

UNIVERSIDADE FEDERAL DO RECÔNCAVO DA BAHIA  
CENTRO DE CIÊNCIAS AGRÁRIAS, AMBIENTAIS E BIOLÓGICAS

USO DE FILOGENIA MOLECULAR, GENÔMICA COMPARATIVA E REDES  
COMPLEXAS PARA RESOLVER RELAÇÕES EVOLUTIVAS ENTRE FtsHs DO FILO  
CYANOBACTERIA

LENON LIMA DE SANTANA  
Bacharel em Biologia

CRUZ DAS ALMAS  
BAHIA - BRASIL  
2016

LENON LIMA DE SANTANA

USO DE FILOGENIA MOLECULAR, GENÔMICA COMPARATIVA E REDES  
COMPLEXAS PARA RESOLVER RELAÇÕES EVOLUTIVAS ENTRE FtsHs DO FILO  
CYANOBACTERIA

Trabalho de Conclusão de Curso  
apresentado à Universidade Federal do  
Recôncavo da Bahia, como parte das  
exigências do Curso de Graduação de  
Bacharelado em Biologia, para obtenção  
do título de Bacharel em Biologia.

CRUZ DAS ALMAS  
BAHIA - BRASIL  
2016

LENON LIMA DE SANTANA

USO DE FILOGENIA MOLECULAR, GENÔMICA COMPARATIVA E REDES  
COMPLEXAS PARA RESOLVER RELAÇÕES EVOLUTIVAS ENTRE FtsHs DO FILO  
CYANOBACTERIA

Trabalho de Conclusão de Curso  
apresentado à Universidade Federal do  
Recôncavo da Bahia, como parte das  
exigências do Curso de Graduação de  
Bacharelado em Biologia, para obtenção  
do título de Bacharel em Biologia.

APROVADO: 08 de agosto de 2016

---

Flora Maria de Campos Fernandes – DS  
UNIVERSIDADE FEDERAL DA BAHIA

---

Patrícia Luz Ribeiro – DS UNIVERSIDADE  
FEDERAL DO RECÔNCAVO BAHIA

---

Phellippe Arthur Santos Marbach – DS  
Orientador  
UNIVERSIDADE FEDERAL DO RECÔNCAVO DA BAHIA

*Dedico aos meus pais: Joseval Santana e Miriam Santana, pois sem eles eu não seria nada e a minha avó Augusta Lima (in memoriam).*

## AGRADECIMENTOS

Agradeço primeiramente a Universidade Federal do Recôncavo da Bahia, por me proporcionar a oportunidade de ampliar meus conhecimentos, a toda administração e corpo docente que colaboraram com o meu desenvolvimento na instituição.

Ao meu querido orientador Phellippe Arthur Santos Marbach, por esses anos de convívio, paciência, por ter acreditado em mim, pelos conselhos e ensinamentos que me ajudaram muito, por todo esforço e dedicação, pela amizade, sei que levarei tudo isso e muito mais comigo para o resto da vida.

Aos meus pais, Joseval Santana e Miriam Santana, por estarem sempre comigo me ensinando a cada dia que podemos melhorar, por terem a paciência necessária, o carinho e por serem meus principais alicerces na vida, todo esforço para me proporcionar não o melhor nem o pior, mas o necessário.

Aos meus irmãos, por todo suporte e pelas longas conversas que tivemos além dos conselhos e da amizade.

Aos meus amigos, que sempre me auxiliaram e me deram ânimo para continuar todas as vezes que encontrei um obstáculo, pelos risos, choros, e mesmo com as reviravoltas que a vida pode dar guardarei vocês sempre no meu coração.

Aos meus colegas/amigos de trabalho em especial a Dennifier e o Marcio que sempre estava lá para me ajudar quando “o bixo pegava”, contribuindo no meu crescimento profissional e pessoal ao longo da minha vida acadêmica.

Agradeço ao Daniel e toda equipe da UFBA que me ajudaram na construção deste trabalho  
A todos o meu muito obrigado!

## LISTA DE SÍMBOLOS E ABREVIATURAS OU LISTA DE FIGURAS E TABELAS

NCBI – *National Center for Biotechnology Information*

FtsH - Filamentous Temperature Sensitive H

BLAST – Basic Local Alignment Search Tool

THG – Transferência Horizontal de Genes

SRH – Segunda Região de Homologia

MEGA - *Molecular Evolutionary Genetics Analysis*

ARL – Atração de Ramos Longos

## LISTA DE FIGURAS E TABELAS

Figura 1. Número de cópias de ftsH em espécies do Filo Cyanobacteria .....	11
Figura 2. Resultados da árvore filogenética e da abordagem de Redes Complexas para inferir grupos de ortólogos da FtsH em cianobactérias.....	12
Figura 3. Árvore filogenética não enraizada de FtsH bacterianas.....	15
Figura 4. Padrão de <i>indels</i> indicado nas diferentes isoformas da FtsH de cianobactérias.....	17
Figura 5. Árvore filogenética enraizada das FtsHs do Filo Cyanobacteria contendo apenas espécies com o genoma completamente sequenciado.....	18
Figura 6. Análises filogenéticas das FtsHcb1.....	23
Quadro 1. Distribuição dos parálogos da FtsH nos táxons do Filo Cyanobacteria.....	24
Figura 7. Análises filogenéticas das FtsHcb1.....	25
Tabela 1: FtsHs codificadas no genoma de <i>Synechocystis</i> sp. PCC 6803.....	26
Figura 8. Gráfico de identidade média (%) entre as isoformas do gene que codificam a FtsH em Cyanobacteria.....	27
Figura 9. Conservação dos motivos das FtsHs do Filo Cyanobacteria.....	29
Figura 10. História evolutiva proposta para as FtsHs do Filo Cyanobacteria.....	31

## RESUMO

LIMA DE SANTANA, LENON, Bacharel em Biologia, Universidade Federal do Recôncavo da Bahia, fevereiro de 2016. Evolução das FtsHs do Filo Cyanobacteria. Orientador: Phellippe Arthur Santos Marbach.

**Palavras-chave:** Evolução de proteína, ortólogos, Fotossíntese, HflB, análise genômica

**Introdução.** Espécies do Filo Cyanobacteria possuem duas ou mais cópias do gene que codifica a FtsH (*filamentation temperature sensitive*). Estas proteínas estão envolvidas com o metabolismo fotossintético e estudos anteriores mostraram que elas são filogeneticamente e funcionalmente relacionada com as FtsHs plastidiais. Apesar da sua importância, a história evolutiva das FtsHs de cianobactérias ainda deve ser elucidada.

**Objetivo geral.** Compreender as relações evolutivas entre as FtsHs codificadas nos genomas das cianobactérias.

**Métodos.** As relações filogenéticas entre as FtsHs codificadas em 66 genomas de cianobactérias completamente sequenciados disponíveis no NCBI (*National Center for Biotechnology Information*) foram investigadas usando a filogenia molecular, genômica comparativa e construção de redes complexas a partir da matriz de identidade de proteínas, um método recentemente desenvolvido para fazer inferências filogenéticas.

**Resultados.** O número de cópias de genes que codificam FtsH nos genomas de cianobactérias analisados variou de duas a nove cópias. Os resultados apresentados neste trabalho indicam que existem pelo menos sete parálogos da FtsH nas cianobactérias. Os parálogos FtsHcb1A, FtsHcb1B e FtsHcb1C derivam de uma FtsH ancestral oriunda de um evento de transferência horizontal (THG) de genes no ancestral das cianobactérias, em contraste com os parálogos FtsHcb2A, FtsHcb2B, FtsHcb3 e FtsHcb4, que derivam de eventos de duplicação gênica da FtsH nativa do ancestral das cianobactérias. Os genes que codificam FtsHcb2A, FtsHcb2B e FtsHcb3 estão amplamente distribuídas nos genomas das cianobactérias, contudo, a FtsHcb3 é o único parálogo que está presente em todos os genomas de cianobactérias analisados. Foi observado uma coocorrência de genes que codificam as FtsHcb1C, FtsHcb2A, FtsHcb3 e FtsHcb4, isto pode indicar algum tipo de interação funcional entre estes parálogos. A FtsHcb3 é o parálogo mais conservado, em contraste com a FtsHcb4. Os parálogos da FtsHcb4 codificados nos genomas de cepas de *Prochlorococcus marinus* possuem um motivo de ligação a zinco atípico, NEVGR, não descrito ainda para FtsHs, o que sugere uma divergência funcional deste parálogo nesta espécie.

**Conclusão.** Os resultados apresentados aqui mostram que ambos, a THG e eventos de duplicação gênica, modelaram a diversidade das FtsHs no Filo Cyanobacteria. Adicionalmente, estes resultados mostram que redes complexas obtidas a partir da matriz de identidade de genes e proteínas são ferramentas úteis na definição das relações de homologia entre genes ou proteínas.

## ABSTRACT

LIMA DE SANTANA, LENON, Bacharel em Biologia, Universidade Federal do Recôncavo da Bahia, fevereiro de 2016. Evolution of the FtsH from phylum Cyanobacteria. Advisor: Phellippe Arthur Santos Marbach.

**Key-words:** protein evolution, orthologous, photosynthesis, HfIB, genomic analysis

**Introduction.** Species of the Phylum Cyanobacteria possessing two or more copies of genes encoding FtsH (filamentation temperature sensitive). These proteins are involved in photosynthetic metabolism and previous studies showed that they are phylogenetically and functionally related to plastidial FtsHs. Despite its importance, the evolutionary history of cyanobacteria FtsHs remain to be elucidated.

**General objective.** Understanding the evolutionary relationships between FtsHs encoded in the genomes of cyanobacteria.

**Methods.** Phylogenetic relationships between FtsHs encoded in 66 completely sequenced cyanobacterial genomes available at NCBI (National Center for Biotechnology Information) they were investigated using molecular phylogeny, comparative genomics and construction of complex networks from the identity matrix proteins, a method recently developed to make phylogenetic inferences.

**Results.** The number of gene copies encoding the cyanobacterial FtsH genomes examined varies from two to nine copies. The results presented here indicate that there are at least seven paralogous of FtsH in cyanobacteria. The paralogous FtsHcb1A, FtsHcb1B and FtsHcb1C derive from an ancestral FtsH originating from a horizontal transfer event (THG) genes in the ancestor of cyanobacteria, in contrast to the paralogous FtsHcb2A, FtsHcb2B, FtsHcb3 and FtsHcb4, that derive from gene duplication events of native ancestral FtsH of cyanobacteria. The genes encoding FtsHcb2A, FtsHcb2B and FtsHcb3 are widely distributed in the cyanobacterial genomes, however, FtsHcb3 is the only paralogue that is present in all the genomes analyzed. It was observed a co-occurrence of genes encoding FtsHcb1C, FtsHcb2A, FtsHcb3 and FtsHcb4, this may indicate some sort of functional interaction between these paralogous. The FtsHcb3 is the most conserved paralogue, in contrast to FtsHcb4. The paralogous of FtsHcb4 encoded in *Prochlorococcus marinus* genomes have an atypical zinc-binding motif, NEVGR, it was not described yet for FtsH, suggesting a functional divergence of the paralogue in this species.

**Conclusion.** The results presented here show that both THG and gene duplication events have shaped the diversity of FtsHs in Filo Cyanobacteria. Additionally, these results show that complex networks obtained from the identity matrix genes and proteins are useful tools in defining the homology relationships between genes or proteins.



## ÍNDICE

Artigo

### **USO DE FILOGENIA MOLECULAR, GENÔMICA COMPARATIVA E REDES COMPLEXAS PARA RESOLVER RELAÇÕES EVOLUTIVAS ENTRE FtsHS DO FILO CYANOBACTERIA**

Resumo .....	01
Abstract .....	02
Introdução .....	03
Material e métodos .....	06
Resultados e Discussão.....	10
Considerações Finais.....	32
Referências.....	33

## **Uso de filogenia molecular, genômica comparativa e redes complexas para resolver relações evolutivas entre FtsHs do Filo Cyanobacteria.**

Lenon Lima de Santana, Dennifier Costa Brandão Cruz, Daniel Santana de Carvalho, Roberto Fernandes Silva Andrade, Charbel Ninõ El-Hani, Phellippe Arthur Santos Marbach

**Introdução.** Espécies do Filo Cyanobacteria possuem duas ou mais cópias do gene que codifica a FtsH (*filamentation temperature sensitive*). Estas proteínas estão envolvidas com o metabolismo fotossintético e estudos anteriores mostraram que elas são filogeneticamente e funcionalmente relacionada com as FtsHs plastidiais. Apesar da sua importância, a história evolutiva das FtsHs de cianobactérias ainda deve ser elucidada.

**Objetivo geral.** Compreender as relações evolutivas entre as FtsHs codificadas nos genomas das cianobactérias.

**Métodos.** As relações filogenéticas entre as FtsHs codificadas em 66 genomas de cianobactérias completamente sequenciados disponíveis no NCBI (*National Center for Biotechnology Information*) foram investigadas usando a filogenia molecular, genômica comparativa e construção de redes complexas a partir da matriz de identidade de proteínas, um método recentemente desenvolvido para fazer inferências filogenéticas.

**Resultados.** O número de cópias de genes que codificam FtsH nos genomas de cianobactérias analisados variou de duas a nove cópias. Os resultados apresentados neste trabalho indicam que existem pelo menos sete parálogos da FtsH nas cianobactérias. Os parálogos FtsHcb1A, FtsHcb1B e FtsHcb1C derivam de uma FtsH ancestral oriunda de um evento de transferência horizontal (THG) de genes no ancestral das cianobactérias, em contraste com os parálogos FtsHcb2A, FtsHcb2B, FtsHcb3 e FtsHcb4, que derivam de eventos de duplicação gênica da FtsH nativa do ancestral das cianobactérias. Os genes que codificam FtsHcb2A, FtsHcb2B e FtsHcb3 estão amplamente distribuídas nos genomas das cianobactérias, contudo, a FtsHcb3 é o único parálogo que está presente em todos os genomas de cianobactérias analisados. Foi observado uma coocorrência de genes que codificam as FtsHcb1C, FtsHcb2A, FtsHcb3 e FtsHcb4, isto pode indicar algum tipo de interação funcional entre estes parálogos. A FtsHcb3 é o parálogo mais conservado, em contraste com a FtsHcb4. Os parálogos da FtsHcb4 codificados nos genomas de cepas de *Prochlorococcus marinus* possuem um motivo de ligação a zinco atípico, NEVGR, não descrito ainda para FtsHs, o que sugere uma divergência funcional deste parálogo nesta espécie.

**Conclusão.** Os resultados apresentados aqui mostram que ambos, a THG e eventos de duplicação gênica, modelaram a diversidade das FtsHs no Filo Cyanobacteria. Adicionalmente, estes resultados mostram que redes complexas obtidas a partir da matriz de identidade de genes e proteínas são ferramentas úteis na definição das relações de homologia entre genes ou proteínas.

**Palavras-chave:** Evolução de proteína, ortólogos, Fotossíntese, HflB, análise genômica, parálogos.

## Using of molecular phylogeny, comparative genomics and complex networks to resolve evolutionary relationship between FtsHs of the Phylum Cyanobacteria.

### Abstract

**Introduction.** Species of the Phylum Cyanobacteria possessing two or more copies of genes encoding FtsH (filamentation temperature sensitive). These proteins are involved in photosynthetic metabolism and previous studies showed that they are phylogenetically and functionally related to plastidial FtsHs. Despite its importance, the evolutionary history of cyanobacteria FtsHs remain to be elucidated.

**General objective.** Understanding the evolutionary relationships between FtsHs encoded in the genomes of cyanobacteria.

**Methods.** Phylogenetic relationships between FtsHs encoded in 66 completely sequenced cyanobacterial genomes available at NCBI (National Center for Biotechnology Information) they were investigated using molecular phylogeny, comparative genomics and construction of complex networks from the identity matrix proteins, a method recently developed to make phylogenetic inferences.

**Results.** The number of gene copies encoding the cyanobacterial FtsH genomes examined varies from two to nine copies. The results presented here indicate that there are at least seven paralogous of FtsH in cyanobacteria. The paralogous FtsHcb1A, FtsHcb1B and FtsHcb1C derive from an ancestral FtsH originating from a horizontal transfer event (THG) genes in the ancestor of cyanobacteria, in contrast to the paralogous FtsHcb2A, FtsHcb2B, FtsHcb3 and FtsHcb4, that derive from gene duplication events of native ancestral FtsH of cyanobacteria. The genes encoding FtsHcb2A, FtsHcb2B and FtsHcb3 are widely distributed in the cyanobacterial genomes, however, FtsHcb3 is the only paralogue that is present in all the genomes analyzed. It was observed a co-occurrence of genes encoding FtsHcb1C, FtsHcb2A, FtsHcb3 and FtsHcb4, this may indicate some sort of functional interaction between these paralogous. The FtsHcb3 is the most conserved paralogue, in contrast to FtsHcb4. The paralogous of FtsHcb4 encoded in *Prochlorococcus marinus* genomes have an atypical zinc-binding motif, NEVGR, it was not described yet for FtsH, suggesting a functional divergence of the paralogue in this species.

**Conclusion.** The results presented here show that both THG and gene duplication events have shaped the diversity of FtsHs in Filo Cyanobacteria. Additionally, these results show that complex networks obtained from the identity matrix genes and proteins are useful tools in defining the homology relationships between genes or proteins.

**Key-words:** protein evolution, orthologous, photosynthesis, HflB, genomic analysis, paralogous.

## INTRODUÇÃO

A FtsH é uma metaloprotease de membrana dependente de ATP que pertence à Família AAA (ATPases Associadas a diversas Atividades celulares). Os membros dessa família possuem o domínio AAA que contém os motivos conservados Walker A, Walker B e SRH (Second Region of Homology) [1]. O motivo Walker A possui um resíduo de lisina que interage com um nucleotídeo trifosfatado, normalmente o ATP ou GTP, e tem como característica o padrão GxxxxGK (T/S). O Walker B consiste em dois resíduos de ácidos precedidos por quatro resíduos hidrofóbicos uuuu (D/E) (D/E) [2].

A FtsH foi caracterizada inicialmente em um mutante *fts* (*Filamentation temperature sensitive*) de *Escherichia coli*. Mutantes *fts* formam filamentos celulares em altas temperaturas [3]. Ela possui na região N-terminal um ou dois segmentos transmembrana e o domínio AAA e, na região C-terminal, o domínio proteásico M41 peptidase, com o motivo de ligação de zinco HExxH [4, 5].

A FtsH é usada como assinatura molecular em estudos sobre relações filogenéticas entre taxas bacterianas devido a sua ampla distribuição nas espécies desse Domínio [6, 7]. Esta proteína também é encontrada em eucariotos, e análises filogenéticas indicam que elas são oriundas de eventos de transferência horizontal de genes (THG) entre os endossimbiontes que originaram as mitocôndrias e os plastídeos e a célula eucarionte primitiva [8]. Até o momento, não foram encontrados genes que codificam FtsH em espécies do Domínio Archaea [9].

A estrutura quaternária das FtsHs é hexamérica e pode ser composta pelas mesmas subunidades ou subunidades diferentes, formando um homo ou hetero-hexamero, respectivamente [10, 11, 12, 13, 14]. Em geral, as FtsHs são proteínas essenciais e destacam-se por serem as únicas proteínas de membrana conhecidas que possuem atividades de chaperona e de protease, as demais proteínas conhecidas que desempenham ambas funções são citosólicas [15]. Elas estão envolvidas em diversas atividades celulares como resposta ao estresse, divisão e diferenciação celular, fotossíntese, degradação de fatores transcricionais e no controle da biossíntese de lipídios de membrana [1, 16].

A maioria das espécies procarióticas cuja FtsH foi caracterizada possui apenas uma cópia do gene que codifica esta proteína no genoma [17]. Entretanto existem exceções, por exemplo, *Synechocystis* sp. PCC 6803, que possui quatro FtsHs codificadas no seu genoma.

Elas diferem quanto à localização subcelular e desempenham funções relacionadas com a fotossíntese oxigênica [18, 19, 20, 21, 22, 23, 24]. As FtsHs plastidiais estão envolvidas na gênese dos plastídeos e fotossíntese. Além disso, análises evolutivas indicam que elas foram adquiridas por eventos de transferência horizontal da cianobactéria ancestral endossimbionte que originou os plastídeos [9, 25, 26].

Análises de genômica comparativa realizadas previamente (dados não mostrados) revelaram que ter múltiplas cópias de genes que codificam FtsH no genoma não é uma característica apenas de *Synechocystis* sp. PCC 6803. Isto parece ser uma característica das espécies do Filo Cyanobacteria, apesar do número de cópias por genoma ser variável entre as espécies analisadas (dados não mostrados).

A maioria das FtsHs de *Synechocystis* sp. PCC 6803 está funcionalmente relacionada com a fotossíntese oxigênica, uma invenção evolutiva desse grupo de bactérias que modelou a vida na terra no passado. Portanto, o conhecimento sobre a evolução das FtsHs deste filo bacteriano pode ajudar a expandir o conhecimento sobre a evolução desse metabolismo fotossintético. A elucidação da história evolutiva das FtsHs de cianobactérias ajudaria a definir os eventuais grupos de ortólogos, parálogos e xenólogos das FtsHs deste filo bacteriano [27].

Genes ortólogos tendem a ser funcionalmente relacionados, portanto, esse conhecimento é valioso do ponto de vista evolutivo e aplicado. Por exemplo, permitiria inferir quais FtsHs codificadas nos genomas sequenciados de cianobactérias podem ser funcionalmente relacionadas com as FtsHs de *Synechocystis* sp. PCC 6803. Esta informação é útil para uma anotação mais acurada das FtsHs codificadas nos genomas sequenciados das espécies desse filo bacteriano.

Infelizmente, as relações evolutivas entre as FtsHs codificadas nos genomas das cianobactérias ainda não foram elucidadas e isso pode ser uma tarefa difícil por dois motivos. Primeiro por que as relações evolutivas entre os filios bacterianos ainda não foram completamente resolvidas e, segundo, por que ainda não é conhecida a extensão da influência da THG na diversidade das FtsHs codificadas nos genomas bacterianos. Isso dificulta a escolha de modo não arbitrário um grupo taxonômico bacteriano cujas FtsHs possam ser usadas como grupo externo na filogenia molecular das FtsHs do Filo Cyanobacteria.

Recentemente Andrade et al. [28] propuseram um novo método para inferir sobre relações evolutivas entre proteínas usando redes complexas construídas a partir da matriz de identidade das proteínas. Redes complexas são conjuntos de grafos constituídos por nós conectados por arestas em que os nós representam os objetos em estudo e as arestas as relações entre eles. Essa abordagem é utilizada para definir relações entre objetos de estudos variados, indo desde relações sociais até estudos de relações biológicas [29]. Nas redes complexas obtidas a partir da matriz de identidade das proteínas os nós representam as proteínas e as arestas a identidade entre elas.

Este método de inferência filogenética possui baixo custo computacional e representa relações evolutivas coerentes [28, 30]. Carvalho et al. [30] usaram essa abordagem para inferir sobre a origem evolutiva das mitocôndrias, usando redes complexas geradas a partir da matriz de identidade de subunidades da ATP sintase de eucariotos e seus homólogos de espécies da Classe Alphaproteobacteria. O modo como as redes complexas são estruturadas indica que elas também podem ser úteis para definir relações de homologia entre genes e proteínas, porém o seu potencial para isso ainda não foi explorado.

Diante do exposto, o objetivo deste trabalho foi inferir sobre a história evolutiva das FtsHs do Filo Cyanobacteria. Para atingir este objetivo foram realizadas análises genômicas, filogenia molecular, análises de redes complexas e análises da estrutura primária das FtsHs de cianobactérias. Em conjunto, estes resultados permitiram elucidar eventos evolutivos que modelaram a evolução das FtsHs no Filo Cyanobacteria.

## MATERIAL E MÉTODOS

### Busca por genes que codificam a FtsH nos genomas procarióticos

Inicialmente foi criada uma lista das espécies com genomas completamente sequenciados até janeiro de 2015, disponíveis na base de dados do *National Center for Biotechnology Information* (NCBI). Cepas identificadas apenas em nível de gênero foram consideradas como espécies. Posteriormente, foram feitas buscas por genes que codificam homólogos da FtsH nos genomas dessas espécies utilizando os algoritmos BLASTP e TBLASTN, e a sequência de aminoácidos da FtsH de *Escherichia coli* (GI: 146028) como *query*. Nas análises com o TBLASTN foram usados os parâmetros de pontuação padrão do programa e nas análises com o BLASTP foi usada a matriz de substituição BLOSUM62.

Durante as análises com o BLASTP e TBLASTN foram recuperadas todas as sequências que possuísem o domínio AAA com os motivos Walker A, Walker B e SHR e o domínio proteásico M41 peptidase. Esta análise foi feita em cada genoma, individualmente, para evitar qualquer erro como recuperar sequências oriundas de diferentes anotações do mesmo gene. Durante as buscas realizadas com BLASTP e TBLASN também foi registrado o número de cópias de genes codificando homólogos da FtsH em cada genoma analisado. Essa informação foi utilizada, posteriormente, para analisar a distribuição taxonômica do número de cópias dos genes que codificam FtsH nos genomas bacterianos.

### Obtenção e edição dos alinhamentos múltiplos

O alinhamento múltiplo das sequências de nucleotídeos foi feito no GUIDANCE SERVER [31], com o algoritmo MAFFT, e os parâmetros padrões do programa. Posteriormente, o alinhamento múltiplo de nucleotídeos foi traduzido para aminoácidos e corrigido manualmente no programa MEGA 6.06 [32].

### Análises filogenéticas

O programa MEGA 6.06 [32] foi usado para realizar a escolha do melhor modelo evolutivo e da matriz de substituição empregados nas análises filogenéticas. Foram escolhidos modelos e matrizes de substituição com menor valor de escore BIC (*Bayesian Information Criterion*) e AICc (*Akaike Information Criterion, corrected*). Todas as análises filogenéticas foram realizadas no programa MEGA 6.06 com posições não ambíguas dos alinhamentos múltiplos das sequências de nucleotídeos e de aminoácidos das FtsHs usando o método da máxima verossimilhança. A confiabilidade da topologia das árvores obtidas foi avaliada

usando o método de *bootstrap*, com 1000 reamostragens, e o método heurístico *Nearest-Neighbor-Interchange* foi usado na busca pela árvore filogenética mais verossímil. Todas as árvores filogenéticas obtidas foram visualizadas no programa Archaeopterix [33].

### **Busca por *indels* e motivos conservados nas sequências de aminoácidos das FtsHs**

As sequências do alinhamento múltiplo das FtsHs de cianobactérias foram reordenadas de acordo com a topologia dos clados da árvore filogenética não enraizada obtida com esse alinhamento múltiplo. Em seguida, esse alinhamento múltiplo foi usado para realizar buscas por *indels* e motivos compartilhados pelas FtsHs de cada grupo filogenético definido nas análises filogenéticas.

### **Cálculo das matrizes de identidades das FtsHs de cianobactéria**

Inicialmente, foi calculada a matriz par a par de distância  $p$  a partir dos alinhamentos múltiplo das sequências de nucleotídeos e de aminoácidos das FtsH de cianobactérias no programa MEGA 6.06. A distância  $p$  é baseada na proporção de posições em que duas sequências diferem. As matrizes de distância genética foram convertidas em matrizes de identidades usando a fórmula:

$ID_{(ij)} = 1 - DG_{(ij)} \times 100$ ; em que  $ID_{(ij)}$  é a identidade entre as sequências  $i$  e  $j$ ,  $DG_{(ij)}$  é a distância genética entre as sequências  $i$  e  $j$  e o valor de identidade obtido é expresso em porcentagem.

### **Construção das redes complexas das FtsHs de cianobactérias**

Depois de gerar uma matriz de identidade simétrica, foi construído um conjunto de redes de acordo com o parâmetro  $\sigma$ , chamado de limiar de similaridade, em que cada nó representa uma sequência do conjunto de dados, enquanto as arestas são baseadas nos índices de semelhança entre todos os pares de sequências. Para um dado valor de  $\sigma$ , dois nós da rede (por exemplo,  $i$  e  $j$ ) foram ligados por uma aresta apenas quando as sequências correspondentes possuíam um índice de semelhança maior do que um limiar de identidade crítico  $\sigma S_{ij}$ . Estas ligações foram representadas por uma matriz de adjacência  $M(\sigma)$  de tal modo que os elementos de matriz  $M_{ij}$  são 1 ou 0 dependendo se os nós  $i$  e  $j$  são ligados por uma aresta (representando 1 na  $M(\sigma)$ ) ou não (representando 0 na  $M(\sigma)$ ). As redes foram construídas para valores  $\sigma$  variando de 0 a 100% de identidade. Em seguida, as matrizes de



vizinhança (NM), indicadas pelo símbolo  $\hat{M}(\sigma)$ , foram construídas para cada  $M(\sigma)$  [28, 29]. As matrizes  $M(\sigma)$  foram utilizadas para calcular a distância de rede de  $\delta(\sigma, \sigma + \Delta\sigma)$  entre os 101 pares de redes sucessivas (i.e.,  $\Delta\sigma = 1$ ). Aqui, a função  $\delta$  é definida para quaisquer duas redes  $\alpha$  e  $\beta$  com o mesmo número  $N$  de nós, respectivamente representadas pelas NM's  $\hat{M}(\alpha)$  e  $\hat{M}(\beta)$ , pela expressão

$$\delta(\alpha, \beta) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \left( \frac{\hat{m}_{i,j}(\alpha)}{D(\alpha)} - \frac{\hat{m}_{i,j}(\beta)}{D(\beta)} \right)^2,$$

em que  $D$  indica o diâmetro da rede, a maior distância mínima entre dois nós da rede [30].

Este procedimento revela, pela presença de picos, os valores de  $\sigma$  no qual a rede sofre uma mudança estrutural relevante. Em tais valores, uma suposta estrutura modular pode ser identificada com a maior quantidade de informações relativas recuperáveis, as relações evolutivas entre os organismos representadas no conjunto de dados. Para este estudo, os resultados foram baseados principalmente na rede no valor  $\sigma_{\max}$ , correspondente ao valor de  $\sigma$  para os quais  $\delta(\sigma, \sigma + \Delta\sigma)$  atinge o seu maior valor, porque neste valor será formada a estrutura modular mais informativa para o desenvolvimento do trabalho [34].

Para se obter um melhor entendimento da estrutura modular da rede, é importante investigar as redes geradas para valores  $\sigma$  imediatamente antes de  $\sigma_{\max}$ , pois essas estruturas podem ser bem informativas, bem como as redes associadas a outros picos significativos de  $\sigma$ . As redes foram visualizadas com a ajuda do pacote GePhi [35]. Os valores de similaridade são analisados próximos ao pico crítico, pois valores muito baixos podem não revelar estrutura modular e valores muito altos podem formar redes com pouca interação entre os nós, tornando-se pouco informativo.

### **Conservação dos grupos de FtsHs ortólogas das cianobactérias**

A conservação dos grupos de FtsHs de cianobactérias foi analisada calculando a identidade média entre as sequências de FtsHs ortólogas e entre os grupos de FtsHs parálogas definidos previamente nas análises evolutivas. Os valores de identidades usados nesta análise foram obtidos conforme a metodologia descrita anteriormente para o cálculo da matriz de identidade. A identidade média entre as sequências de FtsHs ortólogas foi calculada a partir

das identidades entre todos os pares de sequências de cada grupo de FtsHs ortólogas. A identidade média entre cada dois grupos de parálogos foi calculada a partir apenas das identidades entre os pares de sequências parálogas.

### **Obtenção dos LOGOS dos motivos das FtsHs ortólogas**

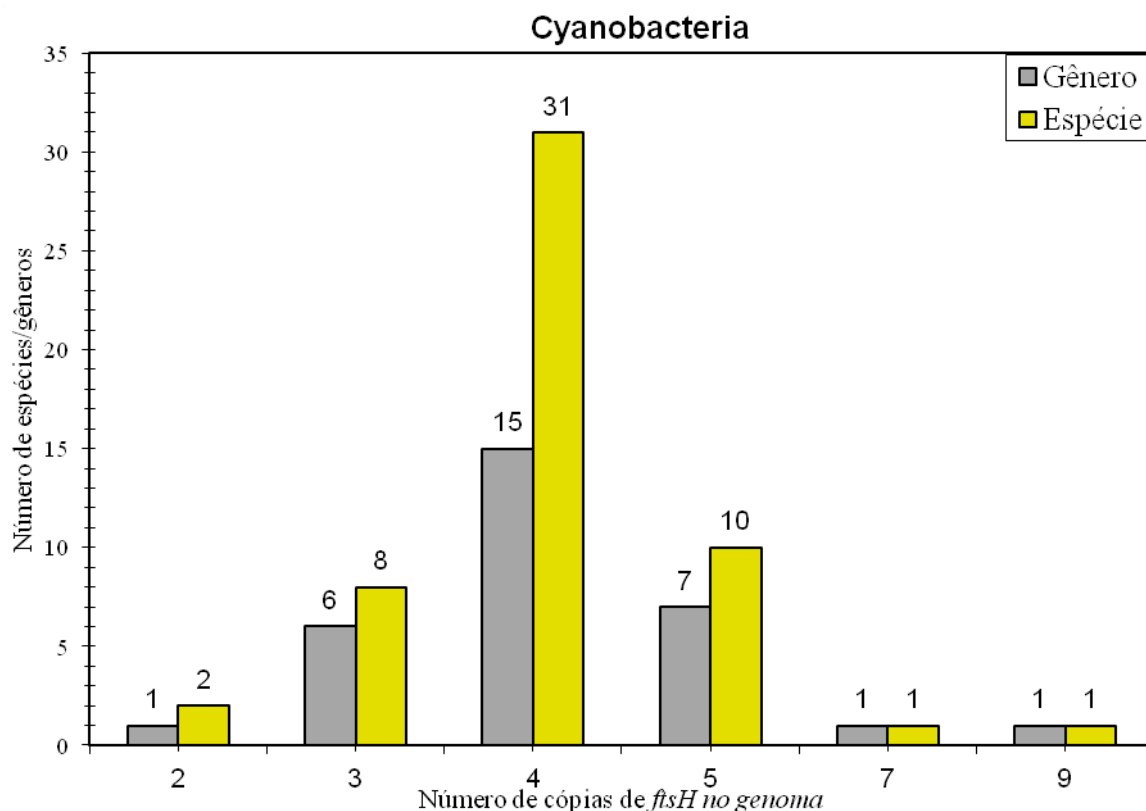
O programa WEB LOGO foi utilizado para analisar os padrões e a conservação dos aminoácidos nos motivos das FtsHs de cianobactéria. As sequências logo representam graficamente a frequência de nucleotídeos ou de aminoácidos em um determinado sítio do alinhamento múltiplo. Essa frequência varia numa escala de 0 a 4, em que o 4 representa 100% de frequência.

## RESULTADOS E DISCUSSÃO

Foram recuperadas sequências de nucleotídeos de 267 homólogos da *ftsH* em 66 genomas completamente sequenciados de 55 espécies do Filo Cyanobacteria, que representam seis ordens do filo. O número de cópias de genes que codificam FtsH nos genomas de espécies do Filo Cyanobacteria analisados variou de duas, no genoma de *Synechococcus* sp. JA-2-3B'a (2-13), a nove cópias encontradas no genoma de *Acaryochloris marina* MBIC11017, porém a maioria das espécies analisadas possui quatro cópias do gene que codifica FtsH no genoma (Figura 1).

A FtsH é uma protease com atividade de chaperona que fica ancorada em membranas. [15]. Em eucariotos existem várias isoformas de FtsHs mitocondriais e plastidiais que diferem quanto ao sub-compartimento organelar em que estão presentes. Por exemplo, os domínios ATPásicos e proteólíticos da FtsH mitocondrial YME1 estão localizados no espaço intermembrana enquanto que os domínios ATPásicos e proteólíticos da YTA10 e YTA 11 estão localizados na matriz mitocondrial [10]. Além disso, YTA10 e YTA11 formam estruturas quaternárias homohexaméricas ou heterohexaméricas [11].

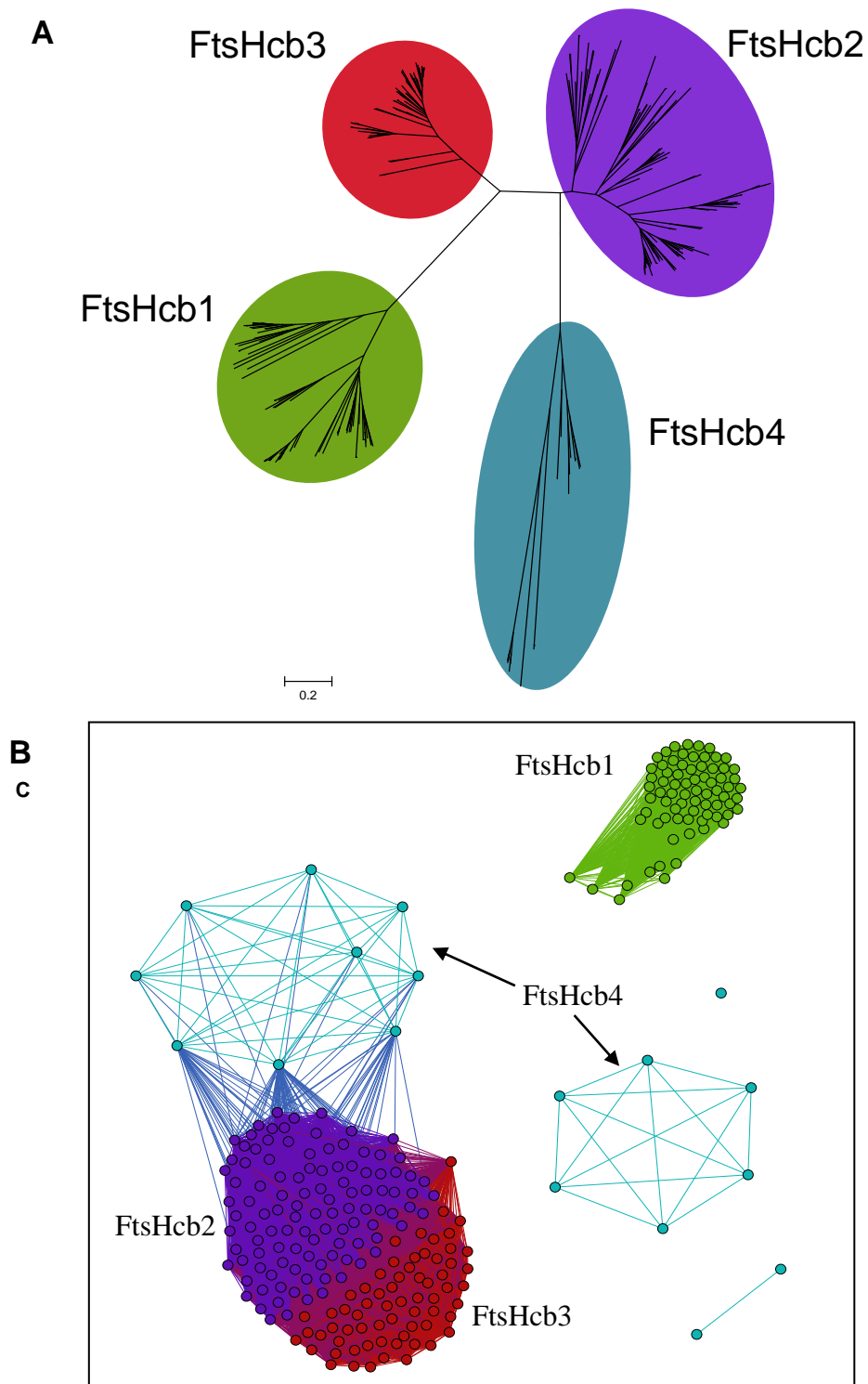
As endomembranas presentes no citoplasma de *Synechocystis* sp. PCC 6803 não possuem comunicação com a membrana plasmática e formam sáculos análogos aos tilacóides plastidiais [36, 37]. Portanto, em contraste com a maioria dos organismos procarióticos conhecidos, as cianobactérias possuem locais e/ou compartimentos sub-celulares adicionais, como o lúmen dos tilacóides. Logo, uma provável explicação para a existência de múltiplas cópias de FtsH codificadas nos genomas das cianobactérias é que elas desempenham suas funções nos diferentes locais e/ou compartimentos sub-celulares diferentes, formando estruturas quaternárias homo ou heteroxaméricas. Esta hipótese é congruente com os resultados obtidos por Boehm et al. [14] e Barker et al. [38].



**Figura 1. Número de cópias de *ftsH* em espécies do Filo Cyanobacteria.** Todas as espécies de cianobactérias analisadas possuem mais de uma cópia do gene que codifica essa proteína e a maioria possui quatro cópias de *ftsH* no genoma. Os números acima das barras representam o número de gêneros (barras cinza) e espécies (barras amarela).

A árvore filogenética de máxima verossimilhança não enraizada obtida a partir do alinhamento múltiplo das sequências de nucleotídeos das FtsHs de cianobactéria possui quatro grupos filogeneticamente distintos nomeados FtsHcb1, FtsHcb2, FtsHcb3, FtsHcb4, todos com *bootstrap* igual ou superior a 95 (Figura 2A). A mesma topologia foi observada na árvore filogenética obtida com as sequências de nucleotídeos das FtsHs, com suporte estatístico similar (dados não mostrados).

Nestas análises não foram incluídas duas sequências de FtsH pertencentes às espécies *Gloeobacter violaceus* e *Gloeobacter kilaueensis* (GIs 37522218 e 554638140) para evitar eventuais vieses nas análises filogenéticas, como atração de ramos longos, devido à alta divergência destas sequências de *Gloeobacter* em relação as demais sequência de FtsH usadas no trabalho.



**Figura 2. Resultados da árvore filogenética e da abordagem de redes complexas para inferir grupos de ortólogos da FtsH em cianobactérias.** **A.** Árvore filogenética não enraizada da FtsH de cianobactérias contendo apenas espécies com genoma completamente sequenciado, obtida através do método da Máxima Verossimilhança com o modelo LG+  $\Gamma$ +I e 1000 repetições de *bootstrap*. **B.** Representação da análise de redes complexas utilizando sequências das FtsHs de cianobactérias com pico crítico em 66% de identidade; as cores indicam os diferentes clados formados na árvore filogenética (Figura 2A), FtsHcb1: verde; FtsHcb2: violeta; FtsHcb3: vermelho; e FtsHcb4: azul.

A árvore filogenética da figura 2A não está enraizada, logo, não é possível inferir sobre a origem e relações evolutivas entre os grupos FtsHcb1, FtsHcb2, FtsHcb3, FtsHcb4, por exemplo, se eles são grupos de genes ortólogos, parálogos ou se são oriundos de eventos de transferência horizontal de genes (THG). Definir relações de homologia entre genes é um passo importante não apenas para a compreensão da história evolutiva gênica, este conhecimento é uma ferramenta útil na anotação funcional de genes em genomas sequenciados [39].

Enraizar de modo não arbitrário árvores filogenéticas de genes cuja história evolutiva não é conhecida pode ser uma missão difícil, pois em princípio não é possível definir qual gene ou grupo de genes é basal, isto é, divergiu primeiro em relação aos demais genes. Optar por enraizar a árvore filogenética no ponto médio, além de ser arbitrário, pode produzir uma árvore filogenética incongruente com a história evolutiva gênica devido a vieses como atração de ramos longos (ARL) e transferência horizontal de genes [40, 41].

Andrade et al. propuseram um novo método para identificar comunidades em redes complexas construídas a partir da matriz de similaridade entre sequências de proteínas e usar esta informação para fazer inferências filogenéticas [29]. Esta abordagem para inferir sobre a origem das mitocôndrias usando redes complexas geradas a partir da matriz de similaridade de três subunidades de ATP sintase eucariótica (4, 6 e 9) e seus respectivos homólogos bacterianos (b, a e c) [30].

Esta nova abordagem para inferir relações filogenéticas implica que as proteínas que estão em uma comunidade de uma rede complexa devem ser mais relacionadas filogeneticamente entre si do que com proteínas que estão em outra comunidade. Então, as comunidades de uma rede complexa obtida a partir de uma matriz de identidade de um grupo de proteínas que contém ortólogos, parálogos e xenólogos devem reproduzir a história evolutiva dessas proteínas.

Para testar essa hipótese foi obtida uma rede complexa a partir da matriz de identidade das sequências de aminoácidos das FtsHs de cianobactérias incluídas na análise filogenética da figura 2. A rede complexa obtida possui duas comunidades principais (Figura 2B). A primeira comunidade é formada por todas as sequências dos grupos FtsHcb2, FtsHcb3 e por 9 das 18 sequências do grupo FtsHcb4, enquanto que a segunda comunidade é formada por todas as sequências do grupo FtsHcb1. As demais comunidades obtidas são constituídas por

proteínas altamente divergentes do grupo FtsHcb4 (Figura 2A) e não possuem conexões com as outras comunidades descritas acima (Figura 2B).

A atração de ramos longos (ARL) é um viés metodológico em que linhagens muito divergentes em relação às demais linhagens incluídas na análise filogenética formam um clado basal na árvore filogenética. As linhagens formam esse clado não porque possuem uma relação filogenética estreita, mas porque acumularam uma grande quantidade de modificações em relação às demais linhagens incluídas na análise filogenética. Métodos de reconstrução filogenética baseados em distância genética, como o *neighbor joining*, são mais sensíveis a ARL do que métodos probabilísticos, tais como o de máxima verossimilhança [41].

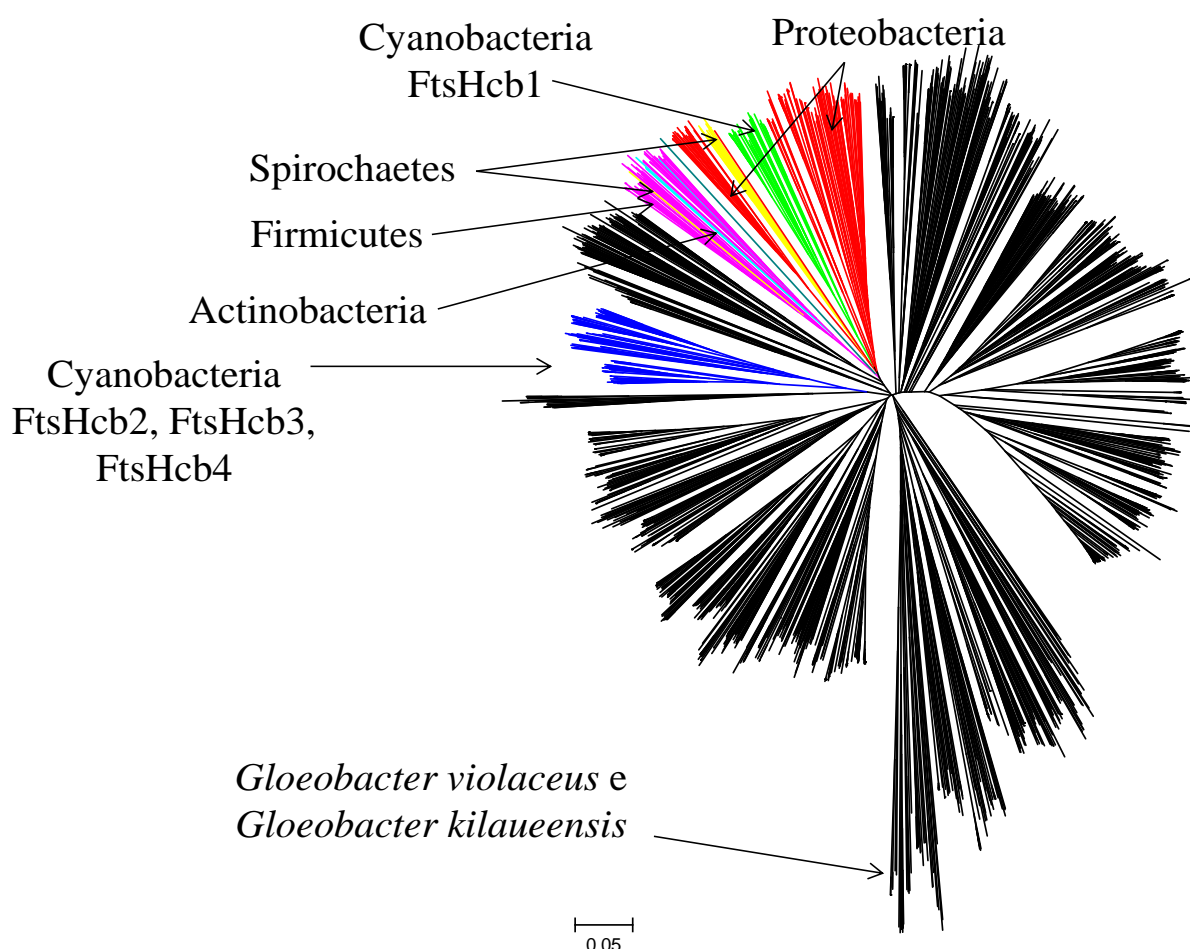
A árvore filogenética das FtsHs de cianobactéria (Figura 2A) foi obtida pelo método da máxima verossimilhança. Por sua vez, a rede complexa foi construída a partir da matriz de similaridade das sequências de aminoácidos das FtsHs, um tipo de informação similar a utilizada em análises filogenéticas realizadas pelo método de *neighbor joining*. Portanto, o fato das FtsHs altamente divergentes do grupo FtsHcb4 (figura 2B) formarem comunidades sem conexão com a comunidade contendo as demais FtsHcb4 pode ser um viés semelhante ao viés observado devido a ARL em árvores filogenéticas obtidas por métodos baseados em distância genética.

Os resultados da rede complexa e da filogenia molecular das FtsHs indicaram que as sequências do grupo FtsHcb1 são filogeneticamente distintas das demais FtsHs das cianobactérias incluídas nas análises. Uma hipótese que poderia explicar esses resultados é que as FtsHcb1 tiveram origem em um evento de transferência horizontal de genes que ocorreu no ancestral das atuais cianobactérias.

Para testar essa hipótese foi realizada uma análise filogenética incluindo as FtsHs de cianobactérias e as FtsHs recuperadas em 1294 genomas bacterianos completamente sequenciados, incluindo as duas FtsHs altamente divergentes de *Gloeobacter violaceus* e *Gloeobacter kilaueensis*. Estes genomas representam 25 Filos bacterianos.

Nesta análise filogenética, as FtsHs das cianobactérias ficam divididas em três grupos distintos. O primeiro grupo contém todas as FtsHs do grupo FtsHcb1 e é filogeneticamente relacionado com as FtsHs de espécies do Filo Proteobacteria (Figura 3) enquanto que segundo grupo contém todas as FtsHs dos grupos FtsHcb2, FtsHcb3 e FtsHcb4. As duas FtsHs altamente divergentes de *Gloeobacter violaceus* e *Gloeobacter kilaueensis*,

formam um terceiro grupo e são filogeneticamente relacionadas com FtsHs de espécies do Filo Actinobacteria (Figura 3). Estas FtsHs possuem alta taxa de divergência e foram excluídas das análises posteriores.



**Figura 3. Árvore filogenética não enraizada de FtsH bacterianas.** Árvore filogenética contendo 1703 FtsHs codificadas em 1294 genomas bacterianos completamente sequenciados de espécies de 25 filos. Em verde o grupo que contém todas as FtsHs do grupo FtsHcb1; em azul o grupo que contém todas as FtsHs dos grupos FtsHcb2, FtsHcb3 e FtsHcb4.

Os resultados apresentados acima são congruentes com a hipótese de que as FtsHcb1 tem uma origem distinta das demais FtsHs das cianobactérias. Esta hipótese foi formulada a partir dos resultados da rede complexa obtida a partir da matriz de identidade das sequências de aminoácidos das FtsHs. Isto mostra que essa nova abordagem para fazer inferências filogenéticas pode ser usada para distinguir as relações de homologia entre genes ou proteínas.



A hipótese para a origem da FtsHcb1 apresentada acima também é corroborada pelo padrão de *indels* destas proteínas. *Indels* são inserções ou deleções conservadas de nucleotídeos. A ocorrência e principalmente a fixação de *indels* são eventos raros, de modo que o compartilhamento de *indels* por genes ou proteínas indica relação de ancestralidade comum entre eles. Portanto, os *indels* podem ser usados como assinaturas moleculares desde em estudos taxonômicos até na identificação de isoformas de proteínas [42].

As FtsHcb1 compartilham um *indel* que não está presente nas demais FtsHs (Figura 4A) enquanto que as FtsHcb2, FtsHcb3 e FtsHcb4 compartilham um *indel* que não está presente nas FtsHcb1 (Figura 4B) indicando que as FtsHs destes três grupos são filogeneticamente relacionadas e distintas das FtsHcb1.

Considerando as evidências apresentadas acima para a origem das FtsHcb1, a árvore filogenética das FtsHs de cianobactérias foi enraizada no ramo que conecta a FtsHcb1 às demais FtsHs das cianobactérias (Figura 5). De acordo com a topologia da árvore filogenética enraizada, as FtsHcb2 e FtsHcb4 compartilham um ancestral comum enquanto que o grupo das FtsHcb3 é basal (Figura 5). Esta topologia é congruente com dois *indels* compartilhados apenas pelas FtsHcb2 e FtsHcb4 (Figuras 4C-D), que indica que elas são filogeneticamente mais relacionadas entre si.

Todas as espécies de cianobactérias analisadas possuem apenas uma cópia da FtsHcb3, isto sugere que a função dessa FtsH é conservada nas espécies desse filo bacteriano (Quadro 1). Contudo, algumas espécies possuem duas ou mais cópias das FtsHcb1 e FtsHcb2 (Quadro 1). Para investigar a origem evolutiva destas cópias foram realizadas novas análises de redes complexas apenas com as FtsHcb1 e FtsHcb2 bem como analisadas as relações evolutivas das cópias dessas isoformas codificadas no mesmo genoma (Figura 6 e7).

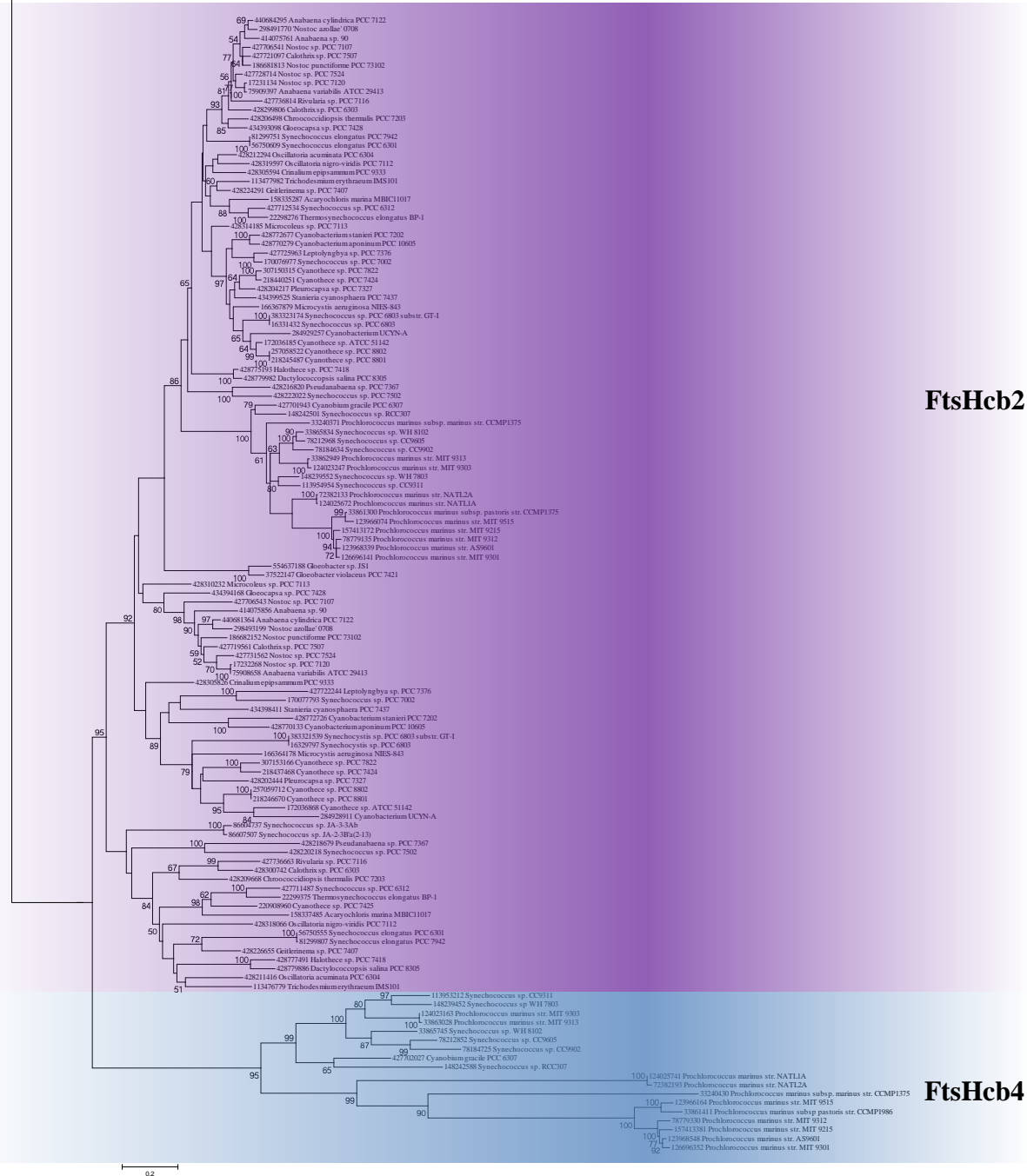




**FtsHcb1**



**FtsHcb3**



**Figura 5. Árvore filogenética enraizada das FtsHs do Filo Cyanobacteria.** A árvore foi obtida através do método da Máxima Verossimilhança usando o modelo LG+  $\Gamma$  +I e com 1000 repetições de *bootstrap*. O enraizamento foi feito utilizando a FtsHcb1 como grupo externo.

A maioria das espécies de cianobactéria analisadas possuem apenas uma cópia no clado FtsHcb1 e dez espécies não possuem nenhuma FtsH neste clado (Figura 6A e Quadro 1). Contudo, doze espécies possuem duas ou mais cópias da FtsHcb1, que estão distribuídas nos subclados FtsHcb1A, FtsHcb1B e FtsHcb1C, todos com altos valores de *bootstrap*, exceto o subclado FtsHcb1C (Figura 6A). A topologia da árvore filogenética indica que os subclados FtsHcb1B e FtsHcb1C são filogeneticamente relacionados e derivados em relação ao subclado FtsHcb1A. Esta topologia é corroborada por *indels* compartilhados pelas FtsHs destes subclados (Figura 4E-F). Apesar do *bootstrap* do clado FtsHcb1C ser apenas 52 o *bootstrap* do nó que conecta este clado com o clado FtsHcb1B é de 100. Isto sugere que o *bootstrap* baixo do clado FtsHcb1C pode ser resultante de relações filogenéticas não resolvidas entre as FtsHs deste clado.

Em geral, as FtsHcb1 codificadas no mesmo genoma ficam distribuídas em dois destes subclados. Por exemplo, *Cyanothece* sp. PCC 8801 e *Cyanothece* sp. PCC 8802 possuem FtsHs que ficam nos subclados FtsHcb1A e FtsHcb1B (Figura 6A e Quadro 1). Já as espécies *Gloeocapsa* sp. PCC 7428 e *Stanieria cyanosphaera* PCC 7437 possuem FtsHs nos subclados FtsHcb1B e FtsHcb1C (Figura 6A e Quadro 1). As exceções são as FtsHcb1 de *Pleurocapsa* sp. PCC 7327, que estão distribuídas nos três subclados, sendo duas no subclado FtsHcb1A, e as FtsHcb1 de *Halothece* sp. PCC 7418, *Nostoc* sp. PCC 7524 e *Rivularia* sp. PCC 7116, que estão apenas no subclado FtsHcb1A (Figura 6A e Quadro 1).

As seis FtsHcb1 de *Acaryochloris marina* MBIC11017 estão distribuídas em dois subclados, contudo, cinco cópias estão no subclado FtsHcb1C (Figura 6A e Quadro 1). Estas cópias ficam em um único clado, indicando que são oriundas de eventos de duplicação gênica que ocorreram nesta espécie (Figura 6A e Quadro 1). Em contraste, as FtsHcb1A de *Halothece* sp. PCC 7418, *Nostoc* sp. PCC 7524 e *Rivularia* sp. PCC 7116 e *Pleurocapsa* sp. PCC 7327 não são filogeneticamente relacionadas (Figura 6A). Isto sugere que, pelo menos uma delas pode ter origem em um evento de transferência horizontal de genes.

As espécies que possuem cópias da FtsHcb1 em dois ou mais subclados pertencem a diferentes ordens do Filo Cyanobacteria (Quadro 1). Portanto, as relações evolutivas dessas cópias e sua distribuição taxonômica são congruentes com a hipótese de que os subclados FtsHcb1A, FtsHcb1B e FtsHcb1C são parálogos ocultos antigos, oriundos de eventos de duplicação gênica que ocorreram, pelo menos, no ancestral comum dessas ordens.

Estes parálogos não estão igualmente distribuídos nos genomas de cianobactérias analisados, indicando que eles foram seletivamente mantidos ou perdidos de modo

independente nas linhagens do Filo Cyanobacteria. Além disso, dez das espécies de cianobactérias analisadas neste trabalho não possuem FtsHcb1A, cb1B ou cb1C codificadas no genoma e, a FtsHcb1C de *Synechocystis* sp. PCC 6803 não é essencial [25]. Isto sugere que a função destas proteínas pode não ser conservada nas espécies do Filo Cyanobacteria.

A rede complexa obtida com a matriz de identidade das FtsHcb1 possui duas comunidades, a primeira é formada pelas sequências do clado FtsHcb1A (Figura 6B). A segunda comunidade contém as FtsHcb1B e FtsHcb1C, mas, apesar das FtsHcb1B terem conexões com as FtsHcb1C, elas possuem mais conexões entre elas formando um subgrupo dentro dessa comunidade (Figura 6B). Estes resultados são congruentes com a topologia do clado das FtsHcb1 e mostra que a estrutura das comunidades da rede complexa reflete as relações de paralogia inferidas a partir dos resultados da filogenia molecular, padrão de *indels* compartilhados e distribuição taxonômica das FtsHcb1A, cb1B e cb1C.

O clado das FtsHcb2 possui dois subclados distintos denominados FtsHcb2A e FtsHcb2B (Figura 7A e Quadro 1). O suporte estatístico desses clados foi baixo, com valores de *bootstrap* inferiores a 50. Contudo, estes subclados são corroborados pelos resultados da rede complexa obtida com a matriz de identidade das FtsHcb2 (Figura 7B), em que as FtsHcb2B apresentam mais conexões entre elas do que com as FtsHcb2A. Adicionalmente, sessenta e três dos sessenta e seis genomas de cianobactérias analisadas possuem uma cópia no subclado FtsHcb2A, destas espécies, quarenta e seis também possuem uma cópia do subclado FtsHcb2B (Figura 7A). Estes resultados são congruentes com a hipótese de que as FtsHcb2A e FtsHcb2B são parálogos oriundos de um evento de duplicação gênica que ocorreu no ancestral das atuais cianobactérias.

As espécies analisadas neste trabalho representam todas as ordens conhecidas do Filo Cyanobacteria (Quadro 1). A distribuição destas espécies na árvore filogenética enraizada das FtsHs indica que os clados FtsHcb1, FtsHcb1B, FtsHcb1C, FtsHcb2A, FtsHcb2B, FtsHcb3 e FtsHcb4 representam grupos de parálogos, e não diferentes grupos taxonômicos do Filo Cyanobacteria.

A análise de genômica comparativa mostra que existem padrões de distribuição dos parálogos das FtsHs nos genomas das cianobactérias (Quadro 1). Por exemplo, as FtsHcb3, FtsHcb2A e FtsHcb2B são os parálogos mais amplamente distribuídos, todas as espécies possuem uma cópia de FtsHcb3 e, a maioria delas possuem uma cópia dos parálogos FtsHcb2A e FtsHcb2B, ou pelo menos um deles (Quadro 1). Por último, todas as espécies que

possuem uma cópia da FtsHcb4 também possuem uma cópia da FtsHcb1C, FtsHcb2A e FtsHcb3 (Quadro 1).

A FtsH forma homo ou heterohexâmeros, ou seja, diferentes isoformas desta proteína podem interagir para desempenhar suas funções [43]. *Synechocystis* sp. PCC 6803 possui quatro FtsHs codificadas no seu genoma, FtsH1 (FtsHcb2B), FtsH2 (FtsHcb2A), FtsH3 (FtsHcb3) e FtsH4 (FtsHcb1C) (Ver tabela 1 para informações adicionais) [25]. Boehm et al. [14] mostraram que a FtsHcb3 dessa cianobactéria forma heterohexâmeros com a FtsHcb2A e com a FtsHcb2B enquanto que a FtsHcb1C forma homohexâmeros.

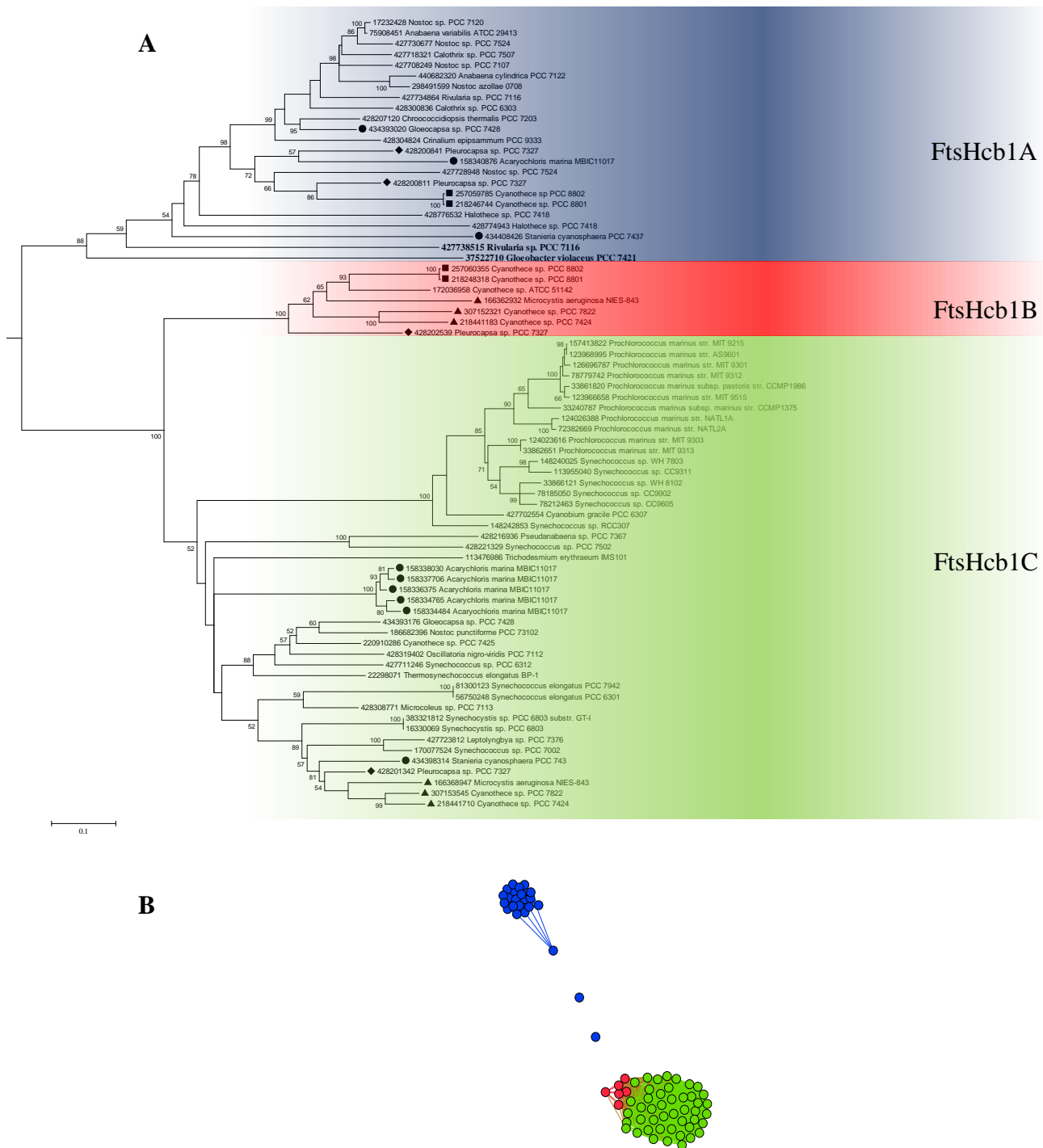
O *knockout* da FtsHcb2B e FtsHcb3 de *Synechocystis* sp. PCC 6803 é letal, em contraste, o *knockout* das FtsHcb1C e FtsHcb2A não é letal. Porém, o *knockout* da FtsHcb2A compromete o funcionamento do fotossistema I diminuindo a atividade fotossintética enquanto que o *knockout* da FtsHcb1C não resultou em alteração significativa nas funções da célula [25].

Ou seja, em *Synechocystis* sp. PCC 6803 os parálogos de FtsH com maior distribuição nas espécies do Filo Cyanobacteria analisadas neste trabalho são proteínas essenciais que fazem interações funcionais, ou, se não é essencial, está envolvida no metabolismo fotossintético da bactéria, como a FtsHcb2A. Portanto, os padrões de distribuição dos parálogos das FtsHs de cianobactérias descritos acima podem indicar parálogos que possuem funções conservadas nas cianobactérias, por exemplo, a FtsHcb3, bem como a existência de associações funcionais entre eles.

Recentemente, Sacharz et al. [39] mostraram que as FtsHs de *Synechocystis* sp. PCC 6803 diferem quanto a localização sub-celular. As FtsHcb1C e FtsHcb3 estão localizadas na extremidade distal dos tilacóides, adjacente a membrana plasmática e, possivelmente existe uma pequena população na membrana plasmática em condições de baixa luminosidade. A FtsHcb2A está localizada apenas nos tilacóides e a FtsHcb2B na membrana plasmática com uma possível uma pequena população nos tilacóides. Resultados similares para a localização subcelular das FtsHcb2A, FtsHcb2B e FtsHcb3 foram encontrados por outros autores (Srivastava et al., 2005; Komenda et al., 2006; Pisavera et al., 2007; Krynicka et al., 2014) [20, 21, 44, 45].

Os resultados apresentados neste trabalho mostram quais FtsHs codificadas nos genomas de outras cianobactérias poderiam ter localização subcelular e função semelhantes as descritas para as FtsHs de *Synechocystis* sp. PCC 6803. Seria interessante verificar se a

localização subcelular dos parálogos das FtsHs de *Synechocystis* sp. PCC 6803 é conservada em outras espécies de cianobactéria.



**Figura 6. Análises filogenéticas das FtsHcb1.** A. Árvore filogenética da FtsHcb1, indicando a existência de três isoformas derivadas. Valores de *bootstrap* exibidos na árvore são acima de 50. Foram detectados quatro padrões de presença de espécies em mais de um clado da FtsHcb1, FtsHcb1A e FtsHcb1B (■), FtsHcb1A e FtsHcb1C (●), FtsHcb1B e FtsHcb1C (▲)

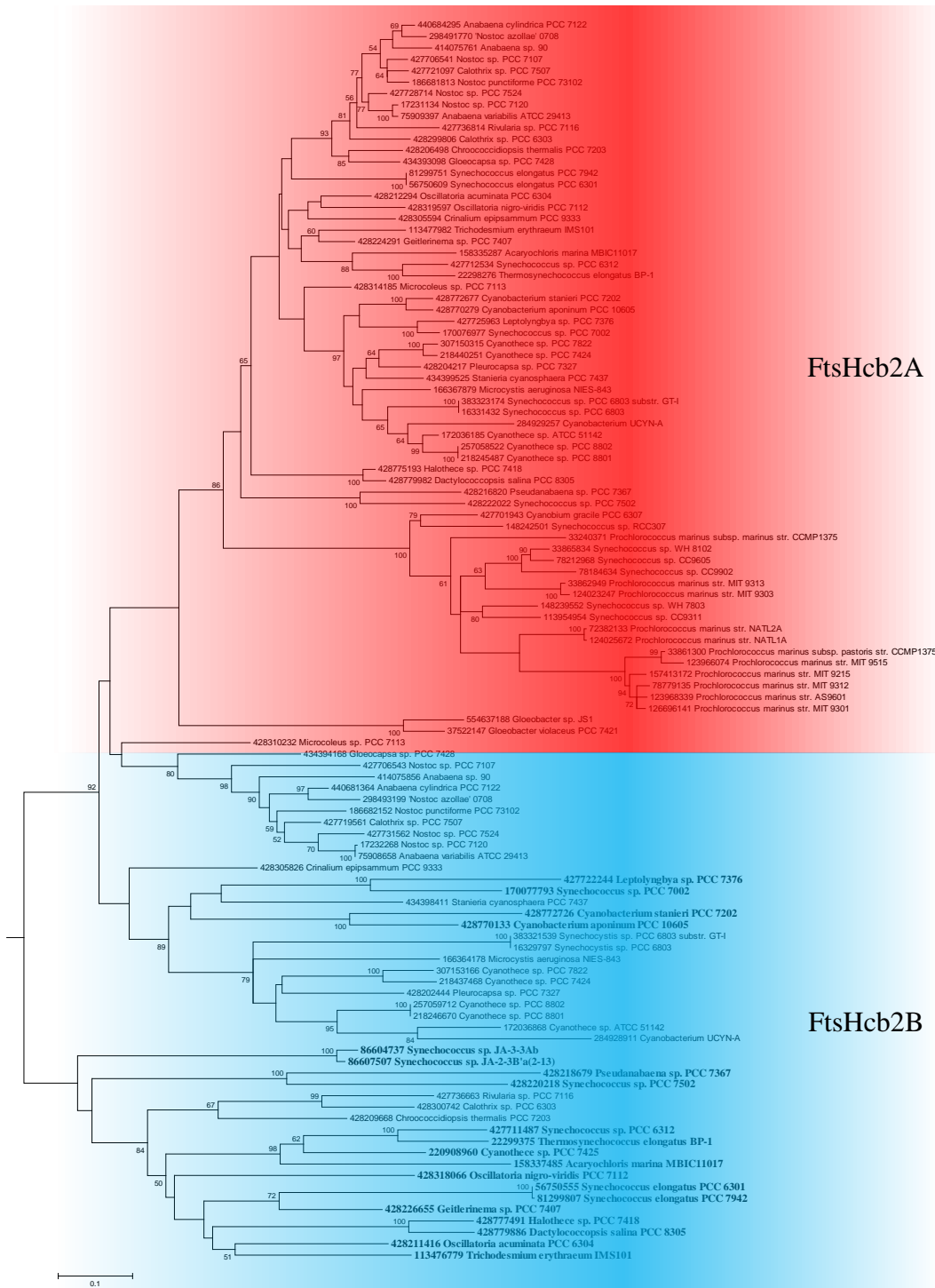


e FtsHcb1A, FtsHcb1B e FtsHcb1C (♦). **B.** Análise de redes complexas baseada nas sequências pertencentes ao grupo da FtsHcb1 com pico crítico em 73%. As sequências que estão desconectadas do grupo ao qual pertence na rede complexa é representada na árvore filogenética em **negrito**.

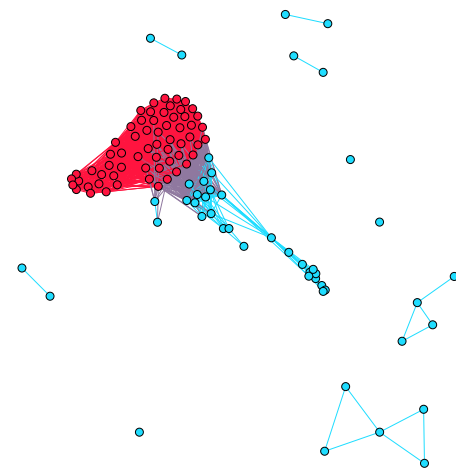
**Quadro 1. Distribuição dos parálogos da FtsH nos táxons do Filo Cyanobacteria. A numeração dentro de cada célula indica o GI da sequência.**

Esquema taxonômica										
CLASSE	ORDEM	FAMÍLIA	ESPÉCIES	FtsHcb1A	FtsHcb1B	FtsHcb1C	FtsHcb2A	FtsHcb2B	FtsHcb3	FtsHcb4
?	Chroococcales	?	<i>Acaryochloris marina</i> MBIC11017	158340876	-	158337706 158336375 158338030 158334484 158334765	158335287	158337485	158335586	-
?	Chroococcales	?	<i>Cyanobacterium apoinum</i> PCC 10605	-	-	-	428770279	428770133	428771133	-
?	Chroococcales	?	<i>Cyanobacterium stanieri</i> PCC 7202	-	-	-	428772677	428772726	428774596	-
?	Chroococcales	?	<i>cyanobacterium</i> UCN-A	-	-	-	284929257	284928911	284929519	-
?	Chroococcales	?	<i>Cyanobium gracile</i> PCC 6307	-	-	427702554	427701943	-	427703137	427702027
?	Chroococcales	?	<i>Cyanothece</i> sp. ATCC 51142	-	172036958	-	172036185	172036688	172038492	-
?	Chroococcales	?	<i>Cyanothece</i> sp. PCC 7424	-	218441183	218441710	218440251	218437468	218440504	-
?	Chroococcales	?	<i>Cyanothece</i> sp. PCC 7425	-	-	-	220910286	-	220908960	-
?	Chroococcales	?	<i>Cyanothece</i> sp. PCC 7822	-	307152321	307153545	307150315	307153166	307150022	-
?	Chroococcales	?	<i>Cyanothece</i> sp. PCC 8801	218246744	218248318	-	218245487	218246670	218245163	-
?	Chroococcales	?	<i>Cyanothece</i> sp. PCC 8802	257059785	257060355	-	257058522	257059712	257058189	-
?	Chroococcales	?	<i>Dactylococcopsis salina</i> PCC 8305	-	-	-	428779982	428779886	428781522	-
?	Chroococcales	?	<i>Gloeoecapsa</i> sp. PCC 7428	434393020	-	434393176	434393098	434394168	434392922	-
?	Chroococcales	?	<i>Halothece</i> sp. PCC 7418	428774943 428776532	-	-	428775193	428777491	428778106	-
?	Chroococcales	?	<i>Microcystis aeruginosa</i> NIES-843	-	166362932	166368947	166367879	166364178	166363123	-
?	Chroococcales	?	<i>Synechococcus elongatus</i> PCC 6301	-	-	56750248	56750609	56750555	56751224	-
?	Chroococcales	?	<i>Synechococcus elongatus</i> PCC 7942	-	-	81300123	81299751	81299807	81299108	-
?	Chroococcales	?	<i>Synechococcus</i> sp. CC9311	-	-	113955040	113954954	-	113953333	113953212
?	Chroococcales	?	<i>Synechococcus</i> sp. CC9605	-	-	78212463	78212968	-	78211853	78212852
?	Chroococcales	?	<i>Synechococcus</i> sp. CC9902	-	-	78185050	78184634	-	78185610	78184725
?	Chroococcales	?	<i>Synechococcus</i> sp. JA-2-3Ba(2-13)	-	-	-	-	86607507	86608120	-
?	Chroococcales	?	<i>Synechococcus</i> sp. JA-3-3Ab	-	-	-	-	86604737	86607354	-
?	Chroococcales	?	<i>Synechococcus</i> sp. PCC 6312	-	-	427711246	427712534	427711487	427712430	-
?	Chroococcales	?	<i>Synechococcus</i> sp. PCC 7002	-	-	170077524	170076977	170077793	170076675	-
?	Chroococcales	?	<i>Synechococcus</i> sp. PCC 7502	-	-	428221329	428222022	428220218	428220339	-
?	Chroococcales	?	<i>Synechococcus</i> sp. RCC307	-	-	148242853	148242501	-	148243292	148242588
?	Chroococcales	?	<i>Synechococcus</i> sp. WH 7803	-	-	148240025	148239552	-	148238691	148239452
?	Chroococcales	?	<i>Synechococcus</i> sp. WH 8102	-	-	33866121	33865834	-	33864841	33865745
?	Chroococcales	?	<i>Synechocystis</i> sp. PCC 6803	-	-	16330069	16331432	16329797	16329602	-
?	Chroococcales	?	<i>Synechocystis</i> sp. PCC 6803 subsp. GT-1	-	-	383321812	383323174	383321539	383321343	-
?	Chroococcales	?	<i>Thermosynechococcus elongatus</i> BP-1	-	-	22298071	22298276	22299375	22297675	-
?	Nostocales	Nostocaceae	<i>Anabaena cylindrica</i> PCC 7122	440682320	-	440684295	440681364	440681685	-	-
?	Nostocales	Nostocaceae	<i>Anabaena</i> sp. 90	-	-	-	414075761	414075856	414076877	-
?	Nostocales	Nostocaceae	<i>Anabaena variabilis</i> ATCC 29413	75908451	-	-	75909397	75908658	75906798	-
?	Nostocales	Nostocaceae	<i>Nostoc azolae</i> 0708	298491599	-	-	298491770	298493199	298489876	-
?	Nostocales	Nostocaceae	<i>Nostoc punctiforme</i> PCC 73102	-	-	186682396	186681813	186682152	186684974	-
?	Nostocales	Nostocaceae	<i>Nostoc</i> sp. PCC 7107	427708249	-	-	427706541	427706543	427709089	-
?	Nostocales	Nostocaceae	<i>Nostoc</i> sp. PCC 7120	17232428	-	-	17231134	17232268	172328756	-
?	Nostocales	Nostocaceae	<i>Nostoc</i> sp. PCC 7524	427730677 427728948	-	-	427728714	427731562	427730489	-
?	Nostocales	Rivulariaceae	<i>Calothrix</i> sp. PCC 6303	428300836	-	-	428299806	428300742	428298512	-
?	Nostocales	Rivulariaceae	<i>Calothrix</i> sp. PCC 7507	427718321	-	-	427721097	427719561	427718379	-
?	Nostocales	Rivulariaceae	<i>Rivularia</i> sp. PCC 7116	427734864 427738515	-	-	427736614	427736663	427733813	-
?	Oscillatoriales	?	<i>Criminalium epissumum</i> PCC 9333	428304824	-	-	428305594	428305826	428306786	-
?	Oscillatoriales	?	<i>Geitlerinema</i> sp. PCC 7407	-	-	-	428224291	428226655	428226442	-
?	Oscillatoriales	?	<i>Leptolyngbya</i> sp. PCC 7376	-	-	427723812	427725963	427722244	427723584	-
?	Oscillatoriales	?	<i>Microcoleus</i> sp. PCC 7113	-	-	428308771	428314185	428310232	428309071	-
?	Oscillatoriales	?	<i>Oscillatoria acuminata</i> PCC 6304	-	-	-	428212294	428211416	428210836	-
?	Oscillatoriales	?	<i>Oscillatoria nigro-viridis</i> PCC 7112	-	-	428319402	428319597	428318066	428318597	-
?	Oscillatoriales	?	<i>Pseudanabaena</i> sp. PCC 7367	-	-	428216936	428216820	428218679	428218897	-
?	Oscillatoriales	?	<i>Trichodesmium erythraeum</i> IMS101	-	-	113476986	113477982	113476779	113477219	-
?	Pleurocapsales	?	<i>Chroococcidiopsis thermalis</i> PCC 7203	428207120	-	-	428206498	428209668	428209090	-
?	Pleurocapsales	?	<i>Pleurocapsa</i> sp. PCC 7327	428200841 428200811	428202539	428201342	428204217	428202444	428202589	-
?	Pleurocapsales	?	<i>Staniaeria cyanosphaera</i> PCC 7437	434408426	-	434398314	434399525	434398411	434396728	-
?	Prochlorales	Prochlorococcaceae	<i>Prochlorococcus marinus</i> str. AS9601	-	-	123968995	123968339	-	123967784	123968548
?	Prochlorales	Prochlorococcaceae	<i>Prochlorococcus marinus</i> str. MIT 9215	-	-	157413822	157413172	-	157412585	157413381
?	Prochlorales	Prochlorococcaceae	<i>Prochlorococcus marinus</i> str. MIT 9301	-	-	126696787	126696141	-	126695586	126696352
?	Prochlorales	Prochlorococcaceae	<i>Prochlorococcus marinus</i> str. MIT 9303	-	-	124023616	124023247	-	124024074	124023163
?	Prochlorales	Prochlorococcaceae	<i>Prochlorococcus marinus</i> str. MIT 9312	-	-	78779742	78779135	-	78778614	78779330
?	Prochlorales	Prochlorococcaceae	<i>Prochlorococcus marinus</i> str. MIT 9313	-	-	33862651	33862949	-	33864065	33863028
?	Prochlorales	Prochlorococcaceae	<i>Prochlorococcus marinus</i> str. MIT 9515	-	-	123966658	123966074	-	123965493	123966164
?	Prochlorales	Prochlorococcaceae	<i>Prochlorococcus marinus</i> str. NATL1A	-	-	124026388	124025672	-	124025019	124025741
?	Prochlorales	Prochlorococcaceae	<i>Prochlorococcus marinus</i> str. NATL2A	-	-	72382669	72382133	-	72383431	72382193
?	Prochlorales	Prochlorococcaceae	<i>P. marinus</i> subsp. <i>marinus</i> str. CCMP1375	-	-	33240787	33240371	-	33239707	33240430
?	Prochlorales	Prochlorococcaceae	<i>P. marinus</i> subsp. <i>pastoris</i> str. CCMP1986	-	-	33861820	33861300	-	33860785	33861411
Gloeobacteria	Gloeobacterales	?	<i>Gloeobacter kilauensis</i> JS1	-	-	-	554637188	-	554637072	-
Gloeobacteria	Gloeobacterales	?	<i>Gloeobacter violaceus</i> PCC 7421	37522710	-	-	37522147	-	37521486	-

A



B



**Figura 7. Clado FtsHcb2 da árvore filogenética das FtsHs de cianobactérias. A.** Clado da FtsHcb2 da árvore de cianobactérias, indicando que duas isoformas originaram a partir da FtsHcb2. Valores de *bootstrap* exibidos na árvore são acima de 50 **B.** Análise de redes complexas baseadas nas sequências constituintes do grupo FtsHcb2 com pico crítico em 77%. As sequências que estão desconectadas do grupo ao qual pertence na rede complexa é representada na árvore filogenética em **negrito**.

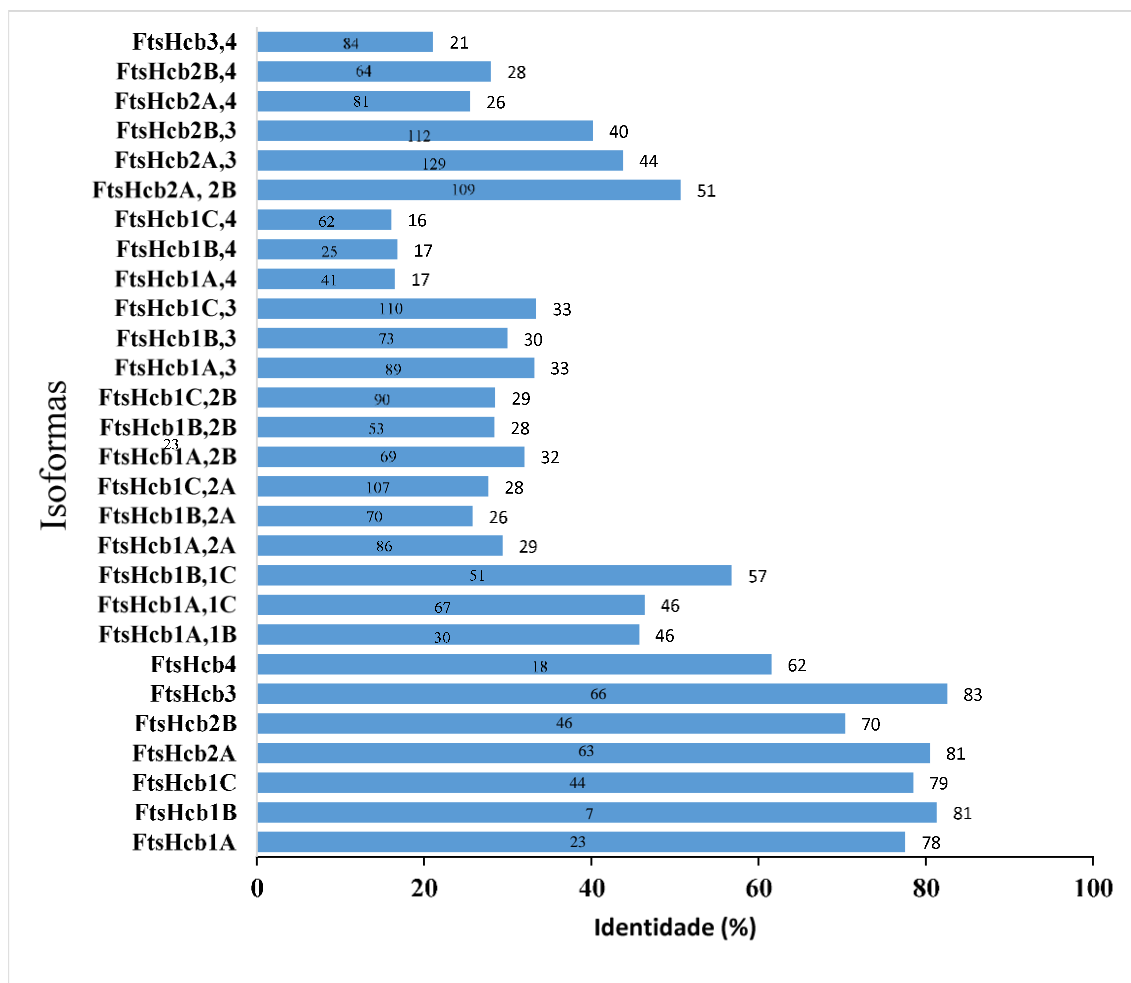
A conservação dos parálogos das FtsHs de cianobactérias definidos nas análises anteriores foi avaliada por meio do cálculo da identidade média das sequências de cada grupo de parálogos e entre todas as combinações entre eles (Figura 8). A identidade média das FtsHcb1A, FtsHcb1B e FtsHcb1C variou de 78 a 81%, uma variação menor do que a observada entre as FtsHcb2A, FtsHcb2B, que foi de 80 e 70%, respectivamente (Figura 8). As FtsHcb3, essencial em *Synechocystis* sp. PCC 6803, possuem a maior identidade média dentre os grupos de parálogos enquanto que as FtsHcb4 possuem a menor identidade média (Figura 8). A identidade média de todas as combinações de grupos de parálogos foi menor do que a observada para cada grupo, variando entre 16 e 57% de identidade (Figura 8).

Estes resultados mostram que embora os parálogos das FtsHs de cianobactérias sejam conservados eles variam consideravelmente entre eles. Essa divergência pode ser resultante de uma diferenciação funcional entre elas, conforme descrito para os parálogos codificados no genoma de *Synechocystis* sp. PCC 6803, mas parece não ser impeditivo para eventuais interações funcionais. Por exemplo, identidade média das FtsHcb3 é de 83%, no entanto ela forma hetero-oligômeros com as FtsHcb2A e FtsHcb2B, com que tem apenas 44 % e 40 % de identidade média (Figura 8).

Tabela 1: FtsHs codificadas no genoma de *Synechocystis* sp. PCC 6803.

Parálogos de FtsH de cianobactérias identificados neste trabalho	Loco e designação da proteína [14, 25]	Organização do complexo oligomérico [14]	Localização subcelular [20, 21, 22, 44, 45]
FtsHcb1A	-	-	-
FtsHcb1B	-	-	-
FtsHcb1C	<i>sll1463</i> - FtsH 4	Forma homo-oligômero	Extremidade distal dos tilacóides adjacente a membrana plasmática e uma possível pequena quantidade na membrana plasmática em LL.
FtsHcb2A	<i>slr0228</i> - FtsH2	Forma hetero-oligômero com a FtsH3	Apenas nos tilacóides
FtsHcb2B	<i>slr1390*</i> - FtsH1	Forma hetero-oligômero com a FtsH3	<b>Membrana plasmática</b> e uma possível pequena quantidade nos tilacóides.
FtsHcb3	<i>slr1604*</i> - FtsH3	Forma hetero-oligômero com a FtsH1 e FtsH2	Extremidade distal dos tilacóides adjacente a membrana plasmática e uma possível pequena quantidade na membrana plasmática em LL.
FtsHcb4	-	-	-

\* Indica *ftsHs* essenciais de acordo com Mann et al. [25].



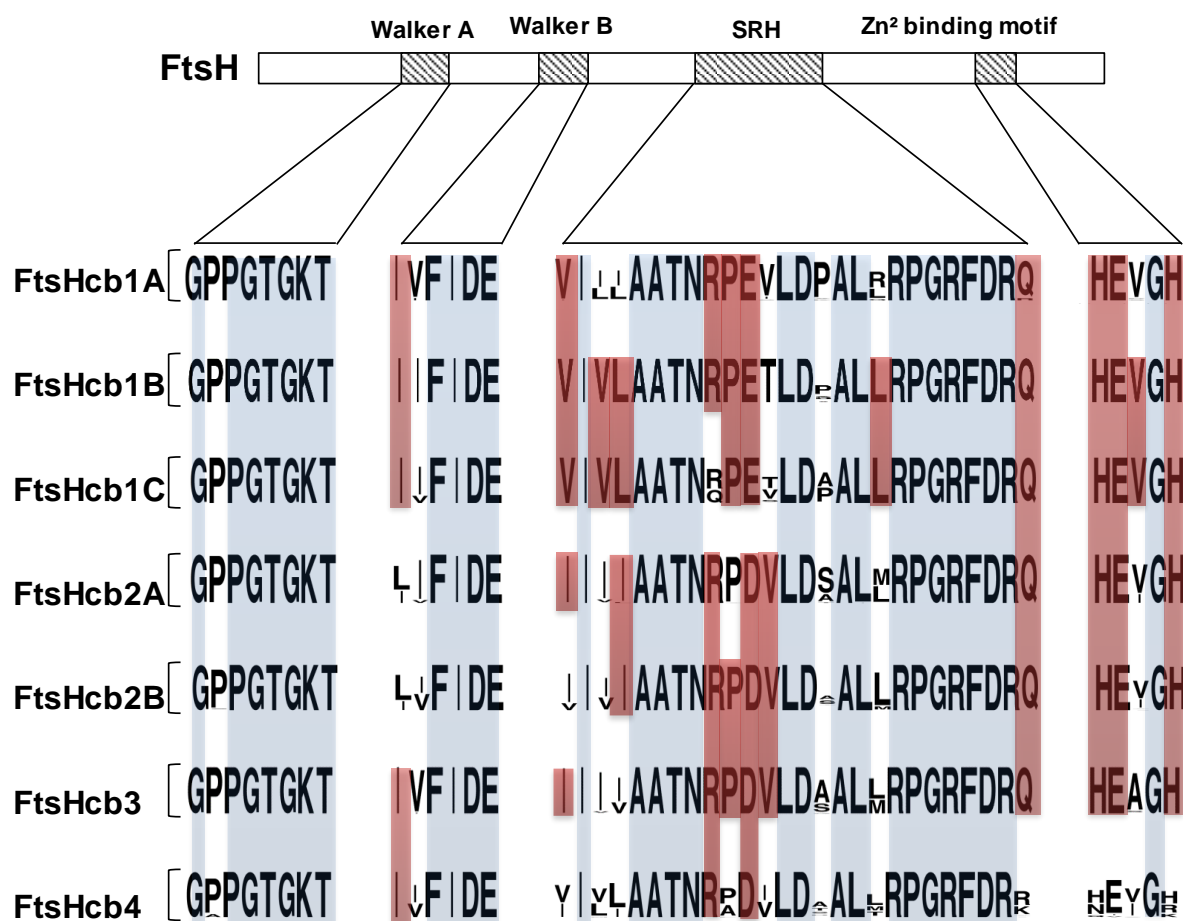
**Figura 8. Gráfico de identidade média (%) entre as isoformas do gene que codificam a FtsH em Cyanobacteria.** Os valores dentro das barras indicam o número de seqüências presentes nos parálogos e suas respectivas combinações. Os valores acima da barra indicam as identidades de cada grupo. Para definir a média entre as seqüências foi gerada uma matriz de distância que varia de zero à um, e subtraiu um desse valor obtendo a identidade, depois somou todos os valores e dividiu pelo número de sequencias, adquirindo a identidade média entre as seqüências.

Os parálogos das FtsHs das cianobactérias descritos neste trabalho divergem consideravelmente. Para verificar se a divergência observada entre os parálogos se estende às regiões relacionadas com as suas atividades ATPásica e proteolítica foram feitas análises no software LOGO dos motivos Walker A, Walker B e SRH do domínio AAA e do sítio de ligação a zinco do domínio Peptidase M41 destas proteínas.

A maioria dos aminoácidos dos motivos do domínio AAA são conservados em todos os parálogos das FtsHs de cianobactérias, sendo o motivo Walker A o mais conservado, com variação apenas na segunda posição nas FtsHcb2B e FtsHcb4 (Figura 9). Na primeira posição do Walker B os parálogos FtsHcb2A e 2B possuem isoleucina ou leucina (Figura 9).

Esta variação foi encontrada nas FtsHcb2A e cb2B de espécies que pertencem a diferentes ordens, portanto não parece ser uma variação taxonômica. Nos demais parálogos foi observado apenas isoleucina nesta posição (Figura 9). A segunda posição do Walker B também é variável, as FtsHcb1A e FtsHcb3 possuem valina nessa posição enquanto que todas as FtsHcb1B possuem isoleucina, nos demais parálogos foi encontrado valina ou isoleucina.

A maior variação de aminoácidos por posição foi observada no motivo SRH das FtsHs das cianobactérias, tanto entre os parálogos como entre ortólogos (Figura 9). O número de posições variáveis variou de uma, nas FtsHcb1B, a oito, nas FtsHcb4. As FtsHcb1B também não apresentaram variação nos motivos Walker A e B. Dentre as FtsHs de cianobactérias analisadas nesse trabalho apenas sete pertencem ao grupo de ortólogos FtsHcb1B, sendo seis de espécies de dois gêneros da Ordem *Chroococcales* e uma de *Pleurocapsa* sp. PCC 7327, da Ordem *Pleurocapsales* (Quadro 1 e Figura 8). Portanto, o alto nível de conservação dos motivos do domínio AAA nas FtsHcb1B pode ser resultante de um viés amostral. O padrão de conservação do décimo primeiro aminoácido do motivo SRH parece refletir as relações evolutivas das FtsHs de cianobactérias, pois enquanto FtsHcb1A, cb1B e cb1C possuem ácido glutâmico nesta posição os demais parálogos possuem ácido aspártico (Figura 9).



**Figura 9. Conservação dos motivos das FtsHs do Filo Cyanobacteria.** Os sete grupos de parálogos foram comparados em regiões conservadas presentes nos dois domínios da FtsH. Em azul estão marcados todos os aminoácidos conservados em todos os parálogos. Marcado na cor vermelha os aminoácidos conservados em dois ou mais parálogos.

O consenso do motivo de ligação a zinco das FtsHs, HExxH [5, 46], está presente em todos os parálogos da FtsHs de cianobactérias, a exceção são as FtsHcb4 das espécies de *Prochlorococcus*, que possuem um motivo distinto, (Q/N) xuG (K/R) (Figura 9). Esta variação no motivo de ligação a zinco não foi descrita ainda para as FtsHs e não foi encontrada nas FtsHs de 1694 espécies bacterianas analisadas neste trabalho (Figura 3). Esse consenso atípico pode refletir uma especificidade funcional desses parálogos nas espécies do gênero *Prochlorococcus*, um gênero cuja espécies são as principais responsáveis pela biomassa fotossintética nos oceanos [47].

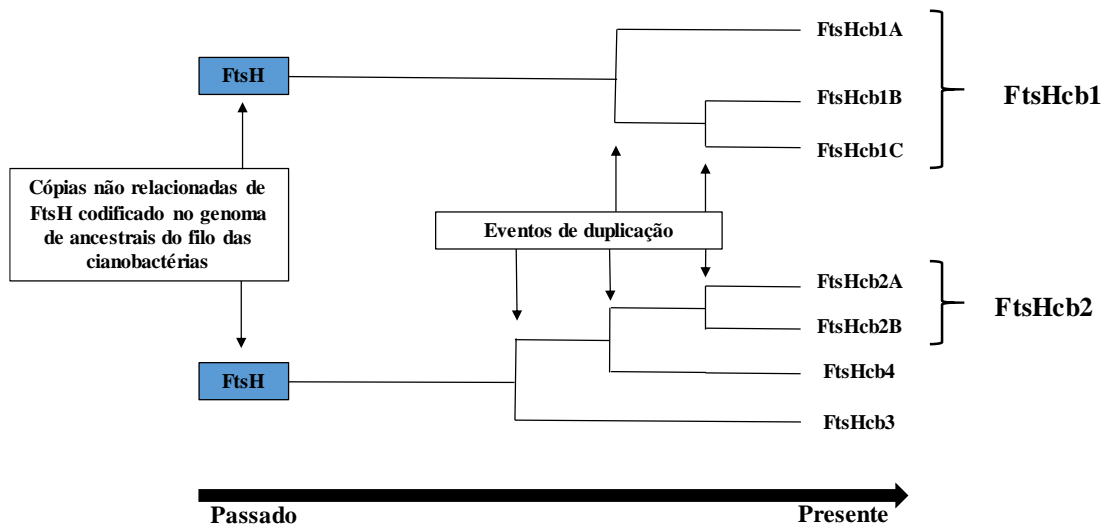
Os resultados das análises filogenéticas, genômicas, de redes complexas e da estrutura primária das FtsHs de cianobactérias apresentados neste trabalho são congruentes, complementares e permitem inferir sobre a história evolutivas das FtsHs do Filo Cyanobacteria. Os resultados das análises filogenéticas e da rede complexa obtida a partir da matriz de identidade das FtsHs de cianobactérias mostram claramente que as FtsHs do grupo

FtsHcb1 não são filogeneticamente relacionadas com os demais FtsHs desse filo bacteriano. Estes resultados indicam também que este grupo é constituído por 3 parálogos que possuem ampla distribuição taxonômica no Filo Cyanobacteria.

Portanto, nós propomos que o ancestral das FtsHcb1A, cb1B e cb1C foi adquirido em um evento de THG no ancestral comum das cianobactérias e, posteriormente dois eventos de duplicação originaram as FtsHcb1A, FtsHcb1B e FtsHcb1C (Figura 10). As demais FtsHs originaram-se de três eventos de duplicação da FtsH nativa do ancestral das cianobactérias (Figura 10).

Os resultados apresentados neste trabalho podem ser usados para classificar de modo não arbitrário as FtsHs das cianobactérias, com base na história evolutiva delas. Geralmente, proteínas ortólogas desempenham a mesma função ou funções similares, assim, este sistema de classificação ajudaria na anotação das novas FtsHs depositadas nas bases de dados de sequências oriundas do sequenciamento de genomas de cianobactérias. Adicionalmente, este sistema de classificação ajudaria fazer inferência funcionais de FtsHs de cianobactérias não caracterizadas que poderiam ser posteriormente testadas.

Por exemplo, *Synechocystis* sp. PCC 6803 possui quatro FtsHs codificadas no seu genoma que estão parcialmente caracterizadas funcionalmente [22, 25], e pertencem aos grupos de ortólogos FtsHcb1C, FtsHcb2A, FtsHcb2B e FtsHcb3 que estão amplamente distribuídos nas cianobactérias. Com base na classificação filogenética das FtsHs da cianobactérias proposta neste trabalho é possível fazer a predição da função das FtsHs que pertencem a estes grupos de ortólogos. Esta informação auxiliaria no planejamento de estudos envolvendo a caracterização funcional dessas proteínas.



**Figura 10. História evolutiva proposta para as FtsHs do Filo Cyanobacteria.** Neste modelo o ancestral das FtsHcb1A, FtsHcb1B e FtsHcb3C foi adquirido em um evento de TLG no ancestral comum do Filo Cyanobacteria. Por sua vez, as FtsHcb2A FtsHcb2B, FtsHcb3 e FtsHcb4 são parálogos que tem como ancestral comum a FtsH nativa do ancestral do Filo Cyanobactéria.



## CONSIDERAÇÕES FINAIS

Neste trabalho foram apresentados resultados que permitem expandir o conhecimento sobre a história evolutiva das FtsHs de cianobactérias, fornecendo um ponto de partida para a formulação de novas hipóteses e desenvolvimento de novos estudos sobre as relações evolutivas e funcionais das FtsHs deste filo bacteriano. Por exemplo, a maior parte dos estudos sobre caracterização funcional das FtsHs de cianobactérias foram desenvolvidos com as FtsHs de *Synechocystis* sp. PCC 6803, que pertencem aos grupos de ortólogos FtsHcb1C, FtsHcb2A, FtsHcb2B e FtsHcb3. Proteínas ortólogas tendem a ser funcionalmente relacionadas, logo, o conhecimento adquirido sobre a função das FtsH1, FtsH2, FtsH3 e FtsH4 de *Synechocystis* sp. PCC 6803 podem ser extrapolados e testados para as demais FtsHs pertencentes a esses grupos de ortólogos.

Estas informações podem ser usadas para formular hipóteses testáveis sobre a função de FtsHs não caracterizadas funcionalmente em cianobactérias. Por exemplo, as FtsHcb1A, FtsHcb1B, FtsHcb4.

As FtsHs das cianobactérias estão relacionadas com a fotossíntese oxigênica, uma invenção evolutiva desse filo procariótico. Os resultados deste trabalho mostram que os eventos que originaram os diferentes grupos de ortólogos das FtsHs de cianobactéria ocorreram no ancestral desse filo, o que sugere que a origem das FtsHs de cianobactérias pode estar relacionada com a evolução desse metabolismo fotossintético.

Durante a elaboração desse manuscrito estavam disponíveis nas bases de dados do NCBI 66 genomas de cianobactérias completamente sequenciados e 349 parcialmente sequenciados ou sendo sequenciados. Nesse cenário os resultados aqui apresentados também possuem uma importância prática, pois permitirão fazer uma anotação mais precisa das FtsHs encontradas nos genomas de cianobactérias que estão sendo sequenciados. Adicionalmente, os resultados deste trabalho demonstraram que redes complexas obtidas a partir da matriz de identidade de proteínas é uma ferramenta útil na definição das relações de homologia entre as sequências de proteínas. Essa abordagem para o uso das redes complexas ainda não tinha sido explorada até então.

## REFERÊNCIAS

1. Tomoyasu T, Yuki T, Morimura S, Mori H, Yamanaka K, Niki H, Hiraga S, Ogura T. The *Escherichia coli* FtsH protein is a prokaryotic member of a protein family of putative ATPases involved in membrane functions, cell cycle control, and gene expression. *Journal of bacteriology*. March 1993; 175:1344-1351.
2. Schneider E, Hunke S. ATP-binding-cassette (ABC) transport systems: Functional and structural aspects of the ATP-hydrolyzing subunits/domains. *FEMS Microbiology Reviews* 22. 1998; 1-20.
3. Herman C, Ogura T, Tomoyasu T, Hiraga S, Akiyama Y, Ito K, Thomas R, Dauri R, Bouloc P. Cell growth and A phage development controlled by the same essential *Escherichia coli* gene, ftsH/hflB. *PNAS*. 1993; 90: 10861-10865.
4. Rawlings ND, Barrett AJ. Evolutionary families of metallopeptidases. *Methods Enzymol*. 1995; 248:183-228.
5. Shotland Y, Teff D, Koby S, Kobiler O, Oppenheim AB. Characterization of a conserved  $\alpha$ -helical, coiled-coil motif at the C-terminal domain of the ATP-dependent FtsH (HflB) protease of *Escherichia coli*. 2000: doi:10.1006/jmbi.2000.3767.
6. Gupta RS. Molecular signatures (unique proteins and conserved indels) that are specific for the epsilon proteobacteria (Campylobacterales). 2006: doi:10.1186/1471-2164-7-167.
7. Lechner M, Findeiß S, Stein L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: Detection of (Co-) orthologs in large-scale analysis. *BMC Bioinformatics*. 2011; 12:124.
8. Cavalier-Smith T. The number of symbiotic origins of organelles. *BioSystems*. 1992; 28:91-106
9. Marbach PAS, Coelho ASG, Filho MCS. Mitochondrial and chloroplast localization of FtsH-like proteins in sugarcane based on their phylogenetic profile. *Genetics and Molecular Biology*. 2001; 24:183-190.
10. Juhola MK, Shah ZH, Grivell LA, Jacobs HT. The mitochondrial inner membrane AAA metalloprotease family in metazoans. *FEBS Letters*. 2000: 481;91-95.

11. Leonhard K, Herrmann M, Stuart RA, Mannhaupt G, Neupert W, Langer T. AAA proteases with catalytic sites on opposite membrane surfaces comprise a proteolytic system for the ATP-dependent degradation of inner membrane proteins in mitochondria. *The EMBO Journal*. 1996; 15:4218-4229.
12. Langklotz S, Baumann U, Narbearhaus F. Structure and function of the bacterial AAA protease FtsH. *Biochimica et Biophysica Acta*. 2012; 1823:40-48.
13. Moldavski O, Levin-Kravets O, Ziv T, Adam Z, Prag G. The hetero-hexameric nature of a chloroplast AAA+ FtsH protease contributes to its thermodynamic stability. *PLoS ONE*. april 2012; 7.
14. Boehm M, Yu J, Krynicka V, Barker M, Tichy M, Komenda J, Nixon PJ, Nield Jon. Subunit organization of a *Synechocystis* Hetero-Oligomeric thylakoid FtsH complex involved in photosystem II repair. *The Plant Cell*. 2012; 24:3669-3683.
15. Tomoyasu T, Gamer J, Bukau B, Kanemori M, Mori H, Rutman A J, Oppenheim A B, Yura T, Yamanaka K, Niki H, Hiraga S, Ogura T. *Escherichia coli* FtsH is a membrane-bound, ATP-dependent protease which degrades the heat-shock transcription factor  $\sigma^{32}$ . *The EMBO Journal*. 1995; 14:2551-2556.
16. Yoshioka M, Nakayama Y, Yoshida M, Ohashi K, Morita N, Kobayashi H, Yamamoto Y. Quality control of photosystem II: FtsH hexamers are localized near photosystem II at grana for the swift repair of damage. *J. Biol. Chem.* December 2010; 285:41972-41981.
17. Schumann W., FtsH – a single-chain charonin?. *FEMS Microbiology Reviews*. 1999; 23:1-11.
18. Clarke AK, Hurry VM, Gustafsson P, Oquist G. Two functionally distinct forms of the photosystem II reaction-center protein D1 in the cyanobacterium *Synechococcus* sp. PCC: Biochemistry. 1993; 7942; Vol. 90:11985-11989.
19. Bailey S, Silvat P, Nixont P, Mullineauxf C, Robinson C, Mann N. Auxiliary functions in photosynthesis: the role of the FtsH protease. 2001; 29:455-459.
20. Komenda J, Barker M, Kuvikova S, de Vries R, Mullineaux CW, Tichy M, Nixon PJ. The FtsH protease slr0228 is important for quality control of Photosystem II in the

- thylakoid membrane of *Synechocystis* sp. PCC 6803. *J Biol Chem*. 2006; 281:1145–1151.
21. Pisareva T, Shumskaya M, Maddalo G, Ilag L, Norling B. Proteomics of *Synechocystis* sp. PCC 6803 identification of novel integral plasma membrane proteins. *FEBS J*. 2007; 274:791–804.
  22. Schwartz J, Bryan SJ, Yu J, Burroughs NJ, Spence EM, Nixon PJ, Mullineaux CW. Sub-cellular location of FtsH proteases in the cyanobacterium *Synechocystis* sp. PCC 6803 suggests localised PSII repair zones in the thylakoid membranes. 2015; 96:448-462.
  23. Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, Miyajima N, Hiroshiwa M, Sugiura M, Sasamoto S, Kimura T, Hosouch T, Matsuno A, Muraki A, Nakazaki N, Naruo K, Okumura S, Shimpo S, Takeuchi C, Wada T, Watanabe A, Yamada M, Yasuda M, Tabata S. Sequence Analysis of the Genome of the Unicellular Cyanobacterium *Synechocystis* sp. Strain PCC6803. II. Sequence Determination of the Entire Genome and Assignment of Potential Protein-coding Regions .*DNA Research*. 1996; 3: 109-136.
  24. Sacharz J, Bryan SJ, Yu J, Burroughs NJ, Mullineaux CW. Sub-cellular location of FtsH proteases in the cyanobacterium *Synechocystis* sp. PCC 6803 suggests localised PSII repair zones in the thylakoid membranes. *Molecular Microbiology*. 2015; 96: 448-462
  25. Mann NH, Novac N, Mullineaux CW, Newman J, Bailey S, Robinson C. Involvement of an FtsH homologue in the assembly of functional photosystem I in the cyanobacterium *Synechocystis* sp. PCC 6803. *FEBS Letters*. 2000; 479:72-77.
  26. Cavalier-Smith T. The number of symbiotic origins of organelles; *BioSystems*. 1992; 28:91-106.
  27. Li C, Ding X-Q, O'Brien J, Al-Ubaidi MR, Naash M I. Molecular Characterization of the Skate Peripherin/rds Gene: Relationship to Its Orthologues and Paralogues .*Invest Ophthalmol Vis Sci*. June 2003; 44(6): 2433–2441.
  28. Andrade RFS, Rocha-Neto IC, Santos LBL, Santana CN, Diniz MVC, Lobão TP, Goés-Neto A, Pinho STR, El-Hani CN. Detecting Network Communities: An

- Application to Phylogenetic Analysis. *PLoS Comput Biol.* 2011 May 05; 7(5): e1001131. doi:10.1371.
29. Góes-Neto A, Diniz MVC, Santos LBL, Pinho STR, Miranda JGV, Lobão TP. Comparative protein analysis of the chitin metabolic pathway in extant organisms: A complex network approach. *BioSystems.* 2010; 101:59-66.
  30. Carvalho DS, Andrade RFS, Pinho STR, Góes-Neto A, Lobão TCP, Bomfim GC. What are the Evolutionary Origins of Mitochondria? A Complex Network Approach. *PLoS ONE* 10(9): e0134988. doi:10.1371/journal.pone.0134988.
  31. Penn O, Privman E, Ashkenazy H, Landan G, Graur D, Pupko T. GUIDANCE: a web server for assessing alignment confidence scores. *Nucleic acids research.* 2010; doi:10.1093/nar/gkq443.
  32. Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular biology and evolution.* 2013; 30(12):2725–9.
  33. Han MV, Zmasek CM. phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics.* 2009; 10:356.
  34. Andrade RFS, Pinho STR, Lobão TP. Identification of community structure in networks using higher order neighborhood concepts. *Int J Bifurc Chaos.* 2009; 19: 2677–2685.
  35. Bastian M, Heymann S. Gephi: An Open Source Software for Exploring and Manipulating Networks. *International AAAI Conference on Weblogs and Social Media.* 2009.
  36. Schneider D, Fuhrmann E, Scholz I, Hess WR, Graumann P L. Fluorescence staining of live cyanobacterial cells suggest non-stringent chromosome segregation and absence of a connection between cytoplasmic and thylakoid membranes. *BMC Cell Biology.* 2007: 8-39
  37. Liberton M, Howard BR, Heuser J, Roth R, Pakrasi HB. Ultrastructure of the membrane systems in the unicellular cyanobacterium *Synechocystis* sp. strain PCC 6803. *Protoplasma.* 2006: 227;129–138.

38. Barker M., Boehm M., Nixon P.J., Nield J. Structural analysis of an FtsH2/FtsH3 complex isolated from *Synechocystis* sp. PCC 6803. *Photosynthesis. Energy from the Sun*.2008: 737-740.
39. Schwartz E, Kroger M, Rak B. IS150: distribution, nucleotide sequence and phylogenetic relationships of a new *E.coli* insertion element. *Nucleic Acids Research*. 1988: 16;6789-6802.
40. Bromberg R, Grishin NV, Otwinowski Z. Phylogeny Reconstruction with Alignment-Free Method That Corrects for Horizontal Gene Transfer. *Plos Comput Biol*. 2016: 12(6): e1004985.doi:10.1371/journal.pcbi.1004985.
41. Parks SL, Goldman N. Maximum likelihood inference of small trees in the presence of long branches. *Syst. Biol*. 2014: 63;798-811.
42. Gao B, Gupta RS. Conserved indels in protein sequences that are characteristic of the phylum Actinobacteria; *International Journal of Systematic and Evolutionary Microbiology*; 2005; 55: 2401–2412.
43. Thorsness PE, White KH, Fox TD. Inactivation of YME1, a member of the ftsH-SEC18-PAS1-CDC48 family of putative ATPase-encoding genes, causes increased escape of DNA from mitochondria in *Saccharomyces cerevisiae*. *Mol. Cell. Biol*. 1993: 13;5418-5426.
44. Srivastava R, Pisareva T, Norling B. Proteomic studies of the thylakoid membrane of *Synechocystis* sp. PCC 6803. *Proteomics*.2005: 5;4905–4916.
45. Krynická V, Tichy M, Krafl J, Yu J, Kaňa R, Boehm M. Two essential FtsH proteases control the level of the Fur repressor during iron deficiency in the cyanobacterium *Synechocystis* sp. PCC6803. *Mol Microbiol*. 2014: 94;609–624.
46. Hooper NM. Families of zinc metalloproteases. *FEEBS Lett*. 1994: 354;1-6.
47. Kent AG, Dupont CL, Yooseph S, Martiny AC. Global biogeography of *Prochlorococcus* genome diversity in the surface ocean. *The ISME Journal*. 2016: 1-10.