610 16620 16630 16640 16650 16660  
 Gt GCTACTAAATAAAATCGAATATCAAACGTTGCTCTAGCATCATCAAACGGATTGAAAGCCGCAATTCTAAAG  
 Cc GCAACTAAATAAAATCTGATCTCAAATGTTGCTCGAGCATCATCGAACGGATTAAATCCGCACTCATAAAG  
 \*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
 16670 16680 16690 16700 16710 16720 16730  
 Gt CACTTATTTCTTGTACGCTTTGGAAATTAAAAGTAAGATAAAACTAAAAATTAGCACT-GATA  
 Cc CACTAACCTTTCTTGTACGCTGGTTTGAGGAATTAAAATTAAGA-AAGTAAAATATCACAGAAGAGA  
 \*\*\*\*.\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.











20240 20250 20260 20270 20280 20290 20300  
 Gt AATAAATTGATGCCTCCGGAAACATAGGTATAGAAAAACGTAAGAAACCAAACCCCCATTTTTAAATA  
 Cc AAAAATTGATGCTACTGGAAACATAGGTATAGAAATCGTAAAACCCGTATCTCCCATTAAAGAAG  
 \*\*\*.\*\*\*\*\*.\*\*\*.\*\*\*\*\*.\*\*\*.\*\*\*\*\*.\*\*\*.\*\*\*\*\*.\*\*\*.\*\*\*\*\*.\*\*\*.\*\*\*\*\*.\*\*\*.  
  
 20310 20320 20330 20340 20350 20360 20370  
 Gt TATACCTGCTAAAATTACAGATCCC GCC GTT GGAG CCT CT GC AT GT GCT T CT GG TAAT CAT AT AT GAAAT  
 Cc TATTCCAGCTAAAATACAGAACCTGCAGTCGGTGCTCTGC GT CT GCT CAGGAAGTCATATGTAAAAA  
 \*\*\*.\*\*.\*\*\*\*\*.\*\*\*.\*\*.\*\*.\*\*\*.\*\*\*.\*\*\*\*\*.\*\*\*.\*\*\*\*\*.\*\*\*.\*\*\*\*\*.\*\*\*.  
  
 20380 20390 20400 20410 20420 20430 20440  
 Gt GGAATC ATAGGTATTTTACAGAAACCTGAAAGAAAGCAAGCCAAAAATAATTGCTTTATCTCAG  
 Cc GGAATC ATAGGAATTTTACGGCAAAACTGCAAAGAAAGAAACTCAAAGGAAAATTGACTAATTCAG  
 \*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
  
 20450 20460 20470 20480 20490 20500 20510  
 Gt TAAAATTATAATATCATAAAAC TTG TAAATCTGTAGTAGCTACTC AAAAATAATTGCAAGTAGCGCTAA  
 Cc AAAATTGAATATATCATAGTATTTGTAAGTCAGTAGTCCAGTTGGAAATAATTAAGTAGTGC TAA  
 .\*\*\*.\*\*.\*\*\*\*\*.\*\*\*.\*\*\*\*\*.\*\*\*.\*\*\*\*\*.\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
  
 20520 20530 20540 20550 20560 20570 20580  
 Gt TAACATTAATAATGATCCAATTAGTG TATATAAAAAAAATTGATAAGCCGCTCTATTTCAGTTCTCTA  
 Cc TAACATTAGTAAAGACCCAAATTAAGATATATAAAAAAAATTGGTAAAGCAGCTCGAATTTTCTGTTGCTA  
 \*\*\*.\*\*\*\*\*.\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
  
 20590 20600 20610 20620 20630 20640 20650  
 Gt GATCTCATATACCAACAATTAAAACATGGTATTAAAACACTTCAAAAAAGATATAAAATCATAATA  
 Cc GAACCTCACACGCC TACAATTAGAACATGGATTAAAACACTTTCGAAATAATATAAAATAAAAGCA  
 \*\*\*.\*\*\*\*\*.\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
  
 20660 20670 20680 20690 20700 20710 20720  
 Gt AATCTAAAACACAAAAGACTTG AATTAAAAAAATTCTAATAATAAAAACAAATTAAATACTCTTAAAC  
 Cc AATCTAGAACTGAAAATACTTG AATTAAAAGAAATTCAAGAAATTAAAAGCAAAATTAAAGTAACTTTAA  
 \*\*\*.\*\*\*\*\*.\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
  
 20730 20740 20750 20760 20770 20780 20790  
 Gt TAAATCATGTACAGACCCCCAAC TACACCAAAATACAGAAAATAGTTAATAATGTTGTTAACAAAAGAAAA  
 Cc ATTGTTGGATAGAACCTCAACTAAC TAAAATACAAGAAAATTATAATAAAAGTGTGTTAACAAATATGAAA  
 .\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
  
 20800 20810 20820 20830 20840 20850 20860  
 Gt AACAAACGAAATACCAATCAATACCAATAGTATAATACAAATTAACTGGAAGAAAATTCAATCGTATGTACAC  
 Cc AACAAAGAAAATTCCGTCATTCACATTG TATAATAGAGATTGAAAAAGGAAATCAAGAAAAGTAGTTG  
 \*\*\*.\*\*\*\*\*.\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
  
 20870 20880 20890 20900 20910 20920 20930  
 Gt AAAACTGAAAATAAAGATGTACTATTATCAAAATTGAACTCACAAGAAATAAAGAAAATAAAAAAAGTTAAACCA  
 Cc AAAACTGAAAAAGGATGTTCCCTGAATTAAC TGAATCCACAAAAGTAAAGAAAAAGAAA TGTTAACAGCA  
 \*\*\*.\*\*\*\*\*.\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.



21640 21650 21660 21670 21680 21690 21700  
 Gt TACAAATCATGTCACACGCGTACGCTCTAACCTTTAAGCTATAAGACCCCAAGATAATTAAATTACTTT  
 Cc TATCAA-CATGTCACACGCGTACGCTCTAACCTTTAAGCTATAAGACTCTATGTAATTAACTAGATG  
 \*\*\*.\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
  
 21710 21720 21730 21740 21750 21760 21770  
 Gt TTTCTTATA---TGAAAAAAATTAAATACTGAATAAAAACAACTAATTGATTGATTACATGT---AA  
 Cc AAAAAAAAAGAGTGAAAAAGACTAATAAAATTAAAAGTAA---GATTA-ATATAACTATACACTAAAA  
 .....\*.\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
  
 21780 21790 21800 21810 21820 21830 21840  
 Gt TCATTAAGTAAACAGACACCTAACAGACAATAAAAATAAGTAATGATTAAATTGACCAGTTGTAACCT  
 Cc TAATTGAAAAGGAACAAAAGTAAAAATCATAAAAGAAAAGATAATGAGTTATTGACCAGTTGTTAGTT  
 \*.\*.\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
  
 21850 21860 21870 21880 21890 21900 21910  
 Gt TGAAAATAATTGATCATATAACTATTTTGATAAAACCATACGGACCAATAATTCAATAATCCT  
 Cc ATTAATAATTGAGACCAAAAGAAACAAATTAGTAAATCCGTAAAGGTCACACTCAATTCAATAAAACCA  
 .....\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
  
 21920 21930 21940 21950 21960 21970 21980  
 Gt CTATCTAAATTAAAGAAATAGTATAACCAAAATAAGAAATTAGGATATACAAAAATCTATTATATA  
 Cc CGATCTAAATTAAATGATATTGATAACCAAAAGTCAAAATTGGTAAACGATTAAACGATTATACA  
 \*.\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
  
 21990 22000 22010 22020 22030 22040 22050  
 Gt ATACATCGAAATATCACTTTTACTAATTAAAAACATATTTCAGATAAAAAT---GTATAT-AAAT-  
 Cc GCACATCCCATAACCACTTTTATTTATCAAGAAG-GTAAAAAACGATAATAAGTTGTAATTAGATT  
 ..\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
  
 22060 22070 22080 22090 22100 22110 22120  
 Gt ---AAGTATGCACTTT---GACTTATTAAACAATGGTGCTAAA-TAATCCCAATACTTAATGTAAA  
 Cc TAAATGAAAAAGTTTAAATTTGATACAAAACCTGCACAAAATATTCTAGTAAACCTTAATAAATAAA  
 ..\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
  
 22130 22140 22150 22160 22170 22180 22190  
 Gt CGGTAAATCATTTAAATTAACTTAACTCACTCAGCTCAATAAGTAAGAATTGTTAGGTAAAATAAGA  
 Cc AGGTAAATCATTTAAACTACAATAGGTAAATTCTTCAGCTCAATAAAAGTTGAATGTTAGGTAAAATAAA  
 ..\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
  
 22200 22210 22220 22230 22240 22250 22260  
 Gt ATAGCAGATTTCAAAAATCGGTACCTAAGCTATAAAATAATCTTTCCATAATAACCTATAAAAATAC  
 Cc ATTGCTGCACCTCAAAAATCGCTTCCGGAACCCACAAATAATCTCGTATTAGATAACCAGCATAATAC  
 \*.\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
  
 22270 22280 22290 22300 22310 22320 22330  
 Gt TTCTAAAGCTAAAAATAATTATGGTATTAAACATTAACAAAAGAAGGCTCATGACTTTTTAAATATT  
 Cc TACCAAATACTTAGTATAATTAGGAAAAATCATGAGTAACGAAGATTGAACTAAATTAAAGAGAACT  
 \*.\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.



	23040	23050	23060	23070	23080	23090	23100
Gt	CCTATTAAATGTTAACCTGAAAAATAAAAATCC	TGCAAAATTAAAAAATATTAGTAATAAGGAGCTA					
Cc	GAAATTAGAGTTAAAACATGTAATTAAAGCC	AAAAATGTTAGAAATTGATTATCGAATAAGGAAC					
	23110	23120	23130	23140	23150	23160	23170
Gt	AAGAAAAATACAGTAAAATACTACAGATTGAA	AAAATGTTGTAAGTATACTAAACTTAATC					
Cc	AGAAAAAAATTACAGAAATACTAACGGATTGAA	AAACTCAGAATACTCAAGGAATTGAAACTCAGA					
	23180	23190	23200	23210	23220	23230	23240
Gt	AAAATCTCCAATTCTATTAACTACTAACGCTT	AAATAGCTGATTGATTGCGGCCAGTC					
Cc	AAAATCACCACACTATTAAACAATTGCTTAA	TTGCTGATTAGCTGCTAACCGAGTATATCAA					
	23250	23260	23270	23280	23290	23300	23310
Gt	AAATTTATTAAACAAAAGATGCTAAACCTAC	ACCCCTCTCATCCTACAAACATTGAAACAGATT					
Cc	AGTTAAAGAAGATAAGATGCAAGACCAAC	TCCTCTCACCCCTAAACACATTGTATCAAATT					
	23320	23330	23340	23350	23360	23370	23380
Gt	CAGTTACTAAAACCAACATAAAAAGGTA	AAATTTCTAAAGGCCATAAAACGTGGTAAATGAG					
Cc	CTGTTACTAGTATTACATAAAAGTAGAAACTA	AACTAGAAACTAAATGTAATAACACTAGCAT					
	23390	23400	23410	23420	23430	23440	23450
Gt	ATTTTGCAATATTCTAAAGAGTAAATGTA	CTAAACTCGATATTACCGTAAACTACTAACATT					
Cc	ATTTTCCATGTATTATGGAATAAAATGAA	ACTAAACTAGAAACTAAATGTAATAACACTAGC					
	23460	23470	23480	23490	23500	23510	23520
Gt	ACAGTAAGACTATCGAAATAAAAATTC	CAAGAAACTTTAAATACCTACATTAACTCAAGAAG					
Cc	ACTGTAAGACTATCAAATAAAATCCTCAT	GAACATAAAAGCTCCAGATTAACTCAAGATAC					
	23530	23540	23550	23560	23570	23580	23590
Gt	TAATATAATGTTACTCCCATTAAACCTACT	TCATAAAATGCAGCGAGAGATAAAGCCATTGAG					
Cc	AAATAGTACAAGTAGTATTGTTAACCAATT	TCAAAATGCAAGCAAGAAAATAATGACGAAAC					
	23600	23610	23620	23630	23640	23650	23660
Gt	AACACAAACTATCGAAAATACATAGATCC	CTTGACCTACAAATCGGCCATAAAACCCAGAA					
Cc	TACACAAAGAGTTGAAAAGTGTGTTGAC	ACACGACAACCTAACCAACGCGACCTCCAAA					
	23670	23680	23690	23700	23710	23720	23730
Gt	GATCCTAGTAATGGTAACCAATAAGTAAG	TACATAAT-----AATTATTT-----ATTAAT					
Cc	GATCCCATAAAAGGTAAAAATAAGTA	AACTACATAATTATGAATTAGTTACTGGTTAAAGT					

23740 23750 23760 23770 23780 23790 23800  
**Gt** **AAATTC****AAATGATTAGGATACACTA**TATTGAATTAAAGAGTTTAAATCACA  
**Cc** **AAATTTAAGAGTTAGGATAAAAGAAA**TACAATTAAATTGAACATCA  
 .....|.....|.....|.....|.....|.....|.....|.....|.....|  
 23810 23820 23830 23840 23850 23860 23870  
**Gt** **AAAAATCGCAAAAC**TAATTTTTGATCAATTGATAAGTATCAATTTCCTCTATTCTATTAAATTGAG  
**Cc** **CGACTAACTGTACTT**ATTGGTATCAATTGGTAGATGTCTAGTT-TGTTAAAGTATTAGATATTGAA  
 .....|.....|.....|.....|.....|.....|.....|.....|  
 23880 23890 23900 23910 23920 23930 23940  
**Gt** **TTAAATGATCTGTTA**ATAAATTGAACTTAATTCAAGATTGGATTAAAGTTTCCTCAATAAAATT  
**Cc** **AATTCA**TGGCCTAACGAAATATTACCAACT--AAATTGTTAACAAAAGTTGTTAACAAAAGTT  
 .....|.....|.....|.....|.....|.....|.....|  
 23950 23960 23970 23980 23990 24000 24010  
**Gt** **TTTTACTTAATGTTG**CTTGTTTCTAAACTTCTTGAGTATTATATCTATAATTGTTACGTAT  
**Cc** **GCTTGT**TTATAATTAGAAGTTAATTAGAAAGCTCAGTGTCAATTGCCTCAATAATTGTTTCGAGA  
 .....|.....|.....|.....|.....|.....|.....|  
 24020 24030 24040 24050 24060 24070 24080  
**Gt** **TTTTAATGATTAAAGAAACATTGG**TAAGAAAATAATGTTAACACTACATAAAATAGTAAAAAAATTAA  
**Cc** **CTTC**AAAGATTTATAAAACTTTGGTAAAAAAATGAGTTAAAACAAATTATATTAAACAAAATAATA  
 .....|.....|.....|.....|.....|.....|.....|  
 24090 24100 24110 24120 24130 24140 24150  
**Gt** **AACAA**TCAAAAAATTGGAAAATACAATGATACGATCTAATTGAGGCATTCTTATTATATTAA  
**Cc** **AATAATC**AAAAAATTGAGAGAACTATAATACGGTCAATTGTTGGCATTTTATATT-TTGATTAAAT  
 .....|.....|.....|.....|.....|.....|.....|  
 24160 24170 24180 24190 24200 24210 24220  
**Gt** **GGATTCTAAACATCGTTAGATAAA**TACATGTTAAACGTAAAAACATAACGTTGAAACCTGCTAT  
**Cc** **GCATATCTAGAACATCGTTCAAGTAAATG**CATGTTAAAGGTAAAACGTAAGCTTGAAGACCAGCAAT  
 .....|.....|.....|.....|.....|.....|.....|  
 24230 24240 24250 24260 24270 24280 24290  
**Gt** **TGCTAATTCAAGACCTA**AAAGTGTTAATAGAAGACCTAAGGGTATTAAAGTAAAATTGCAAATAAAACCG  
**Cc** **GGCAAGTTCAAGTCCGGTAAAGAGCTAA**AAAAGAGCCAAAGGAATTAAATGAAAAATTGCTAATAAACCA  
 .....|.....|.....|.....|.....|.....|.....|  
 24300 24310 24320 24330 24340 24350 24360  
**Gt** **CCGGCCGATAACATGGTTCAAGC**AAAACCTGCAATAATTAAATATGTATGACCTGGAAGTCATATTG  
**Cc** **CCTGCTAAAGACATCGTTCAAGCAA**TCTGCATGTTTAATAAAAGTGTGGCCAGAAGTCATATTG  
 .....|.....|.....|.....|.....|.....|.....|  
 24370 24380 24390 24400 24410 24420 24430  
**Gt** **CAAAATACGTATAGATAAAAGTAAACACCTTAACTATATAAAGAAAAGAAA**TTCAATTGTAATTAAATAATTG  
**Cc** **CAAAATCTAATGGATAAGGTAAAAC**TTTTACGATATAAGATAGAAAATTCAATTGTTATAAGAAGAGG  
 .....|.....|.....|.....|.....|.....|.....|



25140 25150 25160 25170 25180 25190 25200  
 Gt GAAAAAAATCGAAAGA-----TGGAACTTTAGTTTGTAAATTATTTAGA-CTGTTGCA-----  
 Cc AATTGATTCGAAGAAAAATAGGAGTTTTTATTTAATTTCTGATTTAAAACACTAGTGAAGAAAATA  
 \*.\*.\*.\*\*\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.  
  
 25210 25220 25230 25240 25250 25260 25270  
 Gt -----TTAAAAGTACACTATGTAATAAAGTTCTAAAAGTAAAAAAATTATTGTCTCGAAT  
 Cc ATTTAATTGGGTTTAAAAAAATAGTAGGTTAAATAAGGGTCCAAAATGCTAAAATTATCCTTGCGAT  
 \*\*\*\*\*.\*\*\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.  
  
 25280 25290 25300 25310 25320 25330 25340  
 Gt AGTTACTTTAGTAATAAAGAACCTTTTTTTAGAGTTTCTGTTATGAATAAAAAAACTTTGGCA  
 Cc AGTTACTTTA-AACTATAAAATTGAAAATCTAGATTTACTTGAGCTTGTTAAATAAGATCGTTCTTG  
 \*\*\*\*\*.\*\*\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.  
  
 25350 25360 25370 25380 25390 25400 25410  
 Gt ACTTAAAGCTTCAGAAGTAAACTGTTAGCTATAGTTATTGAAAATTTC  
 Cc CGTTAAAAGTTC-AAATAAAATTGTTAGTGTAAATAATCATAATTGGAGGAAGAATTAAAGTTTAAATT  
 .\*\*\*\*\*.\*\*\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.  
  
 25420 25430 25440 25450 25460 25470 25480  
 Gt --ATAAAATAGTAAATAATTAACTTTAAATTAAAGATTTGATCTCTGAAAAATTATTGTCTA  
 Cc GGATAAAAATTCATTTAAATCTAAATGTATTAAATTGATAAGAAATTG-TACTCTAAAAATTGGTTGTCAA  
 .\*\*\*\*\*.\*\*\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.  
  
 25490 25500 25510 25520 25530 25540 25550  
 Gt AAGAACAAACTTCTATAAAAT-----TGCATTGATTATA-TAATCTAAGTTGAT-----TCAT  
 Cc CACCATCAAATTGTTAAATGATAAGACGAATTAGCTCTTGCTGGACATTGGATAAGTTTAT  
 .\*\*\*\*\*.\*\*\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.  
  
 25560 25570 25580 25590 25600 25610 25620  
 Gt TATTAATGTTAACGTTTGGAAAATAATAGGACTTGAACCTATAATTATGTATGCAAAACATATGTTT  
 Cc CATAATTGTTAAATGACTGGAAGTAATTGGACTTGAACCAATAATTATGTATGCAAAACATATGTTT  
 .\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
  
 25630 25640 25650 25660 25670 25680 25690  
 Gt ACCAAAATTAAACTATATTCCCT-TTTTCATTTGTGTGGAAAGAGAGTAGGATTGAAAC-CTACGAT  
 Cc ACCAAA-TTAAACTATACCCCCCTGTTTTTGTGGATTGATAAAATTACGTTAAAATATTTCAGT  
 .\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.  
  
 25700 25710 25720 25730 25740 25750 25760  
 Gt T-ATTCTTAATAAAATAGATT-ACAGTCTACCGCTTAAGGATTGAACCTACAATAATCAATTAGAA  
 Cc TTGAATCTAAATCCTTAAATTAGATAACTTGCCTTGAAGGGTTCGAACCTCAATATTCAATTAGAA  
 .\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*.  
  
 25770 25780 25790 25800 25810 25820 25830  
 Gt GATTGATGTTTATCCG-TTAAACTAAAAACGC-TCACA-AGCTTGAGAGATCATAAGAACT-  
 Cc GATTGACGTTTATCCCTAAATTAAACTAAAAACGCAGTTATATAATCTGAAGAGAGTAGGATTGCG  
 .\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.

