UNIVERSIDADE FEDERAL DO PAMPA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS

EVELISE LEIS CARVALHO

A ALGA ANTÁRTICA *Prasiola crispa*: UM MODELO PARA ENTENDER A VIDA EM CONDIÇÕES EXTREMAS

São Gabriel
2023

EVELISE LEIS CARVALHO

A ALGA ANTÁRTICA *Prasiola crispa*: UM MODELO PARA ENTENDER A VIDA EM CONDIÇÕES EXTREMAS

Tese apresentada ao Programa de Pós-Graduação *Stricto sensu* em Ciências Biológicas da Universidade Federal do Pampa, como requisito parcial para obtenção do Título de Doutora em Ciências Biológicas.

Orientador: Prof. Dr. Juliano Tomazzoni Boldo
Co-orientador: Prof. Dr. Paulo Marcos Pinto

São Gabriel
2023

C331a Carvalho, Evelise Leis
A alga Antártica Prasiola crispa: um modelo para entender a vida em condições extremas / Evelise Leis Carvalho.
102 p.

Tese(Doutorado)-- Universidade Federal do Pampa, DOUTORADO EM CIÊNCIAS BIOLÓGICAS, 2023.
"Orientação: Juliano Tomazzoni Boldo".

1. filogenômica. 2. transcriptômica. 3. genoma plastidial. 4. genoma mitocondrial . 5. sequenciamento de RNA. I. Título.

EVELISE LEIS CARVALHO

A ALGA ANTÁRTICA *Prasiola crispa*: UM MODELO PARA ENTENDER A VIDA EM CONDIÇÕES EXTREMAS

Tese apresentada ao Programa de Pós-Graduação *Stricto sensu* em Ciências Biológicas da Universidade Federal do Pampa, como requisito parcial para obtenção do Título de Doutora em Ciências Biológicas.

Tese defendida e aprovada em 17 de maio de 2023.

Banca examinadora:

_____

Prof. Dr. Juliano Tomazzoni Boldo

Orientador

Unipampa

_____

Profª. Drª. Lívia Kmetzsch Rosa e Silva

UFRGS

_____

Profª. Drª. Lucia Helena do Canto Vinadé

PPGCB - Unipampa

_____

Profª. Drª. Marícia Fantinel D'Ávila

Unipampa

Assinado eletronicamente por **JULIANO TOMAZZONI BOLDO**, **PROFESSOR DO MAGISTERIO SUPERIOR**, em 22/05/2023, às 10:48, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.

Assinado eletronicamente por **LUCIA HELENA DO CANTO VINADE**, **PROFESSOR DO MAGISTERIO SUPERIOR**, em 22/05/2023, às 11:06, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.

Assinado eletronicamente por **Lívia Kmetzsch Rosa e Silva**, **Usuário Externo**, em 12/06/2023, às 16:08, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.

Assinado eletronicamente por **MARICIA FANTINEL DAVILA**, **BIOLOGO**, em 13/06/2023, às 14:31, conforme horário oficial de Brasília, de acordo com as normativas legais aplicáveis.

A autenticidade deste documento pode ser conferida no site https://sei.unipampa.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1131427** e o código CRC **BB8B1257**.

Ao Paulo, Sofia e Alice, cujas as palavras são insuficientes para expressar o amor e a gratidão de tê-los em minha vida.

AGRADECIMENTOS

Agradeço aos meus pais, Indiara e Evaldo, que muitas vezes renunciaram seus sonhos para que eu pudesse realizar os meus. O amor, o apoio e a dedicação que sempre tiveram por mim, foram a base que eu pudesse alcançar minhas conquistas. Muito obrigada por serem meu apoio e meu refúgio! Eu amo vocês!

Aos meus avós Carlos e Elisa, pelo exemplo de ser humano e pelo apoio de todas as formas. Obrigada por tudo!

Ao meu padrasto, Antônio por me acolher como uma das tuas gurias. Obrigada por todo carinho e apoio.

Ao meu orientador, Prof. Juliano Boldo, meu muito obrigada não só pelos ensinamentos desde a graduação, mas também pela amizade e companheirismo. És um exemplo de competência, dedicação e simplicidade, sempre preocupado, não só com formação acadêmica, mas com ser humano.

Ao meu co-orientador, Prof. Paulo Pinto, por me acolher, desde a iniciação científica até o doutorado. Durante estes 12 anos de convívio pude aprender diariamente com um profissional brilhante, um professor e um cientista que são referencia para mim. Obrigada por todos os ensinamentos, paciência, pelo apoio constante. Se cheguei até aqui, foi porque o teu incentivo fez a diferença na minha caminhada.

Aos colegas e amigos do Laboratório de Proteômica Aplicada, Darlene, Pedro e Kimberli, pelos momentos de convívio.

Ao amigo e grande cientista Dr. Gabriel Wallau, pela grande contribuição neste trabalho.

Ao Núcleo de Estudos da Vegetação Antártica (NEVA) pelas coletas da alga.

À Universidade Federal do Pampa e aos professores do Programa de Pós-Graduação em Ciências Biológicas.

À comissão examinadora desta dissertação pela disponibilidade.

Aos órgãos financiadores CNPq, CAPES, FAPERGS, INCT-APA.

A todos que contribuíram para a realização deste trabalho.

E por fim, gostaria de agradecer aqueles que são tudo de mais importante na minha vida: Sofia, Alice e Paulo.

Às minhas meninas, Sofia e Alice, minhas princesas que mudaram toda a minha vida pelo simples fato de existirem. Vocês são a minha força, meu incentivo diário e também a minha maior fraqueza... vocês me transformam a cada dia. Obrigada pelos longos abraços, beijinhos e olhares que fazem o meu coração transbordar amor. Como eu canto para vocês, todos os dias "You are my sunshine, my only sunshine/You make me happy when skies are grey/You'll never know, dear, how much I love you…" nunca esqueçam que a mamãe AMA vocês, incondicionalmente!

Ao Paulo, meu amor, acredito que um "muito obrigada" não traduziria o tamanho da minha gratidão, então faço minhas as palavras da música "De Fé" (Engenheiros do Hawaii), "Sempre que eu preciso/Me desconectar/Todos os caminhos/Levam ao mesmo lugar/É meu esconderijo/O meu altar/Quando todo mundo/Quer me crucificar/Quando o tempo fecha/E o céu quer desabar/Perto do limite/Difícil de aguentar/Eu volto pra casa/E te peço pra ficar/Em silêncio/Só ficar/Eu tenho muitos amigos/Tenho discos e livros/Mas quando eu mais preciso/Eu só tenho você/Tenho sorte e juízo/Cartão de crédito/E um imenso disco rígido/Mas quando eu mais preciso/Eu só tenho você/Quando eu mais preciso/Eu só tenho você..." Ik hou van je!

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

“Eu gostaria de aproveitar esse momento para dizer a todas as garotas por aí que sonham em seguir a ciência como profissão: vá atrás. É o melhor trabalho do mundo. E se alguém lhe dizer que você não pode, não escute.”

(Amy Farrah Fowler em The Big Bang Theory 12x24 “A Síndrome de Estocolmo”)

RESUMO

A Antártica, localizada no Polo Sul da Terra, possui padrões de distribuição e biodiversidade relacionados com fatores abióticos. A vegetação é limitada a espécies que possuem estratégias moleculares para sobrevivência, como a expressão diferencial de genes responsivos à baixas temperaturas. Dentre estas, está a alga *Prasiola crispa*, encontrada em áreas de degelo da Antártica. Informações moleculares sobre esta espécie ainda são escassas. Com o avanço das tecnologias de sequenciamento, a geração de novos dados possibilita análises filogenômicas e transcriptômicas. Os estudos filogenômicos consideram processos que atuam em genomas inteiros. Genomas de organelas são ótimas fontes de dados filogenéticos, incluindo sequências de proteínas e informações sobre o conteúdo genético e arquitetura. Já as análises transcriptômicas se valem do conjunto de todos os transcritos expressos. Neste trabalho, nós utilizamos os dados dos genomas do cloroplasto (cpDNA) e mitocondrial (mtDNA) para inferir as relações evolutivas de *P. crispa* e outras espécies de algas verdes, assim como uma análise de genômica estrutural. Também sequenciamos, montamos e anotamos o transcriptoma de *P. crispa*, visando identificar os produtos gênicos relacionados com a capacidade de sobrevivência no continente Antártico. Através da análise filogenômica baseada no cpDNA, pudemos observar a formação do clado Prasiola, composto por *P. crispa, Prasiolopsis sp.* e *Stichococcis bacilaris.* Os resultados da análise com mtDNA demostraram o agrupamento de *P. crispa* e outras espécies de algas da classe Trebouxiophyceae. A análise sintênica de *P. crispa* e espécies de plantas verdes relacionadas evolutivamente apresentou poucos blocos gênicos sintênicos. Na análise transcriptômica, identificamos 17.201 contigs. As informações geradas neste trabalho demonstram que os genomas acessórios são valiosas ferramentas para análises evolutivas e os dados do transcriptoma fornecem os primeiros *insights* sobre a dinâmica molecular de *P. crispa* no ambiente Antártico. Os genes e biomoléculas envolvidas no processo de sobrevivência de *P. crispa* são de grande interesse e potencial na área da Biotecnologia.

Palavras-chave: *Prasiola crispa*, filogenômica, transcriptômica, genoma plastidial, genoma mitocondrial, sequenciamento de RNA.

ABSTRACT

Antarctica, located at the Earth's South Pole, has broad distribution and biodiversity patterns related to abiotic factors. Plants are limited to species that evolved molecular strategies for survival, such as the differential expression of low temperature responsive genes. Among these is the algae *Prasiola crispa*, found in thawed areas of Antarctica. Molecular information about this species is still scarce. With the development of sequencing technologies, producing new data enabled more accurately phylogenomic and transcriptomic analysis. Phylogenetic studies consider processes that act on entire genomes. Organelle genomes are great sources of phylogenetic data, including protein sequences and information about genetic content and architecture. Transcriptomic analysis uses the set of all expressed transcripts. In this work, we used chloroplast (cpDNA) and mitochondrial (mtDNA) genome data to infer the evolutionary relationships of *P. crispa* and other green algae species, as well as a structural genomics analysis. We also sequenced, assembled and annotated the transcriptome of *P. crispa*, aiming to identify gene products related to the ability to survive in the Antarctic continent. Through cpDNA-based phylogenomic analysis, we were able to observe the formation of the Prasiola clade, composed of *P. crispa*, *Prasiolopsis sp.* and *Stichococcis bacilaris.* The results of the mtDNA analysis demonstrated the grouping of *P. crispa* and other species of algae of the class Trebouxiophyceae. Synthenic analysis of *P. crispa* and evolutionarily related green plant species showed few synthetic gene blocks. In the transcriptomic analysis, we identified 17,201 contigs. The information generated in this work demonstrates that accessory genomes are valuable tools for evolutionary analysis and the transcriptome data provided the first insights into the molecular dynamics of *P. crispa* in the Antarctic environment. The genes and biomolecules involved in the survival process of *P. crispa* are of great interest and potencial in the field of Biotechnology.

Keywords: *Prasiola crispa*, phylogenomics, transcriptomics, plastid genome, mitochondrial genome, RNA sequencing.

LISTA DE FIGURAS

INTRODUÇÃO

ARTIGO 1

ARTIGO 2

ANEXO I

ANEXO II

ANEXO III

ANEXO IV

ANEXO V

LISTA DE TABELAS

INTRODUÇÃO

ARTIGO 2

ANEXO I

ANEXO II

ANEXO IV

LISTA DE ABREVIATURAS

BLAST	*Basic Local Alignment Search Tool*
cDNA	Ácido desoxirribonucleico complementar
COG	*Cluster of Orthologous Groups*
cpDNA	Ácido desoxirribonucleio plastidial
ddNTPs	Didesoxinucleotídeos trifosfatados
DNA	Ácido desoxirribonucleico
ESTs	*Expressed Sequence Tag*
GO	*Gene Onthology*
IBPs	Proteínas de ligação ao gelo
IR	Região invertida e repetida
kb	Kilobase
LSC	Região grande de cópia única
mtDNA	Ácido desoxirribonucleio mitocondrial
NCBI	*National Center for Biotechnology Information*
pb	Pares de base
PCR	Reação em cadeia da polimerase
RNA	Ácido ribonucleico
RNA-seq	Sequenciamento de RNA
rRNA	Ácido ribonucleico ribossomal
SOLiD	*Sequencing by Oligonucleotide Ligation and Detection*
SSC	Região curta de cópia única
tRNA	Ácido ribonucleico transportador

SUMÁRIO

1. INTRODUÇÃO E REVISÃO BIBLIOGRÁFICA

1.1. A Antártica

O Continente Antártico, localizado no eixo Polo Sul da Terra é isolado de outras massas de terra pelo Oceano Austral. A Terra do Fogo, uma província no extremo sul da Argentina, é a porção de terra mais próxima da Antártica, localizada a 800 km do extremo norte da Península Antártica (Benninghoff, 1987).

A área total da Antártica é de aproximadamente 14.000.000 km^2, sendo a maior parte coberta por gelo e neve permanentemente, com camadas em média de 1,6 km de espessura e apenas 0,4% do território está livre de gelo, disponíveis para a colonização por plantas e animais (Martínez-Rosales, et al*.,* 2012; Convey, et al., 2008). Grande parte da terra livre de gelo é encontrada ao longo da Península Antártica, nas ilhas associadas e na região costeira ao redor do Continente Antártico (Figura 1). A temperatura média na Estação Comandante Ferraz é -2,8 ºC, chegando à -50 ºC no platô Antártico (Antártica Ocidental) (Simões, 2011).

Na proximidade da costa (Antártica Marítima) as temperaturas diminuem, as máximas diárias no verão estão geralmente entre 0 ºC e 6 ºC e as mínimas estão entre -2 ºC e -4 ºC. A temperatura média no verão é cerca de 2,8 ºC na Ilha Robert (Zúñiga, et al*.*, 1996). No inverno, as temperaturas variam entre -10 ºC e -20 ºC, conforme foi relatado na Estação Casey por Jackson e Seppelt (1995). A temperatura mínima registrada no continente foi de -89 ºC, em julho de 1983, na Estação de Vostok (Phillpot, 1985).

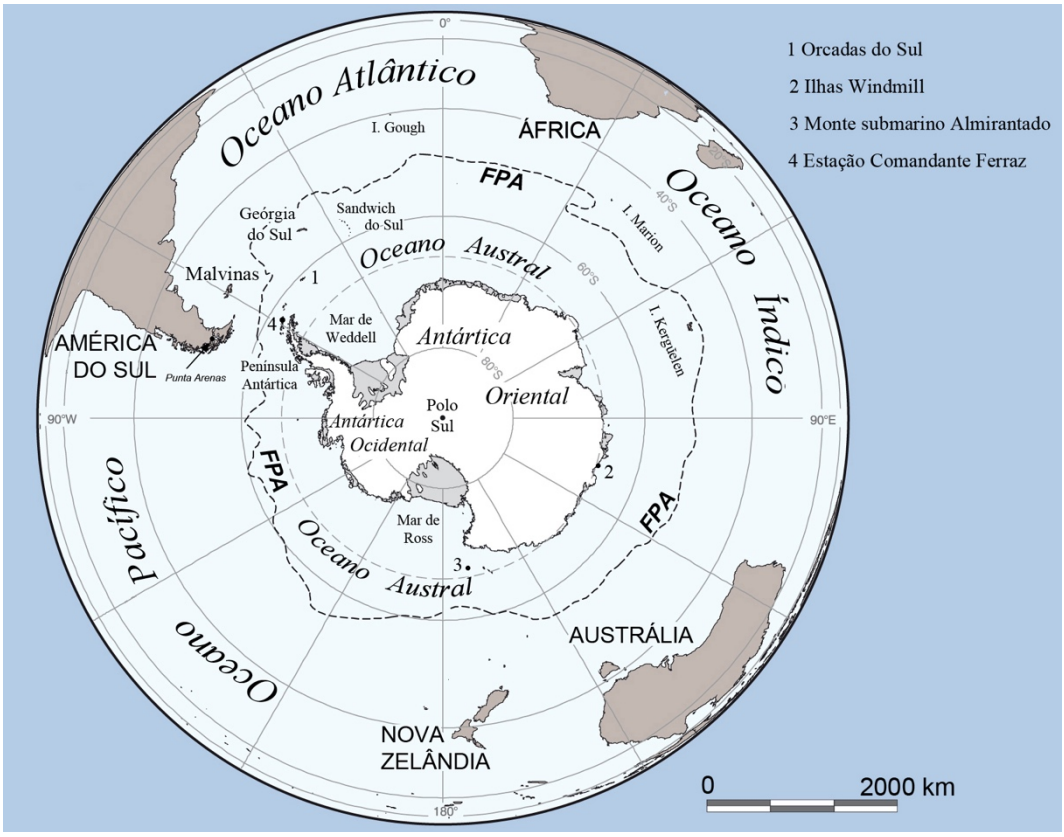
Do interior para a costa do continente, especialmente em direção à costa da península, a temperatura e a precipitação aumentam, resultando na presença de água líquida. A precipitação na Antártica Marítima cai principalmente como neve, embora a chuva seja frequente no verão (Xiong, et al., 2000).

Os solos antárticos, especialmente nas áreas costeiras, são caracterizados por um alto teor de partículas minerais e carbono orgânico, baixa relação C/N e pH ácido, sendo enriquecidos em nutrientes devido à influência do spray marinho e guano das aves marinhas (Beyer, et al., 2000). As condições do *permafrost* e o alto teor de água do solo podem ser restrições importantes para o crescimento de plantas nas regiões antárticas. Após as mudanças

térmicas, a umidade do solo pode passar de um estado de saturação no verão para a aridez fisiológica no inverno (Kappen, 2000).

Pouco ainda se sabe sobre o efeito dos nutrientes no desempenho da vegetação antártica (Kappen, 2000), enquanto o microclima, incluindo fatores como disponibilidade de umidade, baixa temperatura e velocidade do vento ao nível do solo, tem uma grande influência no crescimento das plantas (Beyer, et al., 2000). A cobertura de neve da Antártica é um importante fator microclimático na proteção das plantas do vento, gelo soprado pelo vento, partículas de areia e temperatura extrema (Alberdi, et al., 2002).

Figura 1. Localização e limites da região Antártica.



Fonte: Simões, 2011.

Outros fatores microclimáticos importantes para o desenvolvimento neste habitat são a duração do dia e a irradiância solar. Devido à alta latitude do continente, ocorrem diferenças consideráveis entre os níveis de radiação solar e na duração do dia entre o verão e o inverno. Durante a estação de crescimento (dezembro a fevereiro), a duração do dia é cerca de 20 horas

na Península Antártica. A conjunção da alta irradiação e baixa temperatura pode danificar o aparelho fotossintético, causando uma redução na fotossíntese, fenômeno conhecido como fotoinibição (Demmig-Adams; Adams, 1992).

Os padrões de distribuição e biodiversidade da vegetação estão diretamente relacionados com a temperatura, cobertura de gelo, disponibilidade de luz, oxigênio e água. A biodiversidade vegetal da Antártica é composta por algas, musgos, líquens, plantas hepáticas e duas espécies de plantas com flores (Singh, et al., 2010).

Algas verdes são organismos bem adaptados ao ambiente Antártico, sendo um importante produtor primário, dominante nas condições ambientais extremas (Arrigo, et al., 1997).

1.2. As algas verdes

As algas verdes são um grupo diverso e ecologicamente importante de eucariotos que incluem formas unicelulares e uma grande variedade de formas multicelulares. Juntamente com as algas vermelhas, glaucófitas e as plantas terrestres, as algas verdes são archaeplastida, ou seja, originaram-se através de um processo endossimbiótico, onde uma célula hospedeira eucariótica heterotrófica capturou uma cianobactéria, que se integrou de forma estável, sendo, finalmente, transformado em um plastídio. Acredita-se que esse evento, endossimbiose primária, tenha acontecido entre 1 a 1,5 bilhões de anos atrás, marcando a origem dos primeiros eucariotos fotossintéticos (Zhang, et al., 2019).

As algas verdes são um grupo parafilético em relação às plantas terrestres, e que juntas formam o clado Viridiplantae (Umen, Herron, 2021).

Viridiplantae, ou plantas verdes, é um clado com aproximadamente 500 mil espécies que exibem uma diversidade espantosa de formas de vida (Ruhfel, et al., 2014). Esta linhagem fracionou-se em duas divisões principais: Streptophyta, contendo as algas verdes do filo Charophyta, plantas terrestres (embriófitas) e Chlorophyta, que contém a maioria das algas verdes descritas (Lewis, McCourt, 2004).

As carófitas possuem poucos *taxa* mas que são bem diversificados, exibindo uma ampla gama de formas como organismos unicelulares, filamentosos ou "parenquimatosos" (Lewis, McCourt, 2004), sendo importantes modelos para estudos de estruturas como a parede celular

e outros mecanismos adaptativos que foram importantes para colonização do ambiente terrestre (Sørensen, et al., 2012).

O filo Chlorophyta possui a maior quantidade de espécies de algas verdes descritas, com grande diversidade morfológica e ecológica (Lewis, McCourt, 2004). As espécies constituintes deste táxon foram classificadas com base na ultraestrutura do aparelho flagelar e citocinese durante a mitose (Mattox, Stewart, 1984). Ficologistas reconhecem quatro classes de algas verdes pertencentes ao filo: Chlorophyceae, Trebouxiophyceae, Ulvophyceae e Prasinophyceae, este último o único grupo parafilético (Turmel, et al., 1999).

A classe Trebouxiophyceae abrange as algas das águas marinhas e doces (Servín-Garcideueñas, Martínez-Romero, 2012). Dentre as diversas ordens que compõe esta classe, Prasiolales compreende um dos mais generalizados e versáteis grupos de algas verdes. Esta ordem inclui espécies distribuídas em ambientes marinhos, de água doce, habitats terrestres de regiões polares e frio temperado (Rodríguez, et al., 2007). A morfologia destas algas varia desde filamentos unisseriados para talos em forma de fita, lâminas expandidas ou de pacotes como colônias, que são caracterizadas por uma grande plasticidade fenotípica relacionada a fatores ambientais (Rindi, et al., 2007).

Espécies da família Prasiolaceae são caracterizadas por um talo frondoso de folhas com forma variável, com ou sem uma nítida *holdfast*, células dispostas em linhas verticais ou horizontais e um cloroplasto axial estrelado contendo um pirenóide central (John, 2002).

O gênero *Prasiola* está entre as algas presentes nas áreas de degelo do continente Antártico, onde a espécie mais relatada é *Prasiola crispa* (Lightfoot) Kützing (Figura 2), (Wiencke, Clayton, 2002; Convey, 2007; Kováčik, Pereira, 2001). Possui como característica células individuais com cloroplasto axial estrelado e apenas um pirenoide. Sua reprodução pode ser de modo sexuado (oogamia) ou assexuado (esporos ou fragmentação) e sua distribuição biogeográfica é cosmopolita, presentes do Ártico ao continente Antártico (Kováčik, Pereira, 2001; Moniz, et al., 2012).

No continente Antártico *P. crispa* é um dos organismos mais encontrados e importante produtor primário (Convey, 2006). Esta alga está presente nas regiões supralitorais, formando grandes tapetes verdes, sendo encontrada principalmente em locais próximos a colônias de aves, substrato rico em guano, onde há altas concentrações de nitrogênio e ácido úrico (Kováčik, Pereira, 2001; Graham, et al*.,* 2009).

Por tolerar repetidos ciclos de descongelamento e congelamento durante a primavera e outono, temperaturas negativas durante o inverno e altos níveis de radiação ultravioleta durante o verão (Lud, et al., 2001; Kosugi, et al., 2010), supõe-se que *P. crispa* apresente expressão diferencial de genes responsivos ao frio. Dentre estes, e principalmente por se tratar de um organismo antártico, acredita-se que *P. crispa* possua em seu genoma genes codificadores de proteínas de ligação ao gelo (*ice binding proteins* - IBPs).

IBPs são polipeptídios expressos em uma ampla gama de organismos, que permitem a sobrevivência das células quando expostas à baixas temperaturas (Ustun, et al., 2015; Venketesh, et al.; 2008). A propriedade essencial de uma IBP é a capacidade de adsorção de um ou mais planos do gelo, que tem como consequência natural a alteração do formato dos cristais de gelo, ocasionando uma ligação irreversível (Randy, et al., 2017). A ausência destas proteínas e de outros agentes crioprotetores, em organismos expostos à baixas temperaturas, provocam o congelamento da água e dos fluidos intracelulares, causando desidratação, choque osmótico e rompimento da membrana celular pela formação de cristais de gelo (Kuiper, et al., 2015).

Figura 2. *Prasiola crispa* no continente Antártico.

Fonte: Graciele Alves de Menezes, 2015.

A expressão destes genes durante a exposição à baixas temperaturas ainda permanece pouco estudada, bem como se estes genes também estão envolvidos na proteção da mesma à alta radiação UV do continente Antártico.

Sendo assim, pela sua capacidade de colonizar um ambiente tão extremo, *P. crispa* deve possuir mecanismos adaptativos naturalmente selecionados durante sua evolução. Os genes e biomoléculas envolvidos nestes importantes mecanismos são de grande interesse e potencial na área da Biotecnologia.

1.3. Filogenômica

Durante as últimas décadas, a análise filogenética tornou-se uma ferramenta importante em toda a biologia para comparar informações sobre genes, indivíduos, populações e espécies (Hills, 1997).

As análises filogenéticas fornecem informações sobre as relações sistemáticas entre as espécies, onde, tradicionalmente, caracteres morfológicos ou ultra estruturais são utilizados para a reconstrução filogenética (Behura, 2015). Contudo, o número de marcadores morfológicos confiáveis para análises filogenéticas pode ser limitado (Blair, Murphy, 2011).

Com a introdução do sequenciamento do DNA no início da década de 1970, a utilização de dados moleculares nas reconstruções de filogenias ganhou notoriedade. O gene do RNA ribossômico 18S (rRNA 18S), foi utilizado como referência para muitas construções de árvores filogenéticas devido à sua sequência ser conservada entre as espécies eucarióticas. Subsequentemente, estudos foram conduzidos para combinar múltiplos genes para inferir relações filogenéticas (Legg, et al., 2013).

Com o advento dos métodos de sequenciamento de alta performance, a filogenética tomou um novo rumo nos últimos anos. A filogenômica, a integração da filogenética com os dados do genoma, surgiu como uma poderosa abordagem para estudar a sistemática e a evolução das espécies (Behura, 2015).

O termo filogenômica é usado principalmente para se referir a uma extensão da filogenética que considera não apenas a evolução de sequências específicas de nucleotídeos ou aminoácidos, mas também de processos mais amplos que atuam em genomas inteiros (Duchêne, 2021). O uso de genomas de organelas em estudos filogenômicos tem aumentado

consideravelmente, visto que fornecem inúmeros dados filogenéticos, incluindo as sequências de proteínas e nucleotídeos, bem como informações sobre o conteúdo genético e arquitetura (Letsch, Lewis, 2012).

A caracterização da diversidade biológica e função das algas que colonizam ambientes polares ainda são alvo de pesquisas. Em particular, ainda há muita incerteza quanto à identificação e taxonomia das espécies de algas encontradas na Antártica (Hoham, 2020).

Para *P. crispa*, os dados moleculares disponíveis consistem principalmente de sequências dos genes *rbcL* (Rindi, et al., 2007; Moniz, et al., 2012), rRNA 18S (Rindi, et al., 2004; Friedl; O'Kelly, 2002), *tufA* (Moniz, et al., 2014), *psaB* (Moniz, et al., 2012; Novis, et al., 2013), *rpoB* (Novis, et al., 2013) e *atpB* (Moniz, et al., 2012). Ademais, o genoma do cloroplasto e da mitocôndria de *P. crispa* foram sequenciados pelo nosso grupo (Carvalho, et al., 2015).

O genoma plastidial de *P. crispa* consiste em uma única molécula com um comprimento total de 196.502 pb e conteúdo G+C de 29,32%. Compreende 63 genes codificadores de proteínas, 26 tRNAs e 2 rRNAs. Entre esses, 19 genes codificadores de proteínas relacionadas à fotossíntese, como as proteínas putativas do fotossistema I e II (Carvalho, et al., 2015) (Tabela 1).

O genoma mitocondrial de *P. crispa*, foi o primeiro genoma mitocondrial sequenciado da ordem Prasiolales. Possui um total de 89.819 pb e conteúdo G+C de 29,29%. Apresenta 56 genes, compreendendo 32 genes codificantes de proteínas, 21 tRNAs e 3 rRNAs, onde foram anotados ao menos 17 genes codificadores de proteínas relacionadas ao metabolismo oxidativo mitocondrial, como as proteínas putativas do complexo I, III e IV da cadeia respiratória mitocondrial (Carvalho, et al., 2015) (Tabela 2).

A compreensão da diversidade e biogeografia de algas verdes vem se tornando um importante desafio. Embora alguns aspectos da biologia de *P. crispa* tenham sido completamente investigados, a escassez de dados moleculares impede uma avaliação precisa da sua taxonomia e posição filogenética (Moniz, et al., 2012). As informações geradas nos últimos anos têm reformulado consideravelmente a visão sobre estes organismos, revelando uma diversidade genética maior do que sugere sua morfologia simples (Leliaert, et al., 2012).

Tabela 1. Genes identificados no genoma plastidial de *Prasiola crispa.*

Categoria gênica	**Genes**
Fotossistema I	*psaB, psaC, ycf3*
Fotossistema II	*psbC, psbD, psbE, psbH, psbL, psbN, psbT, psbB, psbA, psbM*
Complexo citocromo b_6f	*petA, petB, petD*
ATP sintase	*atpA, atpB, atpH, atpI*
Ciclo de Calvin	*rbcL*
Proteínas ribossomais - Subunidade maior	*rpl14, rpl16, rpl36*
Proteínas ribossomais - Subunidade menor	*rps12, rps19*
RNA polimerases	*rpoB, rpoC1, rpoC2*
Outras	*accD, clpP, ccsA*
rRNAs	*rrn16S, rrn23S*
tRNAs	*trnC-GCA, trnfM-CAT, trnG-TCC, trnH-GTG, trnI-CAT, trnI-GAT, trnL-CAA, trnM-CAT, trnN-GTT, trnP-TGG, trnQ-TTG, trnR-ACG, trnS-GCT, trnT-TGT, trnV-CAC, trnW-CCA, trnV-TAC, trnA-TGC, trnD-GTC, trnF-GAA, trnK-TTT, trnL-TAA, trnG-GCC, trnS-TGA, trnL-TAG, trnY-GTA, trnR-TCT*

Fonte: do autor

Tabela 2. Genes identificados no genoma mitocondrial de *Prasiola crispa.*

Categoria gênica	Genes
NADH: ubiquinona oxiredutase (complexo I)	*nad1, nad2, nad3, nad4, nad4L, nad5, nad6, nad7*
Complexo citrocromo *bc1* (complexo III)	*cob_a, cob_b*
Citocromo *c* oxidase (complexo IV)	*cox1_0_a, cox1_0_b, cox1_1_a, cox1_1_b*
ATP sintase	*atp1, apt6, atp8, atp9*
rRNAs	*rrnL, rrnSa, rrnSb*
tRNAs	*trnR-ACG, trnN-GTT, trnQ-TTG, trnF-GAA, trnE-TTC, trnG-TCC, trnH-GTG, trnL-TAA, trnK-TTT, trnM-CAT, trnF-AAA, trnP-TGG, trnS-GCT, trnT-TGT, trnV-TAC, trnA-TGC, trnC-GCA, trnL-TAG, trnM-CAT, trnS-TGA, trnW-CCA*

Fonte: do autor

1.3.1. Análise de genomas plastidiais como ferramenta filogenética

Plastídios são uma das principais características distintivas da célula vegetal (Wicke, et al., 2007). Estas organelas fotossintetizantes fornecem a energia essencial para algas, plantas terrestres e alguns protozoários. Além da fotossíntese, outras vias metabólicas estão presentes nos plastídios, incluindo a biossíntese de ácidos graxos, aminoácidos, pigmentos e vitaminas (Wang, et al., 2013). A origem dos plastídios está relacionada com um evento de endossimbiose, pelo qual um organismo protozoário unicelular, através do processo de fagocitose, englobou e manteve uma cianobactéria fotossintetizante (Reyes-Prieto, et al., 2007), permitindo assim a transição de heterotrofia para autotrofia, adquirindo a capacidade de utilizar fotoenergia (Wicke, et al., 2007). Em nível genômico, esta integração envolveu a perda de genes e a transferência de muitos destes genes para o genoma nuclear do hospedeiro, fenômeno semelhante ao mtDNA (deKonin, Keeling, 2006).

Comumente o cloroplasto desenvolve-se a partir de proplastídios subdesenvolvidos (progenitor de todos os plastídios), que contém apenas vesículas, não existindo estruturas

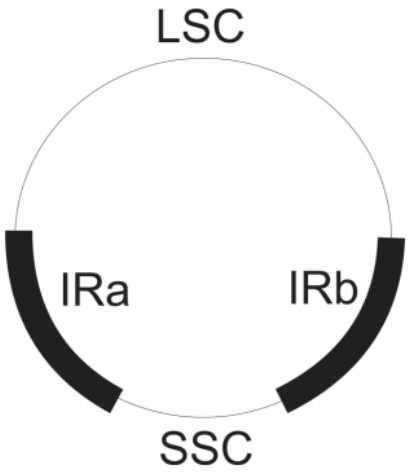
diferenciadas. Durante a diferenciação, tilacóides são formados e organizam-se uns sobre os outros, formando estruturas denominadas grana. Os tilacóides são membranas internas lipídicas entrelaçadas com complexos de proteínas, que fornecem a plataforma para as reações de fotossíntese (Pribil, Labs, Leister, 2014). O espaço interno do cloroplasto é preenchido por um fluido viscoso denominado estroma onde estão presentes o material genético (DNA), enzimas e ribossomos. Possuem membrana dupla, fruto da origem dos plastídios (Mota, 2012). Assim como as mitocôndrias, os plastídios são, em geral, de origem materna (Harrison, Kidner, 2011).

Os plastídios possuem um genoma próprio, circular e independente do DNA nuclear. O cromossomo plastidial é dividido em quatro segmentos principais, sendo duas regiões invertidas e repetidas (IR) que separam a região grande de cópia única (LSC), onde estão presentes a maioria dos genes plastidiais, e a região curta de cópia única (SSC) (Figura 3) (Kolodner, Tewari, 1979).

A organização e a expressão do genoma de plastídios estão sendo exaustivamente estudados. A estrutura e a sequência do genoma do cloroplasto são notavelmente conservadas em Streptophyta, divisão que compreende as algas verdes da classe Charophyceae e as plantas terrestres (Turmel, Otis, Lemieux, 2006). No entanto, estudos com genomas de cloroplastos de Chlorophyta, revelaram que este padrão não se aplica para este grupo de algas (Turmel, Otis, Lemieux, 2009; deCambiaire, et al., 2007). Por exemplo, as algas verdes Prasinophytes *Nephroselmis olivacea*, *Pyramimonas parkeae*, *Prasinoderma coloniale* e *Prasinococcus sp.*, abrigam em seu genoma plastidial genes conservados (*ndhJ*, *rbcR*, *rpl21*, *rps15*, *rps16*, *ycf66*, *rpl22* e *ycf65*) que nunca antes foram descritos para outras Chlorophyta. O genoma do cloroplasto de *N. olivacea* e *P. parkeae* exibe ainda um gene que codifica uma DNA primase, que se supõe ter sido adquirida de um vírus (Lemieux, Otis, Turmel, 2014b).

A Trebouxiophyceae *Leptosira terrestris*, apresenta genoma plastidial circular de 195.081 pb e não possui uma região IR, fato este que se encontra também no genoma do cloroplasto de *Chlorella vulgaris* (150.613 pb) (Wakasugi, et al., 1997), *Chlorella sorokiniana* (109.811 pb) (Orsini, et al., 2014) e *Helicosporidium sp.* (37.454 pb) (deKoning, Keeling, 2006). Por mais que estas espécies apresentem conteúdo gênico semelhante, elas diferem significativamente em densidade gênica, ordem dos genes e quantidade de íntrons.

Figura 3 - Estrutura quadripartida do cromossomo plastidial. LSC, região grande de cópia única; IRa e IRb, regiões invertidas repetidas; SSC, região curta de cópia única.



Fonte: Motta, 2012.

As características semelhantes apresentadas por *L. terrestris* e seus homólogos em Chlorophyceae sugerem que a mesma força evolutiva gerou a ausência da região IR nessas duas linhagens de algas (deCambiaire, et al., 2007).

Existem evidências que as plantas terrestres evoluíram de algas verdes e que durante esta evolução, vários rearranjos ocorreram no genoma dos cloroplastos (Graham, 1996). Portanto, para compreender o processo de evolução do genoma do cloroplasto, informações sobre sequências repetidas, regiões intergênicas e pseudogenes são extremamente úteis (Wakasugi, et al., 1997).

O tamanho e a sintenia dos genomas dos plastídios variam muito entre os grupos de algas (Kim et al., 2014). O genoma do cloroplasto da alga verde *Floydiella terrestris* (Chlorophyceae) com 521.168 pb é o maior genoma de plastídio já sequenciado, apresentando 97 genes conservados, 26 íntrons e teor A+T de 65,5% (Brouard, et al., 2010) (Figura 4A).

N. olivacea exibe um genoma plastidial de 200.799 pb e apresenta o maior repertório gênico relatado até o momento para uma Chlorophyta, 128 genes conservados, enquanto que a maioria dos genomas dos cloroplastos de Chlorophyta apresentam um conjunto reduzido de 86 a 88 genes (Lemieux, Otis, Turmel, 2014a).

A disponibilidade de dados sobre genomas de organelas em Trebouxiophyceae é ainda limitada (Jeong et al., 2014). Contudo, informações retiradas do *Organelle Genome Resources* do NCBI no ano de 2017 (http://www.ncbi.nlm.nih.gov/genomes/GenomesHome) mostram 37

genomas plastidiais de algas verdes Trebouxiophyceae disponíveis (Tabela 3). Adicionalmente, há pelo menos outros três genomas não inclusos na plataforma do NCBI, os de *Chlorella sp.* ArM0029B, (Jeong, et al., 2014), *Prasiolopsis sp.* (Lemieux, Otis, Turmel, 2014b), *Prototheca wickerhamii* (Yan, et al., 2015), (Tabela 4), totalizando-se 40 genomas plastidiais completos sequenciados desta classe de algas verdes.

O conteúdo gênico e número de íntrons de espécies desta classe é bastante diversificado (deCambiaire, et al., 2007). Além disto, estes genomas exibem uma variabilidade de tamanhos de 306.152 pb de *Prasiolopsis sp.* (Lemieux, Otis, Turmel, 2014b) a 37.454 pb da alga verde parasita não fotossintética *Helicosporidium sp.*, que codifica 26 proteínas, 3 rRNAs e 25 tRNA (Figura 4B), carecendo de todos os genes envolvidos com a fotossíntese. Esta última possui apenas um íntron no grupo I do gene tRNALeu (UAA) e espaços intergênicos minúsculos. Ainda assim, a densidade gênica é alta, com apenas 5,1% de DNA não codificante. *Helicosporidium sp.* possui o menor genoma plastidial sequenciado de qualquer Viridiplantae conhecido (deKoning, Keeling, 2006).

Tabela 3 – Exemplares da classe Trebouxiophyceae com genoma plastidial completo sequenciado.

Espécie	Acesso NCBI	Tamanho do genoma (pb)
Auxenochlorella protothecoides	NC_023775	84.576
Botryococcus braunii	NC_025545	172.826
Chlorosarcina brevispinosa	KM462875	295.314
Chlorella sorokiniana	NC_023835	109.811
'Chlorella' mirabilis	NC_025528	167.972
Chlorella variabilis	NC_015359	124.579
Chlorella vulgaris	NC_001865	150.613
Choricystis parasitica	NC_025539	94.206
Coccomyxa subellipsoidea C-169	NC_015084	175.731
Dicloster acuatus	NC_025546	169.201
Dictyochloropsis reticulata	NC_025524	289.394
Elliptochloris bilobata	NC_025548	134.677
Fusochloris perforata	NC_025543	148.459
Geminella minor	KM462883	129.187

Geminella terricola	KM462881	187.843
Gloeotilopsis sterilis	KM462877	132.626
Helicosporidium sp.	NC_008100	37.454
Koliella corcontica	KM462874	117.543
Koliella longiseta	NC_025531	197.094
Leptospira terrestres	NC_009681	195.081
Lobosphaera incisa	NC_025533	156.031
Marvania geminata	NC_025549	108.470
Microthamnion kuetzingianum	NC_025537	158.609
Myrmecia israelensis	NC_025525	146.596
Neocystis brevis	NC_025535	211.747
Oocystis solitaria	FJ968739	96.287
Pabia signiensis	NC_025529	236.463
Parachlorella kessleri	NC_012978	123.994
Paradoxia multiseta	NC_025540	183.394
Parietochloris pseudoalveolaris	KM462869	145.947
Planctonema lauterbornii	NC_025541	114.128
Pseudochloris wilhelmii	NC_025547	109.775
Stichococcus bacillaris	NC_025527	116.952
Trebouxia aggregata	EU123962 / EU124002	8.354 / 765
Trebouxiophyceae sp.	NC_018569	149.707
Watanabea reniformis	NC_025526	201.425
Xylochloris irregularis	NC_025534	181.542

Fonte: http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?opt=plastid&taxid=3041 modificado pelo autor.

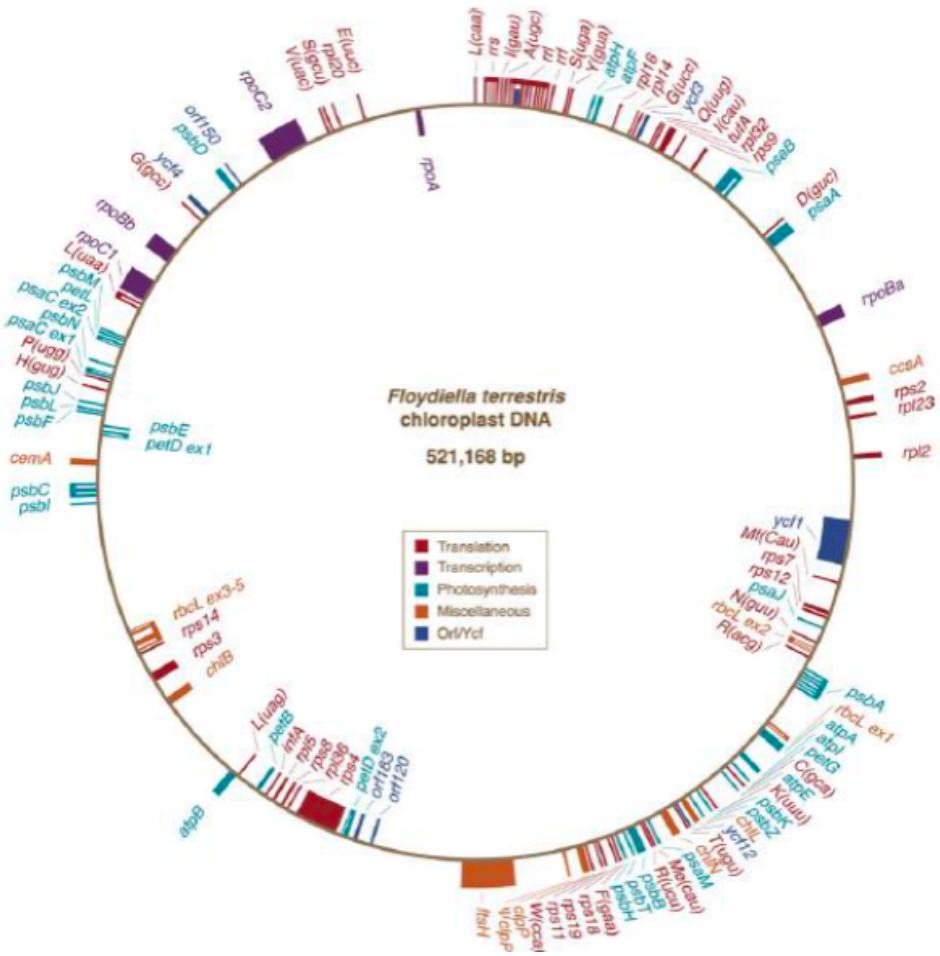
Tabela 4 – Espécies da classe Trebouxiophyceae com genoma plastidial sequenciado não depositado na plataforma *Organelle Genome Resources* (NCBI).

Espécie	Acesso NCBI	Tamanho do genoma (pb)
Chlorella sp. ArM0029B	KF554427.1	119.989
Prasiolopsis sp.	KM462862.1	306.152
Prototheca wickerhamii	KJ001761.1	55.636

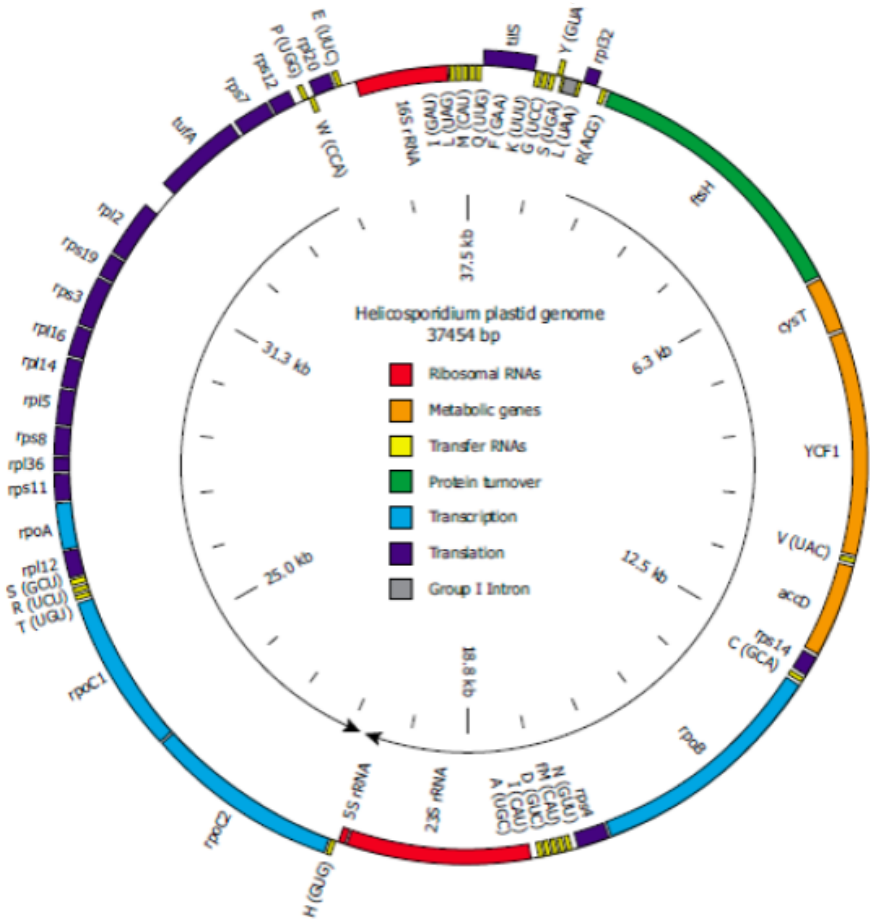
Fonte: do autor

Figura 4 - Mapas representativos dos genomas plastidiais das algas verdes *Floydiella terrestris* (Chlorophyceae) e *Helicosporidium sp.* (Trebouxiophyceae). (A) *F. terrestres* apresenta o maior genoma cloroplastidial já sequenciado até o momento, com 521.168 pb (B) contrariamente ao que é observado em *Helicosporidium sp*., que com 37.454 pb possui o menor genoma de plastídios já sequenciado do Viridiplantae.

A



B



Fonte: Brouard, et al., 2010; deKoning, Keeling, 2006.

A ampla quantidade de informações contida no genoma dos plastídios tem demonstrado que o material genético desta organela pode ser empregado como uma ferramenta adequada e de valor inestimável para a filogenia molecular (Wang, et al., 2013).

Análises filogenéticas inferidas a partir do gene nuclear 18S rRNA muitas vezes são incapazes de desvendar as inter-relações das linhagens de algas verdes. Contudo, filogenias com base na análise de genomas plastidiais são bem sucedidas na resolução de questões distintas referentes às relações de algas (Brouard, et al., 2010). Além de oferecer um grande conjunto de genes para análises filogenômicas, os genomas de cloroplastos revelam as características estruturais do genoma para validação das filogenias (Turmel, et al., 2008).

Em um estudo, Lemieux, Otis e Turmel (2014b) valeram-se das sequências de aminoácidos de 79 genes codificantes de proteínas de 63 espécies de algas verdes para análises filogenômicas entre as principais linhagens de Trebouxiophyceae.

O sequenciamento completo de genomas plastidiais com tamanho relativamente pequeno (aproximadamente 150 kb) foram tecnicamente possíveis desde meados da década de 1980. Com uso das tecnologias de sequenciamento de alta performance o custo e a dificuldade para o sequenciamento de genomas plastidiais foram drasticamente reduzidos o que consequentemente aumentou o número de genomas plastidiais disponíveis (Ruhfel, et al., 2014).

Os dados da sequência do genoma dos plastídios transformaram a sistemática vegetal e contribuíram para visão atual das relações entre plantas (Ruhfel, et al., 2014), oferecendo uma riqueza de dados filogeneticamente informativos que são relativamente fáceis de obter e utilizar (Wicke, et al., 2007), se tornando uma abordagem valiosa para inferir relações entre os eucariontes fotossintetizantes (Lemieux, Otis, Turmel, 2014b).

1.3.2. Análise de genomas mitocondriais como ferramenta filogenética

Mitocôndrias são onipresentes em células eucarióticas e executam um gama de funções celulares essenciais (Douce, 1985). Consideradas a “casa de força” das células, fornecem a energia necessária para as atividades celulares (Hammani, Giege, 2014), mas também estão envolvidas nos processos de sinalização celular, regulação da proliferação celular, diferenciação e sobrevivência (Merrill, Strack, 2014). Nas plantas, além de prover a energia celular e a respiração, estas organelas estão envolvidas em outras vias metabólicas incluindo a assimilação do nitrogênio, fotorrespiração, metabolismo do dióxido de carbono, fotossíntese em plantas C4, metabolismo ácido das crassuláceas, armazenamento de carbono e nitrogênio durante a germinação das sementes (Douce, 1985) e também desempenham um papel na biossíntese de aminoácidos e ácidos graxos (Picault, et al., 2004).

As mitocôndrias originaram-se a partir de um procarioto de vida livre consumidor de oxigênio e que se acredita ter sido capturado, como endossimbionte, por outra célula hospedeira procariótica.Durante a simbiose na célula hospedeira, o genoma mitocondrial de plantas sofreu uma perda massiva de conteúdo gênico, onde parte foi transferido para o núcleo da célula hospedeira, codificando apenas um conjunto parcial dos componentes das membranas de

transdução de energia e componentes da maquinaria de expressão gênica (Hammani, Giege, 2014). Um bom exemplo desta perda é o genoma mitocondrial da Streptophyta *Marchantia polymorpha,* o qual carece de alguns tRNAs, proteínas ribossomais e subunidades do complexo succinato-ubiquinona redutase (complexo respiratório II), o que condiz com a hipótese de transferência dos genes mitocondriais para o núcleo ao longo da evolução (Schuster, Brennicke, 1994).

Dois padrões distintos de evolução do mtDNA foram atribuídas as espécies do filo Chlorophyta (Pombert et al., 2006). Os genomas com um padrão ancestral de evolução mantiveram vestígios claros da sua ancestralidade eubacteriana (Gray, Burger, Lang, 1999) com a estrutura do genoma compacta, repertório gênico substancial e sequências gênicas conservadas. Em contraste, o padrão derivado tem sido atribuído aos mtDNA de Chlorophyta que se afastam radicalmente do padrão ancestral, com pouca ou nenhuma evidência dos traços primitivos, apresentando uma perda extensa de genes, gerando um genoma diminuto em tamanho e conteúdo gênico, divergência acentuada no DNA ribossomal e estrutura do rRNA (manifestada no truncamento da sequência e fragmentação dos genes do rRNA) e acelerada taxa de divergência da sequência, tanto para codificação de proteínas quanto para genes do rRNA (Gray, Burger, Lang, 1999; Pombert, et al., 2006).

O mtDNA da alga verde Prasinophyceae *Nephroselmis olivacea* (Turmel, et al., 1999) e da Trebouxiophyceae *Prototheca wickerhamii* (Wolff, et al., 1994) apresentam o padrão ancestral de evolução, enquanto que o mtDNA da Chlorophyceae *Chlamydomonas reinhardtii* (Michaelis, Vahrenholz, Pratje, 1990) e da Pedinophyceae *Pedinomonas minor* (Turmel, et al., 1999) foram classificados como sendo do padrão derivado (Pombert, et al., 2006). Assim, os dados disponíveis apontam para a existência de padrões distintos da evolução do genoma mitocondrial, não só entre Streptophyta e Chlorophyta, mas também dentro do filo Chlorophyta (Turmel, et al., 1999).

As mitocôndrias de Streptophyta e Chlorophyta adquiriram características únicas que as diferem do seu ancestral procariótico assim como das mitocôndrias de outros reinos (Hammani, Giege, 2014). Em Chlorophyta o mtDNA sofreu mudanças radicais na estrutura, conteúdo gênico, organização dos genes, quantidade de íntrons e são altamente variáveis em tamanho (Lewis, McCourt, 2004; Pombert et al., 2004).

De acordo com dados do *National Center for Biotechnology Information* (NCBI), no ano de 2017, quarenta genomas de mitocôndrias de Chlorophyta foram sequenciados, sendo

sete exemplares da classe Trebouxiophyceae (*Organelle Genome Resources*) (Tabela 5). Além destes, o genoma mitocondrial de *Chlorella sp*. ArM0029B também se encontra sequenciado, apresentando 65.049 pb de tamanho e depositado na plataforma *GenBank* do NCBI sobre o número de acesso KF554428.1 (Jeong, et al., 2014).

Tabela 5 – Exemplares da classe Trebouxiophyceae com genoma mitocondrial sequenciado.

Espécie	Acesso NCBI	Tamanho do genoma (pb)
Auxenochlorella protothecoides	NC_026009	57.274
Chlorella sorokiniana isolate 1230	NC_024626	52.528
Chlorella variabilis isolate NC64A	NC_025413	78.500
Coccomyxa subellipsoidea C-169	NC_015316	65.497
Helicosporidium sp. ex Simulium jonesi	NC_017841	49.343
Prototheca wickerhamii	NC_001613	55.328
Trebouxiophyceae sp. MX-AZ01	NC_018568	74.423

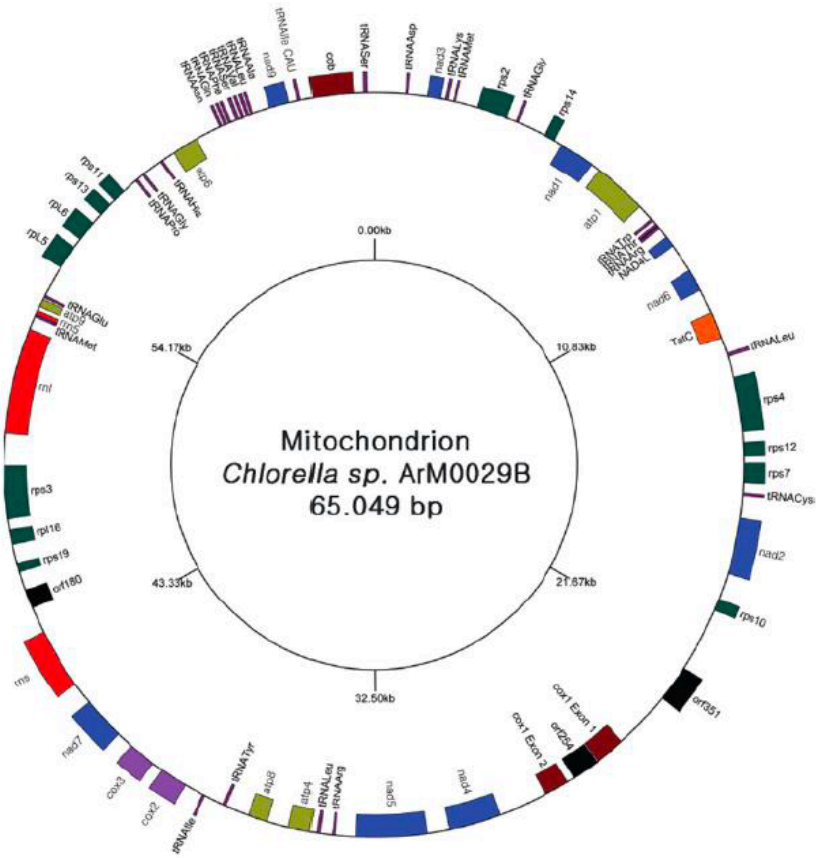
Fonte: http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?opt=organelle&taxid=3041 modificado pelo autor.

Estes genomas variam de 78.500 pb de *Chorella variabilis* (NC_025413) a 49.343 pb de *Helicosporidium sp.* (deKoning, Keeling, 2006). O mtDNA de *Chlorella sp.* ArM0029B (Figura 5) apresenta o maior repertório gênico descrito até o momento para uma Trebouxiophyceae, com 62 genes conservados estando 18 deles envolvidos no metabolismo oxidativo e apenas 1 íntron dentro do gene *cox1*, o menor número de íntrons descrito até o momento para a classe (Jeong, et al., 2014). Contudo, a diferença de tamanho dos genomas não reflete uma capacidade de codificação maior (Turmel, et al., 1999). *Chlorella sorokiniana* possui um genoma mitocondrial de 52.528 pb com 58 genes e apresenta uma densidade codificante de 97,4% (Orsini, et al., 2014) enquanto o mtDNA de *Chlorella sp.* ArM0029B exibe uma capacidade de codificação de 50,2% (Jeong, et al., 2014).

Genomas mitocondriais contém um conjunto limitado de genes que codificam proteínas e RNAs (Gray, Burger, Lang, 1999). As mitocôndrias que possuem os componentes clássicos da fosforilação oxidativa, ou seja, os complexos respiratórios I (NADH-ubiquinona redutase), II (succinato-ubiquinona redutase), III (ubiquinol-citocromo c redutase) e IV (citocromo c oxidase) e ATP sintase F1-F0 (complexo V) que normalmente contém os genes *atp6*, *atp8*, *cob*,

cox1, *cox2*, *cox3*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5* e *nad6* no mtDNA. No entanto, o mtDNA de vários exemplares da linhagem Chlorophyceae e afins normalmente não exibem os genes *atp6*, *atp8*, *cox2*, *cox3*, *nad3* e *nad4L* (Fan, Lee, 2002; Denovan-Wright, et al., 1998; Vahrenholz, 1993). O sistema de tradução das organelas, pelo qual os mRNA mitocondriais são codificados, também é composto em parte por componentes especificados pelo mtDNA. Em plantas, protistas e na maioria dos fungos, algumas proteínas ribossômicas são codificadas no mtDNA (Gray, Burger, Lang, 1999).

Figura 5 - Mapa representativo do genoma mitocondrial de *Chlorella sp.* ArM0029B.



Fonte: Jeong, et al., 2014.

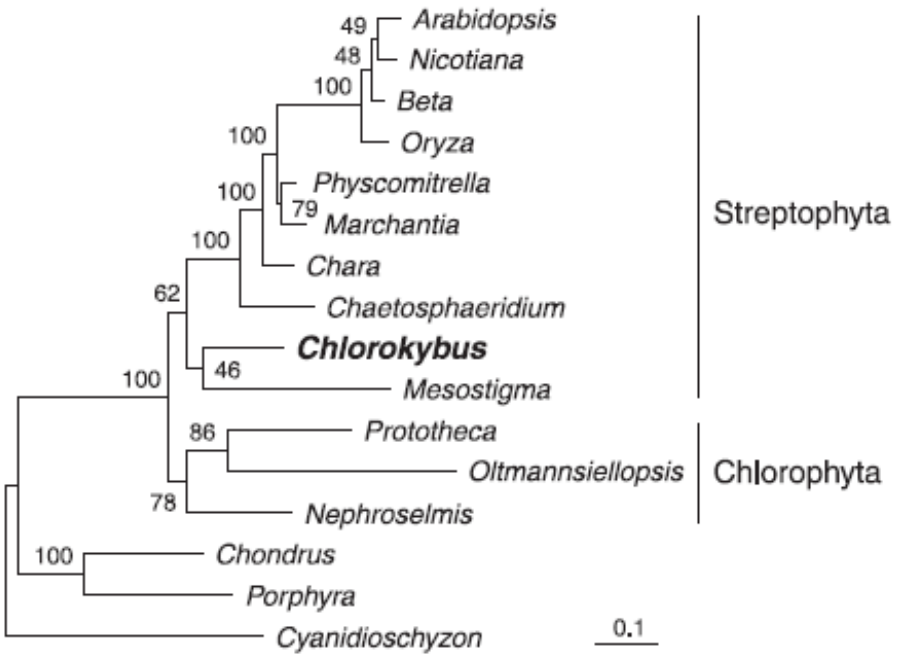
O mtDNA é considerado um marcador molecular útil para análises filogenéticas e identificação de espécies. Atualmente, o mtDNA também tem sido utilizado como DNA *barcoding* para identificação em nível de espécies e populações (Ko, et al., 2013).

Nas últimas três décadas, genes individuais do mtDNA, particularmente o gene que codifica uma citocromo c oxidase (*cox1*), foram utilizados para desvendar as relações filogenéticas (Hajibabaei, et al., 2007). Análises tendo como referência um único gene geraram filogenias atualmente aceitas. No entanto, o poder de resolução da análise de um único gene é limitado pelo pequeno conteúdo de informações obtidas, o que no caso particular das mitocôndrias é ainda mais afetado pelas extremas diferenças na composição gênica e na taxa de divergência da sequência do mtDNA de diferentes linhagens eucarióticas (Gray, Burger, Lang, 1999).

As informações obtidas a partir do sequenciamento do mtDNA completos estão sendo utilizadas por pesquisadores para o desenvolvimento de filogenias. A fim de determinar a posição filogenética de *Chlorokybus atmophyticus,* uma Streptophyta, foi analisado um conjunto de aminoácidos obtidos de dezoito genes que codificam proteínas (*atp4*, *atp6*, *atp8*, *atp9*, *cob*, *cox1*, *cox2*, *cox3*, *mttB*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *rps3* e *rps12*), comuns a dezesseis espécies de algas verdes e plantas terrestres (Figura 6) (Turmel, Otis, Lemieux, 2007). Sequências de aminoácidos traduzidas dos genes codificadores de proteínas *cob*, *cox1*, *nad1*, *nad2*, *nad4*, *nad5* e *nad6*, foram aplicados para estudos filogenéticos da alga verde Trebouxiophyceae *Chlorella sp.* ArM0029B (Jeong, et al., 2014).

As sequências codificantes de proteínas e informações do mtDNA podem auxiliar a desvendar relações filogenéticas que sequências de genes nucleares seriam incapazes de resolver. Genomas mitocondriais compreendem um repositório de genes codificadores de proteínas, cuja origem está bem estabelecida e cuja evolução parece acompanhar a do hospedeiro eucarioto. A determinação de uma maior variedade de sequências de genomas mitocondriais de algas verdes poderá permitir uma reconstrução mais acurada de filogenias embasadas em sequências de aminoácidos de genes codificantes de proteínas (Gary, Burger, Lang, 1999).

Figura 6 - Posição filogenética *Chlorokybus atmophyticus* dentro do reino Viridiplantae. Análise filogenética embasada nos genes mitocondriais codificadores de proteínas *atp4*, *atp6*, *atp8*, *atp9*, *cob*, *cox1*, *cox2*, *cox3*, *mttB*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *rps3* e *rps12*.



Fonte: Turmel; Otis; Lemieux, 2007.

1.4. Análise Transcriptômica

A aplicação de abordagens genômicas às pesquisas com algas, como o sequenciamento e análises da função gênica, causou uma mudança radical em nossa compreensão da biologia, ecologia e evolução das algas. A ascensão de tecnologias de sequenciamento de alta performance a um menor custo, levou à aquisição de dados genômicos significativos sobre algas desde a década de 1990 (Pedrini, 2010). Genomas inteiros de organismos podem ser sequenciados mais facialmente, juntamente com a disponibilidade de um maior número de

ferramentas de bioinformática para montagem e anotação dos dados, facilitando a sequência e a predição de funções biológicas (Khan, et al., 2020).

O transcriptoma é o conjunto de todos os ácidos ribonucleicos (RNAs) expressos por um organismo, sendo o objeto de estudo de uma das áreas da Genômica Funcional, a Transcriptômica (Lu, et al., 2014). O transcriptoma contem apenas as porções transcritas dos genomas, o que simplifica as análises genéticas de eucariotos, removendo elementos genéticos complexos de grandes regiões intergênicas, íntrons e DNA repetitivo (Koid, et al., 2014).

O sequenciamento de RNAs em larga escala é uma abordagem recente, amplamente utilizado para a descoberta de novos genes (Lu, et al., 2014). Esta abordagem, quando comparada a outras empregadas para análises do transcriptoma como, por exemplo, os chips de microarranjos e sequenciamento de pequenas sequências expressas (ESTs), traz grandes vantagens como, bom custo-benefício, alta sensibilidade e acurácia (Wang, et al., 2011).

Para o sequenciamento do transcriptoma, o RNA total é extraído da amostra e convertido em fragmentos de ácido desoxirribonucleico complementar (cDNA), formando uma biblioteca de sequenciamento. Esta biblioteca então é sequenciada, gerando pequenas *reads* que devem ser montadas e anotadas, de maneira muito semelhante aos genomas.

Os primeiros métodos de sequenciamento de ácido desoxirribonucleico (DNA) foram criados na década de 70, quando Sanger e Coulson desenvolveram a técnica “mais e menos” (Sanger, Coulson, 1975) e Maxam e Gilbert a de clivagem química (Maxam, Gilbert, 1977).

Porém, o grande avanço ocorreu quando Sanger desenvolveu o sequenciamento por terminação de cadeia. O mesmo consiste na síntese de uma cadeia nucleotídica utilizando o fragmento a ser sequenciado como molde, sendo a base correspondente para cada posição da cadeia identificada através do interrompimento da síntese, que ocorre a partir da adição de dideoxinucleotídeos (ddNTPs) marcados radioativamente, em quatro reações paralelas (Sanger, et al., 1977). A princípio, a metodologia utilizava a migração em gel de poliacrilamida para identificação das bases. Posteriormente o processo foi automatizado, com a utilização de ddNTPs marcados com fluoróforos e a migração em capilar, tornando-se a metodologia mais aplicada para o sequenciamento de DNA. O sequenciador de Sanger é considerado a primeira geração, e gerava *reads* de ~1 kilobase (kb) (Luckey, et al., 1990).

O sequenciamento de Sanger foi extremamente importante para as descobertas na área de Genômica Estrutural, por exemplo, no Projeto Genoma Humano, porém sua aplicabilidade na área de Transcriptômica era limitada devido à baixa quantidade de dados gerados, alto custo

e dificuldade da avaliação de níveis quantitativos em larga escala (Wang, et al., 2011; Heather, Chain, 2016).

O início do sequenciamento de segunda geração, foi marcado pelo desenvolvimento do pirosequenciador. Esta metodologia também utilizava o método "sequenciamento por síntese" aplicando a DNA polimerase, contudo os nucleotídeos não mais eram marcados e a adição das bases acompanhadas em tempo real, sem a necessidade de migração em capilar (Heather, Chain, 2016).

O pirosequenciamento detecta o pirofosfato liberado durante o processo de formação da ligação fosfodiéster entre nucleotídeos, sendo a detecção realizada através de luminescência. A técnica foi licenciada a 454 *Life Sciences* para a produção de máquinas de pirosequenciamento, que se valiam da técnica de reação em cadeia da polimerase (PCR) em emulsão para amplificação de DNA aderido em *beads* antes do sequenciamento. As *reads* geradas possuem entre 400 e 500 pares de bases (pb) (Ronaghi, et al., 1998).

Após o sucesso dos pirosequenciadores, outras plataformas competidoras foram desenvolvidas, destacando-se as máquinas Solexa/Illumina e SOLiD, todas possuindo como característica o sequenciamento em massa. O sistema SOLiD (*Sequencing by Oligonucleotide Ligation and Detection*) é fundamentado na hibridização e ligação de oligonucleotídeos marcados com fluoróforos, utilizando a enzima DNA ligase (Heather, Chain, 2016).

A plataforma Solexa/Illumina baseia-se na adição de adaptadores nas extremidades dos fragmentos de DNA, formação de *clusters* por PCR em ponte e o sequenciamento por síntese através da incorporação de nucleotídeos "terminadores reversíveis" marcados com fluoróforos (Voelkerding, et al., 2009). As primeiras máquinas produziam *reads* de apenas 35 pb, mas a introdução da estratégia de *paired-end reads* (*reads* geradas no sentido direto e reverso do fragmento sequenciado) permitiram um melhor mapeamento entre as *reads* para a montagem (Heather, Chain, 2016). Os modelos posteriormente lançados, nomeados Illumina HiSeq e Illumina MiSeq, trouxeram como vantagem a geração de *reads* maiores e menor custo.

Os métodos de sequenciamento de alta performance, através do sequenciamento em massa por um preço mais acessível, permitiram o desenvolvimento e maior popularização da área de sequenciamento de transcriptomas, onde a plataforma Illumina é a mais empregada neste tipo de estudos. Porém, como as *reads* são pequenas, surgiram os desafios para a montagem do transcriptoma. Assim, além da evolução das técnicas de sequenciamento, o

desenvolvimento das ferramentas de Bioinformática para a montagem e análise da avalanche de dados gerados também foram essenciais.

1.4.1 Montagem e Anotação de Transcriptomas

A montagem das pequenas *reads* geradas em transcritos pode ser realizada em duas abordagens: montagem com referência e *de novo*. A montagem de referência é escolhida quando há um genoma para guiar a montagem, podendo ser da mesma espécie ou de espécies relacionadas. Neste tipo de abordagem as *reads* são mapeadas contra o genoma de referência utilizando programas de alinhamento de sequências (Marchant, et al., 2016). Porém, a maioria das espécies não-modelo nem sempre possuem um genoma a ser utilizado como referência, sendo necessária a montagem *de novo*. Nesta metodologia, a montagem dos transcritos é realizada através da identificação de sobreposições entre os *reads*, que formam *contigs* pela identificação de uma ou várias sequências consenso (Wang, et al., 2009).

Após a montagem, o último passo é a anotação do transcriptoma. A anotação consiste em agregar de informações biológicas aos transcritos, por meio da busca de sequências homólogas, havendo uma ampla gama de *pipelines* e ferramentas que podem ser aplicadas (Garber, et al., 2011). Esta metodologia de busca baseada em homologia é aplicada não apenas para anotação transcriptomas, mas também genomas, metagenomas e metatranscriptomas (Westreich, et al., 2016). Entre as ferramentas o Blast2GO é um dos mais utilizados, sendo robusto e simples (Bolger, et al., 2017; Conesa, Götz, 2008).

O primeiro passo da anotação feita pelo Blast2GO é a busca utilizando a ferramenta de alinhamento BLAST (*Basic Local Alignment Search Tool*) (Altschul, et al., 1990) por sequências que sejam similares aos transcritos. Os resultados são expressos através do *E-value*, que descreve o número de *hits* que se espera ao realizar um alinhamento de sequências contra um banco de dados de determinado tamanho apenas ao acaso (Kerfeld, Scott, 2011). Também é levado em consideração o tamanho do alinhamento e similaridade. Assim, quanto menor o *E-value*, mais significante é considerado aquele alinhamento entre as sequências.

Após isso, as sequências são mapeadas e anotadas para associação de termos funcionais do banco de dados *Gene Ontology* (GO) de acordo com o resultado do BLAST. O *Gene Ontology Consortium* é um projeto que permite classificar os genes e seus produtos de maneira uniforme, atuando como uma linguagem universal, rotineiramente empregado na pesquisa

durante o processo anotação funcional (Huntley, et al., 2014). Este projeto possui três categorias principais que indicam que o produto gênico possui (I) determinada atividade ou processo a nível molecular (Função Molecular, FM), (II) que ocorre em uma localização específica celular (Componente celular, CC) e que contribui para um efeito biológico (Processo biológico, PB). Além disso, existe a divisão de níveis que vão dos termos mais gerais para os mais específicos.

Outras ferramentas, como o MG-RAST (Kerfeld, Scott, 2011) não possuem anotação tão detalhada quanto o Blast2GO mas tem como vantagem a grande velocidade para trabalhar com números muito grandes de *contigs*. Além disso, outros vocabulários de anotação como o COG (*Cluster of Orthologous Groups*) (Tatusov, et al., 2000) e *Subsystems* (Overbeek, et al. 2005) podem ser aplicados. Ao fim do processo de anotação, o transcriptoma pode ser analisado para a busca por transcritos de interesse.

O sequenciamento do transcriptoma já foi realizado em outras algas da classe *Trebouxiophyceae*, objetivando a identificação de transcritos associados a resistência a estresses bióticos e abióticos, e genes com aplicabilidade na produção de biocombustível (Baba, et al. 2012; Yu, et al., 2016). Contudo, dados genômicos e transcriptômicos de algas verdes psicrofílicas antárticas, como *P. crispa*, são limitados até o momento, de modo que pouco se sabe sobre as adaptações moleculares.

Há na literatura algumas descrições sobre o uso de diferentes extratos de *P. crispa.* Zemolin, et al. (2014) demonstraram a ação bioinseticida de *P. crispa* sobre a mosca da fruta *Drosophila melanogaster* e a barata *Neuphoeta cinerea*, através da modificação de sistemas antioxidantes do organismo. Já o composto químico 7-ceto-estigmasterol, um esterol purificado a partir do extrato de *P. crispa*, apresentou atividade antiviral quando testado contra o Herpesvírus equino 1, o vírus causador de uma doença até o momento sem tratamento eficiente (Marinho, et al., 2017). Por fim, o estudo de Da Silva, et al. (2016) revela o potencial antiveneno do extrato de *P. crispa* contra atividade tóxicas *in vivo* e *in vitro* do veneno de *Bothrops jararacussu.*

O recente aumento de dados de Transcriptomas permite que análises evolutivas em larga escala investiguem a base genética para as adaptações ao ambiente extremo Antártico, permitindo a identificação das forças seletivas que impulsionam a evolução molecular, sendo a base para compreender as estratégias de adaptação à ambientes extremos.

2. OBJETIVOS

2.1. Objetivo geral

Este trabalho tem como objetivos (I) inferir as relações evolutivas de *Prasiola crispa* com outras espécies de plantas verdes, além de analisar a estruturas dos genomas acessórios; (II) analisar o transcriptoma de *P. crispa* e identificar os produtos gênicos diretamente relacionados com a capacidade de sobrevivência desde organismo no continente Antártico.

2.2. Objetivos específicos

- Com base nas sequências genômicas acessórias, realizar uma análise evolutiva (filogenômica) de *Prasiola crispa*;
- A fim de compreender o rearranjo gênico que ocorreu durante a história evolutiva do cpDNA e mtDNA de *P. crispa* e espécies relacionadas, analisar a estrutura dos genomas acessórios do clado *Prasiola*;
- Com o intuito de identificar os transcritos expressos em condições ambientais extremas, realizar uma análise transcriptômica de *P. crispa*;
- Para certificar a qualidade da montagem do transcriptoma, realizar análises de validação e qualidade dos dados obtidos do RNAseq;
- Reconstruir os transcritos a partir das *reads* de alta qualidade;
- Identificar e anotar funcionalmente os transcritos, buscando os transcritos relacionados com a sobrevivência de *P. crispa* na Antártica,
- Comparar as métricas do transcriptoma de *P. crispa* com as de outros organismos da classe Trebouxiophyceae.

3. ARTIGO 1

O artigo intitulado *"Phylogenetic positioning of the Antarctic alga Prasiola crispa (Trebouxiophyceae) using organellar genomes and their structural analysis*" de autoria de **Carvalho E.L.**, Wallau G.L., Rangel D.L., Machado L.C., Pereira A.B., Victoria F.C., Boldo J.T., Pinto P.M. foi publicado no periódico *Journal of Phycology* (ISSN: 1529-8817), no ano de 2017.

Este artigo relata a análise filogenômica utilizando as sequências dos genomas plastidial (cpDNA) e mitocondrial (mtDNA) de *P. crispa*, e outras espécies do filo Chlorophyta e Streptophyta, bem como a análise estrutural dos cpDNA e mtDNA de *P. crispa* e espécies intimamente relacionadas.

A reconstrução filogenética bayesiana foi realizada com MrBayes v. 3.2.1, utilizando o modelo de substituição de aminoácidos LG + I + G e LG + G para os conjuntos de dados plastidiais e mitocondriais, respectivamente. A análise sintênica foi realizado com BLASTn e a ferramenta de comparação Artemis foi utilizada para gerar os blocos sintênicos com mais de 100 bp de tamanho e anotação correspondente.

Os resultados das análises confirmam que *P. crispa* pertence à classe Trebouxiophyceae, e está filogeneticamente relacionada à espécie *Prasiolopsis sp.* SAG 84.81. Na análise sintênica do cpDNA pode-se observar uma baixa sintenia entre *Prasiolopsis sp.* SAG 84.81 e *P. crispa*, contudo, alguns blocos sintênicos podem ser observados. Em relação ao mtDNA, foram encontradas apenas regiões menores de blocos sintênicos entre *P. crispa* e *Coccomyxa subellipsoidea*. Estes resultados demonstram que os genomas, tanto o cpDNA quanto o mtDNA de *P. crispa,* são muito plásticos e sofreram muito rearranjos e inversões.

J. Phycol. **53,** 908–915 (2017)

DOI: 10.1111/jpy.12541

NOTE

PHYLOGENETIC POSITIONING OF THE ANTARCTIC ALGA *PRASIOLA CRISPA* (TREBOUXIOPHYCEAE) USING ORGANELLAR GENOMES AND THEIR STRUCTURAL ANALYSIS[1]

Evelise Leis Carvalho

Applied Proteomics Laboratory, University of Pampa, São Gabriel, RS 97300-000, Brazil

Gabriel Luz Wallau

Departamento de Entomologia Centro de Pesquisas Aggeu Magalhães, Fiocruz, Recife 50740-465, Brazil

Darlene Lopes Rangel, Laís Ceschini Machado

Applied Proteomics Laboratory, University of Pampa, São Gabriel, RS 97300-000, Brazil

Antonio Batista Pereira, Filipe De Carvalho Victoria

NEVA, University of Pampa, São Gabriel, RS 97300-000, Brazil

Juliano Tomazzoni Boldo, and Paulo Marcos Pinto[2]

Applied Proteomics Laboratory, University of Pampa, São Gabriel, RS 97300-000, Brazil

Antarctica is one of the most difficult habitats for sustaining life on earth; organisms that live there have developed different strategies for survival. Among these organisms is the green alga *Prasiola crispa*, belonging to the class Trebouxiophyceae. The literature on *P. crispa* taxonomy is scarce, and many gaps in the evolutionary relationship with its closest relatives remain. The goal of this study was to analyze the evolutionary relationships between *P. crispa* and other green algae using plastid and mitochondrial genomes. In addition, we analyzed the synteny conservation of these genomes of *P. crispa* with those of closely related species. Based on the plastid genome, *P. crispa* grouped with *Prasiolopsis* sp. SAG 84.81, another Trebouxiophyceaen species from the *Prasiola* clade. Based on the mitochondrial genome analysis, *P. crispa* grouped with other Trebouxiophyceaen species but had a basal position. The structure of the *P. crispa* chloroplast genome had low synteny with *Prasiolopsis* sp. SAG 84.81, despite some conserved gene blocks. The same was observed in the mitochondrial genome compared with *Coccomyxa subellipsoidea* C-169. We were able to establish the phylogenetic position of *P. crispa* with other species of Trebouxiophyceae using its genomes. In addition, we described the plasticity of these genomes using a structural analysis. The plastid and mitochondrial genomes of *P. crispa* will be useful for further genetic studies, phylogenetic analysis and resource protection of *P. crispa* as well as for further phylogenetic analysis of Trebouxiophyceaen green algae.

Key index words: **Antarctic algae; chloroplast genome; green algae phylogeny; mitochondrial genome; *Prasiola crispa***

***Abbreviations*: BLASTn, Nucleotide Basic Local Alignment Search Tool; cpDNA, chloroplast genome; mtDNA, mitochondrial genome; NCBI, National Center for Biotechnology Information**

Trebouxiophyceaen green algae of the genus *Prasiola* are among the best-known Antarctic algae and represent the most important primary producers at many terrestrial and supralittoral sites (Kováčik and Pereira 2001, Convey 2007). The species most commonly reported is *Prasiola crispa*. This nitrophilic alga typically grows on moist soils that are fertilized by bird guano (Graham et al. 2009). *Prasiola crispa* is widely distributed throughout continental Antarctica and maritime Antarctica (Moniz et al. 2012), and this species tolerates repeated freeze/thaw cycles in the spring and fall and high levels of UV radiation during the summer (Lud et al. 2001, Kosugi et al. 2010).

Although it is important to the Antarctic ecosystem, investigations on the taxonomy of Antarctic

[1]Received 5 May 2016. Accepted 9 January 2017. First Published Online 10 April 2017. Published Online 12 May 2017, Wiley Online Library (wileyonlinelibrary.com).
[2]Author for correspondence: e-mail paulopinto@unipampa.edu.br.
Editorial Responsibility: C. Lane (Associate Editor)

representatives of the *Prasiola* genus at the molecular level in the literature are scarce (Moniz et al. 2012). The phylogenetic positioning of species that can adapt to different temperatures, such as *P. crispa*, is essential for understanding adaptations to extreme environmental conditions and the moment at which the evolution of such adaptations of green algae emerged (Moniz et al. 2012).

The study of organellar genomes can impact the fields of evolution, biology, and biotechnology (Jeong et al. 2014) and therefore can help us understand which evolutionary processes generated adaptations to extreme environments at the molecular level. In addition, such genetic information also provides many molecular characters (gene order, nucleotide and amino acid sequences) that can be used to estimate phylogenetic trees and to discover the positioning of some species in a broader phylogenetic context (Gray et al. 1999, Wang et al. 2013).

cpDNA and mtDNA provide large sets of genes that have been successfully used in analyses of molecular phylogeny (Turmel et al. 2008, Brouard et al. 2010, Wang et al. 2013). The analysis of the whole or majority of genome data is more likely to provide accurate results than analysis of one or a few gene markers because cpDNA and mtDNA contain increased phylogenetic signal, which reduces the effect of homoplasy, and because cpDNA and mtDNA include increased numbers of characters, reducing stochastic errors (Li et al. 2014). Phylogenomic studies of green algal chloroplasts reported thus far have provided valuable insights into the phylogeny of prasinophytes (Turmel et al. 2009, Lemieux et al. 2014a), streptophytes (Civáň et al. 2014, Zhong et al. 2014), and Chlorophyceae (Turmel et al. 2008, Brouard et al. 2010); however, little information regarding the Trebouxiophyceae class exists (Lemieux et al. 2014b).

Complete mtDNA has been increasingly used to address phylogenetic questions where multigene analyses have been either unresolved or poorly supported (Yang et al. 2015).

Driven by the unanswered questions concerning the biology of *P. crispa*, we have studied the evolutionary relationship of this species (using organellar genomes recently sequenced by our group) and other Trebouxiophyceaen algae using a cpDNA and mtDNA phylogenomic approach. In addition, we performed a structural analysis and evaluated which groups of genes are missing in different green algal clades in order to gain further insights regarding adaptation to different environments.

Prasiola crispa was collected in ice-free areas near the Arctowski Polish Station Region, Admiralty Bay, King George Island (61°50′–62°15′ S and 57°30′–59°00′ W), Antarctica.

Most of the mtDNA and cpDNA sequences were obtained from Turmel et al. (2007) and Lemieux et al. (2014b), respectively. mtDNA and cpDNA sequences of *P. crispa* were obtained from Carvalho et al. (2015). To include as many Trebouxiophyceaen genomes as possible, we performed searches at the NCBI website using both word searches ("Chloroplast genome" AND "Trebouxiophyceae") and mtDNA and cpDNA of *P. crispa* as queries in BLASTn (Altschul et al. 1990). The accession numbers of each genome and species authorities cited in the study are shown in Tables S1–S3 in the Supporting Information.

Phylogenomic analysis was performed with 79 chloroplast coding genes and 18 mitochondrial coding genes. The predicted protein sequences of these genes were aligned individually using MAFFT v. 7.220 (Katoh and Standley 2013) with the default parameters. Following the alignment experiments, each alignment was computationally edited, removing highly variable regions, using Gblocks v. 0.91 (Castresana 2000) with the "Allowed gap positions" selected with the "With Half" option. The remaining positions that passed the Gblocks filter were concatenated using Phyutility v. 2.2.6 (Smith and Dunn 2008).

Bayesian phylogenetic reconstruction was performed with MrBayes v. 3.2.1 (Huelsenbeck and Ronquist 2000) using the best amino acid substitution model evaluated by Prottest (Abascal et al. 2005): LG + I + G and LG + G for the plastid and mitochondrial data sets, respectively. One million generations were evaluated, sampling the most probable tree every 100 generations and discarding 25%.

Synteny analysis of the *P. crispa* genomes and those of closely related species selected based on the reconstructed trees was performed with BLASTn (Altschul et al. 1990) with the default parameters. The Artemis Comparison Tool (Carver et al. 2005) was then used to plot the syntenic blocks with more than 100 bp in size and their corresponding annotation.

Based on the chloroplast phylogenomic analysis, *P. crispa* grouped consistently (posterior probability of 1) as a sister species of *Prasiolopsis* sp. SAG 84.81, a species from the *Prasiola* clade. This clade also includes four more species: *Stichococcus bacillaris*, *Pabia signiensis*, *Koliella longiseta*, and "*Chlorella*" *mirabilis* (Fig. 1).

The phylogenies reported here confirm that *P. crispa* belongs to the Trebouxiophyceae class and that it is phylogenetically related to a sister species, *Prasiolopsis* sp. SAG 84.81, both of the Prasiolaceae family. Species from this family are characterized by thalli that form monostromatic blades, tetragonal or rectangular vegetative cells with few irregular polygons on the surface, and two to four cells usually arranged in the parent cell wall, with a single, stellate chromatophore and a prominent central pyrenoid (Rixiao et al. 2009).

Our results are consistent with the phylogenetic analysis based on 18S rRNA shown by Friedl and O'Kelly (2002), in which species of *Prasiola* were

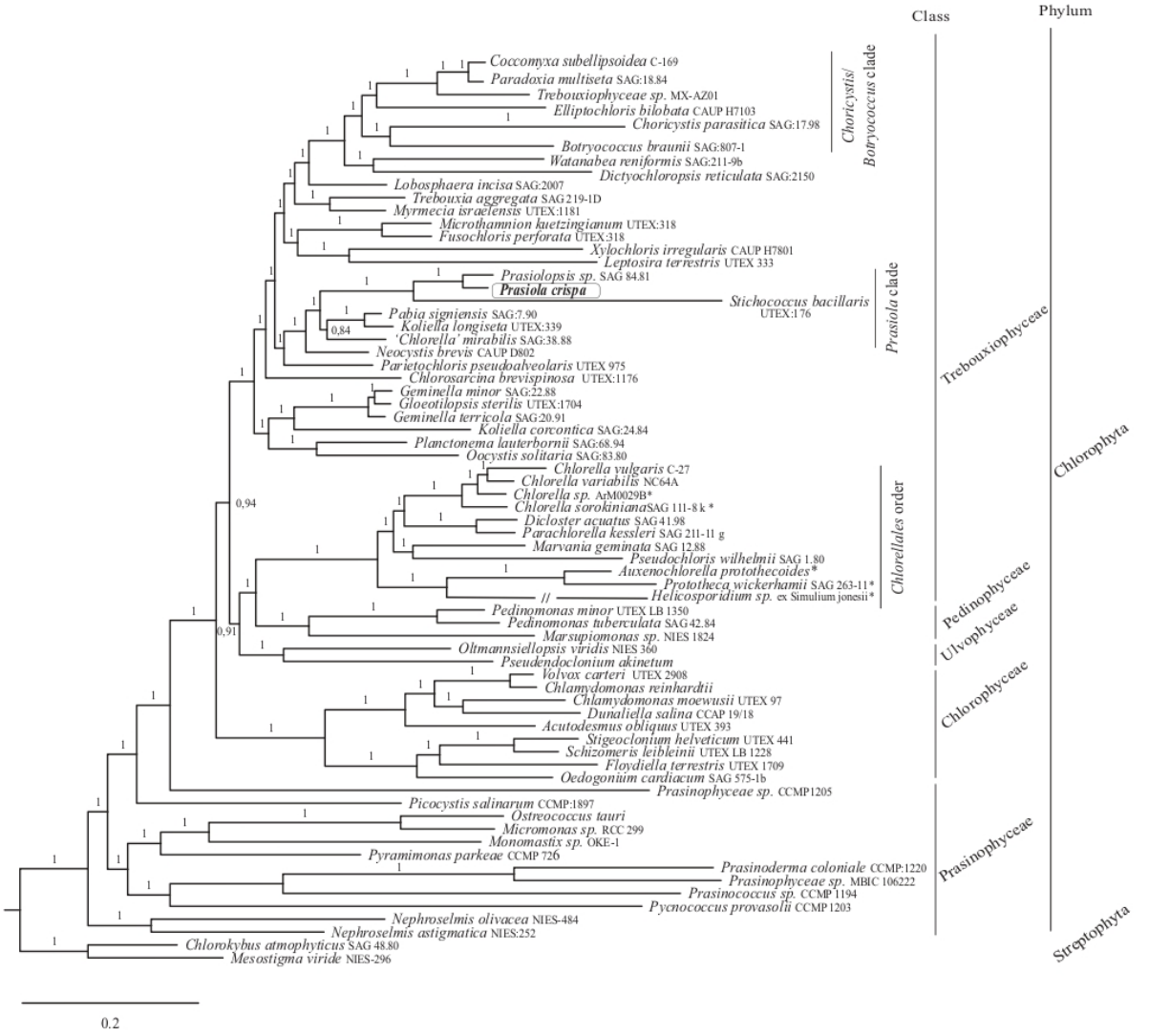


FIG. 1. Phylogenomic analysis of the chloroplast genomes of 69 chlorophytes using 14,469 amino acid positions from 79 cpDNA-encoded proteins. The tree presented was reconstructed using a Bayesian approach with the LG + I + G amino acid substitution model. Branch supports indicate posterior probability values (over each branch). Bars at the right side of the figure represent the taxonomic levels quoted in the text. *Other Trebouxiophyceaen species added to phylogenetic analysis.

shown to be phylogenetically related to other terrestrial green algae with a vegetative morphology different from that of *Prasiola* (i.e., the pseudofilamentous *S. bacillaris* and the coccoid *Chlorella*-like green alga ["*Chlorella*" *mirabilis*; Karsten et al. 2005]).

Handa et al. (2003) also observed a close phylogenetic relationship between a corticolous species of *Stichococcus* and *Prasiola* using 18S rRNA sequences.

To compare the phylogenetic position of *P. crispa*, we performed a complementary phylogenetic analysis with the marker genes *rbcL* and *psaB* (Figs. S1 and S2 in the Supporting Information) with different strains of *P. crispa* and other organisms. These complementary phylogenetic analyses positioned *P. crispa* within the *P. crispa* clade described by Moniz et al. (2012). However, it is not possible to confirm if adaptations at the molecular level enable the survival in extreme environmental conditions due to the lack of data in the literature, which would necessitate more specific studies.

Our mitochondrial phylogenomic analysis positioned *P. crispa* as a basal lineage of the Trebouxiophyceae clade, which includes the species: *Chlorella* sp. ArM0029B, *Chlorella sorokiniana*, *Auxenochlorella protothecoides*, *Prototheca wickerhamii*, *Helicosporidium* sp., *Trebouxiophyceae* sp. MX-AZ0 and *C. subellipsoidea* C-169 (Fig. 2). However, the positioning of *P. crispa*

may be due to the limited numbers of Trebouxiophyceae mtDNA available in the databases.

Even though a much lower number of mtDNA than plastid genomes are available for Trebouxiophyceaen species, we could still reconstruct the phylogenetic positioning of some branches. All of the nodes inside the Trebouxiophyceae clade presented posterior probability supports of 1 (Fig. 2), which supported a phylogenetic position of *P. crispa* as a species belonging to the Trebouxiophyceae class.

Overall, our mitochondrial phylogenomic tree (Fig. 2) matched well with the reconstructed plastid tree (Fig. 1), as can be seen in the Chlorellales order. In addition, we also established novel phylogenetic positioning of other Trebouxiophyceaen algae based on cpDNA and mtDNA (Figs. 1 and 2).

To understand the gene rearrangement that occurred during the evolutionary history of cpDNA and mtDNA of *P. crispa* and related species, we performed a synteny analysis and placed it in an evolutionary context according to the trees obtained (Figs. 3 and 4).

Regarding the chloroplast synteny analysis, we can observe a low synteny between *Prasiolopsis* sp. SAG 84.81 and *P. crispa*. However, some conserved blocks still can be detected in the 5′ (*psbE*, *minD* and *fstH* genes), middle (*tufA* gene), and 3′ (*infA*, *rps8*, *rpl5*, *rpl14*, *rpl16*, *rps3*, *rps19*, *rpl2*, and *rpl23* genes) portions of the *P. crispa* plastid genome. In addition, we also can observe three major inversions involving the blocks of genes *chlN*, *chlB*, *psbK*, *ycf12* and *psaM*; *atpB* and *atpE*; and *rpoB*, *rpoC1*, and *rpoC2* (Fig. 3).

Due to the phylogenetic position of *Prasiolopsis* sp. and *P. crispa* (Fig. 1) more gene rearrangements between these species were expected (Fig. 3), which was not observed. However, species that live in extreme locations, such as *P. crispa*, require a set of physiological and metabolic adaptations, which, in the case of algae, are directly associated with the photosynthetic apparatus. Antarctic algae not only need to be adapted to capture and use high levels of radiation in the summer (radiation levels are extremely low in the winter) but also need adaptations to survive the low temperatures of the Antarctic habitat. At low temperatures, the efficiency of photosynthesis is reduced, as the diffusion of electron carriers and carbon fixation are lower. As a result, reactive oxygen species are formed, which cause oxidative stress (Öquist and Huner 2003, Takahashi and Murata 2008).

As mentioned in our phylogenetics analyses, data concerning *P. crispa* are still scarce in the literature, making it impossible to perform a solid genomic comparison between the strains of *P. crispa* collected at different geographic regions.

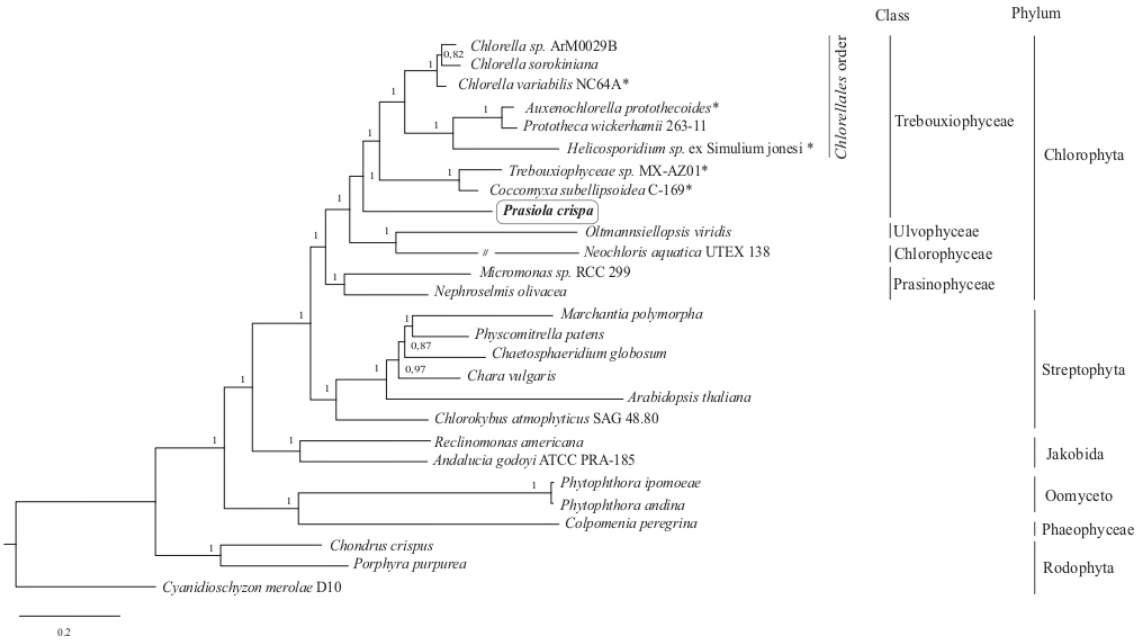


FIG. 2. Phylogenomic analysis of 27 mitochondrial genomes encompassing all nine chlorophyte mtDNA genomes available. The tree presented was reconstructed using a Bayesian approach with the LG + G amino acid substitution model using 2,277 amino acid positions from 18 mtDNA-encoded proteins. Branch supports indicate posterior probability values (over each branch). Bars at the right side of the figure represent the taxonomic levels quoted in the text. *Other Trebouxiophyceaen species added to phylogenetic analysis.

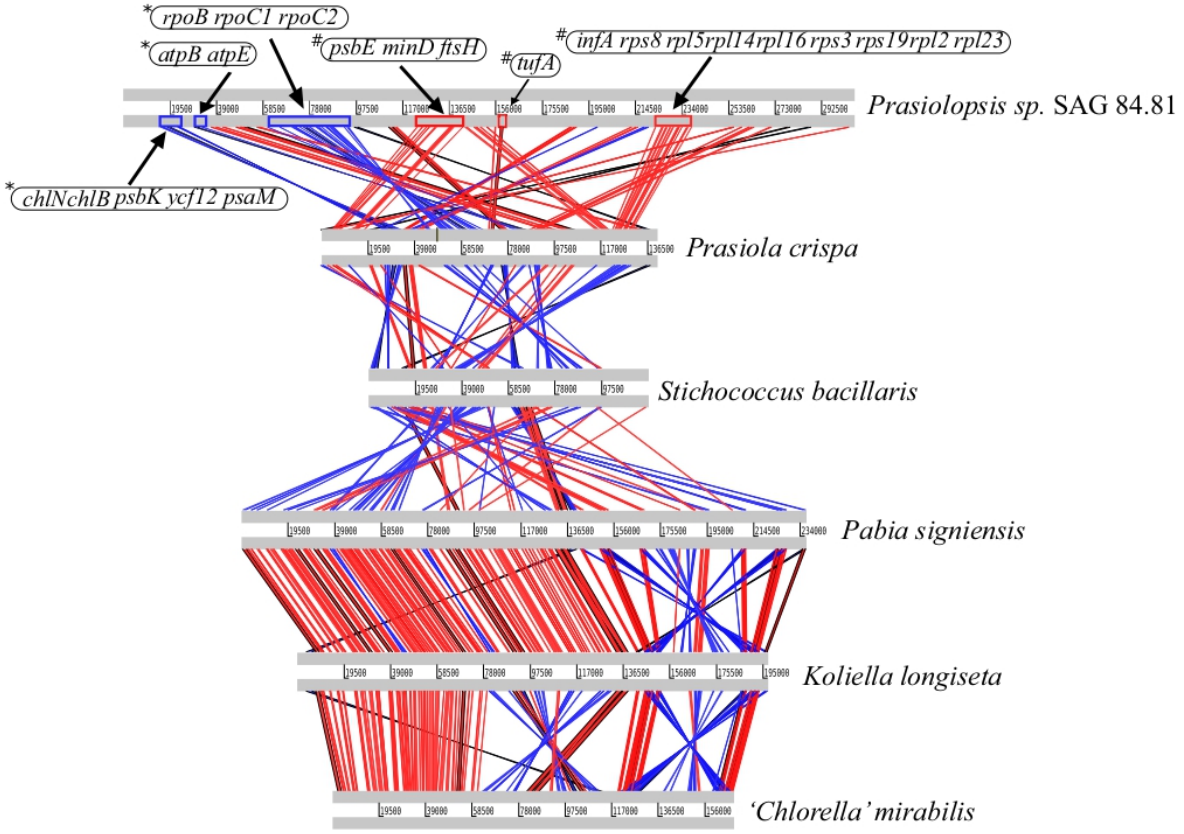


FIG. 3. Chloroplast synteny block analysis performed using BLASTn with the default parameters and Artemis ACT software. Only regions with ≥100 bp are shown. Traces connecting cpDNA genomes represent synteny blocks in one or another strand. *Regions conserved inverted. #Regions conserved.

In a broader phylogenetic context, we could observe a conserved syntenic pattern between "*Chlorella*" *mirabilis, P. signiensis* and *K. longiseta*. These syntenic blocks matched perfectly with the phylogenetic relationship of those species (Fig. 3). Although gene order and gene content are generally conserved among land plant cpDNAs (Palmer 1985), very little synteny is observed between this group and cpDNAs of the chlorophytic green algae *Chlamydomonas reinhardtii* (Boudreau and Turmel 1996, Maul et al. 2002) and *Chlorella vulgaris* (Wakasugi et al. 1997). Jeong et al. (2014) described many conserved gene clusters in the cpDNA of different *Chlorella* species in comparison with cpDNA of one Pedinomonadales and two Chlorellales species.

Our results highlight the dynamic plastid genome of *P. crispa* and its related species (*Prasiolopsis* sp. SAG 84.81 and *S. bacillaris*) compared with that of the *P. signiensis/K. longiseta/*"*Chlorella*" *mirabilis* clade. The results suggest that the ancestor of those species presented a stable cpDNA genome in terms of gene order and that in the ancestral lineage leading to *P. crispa/S. bacillaris* and *Prasiolopsis* sp. SAG 84.81 several genome arrangements occurred, with continuous rearrangements occurring in each lineage after the split from the ancestor.

Regarding the mtDNA, we found only minor regions of syntenic blocks between *P. crispa* and *C. subellipsoidea* C-169 species (Fig. 4). Those regions contain the genes *rrnL* and *cox1* and an inversion of the *rrnS* gene. Moreover, well-conserved syntenic blocks can be observed between *C. subellipsoidea* C-169 and *Trebouxiophyceae* sp. MX-AZ01 and among all species from the Chlorellales order (Fig. 4). The comparison that resulted in only one small syntenic block was between *Helicosporidium* sp. and *Trebouxiophyceae* sp. MX-AZ01, which is in agreement with the phylogenetic positioning of such species (Fig. 2) and with the mtDNA genome reduction in the first species.

Based on the low levels of gene order conservation in other green algal mtDNA, we might expect to see a few blocks of conservation between mtDNA analyzed. A recent split between *Helicosporidium* sp. and *Trebouxiophyceae* sp. MX-AZ01 probably best explains this close resemblance. In the

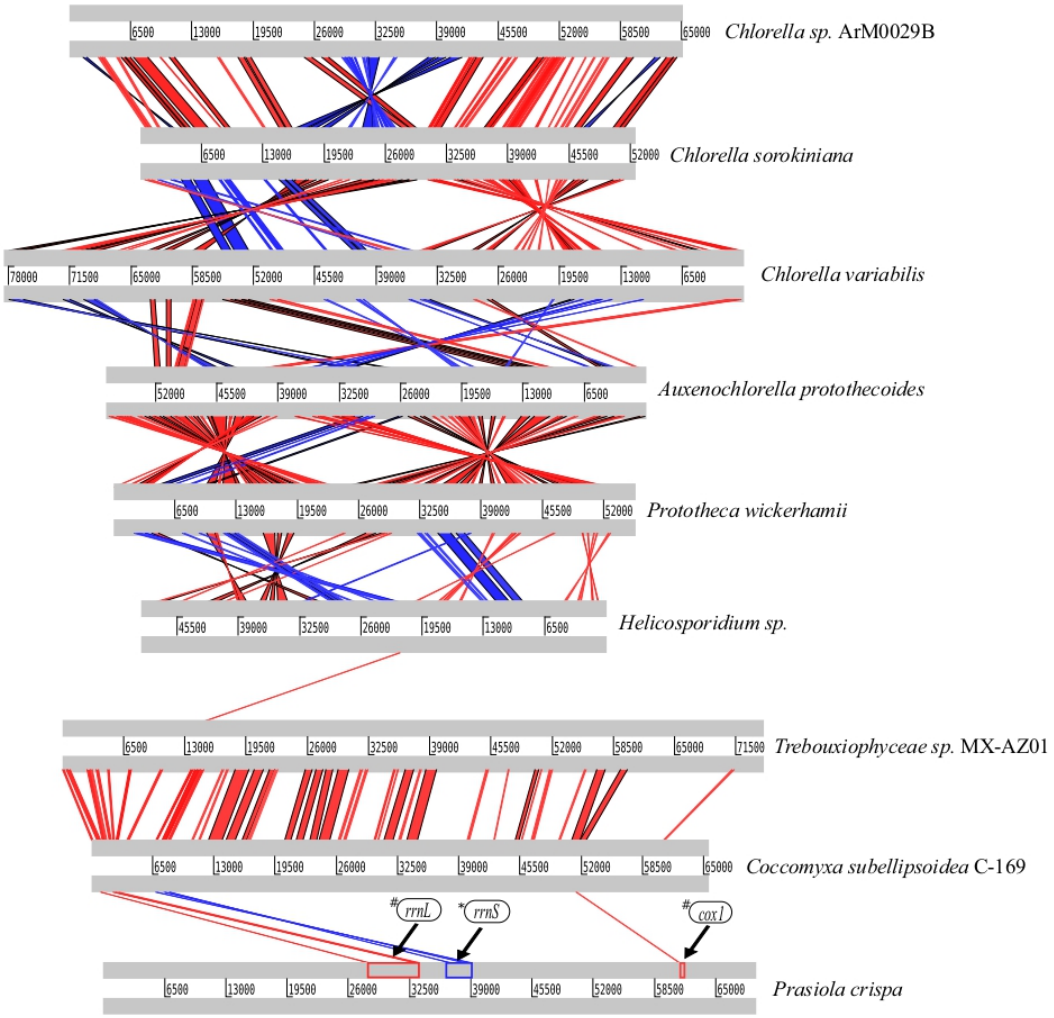


FIG. 4. Mitochondrial synteny block analysis performed using BLASTn with the default parameters and Artemis ACT software. Only regions with ≥100 bp are shown. Traces connecting mtDNA genomes represent synteny blocks in one or another strand. *Regions conserved inverted. #Regions conserved.

Chlorophyceae, the gene-poor mtDNAs of *Chlorogonium elongatum* and *Chlamydomonas eugametos* also display an appreciable level of synteny, with 8 of their 12 genes (66%) located in shared clusters; however, this percentage is still lower than that observed between *Helicosporidium* sp. and *P. wickerhamii* mtDNAs (75%; Pombert and Keeling 2010). Our data disagree with those of Pombert and Keeling (2010), who described a surprisingly high level of similarity of mtDNA between *Helicosporidium* sp. and *P. wickerhamii*. As the first sequenced mitochondrial genome from the Prasiolales order, *P. crispa* mtDNA is expected to have a few conserved blocks, as do other green algae.

In the present work, we established the phylogenetic positioning of *P. crispa* as a sister species of *Prasiolopsis* sp. In addition, we established the phylogenetic positioning of other Trebouxiophyceaen algae based on a phylogenomic cpDNA analysis and compared this analysis with that of the mtDNA tree. Moreover, we observed that *P. crispa* genomes are very plastic, with several inversions and rearrangements, which is in agreement with other findings of Trebouxiophyceaen genomes.

This work was supported by the National Council for Scientific and Technological Development (CNPq-Brazil), the Coordination for the Improvement of Higher Education

Personnel (CAPES-Brazil), the Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS-Brazil) and the National Institute of Science and Technology-Antarctic Environmental Research (INCT-APA). E.L.C. received research fellowships from the Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS-Brazil).

Abascal, F., Zardoya, R. & Posada, D. 2005. Prottest: selection of best-fit models of protein evolution. *Bioinformatics* 21:2104–5.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. 1990. Basic local alignment search toll. *J. Mol. Biol.* 215:403–10.

Boudreau, E. & Turmel, M. 1996. Extensive gene rearrangements in the chloroplast DNA of *Chlamydomonas* species featuring multiple dispersed repeats. *Mol. Biol. Evol.* 13:233–43.

Brouard, J. S., Otis, C., Lemieux, C. & Turmel, M. 2010. The exceptionally large chloroplast genome of the green alga *Floydiella terrestris* illuminates the evolutionary history of the chlorophyceae. *Genome Biol. Evol.* 2:240–56.

Carvalho, E. L., Wallau, G. L., Rangel, D. L., Machado, L. C., Silva, A. F., Pereira, A. B., Victoria, F. C., Boldo, J. T., Dal Belo, C. A. & Pinto, P. M. 2015. Draft plastid and mitochondrial genomes from Antartic alga *Prasiola crispa*. *Genome Announce.* 3:e01151–15.

Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M. A., Barrell, B. G. & Parkhill, J. 2005. ACT: the Artemis comparison tool. *Bioinformatics* 21:3422–3.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use phylogenetic analysis. *Mol. Biol. Evol.* 17:540–52.

Civáň, P., Foster, P. G., Embley, M. T., Seneca, A. & Cox, C. J. 2014. Analyses of charophyte chloroplast genomes help characterize the ancestral chloroplast genome of land plants. *Genome Biol. Evol.* 6:897–911.

Convey, P. 2007. Biogeography. *In* Riffenburgh, B. [Ed.] *Encyclopedia of the Antarctic, vol. 1.* Routledge, New York, pp. 154–60.

Friedl, T. & O'Kelly, C. J. 2002. Phylogenetic relationships of green algae assigned to the genus *Planophila* (Chlorophyta) evidence from 18S rDNA sequence data and ultrastructure. *Eur. J. Phycol.* 37:373–84.

Graham, L. E., Graham, J. M. & Wilcox, L. W. 2009. *Algae*, 2nd edn. Pearson Education, San Francisco, California, 616 pp.

Gray, M. W., Burger, G. & Lang, B. F. 1999. Mitochondrial evolution. *Science* 283:1476–81.

Handa, S., Nakahara, M., Tsubota, H., Deguchi, H. & Nakano, T. 2003. A new aerial alga, *Stichococcus ampulliformis* sp. nov. (Trebouxiophyceae, Chlorophyta) from Japan. *Phycol. Res.* 51:203–10.

Huelsenbeck, J. P. & Ronquist, F. 2000. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–5.

Jeong, H., Lim, J. M., Park, J., Sim, Y. M., Choi, H. G., Lee, J. & Jeong, W. J. 2014. Plastid and mitochondrion genomic sequences from Arctic *Chlorella* sp. ArM0029B. *BMC Genom.* 15:286.

Karsten, U., Friedl, T., Schmann, R., Hoyer, K. & Lembcke, S. 2005. Mycosporine-like amino acids and phylogenies in green algae: *Prasiola* and its relatives from the Trebouxiophyceae (Chlorophyta). *J. Phycol.* 41:557–66.

Katoh, K. & Standley, D. M. 2013. MAFFT Multiple Sequence Alignment Software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30:772–80.

Kosugi, M., Katashima, Y., Aikawa, S., Tanabe, Y., Kudoh, S., Kashino, Y., Koike, H. & Satoh, K. 2010. Comparative study on the photosynthetic properties of *Prasiola* (Chlorophyceae) and *Nostoc* (Cyanophyceae) from Antarctic and non-antarctic sites. *J. Phycol.* 46:466–76.

Kováčik, L. & Pereira, A. B. 2001. Green alga *Prasiola crispa* and its lichenized form *Mastodia tesselatain* Antarctic environment: general aspects. *In* Elster, J., Seckbach, J., Vincent, W. F. & Lhotský, O. [Eds.] *Algae and Extreme Environments. Ecology and Physiology*, Beiheft 123. Nova Hedwigia, Stuttgart, Germany, pp. 465–78.

Lemieux, C., Otis, C. & Turmel, M. 2014a. Six newly sequenced chloroplast genomes from prasinophyte green algae provide insights into the relations ships among prasinophyte lineages and the diversity of streamlined genome architecture in picoplanktonic species. *BMC Genom.* 15:857.

Lemieux, C., Otis, C. & Turmel, M. 2014b. Chloroplast phylogenomic analysis resolves deep-level relationships within the green algal class Trebouxiophyceae. *BMC Evol. Biol.* 14:211.

Li, B., Wolsan, M., Wu, D., Zhang, W., Xu, Y. & Zeng, Z. 2014. Mitochondrial genomes reveal the pattern and timing of marten (*Martes*), wolverine (*Gulo*) and fisher (*Pekania*) diversification. *Mol. Phyl. Evol.* 80:156–64.

Lud, D., Buma, A. G. J., Van De Poll, W., Moerdijk, T. C. W. & Huiskes, A. H. L. 2001. DNA damage and photosynthetic performance in the Antarctic terrestrial alga *Prasiola crispa* ssp. *antarctica* (Chlorophyta) under manipulated UV-B radiation. *J. Phycol.* 37:459–67.

Maul, J. E., Lilly, J. W., Cui, L., de Pamphilis, C. W., Miller, W., Harris, E. H. & Stern, D. B. 2002. *Chlamydomonas* chloroplast chromosome: islands of genes in a sea of repeats. *Plant Cell.* 14:2659–79.

Moniz, M. B. J., Rindi, F., Novis, P. M., Broady, P. A. & Guiry, M. D. 2012. Molecular phylogeny of Antarctic *Prasiola* (Prasiolales, Trebouxiophyceae) reveals extensive cryptic diversity. *J. Phycol.* 48:940–55.

Öquist, G. & Huner, N. P. 2003. Photosynthesis of over wintering evergreen plants. *Annu. Rev. Plant Biol.* 54:329–55.

Palmer, J. D. 1985. Evolution of chloroplast and mitochondrial DNA in plants and algae. *In* Mac Intyre, R. J. [Ed.] *Molecular Evolutionary Genetics.* Plenum Press, New York, pp. 131–40.

Pombert, J. F. & Keeling, P. J. 2010. The mitochondrial genome of the entomoparasitic green alga *Helicosporidium*. *PLoS ONE* 5:e8954.

Rixiao, L., Bingxin, H., Lanping, D. & Shujun, L. 2009. Two new species of family Prasiolaceae (Prasiolales, Chlorophyta) from China Sea. *Acta Oceanol. Sin.* 28:118–22.

Smith, S. A. & Dunn, C. W. 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics* 24:715–6.

Takahashi, S. & Murata, N. 2008. How do environmental stresses accelerate photoinhibition? *Trends Plant Sci.* 13:178–82.

Turmel, M., Brouard, J. S., Gangnon, C., Otis, C. & Lemieux, C. 2008. Deep division in the Chlorophyceae (Chlorophyta) revealed by chloroplast phylogenomic analyses. *J. Phycol.* 44:739–50.

Turmel, M., Otis, C. & Lemieux, C. 2007. An unexpectedly large and loosely packed mitochondrial genome in the charophycean green alga *Chlorokybus atmophyticus*. *BMC Genom.* 8:137.

Turmel, M., Otis, C. & Lemieux, C. 2009. The chloroplast genomes of green algae *Pedinomonas minor, Parachlorella kessleir* and *Oocystis solitaria* reveal a shared ancestry between the Pedinomomadales and Chlorellales. *Mol. Biol. Evol.* 26:2317–31.

Wakasugi, T., Nagai, T., Kapoor, M., Sugita, M., Ito, M., Ito, S., Tsudzuki, J. et al. 1997. Complete nucleotide sequence of the chloroplast genome from the green alga *Chlorella vulgaris*: the existence of genes possibly involved in chloroplast division. *Proc. Natl. Acad. Sci. USA* 94:5967–72.

Wang, L., Mao, Y., Kong, F., Li, G., Ma, F., Zhang, B., Sun, P., Bi, G., Zhang, F., Xue, H. & Cao, M. 2013. Complete sequence and analysis of plastid genomes of two economically important red algae: *Pyropiahaitanensis* and *Pyropiayezoensis*. *PLoS ONE* 8:e65902.

Yang, X., Cameron, S. L., Lees, D. C., Xue, D. & Han, H. 2015. A mitochondrial genome phylogeny of owlet moths (Lepidoptera: Noctuoidea), and examination of the utility of mitochondrial genomes for lepidopteran phylogenetics. *Mol. Phylogenet. Evol.* 85:230–7.

Zhong, B., Xi, Z., Goremykin, V. V., Fong, R., McLaenachan, P. A., Novis, P. M., Davis, C. C. & Penny, D. 2014. Streptophyte algae and the origin of land plants revisited using

heterogeneous models with three new algal chloroplast genomes. *Mol. Biol. Evol.* 31:177–83.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web site:

Figure S1. Phylogenetic analyses of *Prasiola crispa* within gene marker *rbcL*. Phylogenetic analysis of gene marker *rbcL* of 136 species. The tree presented was reconstructed using a Bayesian approach with the LG + I + G amino acid substitution model. Branch supports indicate posterior probability values (over each branch). The *Prasiola* clade is highlighted.

Figure S2. Phylogenetic analyses of *Prasiola crispa* within gene marker *psaB*. Phylogenetic analysis of gene marker *psaB* of 96 species. The tree presented was reconstructed using a Bayesian approach with the LG + I + G amino acid substitution model. Branch supports indicate posterior probability values (over each branch). The *Prasiola* clade is highlighted.

Table S1. GenBank accession numbers for sequences used in the chloroplast phylogenomic analyses.

Table S2. GenBank accession numbers for sequences used in the mitochondrial phylogenomic analyses.

Table S3. Species and authorities cited in the study.

4. ARTIGO 2

O artigo intitulado "*De novo assembly and annotation of the Antartic alga Prasiola crispa transcriptome"* de autoria de **Carvalho E.L**., Maciel L.F., Macedo P.E., Dezordi F.Z., Abreu M.E.T., Victória F.C., Pereira A.B., Boldo J.T., Wallau G.D.L., Pinto P.M. foi publicado no periódico *Frontiers in Molecular Biosciences* (ISSN: 2296-889X), no ano de 2018.

Neste artigo, nosso grupo apresenta os dados de uma análise transcriptômica de *P. crispa*. O número de *contigs* obtido foi igual à 17.201 e o conteúdo GC de 49,66%. O total de *reads* brutas foi de 42.978.976.

O transcriptoma de *P. crispa* é o primeiro disponível de uma espécie da ordem Prasiolales. Estes dados ajudarão a identificar os genes responsáveis pela sobrevivência de desta alga no ambiente Antártico e também podem ser empregues em estudos genéticos, filogenéticos e biotecnológicos de *P. crispa*.

frontiers in Molecular Biosciences

DATA REPORT
published: 08 January 2018
doi: 10.3389/fmolb.2017.00089

De novo Assembly and Annotation of the Antarctic Alga *Prasiola crispa* Transcriptome

Evelise L. Carvalho[1†], Lucas F. Maciel[1†], Pablo E. Macedo[1], Filipe Z. Dezordi[1], Maria E. T. Abreu[1], Filipe de Carvalho Victória[2], Antônio B. Pereira[2], Juliano T. Boldo[1], Gabriel da Luz Wallau[3] and Paulo M. Pinto[1]*

[1] Applied Proteomics Laboratory, Federal University of Pampa, São Gabriel, Brazil, [2] Núcleo de Estudos da Vegetação Antártica, Federal University of Pampa, São Gabriel, Brazil, [3] Department of Entomology, Aggeu Magalhães Institute (IAM), Recife, Brazil

Keywords: RNA-seq, Trebouxiophyceae, Prasiolales, transcriptome, extreme environments, anti-freeze proteins

OPEN ACCESS

Edited by:
Philipp Kapranov, Huaqiao University, China

Reviewed by:
Sergio Verjovski-Almeida, University of São Paulo, Brazil
Peter G. Zaphiropoulos, Karolinska Institute (KI), Sweden

***Correspondence:**
Paulo M. Pinto
paulopinto@unipampa.edu.br

[†]These authors have contributed equally to this work.

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Molecular Biosciences

Received: *26 September 2017*
Accepted: *05 December 2017*
Published: *08 January 2018*

Citation:
Carvalho EL, Maciel LF, Macedo PE, Dezordi FZ, Abreu MET, Victória FdC, Pereira AB, Boldo JT, Wallau GdL and Pinto PM (2018) De novo Assembly and Annotation of the Antarctic Alga Prasiola crispa Transcriptome. Front. Mol. Biosci. 4:89. doi: 10.3389/fmolb.2017.00089

INTRODUCTION

The Antarctic, located on the South Pole of the Earth and isolated from other continents by the Atlantic, Pacific and Indian oceans. It is considered a continent with severe environmental conditions for the development of life, thus limiting the Antarctic fauna and flora to specific organisms that have survival adaptation mechanisms (Jackson and Seppelt, 1997). The average annual precipitation of Antarctic is only 200 mm with winds of 327 km/h and temperatures below −90°C have already been recorded (Martínez-Rosales et al., 2012). The total area is 14,000,000 km², with 98–99.7% covered by snow and ice, with layers averaging 1.6 km thick (Convey et al., 2008; Martínez-Rosales et al., 2012). In addition, the ozone hole over the Antarctic region, first described in the 1980s, causes a high rate of ultraviolet radiation over the region, which is intensified by the ice-generated reflection (Kuttippurath and Nair, 2017; Marizcurrena et al., 2017).

Among the algae present on the Antarctic ice-free areas, *Prasiola crispa* (Lightfoot) Kützing is the most commonly found organism. *P. crispa* is a green macroalga belonging to the Trebouxiophyceae class and is among the most important primary producers in the Antarctic territory. *P. crispa* occurs in hydro-terrestrial soils, in the supralittoral zones of the maritime and continental Antarctica, where they form large and green carpets on the humid soil. *P. crispa* is found close to bird populations, mainly adjacent to penguin colonies, where the soil is rich in guano, a substrate with a high incidence of uric acid and nitrogen compounds (Kováčik and Pereira, 2001).

The morphology of these algae varies from unisserated filaments to stalks in the form of a tape, expanded blades or packages as colonies, which are characterized by a large phenotypic plasticity related to environmental factors (Rindi et al., 2007).

During the course of the seasons, *P. crispa* needs to tolerate extreme environments, such as repeated freeze and thaw cycles, physiological drought, salinity stress, and high levels of UV radiation (Jacob et al., 1992; Jackson and Seppelt, 1997). However, the genes associated with these adaptive characteristics in *P. crispa* remain unknown. Therefore, to better understand the genetic and metabolic adaptations that allow this organism to survive in harsh environments, we sequenced its transcriptome.

Transcriptomes represent all the expressed fractions of genomes and are a viable alternative to understand and characterize genome wide genetic information of organisms since it simplifies genetic analyses, as compared to whole genome sequencing (Riesgo et al., 2012).

High-throughput sequencing of transcriptomes (RNA-Seq) has provided new routes to study the genetic and functional information stored within any organism at an unprecedented scale and speed. Transcriptome approaches have been employed in a large number of the studies involving

non-model organisms, which normally lack reference genomes (Ekblom and Galindo, 2011; Haas et al., 2013).

Among these organisms are the algae group. The available data consists of organisms belonging to different phylum, such as Prymnesiophyte (Koid et al., 2014), Chlorophyta (Rismani-Yazdi et al., 2011), Haptophyta (Talarski et al., 2016), Stramenopiles (Im et al., 2015), and Rhodophyta (Shuangxiu et al., 2014).

P. crispa represents the first organism of the Prasiolales order with an available transcriptome since, until this work, the mitochondrial and plastid genomes were the only molecular data available for this species (Carvalho et al., 2015, 2017). Therefore, the purpose of this study was to sequence the transcriptome of *P. crispa*. The identification of transcripts will help to identify genes that are responsible for organism survival in this environment, as well as assisting in future genetic, phylogenetic, and biotechnological studies of *P. crispa* and other Antarctic organisms.

EXPERIMENTAL DESIGN, MATERIALS, AND METHODS

Algae Collection

P. crispa was collected in areas near the Arctowski Polish Station Region, Admiralty Bay, King George Island (61°50′−62°15′ S and 57°30′−59°00′W), Antartic. The collection was carried out in the Antarctic summer, in January of 2014 Austral summer, with temperature ranging from 0.5 to 2.0°C. The samples were maintained in RNAlater® (Sigma-Aldrich, USA) until RNA extraction.

Total RNA Extraction and RNA-Seq

Total RNA was extracted from three pools of samples using an RNAqueous®-Micro Total RNA Isolation Kit (Thermo Fisher Scientific Inc., USA) according to the manufacturer's instructions. The RNA-Seq library was prepared using random primers. The transcriptome was sequenced by Macrogen Service using the Solexa-Illumina HiSeq 2000 next-generation sequencing platform device according to the manufacturer's instructions. A paired-end reads with a read size of ∼100 bp separated by insert size of 300 bp was employed.

The BioProject ID of our data is PRJNA329112, and the BioSample accession number is SAMN05392062. All raw reads were deposited into the Sequencing Read Archive (SRA) of NCBI with accession number SRR5754271. This Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GFTS00000000.

De novo Transcriptome Assembly

Raw reads from data sets were filtered to remove the adapter sequences, and low quality reads with Fastx-toolkit (quality cut-off = 30) (Gordon and Hannon, 2010) and Trimmomatic v 0.36 using default parameters (Bolger et al., 2014). Next, we used Trinity version r2014-07-17 (Ekblom and Galindo, 2011) as Bruijn graph assembler with 25 kmer size. Due to the sequencing of a complex sample extracted from the Antarctic soil, we expected some amount of bacterial and fungal contamination. Therefore, in order to remove such contaminants, we performed a tblastx with default parameters (Altschul et al., 1990) searching against all of the NCBI nucleotide non-redundant database and recovered all contigs in which the best blast hit occurred with algae and plant sequences. Next, we used Bowtie2 (Langmead and Salzberg, 2012) with default parameters to recover only the reads that mapped against those *P. crispa* contigs.

Functional Annotation

The assembled and recovered contigs were searched against the NCBI protein non-redundant database using the BLASTX algorithm; the *E*-value cut-off was set at 1e-10. Genes were tentatively identified based on the best hits against known sequences. Blast2GO (Conesa and Götz, 2008) was used for mapping and annotation, associating Gene Ontology (GO) terms and predicting their function. Assembled and annotated transcriptome is publicly available on Figshare at: https://figshare.com/s/a60ae8a0445d547b9360.

After a stringent filtering process, the processed reads were assembled into 17,201 contigs. Statistics of the assembly are summarized in **Table 1**. CD-HIT-EST version 4.6.8, 2017-06-21 (Fu et al., 2012) was used for clustering of assembled transcripts with the default parameters with sequence identity threshold set to 95%, in order to indicate the number of unigenes. The clustering reduced the number of transcripts marginally by 3.5%. Our analysis showed that 93% of the unigenes are represented by only one isoform. The metrics of *P. crispa* were compared with others transcriptomes of the organism from the Trebouxiophyceae class, including *Chlorella minutissima* (Yu et al., 2016), *Trebouxia gelatinosa* (Carniel et al., 2016), *Coccomyxa subellipsoidea* (Peng et al., 2016), *Chlorella sorokiniana* (Li et al., 2016), *Botryococcus braunii* (Xu et al., 2015), and was found to have the lowest number of total reads. In relation to the number of contigs, *P. crispa* is in an intermediary position, with *C. sorokiniana* being the organism with the highest number of contigs (63,811) and *C. subellipsoidea* having the lowest number of contigs (9,409). More information on the metrics of transcriptomes is given in Supplementary Table S1.

The search of these contigs against the NCBI protein non-redundant database with BLASTX demonstrates that 8,980 (52.19%) sequences had at least one hit. The mapping of the sequences against the GO database retrieved 7,009 sequences mapped, and all assigned GO terms were classified into three

TABLE 1 | Summary of *Prasiola crispa* assembly.

Attributes	Value
Total raw reads	42,978,976
Total processed reads	5,233,428
Number of contigs	17,201
Total length (pb)	13,127,645
N50	1,036
GC (%)	49.66
Average size (pb)	763.19
Size range (pb)	200–12,802
Contigs > 1 kb	3,973 (22.05%)

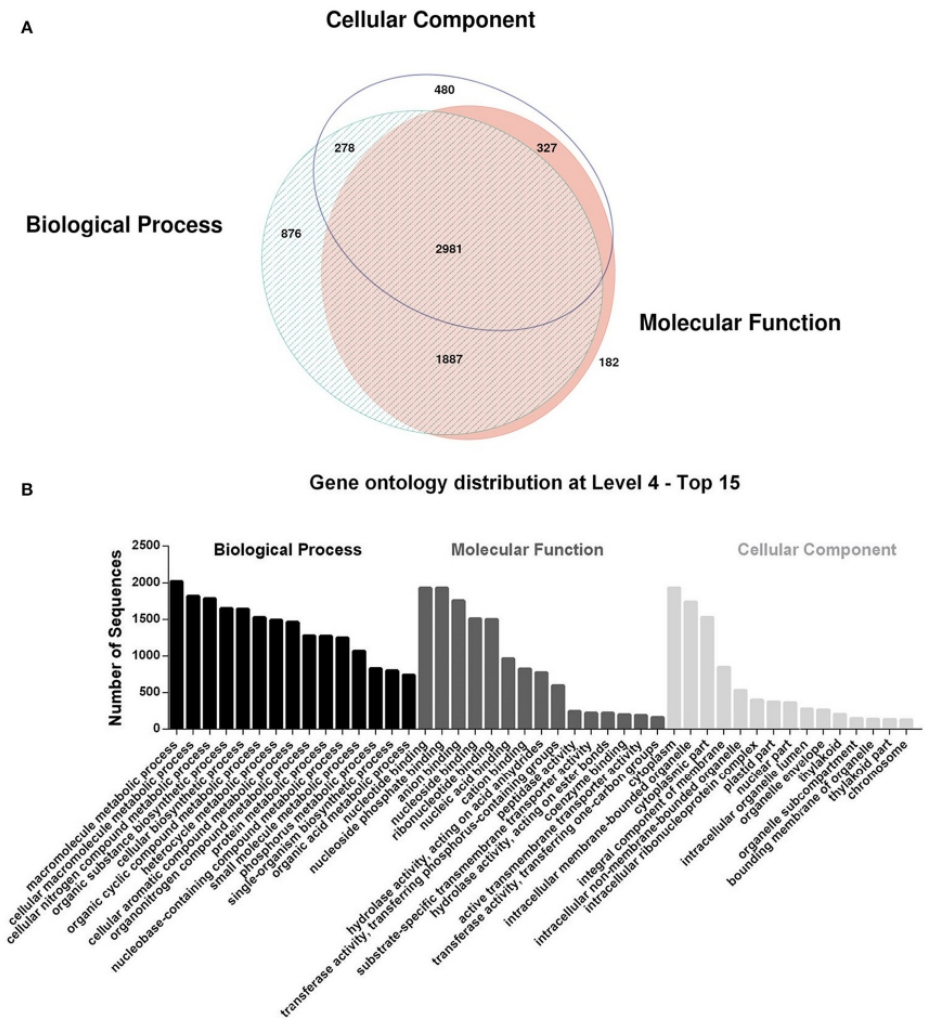


FIGURE 1 | Gene Ontology annotation: **(A)** Venn diagram of the distribution of mapped contigs with Biological Process, Molecular Function and Cellular Component terms and **(B)** distribution of the Top 15 terms at level 4 in the three main categories. The intersection areas indicate the contigs mapped with terms from two or three categories. Venn diagram are performed by software eulerAPE v.3.0.

main categories: cellular component, molecular function, and biological process. The distribution of sequences mapped to the three different categories and the top 15 GO terms are represented in **Figures 1A,B**, respectively.

Thus, taking the comparative numbers of contigs, mean length, the clustering process and sequencing depth and coverage into account its likely that our assembly comprises a representative number of transcripts but which were partially reconstructed. Moreover, it is important to note that although we have used stringent blast parameters to remove contaminations, some contigs can still come from contamination. Therefore, experimental validation to confirm *P. crispa* origin of some contigs is warranted.

Data Validation and Quality Control

The reading quality of the data of this transcriptomic analysis was evaluated through FastQC software (Babraham Bioinformatics) [RRID:SCR_014583]. The paired-end reads results were merged using MultiQC (http://multiqc.info) and are shown in Supplementary Figure S1. Per base quality phred scores range from 32.78 to 40.06, indicating base call accuracies of >99.9% (Supplementary Figure S1A). Per sequence quality

shows that 99.62% of reads had a mean phred score of 30 or above (Supplementary Figure S1B) and per base N content was low, with a maximum value 0.18% (Supplementary Figure S1C).

Re-use Potential

Through the data of this transcriptome, it is possible to perform searches for genes, aiming the heterologous expression of proteins with biotechnological potential, such as antifreeze proteins, which act to inhibit freezing of intracellular fluids (Nath et al., 2013), heat-shock proteins that play an important role in maintain biological activities in algae present in these acclimatization process (Li and Brawley, 2004) and mycosporine-like amino acids responsive to high incidence of ultraviolet radiation (Kováčik and Pereira, 2001). Proteomics approaches may also be employed, aiming at the confirmation of gene expression at the translational level.

AUTHOR CONTRIBUTIONS

EC, LM, GW, PP: Conducted the experiment; EC, LM, PM, FD, MA, GW: Performed analysis on the data; FV, AP, JB, PP: Conceived the project and acquired funding; EC, LM, PM, GW, PP: Wrote the manuscript.

ACKNOWLEDGMENTS

This work was supported by the National Council for Scientific and Technological Development (CNPq-Brazil), the Coordination for the Improvement of Higher Education Personnel (CAPES-Brazil), the Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS-Brazil) and National Institute of Science and Technology—Antarctic Environmental Research (INCT-APA). We thank the Bioinformatic Core at the Aggeu Magalhães Institute for the support with the bioinformatic analysis. This is for Sofia.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2017.00089/full#supplementary-material

REFERENCES

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

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

Carniel, F. B., Gerdol, M., Montagner, A., Banchi, E., De Moro, G., Manfrin, C., et al. (2016). New features of desiccation tolerance in the lichen photobiont Trebouxia gelatinosa are revealed by a transcriptomic approach. *Plant Mol. Biol.* 91, 319–339. doi: 10.1007/s11103-016-0468-5

Carvalho, E. L., Wallau, G. L., Rangel, D. L., Machado, L. C., Pereira, A. B., Victoria, F. C., et al. (2017). Phylogenetic positioning of the Antarctic alga Prasiola crispa (Trebouxiophyceae) using organellar genomes and their structural analysis. *J. Phycol.* 53, 908–915. doi: 10.1111/jpy.12541

Carvalho, E. L., Wallau, G. L., Rangel, D. L., Machado, L. C., da Silva, A. F., da Silva, L. F., et al. (2015). Draft plastid and mitochondrial genome sequences from Antarctic alga Prasiola crispa. *Genome Announc.* 3:e01151-15. doi: 10.1128/genomeA.01151-15

Conesa, A., and Götz, S. (2008). Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* 2008:619832. doi: 10.1155/2008/619832

Convey, P., Gibson, J. A. E., Hillenbrand, C. D., Hodgson, D. A., Pugh, P. J., Smellie, J. L., et al. (2008). Antarctic terrestrial life – challenging the history of the frozen continent? *Biol. Rev. Camb. Philos. Soc.* 83, 103–117. doi: 10.1111/j.1469-185X.2008.00034.x

Ekblom, R., and Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* 107, 1–15. doi: 10.1038/hdy.2010.152

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. doi: 10.1093/bioinformatics/bts565

Gordon, A., and Hannon, G. J. (2010). *Fastx-Toolkit. FASTQ/A Short-Reads Pre-Processing Tools (unpublished)*. Available online at: http://hannonlab.Cshl.Edu/fastx_toolkit/

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–512. doi: 10.1038/nprot.2013.084

Im, S., Choi, S., Hwang, M. S., Park, E. J., Jeong, W. J., and Choi, D. W. (2015). *De novo* assembly of transcriptome from the gametophyte of the marine red algae Pyropia seriata and identification of abiotic stress response genes. *J. Appl. Phycol.* 27, 1343–1353. doi: 10.1007/s10811-014-0406-3

Jackson, A. E., and Seppelt, R. D. (1997). "Physiological adaptations to freezing and UV radiation exposure in Prasiola crispa, an Antarctic terrestrial alga," in *Antarctic Communities: Species, Structure, and Survival*, eds B. Battaglia, J. Valencia, and D. W. H. Walton (Cambridge: University Press), 226–233.

Jacob, A., Wiencke, C., Lehmann, H., and Kirst, G. O. (1992). Physiology and ultrastructure of desiccation in the green alga *Prasiola crispa* from Antarctica. *Botanica Marina* 35, 297–303. doi: 10.1515/botm.1992.35.4.297

Koid, A. E., Liu, Z., Terrado, R., Jones, A. C., Caron, D. A., and Heidelberg, K. B. (2014). Comparative transcriptome analysis of four Prymnesiophyte algae. *PLoS ONE* 9:e97801. doi: 10.1371/journal.pone.0097801

Kováčik, L., and Pereira, A. B. (2001). "Green alga Prasiola crispa and its lichenized form Mastodia tesselata in Antartic," in *Algae and Extreme Environments*, eds J. Elster, J. Seckbach, W. F. Vincent, and O. Lhotský (Czech Republic: Nova Hedwigia 123), 465–478.

Kuttippurath, J., and Nair, P. J. (2017). The signs of Antarctic ozone hole recovery. *Sci. Rep.* 7:585. doi: 10.1038/s41598-017-00722-7

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

Li, L., Zhang, G., and Wang, Q. (2016). *De novo* transcriptomic analysis of Chlorella sorokiniana reveals differential genes expression in photosynthetic carbon fixation and lipid production. *BMC Microbiol.* 16:223. doi: 10.1186/s12866-016-0839-8

Li, R., and Brawley, S. H. (2004). Improved survival under heat stress in intertidal embryos (*Fucus* spp.) simultaneously exposed to hypersalinity and the effect of parental thermal history. *Mar. Biol.* 144, 205–213. doi: 10.1007/s00227-003-1190-9

Marizcurrena, J. J., Morel, M. A., Braña, V., Morales, D., Martinez-López, W., and Castro-Sowinsk, S. (2017). Searching for novel photolyases in UVC-resistant Antarctic bacteria. *Extremophiles* 21, 409–418. doi: 10.1007/s00792-016-0914-y

Martínez-Rosales, C., Fullana, N., Musto, H., and Castro-Sowinski, S. (2012). Antarctic DNA moving forward: genomic plasticity and biotechnological potential. *FEMS Microbiol Lett.* 331, 1–9. doi: 10.1111/j.1574-6968.2012.02531.x

Nath, A., Chaube, R., and Subbiah, K. (2013). An insight in to the molecular basis for convergent evolution in fish antifreeze proteins. *Comput. Biol. Med.* 43, 817–821. doi: 10.1016/j.compbiomed.2013.04.013

Peng, E., Wei, D., Chen, G., and Chen, F. (2016). Transcriptome analysis reveals global regulation in response to CO_2 supplementation in oleaginous microalga Coccomyxa subellipsoidea C-169. *Biotechnol. Biofuels.* 9:151. doi: 10.1186/s13068-016-0571-5

Riesgo, A., Andrade, S. C., Sharma, P. P., Novo, M., Pérez-Porro, A. R., Vahtera, V., et al. (2012). Comparative description of ten transcriptomes of newly sequenced invertebrates and efficiency estimation of genomic sampling in non-model taxa. *Front. Zool.* 9:33. doi: 10.1186/1742-9994-9-33

Rindi, F., McIvor, L., Sherwood, A. R., Friedl, T., Guiry, M. D., and Sheath, R. G. (2007). Molecular phylogeny of the green algal order Prasiolales (Trebouxiophyceae, Chlorophyta). *J. Phycol.* 43, 811–822. doi: 10.1111/j.1529-8817.2007.00372.x

Rismani-Yazdi, H., Haznedaroglu, B. Z., Bibby, K., and Peccia, J. (2011). Transcriptome sequencing and annotation of the microalgae *Dunaliella* tertiolecta: pathway description and gene discovery for production of next-generation biofuels. *BMC Genomics* 12:148. doi: 10.1186/1471-2164-12-148

Shuangxiu, W., Sun, J., Chi, S., Wang, L., Wang, X., Liu, C., et al. (2014). Transcriptome sequencing of essential marine brown and red algal species in China and its significance in algal biology and phylogeny. *Acta Oceanol. Sin.* 33, 1–12. doi: 10.1007/s13131-014-0435-4

Talarski, A., Manning, S. R., and La Claire, J. W. II. (2016). Transcriptome analysis of the euryhaline alga, Prymnesium parvum (Prymnesiophyceae): effects of salinity on differential gene expression. *Phycologia* 55, 33–44. doi: 10.2216/15-74.1

Xu, Z., He, J., Qi, S., and Liu, J. (2015). Nitrogen deprivation-induced de novo transcriptomic profiling of the oleaginous green alga *Botryococcus braunii* 779. *Genom. Data* 6, 231–233. doi: 10.1016/j.gdata.2015.09.019

Yu, M., Yang, S., and Lin, X. (2016). *De-novo* assembly and characterization of Chlorella minutissima UTEX2341 transcriptome by paired-end sequencing and the identification of genes related to the biosynthesis of lipids for biodiesel. *Mar. Genomics* 25, 69–74. doi: 10.1016/j.margen.2015.11.005

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

5. CONSIDERAÇÕES FINAIS

Considerando os dados obtidos no presente estudo pode-se concluir que os genomas de organelas fornecem uma riqueza de dados filogeneticamente informativos, tornando-os uma abordagem valiosa para embasar estudos filogenômicos. Contudo, a determinação de uma maior variedade de sequências de genomas mitocondriais de algas verdes permitirá uma reconstrução mais rigorosa das relações taxonômicas destes organismos. Os genomas plastidiais da maioria das plantas terrestres compartilham genes conservados, ordem e conteúdo gênico semelhantes, fato este que não foi observado em nosso estudo, visto que a baixa sintenia observada entre os genomas plastidiais do clado *Prasiola* não condiz com o encontrado em plantas terrestres. Os dados gerados a partir da análise filogenômica e sintênica, utilizando os genomas acessórios da alga *P. crispa* fornecem um aporte para estudos futuros mais aprofundados.

Quanto ao transcriptoma de *P. crispa*, o RNA foi extraído e sequenciado e as análises de validação e qualidade permitiram a remoção das *reads* de baixa qualidade e contaminantes, trazendo maior confiabilidade aos dados. Ao comparar as métricas de outras algas da mesma classe é possível concluir que o transcriptoma foi montado de maneira satisfatória, onde 52,19% dos transcritos foram identificados e anotados funcionalmente. A identificação dos transcritos poderá contribuir na identificação dos genes responsáveis pela sobrevivência de *P. crispa*, bem como auxiliar em futuros estudos genéticos, filogenéticos e biotecnológicos de *P. crispa* e outros organismos Antárticos.

6. PERSPECTIVAS

- Gerar mais sequências dos genomas das organelas na tentativa de resolver os *gaps* entre os *contigs* de cpDNA e mtDNA.
- Após a conclusão do genoma, reanotar os genomas para confirmar a atual anotação e buscar possíveis novos genes.
- Realizar uma análise e categorização funcional dos genes pertencentes ao cpDNA e mtDNA de *P. crispa*.
- Sequenciar o genoma nuclear de *P. crispa*, em busca de genes relacionados com o metabolismo oxidativo e fotossistemas que possam ter sido transferidos durante a endossimbiose.
- Identificar potenciais *Ice Binding Proteins* (IBPs) por alinhamento contra banco de dados local.
- Modelar a estrutura tridimensional de IBPs identificadas, a partir de sua sequência aminoacídica.
- Realizar a simulação *in silico* de dinâmica molecular das possíveis IBPs identificadas para avaliar o comportamento destas estruturas proteicas em diferentes temperaturas e as interações com cristais de gelo.

7. REFERÊNCIAS BIBLIOGRÁFICAS

Alberdi, M., Bravo, L.A., Gutiérrez, A., Gidekel, M., Corcuera, L.J. **Ecophysiology of Antarctic vascular plants.** Physiologia Plantarum, 115, p. 479-486, 2002.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. Basic Local Alignment Search Tool. Journal of Molecular Biology, 215, p. 403-410, 1990.

Arrigo, K.R.,Worthen, D.L., Lizotte, M.P., Dixon, P., Dieckmann, G. **Primary production in Antarctic sea ice.** Science 276, p. 394–397, 1997.

Baba, M., Ioki, M., Nakajima, N., Shiraiwa, Y., Watanabe, M.M. T**ranscriptome analysis of an oil-rich race A strain of** *Botryococcus braunii* **(BOT-88-2) by *de novo* assembly of pyrosequencing cDNA reads.** Bioresource Technology, 109, p. 282-286, 2012.

Behura, S.K. **Insect phylogenomics.** *Insect Molecular Biology*, 24, p. 403-411, 2015.

Benninghoff, W.S. **The Antarctic Ecosystem.** Environment International, 13, p. 9-14, 1987.

Beyer, L., Bölter, M., Seppelt, R.D. **Nutrient and thermal regime, microbial biomass and vegetation of Antarctic soils in the Windmill Islands Region of east Antarctica (Wilkes Land).** Arctic, Antarctic and Alpine Research, 32, p. 30-39, 2000.

Blair, C., Murphy, R.W. **Recent trends in molecular phylogenetic analysis: where to next?** Journal of Heredity, 102, p. 130-138, 2011.

Bolger, M.E., Arsova, B., Usadel, B. **Plant genome and transcriptome annotations: from misconceptions to simple solutions.** Briefings in Bioinformatics, 12, p. 1-13, 2017.

Brouard, J.S., Otis, C., Lemieux, C., Turmel, M. **The exceptionally large chloroplast genome of the green alga** *Floydiella terrestris* **illuminates the evolutionary history of the chlorophyceae.** Genome Biology and Evolution, 2, p. 240-256, 2010.

Carvalho, E.L., Wallau, G.L., Rangel, D.L., Machado, L.C., da Silva, A.F., da Silva, L.F., Macedo, P.E., Pereira, A.B., Victoria, F.C., Boldo, J.T., Dal Belo, C.A., Pinto, P.M. **Draft Plastid and Mitochondrial Genome Sequences from Antarctic Alga** *Prasiola crispa*. Genome Announcements, 3, e01151-15, 2015.

Conesa, A., Götz, S. **Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics.** International Journal of Plant Genomics, 2008, p. 1-13, 2008.

Convey, P. **Biogeography**. In: Encyclopedia of the Antarctic, 1, AK. Routledge, New York, p. 154-160, 2007.

Convey, P. **Encyclopedia of the Antarctic**. 1ed, Nova Iorque: Routledge, 2006.

Convey, P., Gibson, J.A.E., Hillenbrand, C.D., Hodgson, D.A., Pugh, P.J.A., Smellie, J.L., Stevens, M.I. **Antarctic terrestrial life – challenging the history of the frozen continent?** Biological Reviews, 83, p. 103-117, 2008.

daSilva, A.C.R., Pires, A.M.G., Ramos, C.J.B. Sanchez, F.F., Cavalcanti, D.N., Teixeira, V.L., Fuly, A.F. **The seaweed** *Prasiola crispa* **(Chlorophyta) neutralizes toxic effects of** *Bothrops jararacussu* **snake venom**. Journal Applied Phycology, 29, p. 781–788, 2017.

deCambiaire, J.C., Otis, C., Turmel, M., Lemieux, C. **The chloroplast genome sequence of the green alga** *Leptosira terrestris***: multiple losses of the inverted repeat and extensive genome.** BMC Genomics, 8:213, p. 1-13, 2007.

deKoning, A.P., Keeling, P.J. **The complete plastid genome sequence of the parasitic green alga** *Helicosporidium sp.* **is highly reduced and structured.** BMC Biology, 4, p.1-10, 2006.

Demmig-Adams, B., Adams, W.W. **Photoprotection and other responses of plants to high light stress.** The *Annual Review* of *Plant Biology*,43, p. 599–626, 1992.

Denovan-Wright, E.M., Nedelcu, A.M., Lee, R.W. **Complete sequence of the mitochondrial DNA of** *Chlamydomonas eugametos***.** Plant Molecular Biology, 36, 285-295, 1998.

Douce, R. Mitochondria in Higher Plants. **Structure, Function and Biogenesis.** Nova Iorque, Academic Press, 1985.

Duchêne, D.A. **Phylogenomics**. Current Biology, 31, p. 1177 – 1181, 2021.

Fan, J., Lee, R.W. **Mitochondrial genome of the colorless green alga** *Polytomella parva***: two linear DNA molecules with homologous inverted repeat Termini.** Molecular Biology and Evolution, 19, 999-1007, 2002.

Friedl, T., O'Kelly, C.J. **Phylogenetic relationships of green algae assigned to the genus** *Planophila* **(Chlorophyta) evidence from 18S rDNA sequence data and ultrastructure.** European Journal of Phycology, 37, p. 373–384, 2002.

Garber, M., Grabherr, M.G., Guttman, M., Trapnell, C. **Computational methods for transcriptome annotation and quantification using RNA-seq**. Nature Methods, 8, p. 469-477, 2011.

Graham, L.E. **Green algae to land plants: an evolutionary.** Journal of Plant Research, 109, p. 241-251, 1996.

Graham, L.E., Graham, J.M., Wilcox, L.W. **Algae**. 2. ed., São Francisco: Pearson Education, 2009.

Gray, M.W., Burger, G., Lang, B.F. **Mitochondrial evolution.** Science, 283, p. 1476-1481, 1999.

Hajibabaei, M., Singer, G.A.C, Herbert, P.D.N., Hickey, D.A. **DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics.** Trends IN Genetics, 23, p.167-172, 2007.

Hammani, K., Giege, P. **RNA metabolism in plant mitochondria.** Trends in Plant Science, 19, p. 380-389, 2014.

Harrison, N., Kidner, C.A. **Next-generation sequencing and systematics: what can a billion base of DNA sequence data do for you?** Taxon, 60, p. 1552-1566, 2011.

Heather, J.M., Chain, B. **The sequence of sequencers: The history of sequencing DNA.** Genomics, 107, p. 1-8, 2016.

Hillis, D.M. **Phylogenetic analysis.** Current Biology, 7, p. 129-131, 1997.

Hoham, R.W., Remias, D. **Snow and glacial algae: a review:** Journal of Phycology, 56, p. 264-282, 2020.

Huntley, R.P., Sawford, T., Martin, M.J., O'donovan, C. **Understanding how and why the Gene Ontology and its annotations evolve: the GO within UniProt.** GigaScience, 3, p. 1-9, 2014.

Jackson, A.E., Seppelt, R.D. **The accumulation of proline in** *Prasiola crispa* d**uring winter in Antarctica.** *Plant Physiology, 94, p. 25-30, 1995.*

Jeong, H., Lim, J.M., Park, J., Sim, Y.M., Choi, H.G., Lee, J., Jeong, W.J. **Plastid and mitochondrion genomic sequences from Arctic** *Chlorella sp.* **ArM0029B.** BMC Genomics, 15, p. 2-14, 2014.

John, D.M. **Order Prasiolales**: **The freshwater algal flora of the British Isles.** Cambridge University Press, Cambridge, p. 433-468, 2002.

Khan, A.K., Kausar, H., Jaferi, S.S., Drouet, S., Hano, C., Abbasi, B.H., Anjum, S. **An Insight into the Algal Evolution and Genomics.** Biomolecules, 6, p. 1511 – 1524, 2020.

Kappen, L. **Some aspects of the great success of lichens in the Antarctica.** Antarctic Science, 12, p. 314–324, 2000.

Microbial Environmental Genomics (MEG). Nova Iorque: Springer, p. 207-233, 2016.

Kerfeld, C.A., Scott, K.M. **Using BLAST to Teach "E-value-tionary" Concepts.** PLoS Biology, 9 (2: 1001014), p. 1-14, 2011.

Kim, K.M., Park, J.H., Bhattacharya, D., Yoon, H.S. **Applications of next-generation sequencing to unraveling the evolutionary history of algae.** Internacional Journal of Systematic and Evolutionary Microbiology, 64, p. 333-345, 2014.

Ko, H.L, Wang, Y.T., Chiu, T.S., Lee, M.A., Leu, M.Y., Chang, K.Z., Chen, W.Y., Shao, K.T. **Evaluanting the accuracy of morphological identification of larval fishes by applying DNA barcoding.** PLOS ONE, 8, p. 1-7, 2013.

Koid, A.E., Liu, Z., Terrado, R., Jones, A.C., Caron, D.A., Heidelberg, K.B. **Comparative transcriptome analysis of four prymnesiophyte algae.** PLoS One, 13, e97801, 2014.

Kolodner, R., Tewari, K.K. **Inverted repeats in chloroplast DNA from higher plants**. PNAS, 76, p. 41-45, 1979.

Kosugi, M., Katashima, Y., Aikawa, S., Tanabe, Y., Kudoh, S., Kashino, Y., Koike, H., Satoh, K. **Comparative study on the photosynthetic properties of** *Prasiola* **(Chlorophyceae) and** *Nostoc* **(Cyanophyceae) from Antarctic and non-antarctic sites.** Journal of Phycology, 46, p. 466-476, 2010.

Kováčik, L., Pereira, A.B. **Green alga** *Prasiola crispa* **and its lichenized form** *Mastodia tesselatain* **Antarctic environment: general aspects.** Nova Hedwigia, Beiheft, 123, p. 465-478, 2001.

Kuiper, M.J., Morton, C.J., Abraham, S.E., Gray-Weale, A. **The biological function of an insect antifreeze protein simulated by molecular dynamics.** eLife, 4, p. 1-14, 2015.

Legg, D.A., Sutton, M.D., Edgecombe, G.D. **Arthropod fossil data increase congruence of morphological and molecular phylogenies. Nature Communications**, 4, p. 2485, 2013.

Leliaert, F., Smith, D.R., Moreau, H., Herron, M.D., Verbruggen, H., Delwiche, C.F., deClerck, O. **Phylogeny and molecular evolution of the green algae.** Critical Reviews in Plant Sciences, 31, p. 1-46, 2012.

Lemieux, C., Otis, C., Turmel, M. **Six newly sequenced chloroplast genomes from prasinophyte Green algae provide insights into the relations ships among prasinophyte lineages and the diversity of streamlined genome architecture in picoplanktonic species.** BMC Genomics, 15:857, p. 1-20, 2014a.

Lemieux, C., Otis, C., Turmel, M. **Chloroplast phylogenomic analysis resolves deep-level relationships within the Green algal class Trebouxiophyceae.** BMC Evolution Biology, 14, p. 1-15, 2014b.

Letsch, M., Lewis, L.A. **Chloroplast gene arrangement variation within a closely related group of green algae (Trebouxiophyceae, Chlorophyta).** Molecular Phylogenetics and Evolution, 64, p. 524-532, 2012.

Lewis, L.A., McCourt, R.M. **Green algae and the origin of land plants.** American Journal of Botany, 91, p. 1535-1556, 2004.

Lu, X., Li, J., Yang, J., Liu, X., Ma, J. ***De novo*** **transcriptome of the desert beetle** *Microdera punctipennis* **(Coleoptera:** *Tenebrionidae***) using illumina RNA-seq technology.** Molecular Biology Reports, 41, p. 7293-7303, 2014.

Luckey, J.A., Drossman, H., Kostichka, A.J., Mead, D.A., D'cunha, J., Norris, T.B., Smith, L.M. **High speed DNA sequencing by capillary electrophoresis.** Nucleic Acids Research, 18, p. 4417-4421, 1990.

Lud, D., Buma, A.G.J., Van De Poll, W., Moerdijk, T.C.W., Huiskes, A.H.L. **DNA damage and photosynthetic performance in the Antarctic terrestrial alga** *Prasiola crispa ssp. antarctica* **(Chlorophyta) under manipulated UV-B radiation.** Journal of Phycology, 37, p. 459-467, 2001.

Marchant, A., Mougel, F., Mendonça, V., Quartier, M., Jacquin-Joly, E., Da Rosa, J.A., Petit, E., Harry, M. **Comparing *de novo* and reference-based transcriptome assembly strategies by applying them to the blood-sucking bug** *Rhodnius prolixus*. Insect Biochemistry and Molecular Biology, 69, p. 25-33, 2016.

Marinho, R.S.S., Ramos, C.J.B., Leite, J.P.G., Teixeira, V.L., Paixão, I.C.N.P., Belo, C.A.D., Pereira, A.B., Pinto, A.M.V. **Antiviral activity of 7-keto-stigmasterol obtained from green Antarctic algae** *Prasiola crispa* **against equine herpesvirus 1.** Journal of Applied Phycology, 29, p. 555-562, 2017.

Martínez-Rosales, C., Fullana, N., Musto, H., Castro-Sowinski, S. **Antarctic DNA moving forward: genomic plasticity and biotechnological potential.** FEMS Microbiology Letters, 331, p. 1-9, 2012.

Mattox, K.R., Stewart, K.D. **Classification of the green algae: a concept based on comparative cytology.** In: Irvine DEG, John DM, editors. The systematic of the green algae. London: Academic Press, p. 29-72, 1984.

Maxam, A.M., Gilbert, W. **A new method for sequencing DNA.** Proceedings of the National Academy of Sciences of United States of America, 74, p. 560-564, 1977.

Merrill, R.A., Strack, S. **Mitochondria: a kinase anchoring protein 1, a signaling plataform for mitochondrial form and function.** International Journal of Biochemistry & Cell Biology, 48, p. 92-96, 2014.

Michaelis, G., Vahrenholz, C., Pratje, E. **Mitochondrial DNA of** *Chlamydomonas reinhardtii***: the gene for apocytochrome *b* and the complete functional map of the 15.8 kb DNA.** Molecular and General Genetics, 223, p. 211-216, 1990.

Moniz, M.B.J., Guiry, M.D., Rindi, F. ***tufA*** **phylogeny and species boundaries in the green algal order Prasiolales (Trebouxiophyceae, Chlorophyta).** Phycologia, 53, p. 369-406, 2014.

Moniz, M.B.J., Rindi, F., Novis, P.M., Broady, P.A., Guiry, M.D. **Molecular phylogeny of Antarctic *Prasiola* (Prasiolales, Trebouxiophyceae) reveals extensive cryptic diversity.** Journal of Phycology, 48, p. 940-955, 2012.

Mota, A.P.Z. **Sequência complete do genoma cloroplasmático do feijão-caupi** [*Vigna ungiculata* (L.) Walp] **e diversidade genética de variedades tradicionais brasileiras e africanas.** http://repositorio.unb.br/handle/10482/12969?mode=full, 2012.

Novis, P.M., Smissen, R., Buckley, T.R., Gopalakrishnan, K., Visnovsky, G. **Inclusion of chloroplast gene that have undergone expansion misleads phylogenetic reconstruction in the Chlorophyta.** American Journal of Botany, 100, p. 2194-2209, 2013.

Orsini, M., Costelli, C., Malavasi, V., Cusano, R., Concas, A., Angius, A., Cao, G. **Complete genome sequence of mitochondrial DNA (mtDNA) of** *Chlorella sorokiniana***.** Mitochondrial DNA. No prelo. DOI: 10.3109/19401736.2014.953128, 2014.

Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.-Y., Cohoon, M., De Crécy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E.D., Gerdes, S., Glass, E.M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., Krause, L., Kubal, M., Larsen, N., Linke, B., Mchardy, A.C., Meyer, F., Neuweger, H., Olsen, G., Olson, R., Osterman, A., Portnoy, V., Pusch, G.D., Rodionov, D.A., Rückert, C., Steiner, J., Stevens, R., Thiele, I., Vassieva, O., Ye, Y., Zagnitko, O., Vonstein, V. **The Subsystems Approach to Genome Annotation and its Use in the Project to Annotate 1000 Genomes.** Nucleic Acids Research, 33, p. 5691-5702, 2005.

Pedrini, A.D.G. **Macroalgas**: Uma introdução à taxonomia. 1. ed., Rio de Janeiro: Techincal Books, 2010.

Phillpot, H.R. **Physical geography-climate.** In: Bonner, W.M., Walton, D.W.H. Key Environments Antarctica*,* Pergamon Press, Oxford, p. 23-28, 1985.

Picault, N., Hodges, M., Palmieri, L., Palmieri, F. **The growing family of mitochondrial carriers in Arabidopsis.** Trends Plant Science, 9, p. 138-146, 2004.

Pombert, J.F., Beauchamp, P., Otis, C., Lemieux, C., Turmel, M. **The complete mitochondrial DNA sequence of the green alga** *Oltmannsiellopsis viridis***: evolutionary trends of the mitochondrial genome in the Ulvophyceae.** Current Genetics, 50, p. 137-147, 2006.

Pombert, J.F., Otis, C., Lemieux, C., Turmel, M. **The complete mitochondrial DNA sequence of the Green alga** *Pseudendoclonium akinetum* **(Ulvophyceae) Highlightes distinctive evolutionary trends in the Chlorophyta and suggests a sister-group relationships between the Ulvophyceae and Chlorophyceae.** Molecular Biology and Evolution, 21, p. 922-935, 2004.

Pribil, M., Labs, M., Leister, D. **Structure and dynamics of thylakoids in land plants.** Journal of Experimental Botany, 12 p. 1-18, 2014.

Randy, C.F., C., Tzi, B.N., Jack H.W. **Antifreeze Proteins from Diverse Organisms and their Applications: An Overview.** Current Protein & Peptide Science, 18, p. 262-283, 2017.

Reyes-Prieto, A., Webwe, A.P.M., Bhattacharya, D. **The origin and establishment of the plastid in algae and plants.** Annual Review of Genetics, 41, p.147-168, 2007.

Rindi, F., McIvor, L., Guiry, M.D. **The Prasiolales (Chlorophyta) of the Atlantic Europe: an assessment based on morphological, molecular, and ecological data, including the characterization of** *Rosenvingiella radicans* **(Kützing) comb. nov**. Journal of Phycology, 40, p. 977-997, 2004.

Rindi, F., McIvor, L., Sherwood, A.R., Friedl, T., Guiry, M.D., Sheath, R.G. **Molecular phylogeny of the green algal order Prasiolales (Trebouxiophyceae, Chlorophyta).** Journal of Phycology, 43, p. 811-822, 2007.

Rodríguez, R.R., Jiménez, J.C., Delgado, C.M. **Microhabitat and morphometric variation in two new species of** *Prasiola* **(Prasiolales, Chlorophyta) from streams in central Mexico.** Aquatic Ecology, 41, p. 161 – 168, 2007.

Ronaghi, M., Uhlén, M., Nyrén, P. A **Sequencing Method Based on Real-Time Pyrophosphate.** Science, 281, p. 363-365, 1998.

Ruhfel, B.R., Gitzendanner, M.A., Soltis, P.S., Soltis, D.E., Burleigh, G. **From algae to angiosperms-inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes.** BMC Evolutionary Biology, 14, p. 1-27, 2014.

Sanger, F., Coulson, A.R. **A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase**. Journal of Molecular Biology, 94, p. 441-448, 1975.

Sanger, F., Nicklen, S., Coulson, A.R. **DNA sequencing with chain-terminating inhibitors.** Proceedings of the National Academy of Sciences of United States of America, 74, p. 5463-5467, 1977.

Schuster, W., Brennicke, A. **The plant mitochondrial genome: physical structure, information content, RNA editing, and gene migration to the nucleus.** Annual Review of Plant Physiology and Plant Molecular Biology, 45, p. 61-78, 1994.

Servín-Garcideueñas, L.E., Martínez-Romero, E. **Complete mitochondrial and plastid genomes of the Green Microalga** *Trebouxiophyceae sp.* **strain MX-AZ01 isolated from a Highly Acidic geothermal lake.** Eukaryotic Cell, 11, p. 1417-1418, 2012.

Simões, J.C. **O ambiente antártico: domínio de extremos.** In: Garcia, C.A.E., Evangelista, H., Simões, J.C., Goldemberg, J., Campos, L.S., Mata, M.M., Breme, U.F. Antártica e as Mudanças Globais: um desafio para a humanidade. Série Sustentabilidade. São Paulo, Blucher, p. 15–27, 2011.

Singh, S.M., Pereira, N., Ravindra, R. **Adaptive mechanisms for stress tolerance in Antarctic plants.** Current Science, 99, p. 334-340, 2010.

Sørensen, I., Rose, J.K.C., Doyle, J.J., Domozych, D.S., Willats, W.G.T. **The** *Charophycean* **green algae as model systems to study plant cell walls and other evolutionary adaptations that gave rise to land plants.** Plant Signaling & Behavior, 7, p. 1-3, 2012.

Tatusov, R.L., Galperin, M.Y., Natale, D.A., Koonin, E.V. **The COG database: a tool for genome-scale analysis of protein functions and evolution.** Nucleic Acids Research, 28, p. 33-36, 2000.

Turmel, M., Brouard, J.S., Gangnon, C., Otis, C., Lemieux, C. **Deep division in the Chlorophyceae (Chlorophyta) revealed by chloroplast phylogenomic analyses.** Jounal of Phycology, 44, p. 739-750, 2008.

Turmel, M., Lemieux, C., Burger, G., Lang, B.F., Otis, C., Plante, I., Gray, M.W. **The complete mitochondrial DNA sequences of** *Nephroselmis olivacea* **and** *Pedinomonas minor***: two radically different evolutionary patterns within green algae.** The Plant Cell, 11, p. 1717-1729, 1999.

Turmel, M., Otis, C., Lemieux, C. **An unexpectedly large and loosely packed mitochondrial genome in the charophycean Green alga** *Chlorokybus atmophyticus***.** BMC Genomics, 8:137, p.1-12, 2007.

Turmel, M.; Otis, C.; Lemieux, C. **The chloroplast genome sequence of** *Chara vulgaris* **sheds new light into the closet Green algal relatives of land plants.** Molecular Biology and Evolution, 23, p. 1324-1338, 2006.

Turmel, M.; Otis, C.; Lemieux, C. **The chloroplast genomes of Green algae** *Pedinomonas minor***,** *Parachlorella kessleir* **and** *Oocystis solitaria* **reveal a Shared ancestry between the Pedinomomadales and Chlorellales.** Molecular Biology and Evolution, 26, p. 2317-2331, 2009.

Umen, J., Herron, M.D. **Green Algal Models for Multicellularity.** Annual Review of Genetics, 55, p. 603-632, 2021.

Ustun, N.S., Turhan, S. **Antifreeze Proteins: Characteristics, Function, Mechanism of Action, Sources and Application to Foods.** Journal of Food Processing and Preservation, 39, p. 3189-3197, 2015.

Vahrenholz, C., Riemen, G., Pratje, E., Dujon, B., Michaelis, G. **Mitochondrial DNA of** *Chlamydomonas reinhardtii***: the structure of the ends of the linear 15.8-kb genome suggests mechanisms for DNA replication.** Current Genetics, 24, 241-247, 1993.

Venketesh, S., Dayananda, C. **Properties, Potentials, and Prospects of Antifreeze Proteins.** Critical Reviews in Biotechnology, 28, p. 57-82, 2008.

Voelkerding, K.V., Dames, S.A., Durtschi, J.D. **Next-Generation Sequencing: From Basic Research to Diagnostics.** Clinical Chemistry, 55, p. 641-658, 2009.

Wakasugi, T., Nagai, T., Kapoor, M., Sugita, M., Ito, M., Ito, S., Tsudzuki, J., Nakashima, K., Tsudzuki, T., Suzuki, Y., Hamada, A., Ohta, T., Inamura, A., Yoshinaga, K., Sugiura, M. **Complete nucleotide sequence of the chloroplast genome from the green alga** *Chlorella vulgaris***: the existence of genes possibly involved in chloroplast division.** Proceedings of the National Academy of Sciences, 94, p. 5967-5972, 1997.

Wang, Z., Gerstein, M., Snyder, M. **RNA-Seq: a revolutionary tool for transcriptomics.** Nature Reviews Genetics, 10, p. 57-63, 2009.

Wang, L., Si, Y., Dedow, L.K., Shao, Y., Liu, P., Brutnell, T.P. **A Low-Cost Library Construction Protocol And Data Analysis Pipeline For Illumina-Based Strand-Specific Multiplex Rna-Seq.** Plos One, 6, p. 1-12, 2011.

Wang, L., Mao, Y., Kong, F., Li, G., Ma, F., Zhang, B., Sun, P., Bi, G., Zhang, F., Xue, H., Cao, M. **Complete sequence and analysis of plastid genomes of two economically important red algae:** *Pyropia haitanensis* **and** *Pyropia yezoensis*. PLOS ONE, 8, p. 1-10, 2013.

Westreich, S.T., Korf, I., Mills, D.A., Lemay, D.G. **SAMSA: a comprehensive metatranscriptome analysis pipeline.** BMC Bioinformatics, 17, p. 1-12, 2016.

Wicke, S., Schenneweiss, G., Pamphilis, C, Müller, K., Quandt, D. **The evolution of the plastids chromosome in land plants: gene content, gene order, gene function.** Plant Molecular Biology, 76, p. 237-297, 2007.

Wiencke, C., Clayton, M.N. **Antarctic seaweeds.** In: Synopses of the Antarctic benthos, 9, A.R.G. Gantner Verlag KG, Ruggell, Lichtenstein, p. 239, 2002.

Wolff, G., Plante, I., Lang, B.F., Kück, U., Burger, G. **Complete sequence of the mitochondrial DNA of the chlorophyte alga** *Prototheca Wickerhamii***.** Journal of Molecular Biology, 237, p. 75-86, 1994.

Xiong, F.S., Mueller, E.C., Day, T.A. **Photosynthetic and respiratory acclimation and growth response of Antarctic vascular plants to contrasting temperatures regimes.** American Journal of Botany, 87, p. 700-710, 2000.

Yan, D., Wang, Y., Murakami, T., Shen, Y., Gong, J., Gao, C., Jiang, H., Smith, D.R., Pombert, J.F., Wu, Q., Dai, J. **The plastid genomes of microalgae** *Chlorella protothecoides* **and** *Prototheca wickerhamii* **shed light on the evolutionary transition from autotroph to heterotrophy.** Unpublished. http://www.ncbi.nlm.nih.gov/nuccore/KJ001761.1, 2015

Yu, M., Yang, S., Lin, X. ***De-novo*** **assembly and characterization of** *Chlorella minutissima* **UTEX2341 transcriptome by paired-end sequencing and the identification of genes related to the biosynthesis of lipids for biodiesel**. Marine Genomics, 25, p. 69-74, 2016.

Zemolin, A.P.P., Cruz, L.C., Paula, M.T., Pereira, B.K., Albuquerque, M.P., Victoria, F.C., Pereira, A.B., Posser, T., Franco, J.L. **Toxicity Induced by** *Prasiola crispa* **to Fruit Fly**

Drosophila melanogaster **and Cockroach** *Nauphoeta cinerea***: Evidence for Bioinsecticide Action.** Journal of Toxicology and Environmental Health**,** 77, p. 115-124, 2014.

Zhang, Z., Qu, C., Yao, R., Nie, Y., Xu, C., Miao, J., Zhong, B. **The Parallel Molecular Adaptations to the Antarctic Cold Environment in Two Psychrophilic Green Algae.** *Genome Biology and Evolution,* 11, p. 1897-1908, 2019.

Zúñiga, G.E., Alberdi, M., Corcuera, L.J. **Non structural carbohydrates in** *Deschampsia antarctica* Desv. **from South Shetland Islands, Maritime Antarctic.** Environmental and Experimental Botany, 36, p. 396-399, 1996.

ANEXOS

8. ANEXO I

O Anexo I é o material suplementar do Artigo 1 intitulado *"Phylogenetic positioning of the Antarctic alga Prasiola crispa (Trebouxiophyceae) using organellar genomes and their structural analysis".*

A *Figure S1* apresenta a análise filogenética de *Prasiola crispa* com o gene marcador *rbcL*. O clado *Prasiola* aparece em destaque.

A *Figure S2*, apresenta a análise filogenética de *Prasiola crispa* com o gene marcados *psaB*. O clado *Prasiola* está em destaque.

A *Table S1* e *Table S2,* apresentam os números de acessos GenBank das sequências utilizadas nas análises filogenômicas do cloroplasto e da mitocôndria, respectivamente.

Figure S1. Phylogenetic analyses of *Prasiola crispa* within gene marker *rbcL*. Phylogenetic analysis of gene marker *rbcL* of 136 species. The tree presented was reconstructed using a Bayesian approach with the LG + I + G amino acid substitution model. Branch supports indicate posterior probability values (over each branch). The *Prasiola* clade is highlighted.

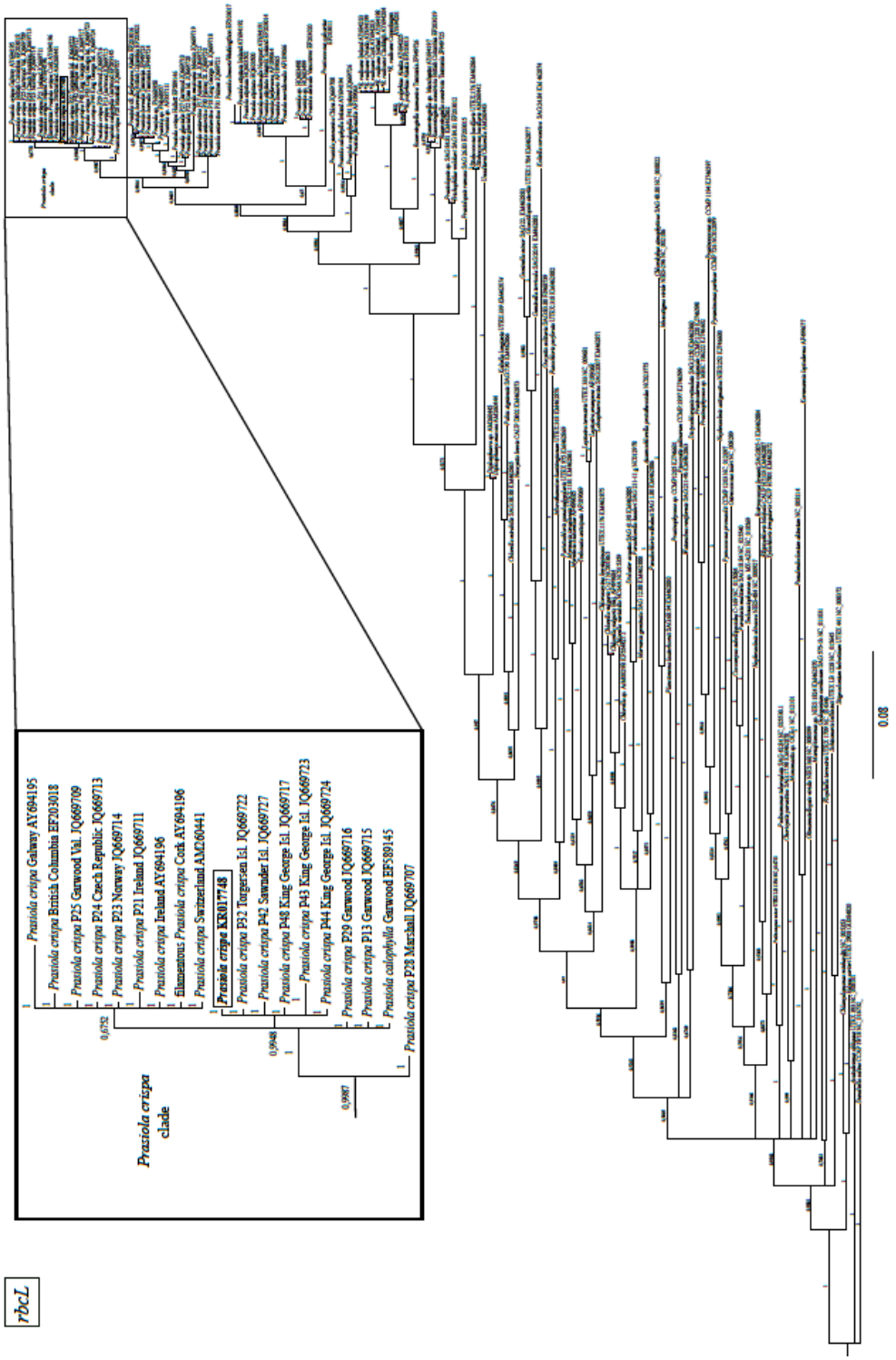


Figure S2. Phylogenetic analyses of *Prasiola crispa* within gene marker *psaB*. Phylogenetic analysis of gene marker *psaB* of 96 species. The tree presented was reconstructed using a Bayesian approach with the LG + I + G amino acid substitution model. Branch supports indicate posterior probability values (over each branch). The *Prasiola* clade is highlighted.

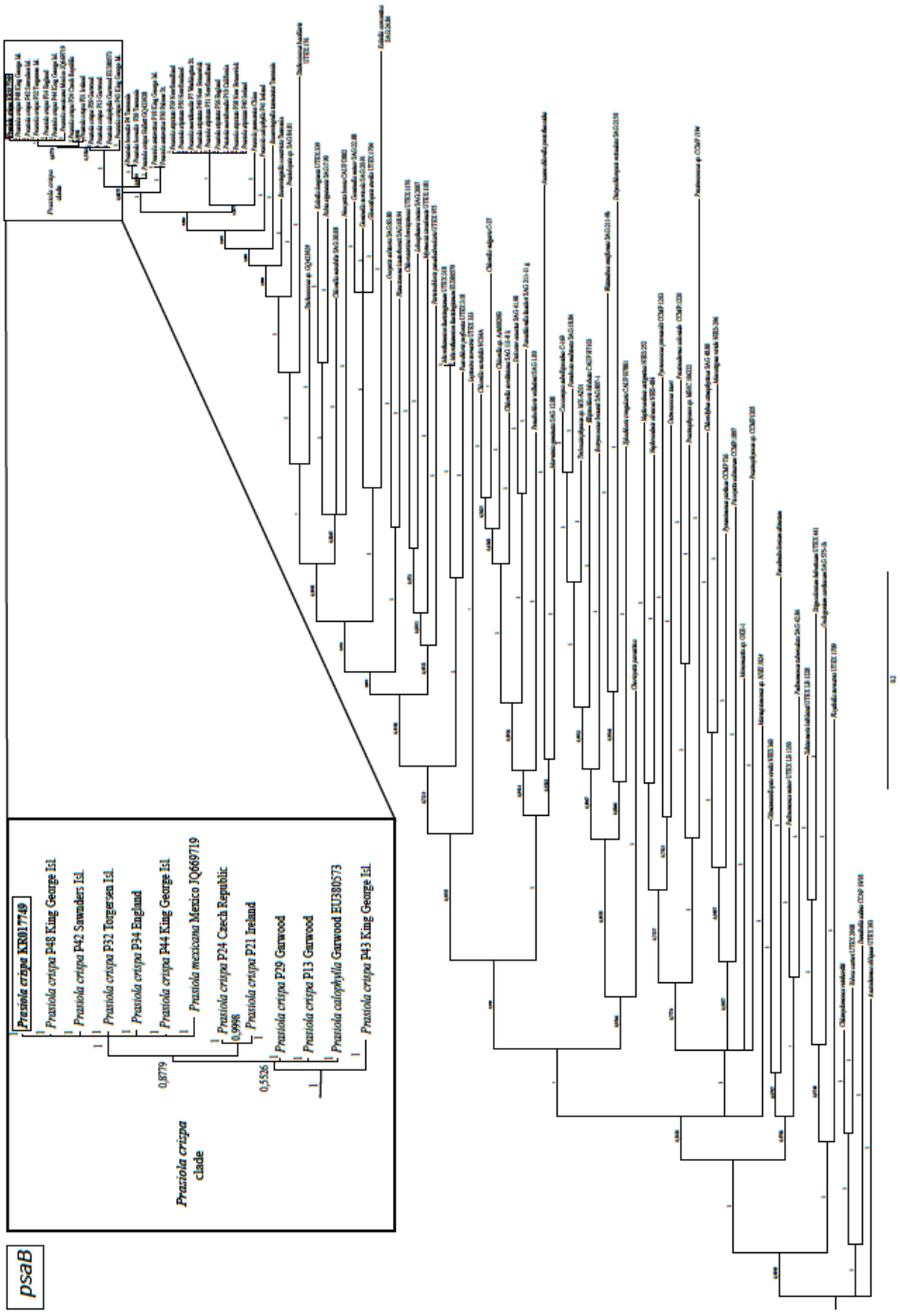


Table S1. GenBank accession numbers for sequences used in the chloroplast phylogenomic analyses.

Species	GenBank access number
Acutodesmus obliquus	NC_008101
Auxenochlorella protothecoides	NC_023775
Botryococcus braunii	KM462884
Chlamydomonas moewusii	EF587443 - EF587503
Chlamydomonas reinhardtii	NC_005353
'Chlorella' mirabilis	KM462865
Chlorella sorokiniana	NC_023835
Chlorella sp. ArM0029B	KF554427.1
Chlorella variabilis	NC_015359
Chlorella vulgaris	NC_001865
Chlorokybus atmophyticus	NC_008822
Chlorosarcina brevispinosa	KM462875
Choricystis parasitica	KM462878
Coccomyxa subellipsoidea	NC_015084
Dicloster acuatus	KM462885
Dictyochloropsis reticulata	KM462860
Dunaliella salina	NC_016732
Elliptochloris bilobata	KM462887
Floydiella terrestris	NC_014346
Fusochloris perforata	KM462882
Helicosporidium sp.	NC_008100
Geminella minor	KM462883
Geminella terricola	KM462881
Gloeotilopsis sterilis	KM462877
Koliella longiseta	KM462868
Koliella corcontica	KM462874
Leptosira terrestris	NC_009681
Lobosphaera incisa	KM462871

Marsupiomonas sp. NIES 1824	KM462870
Marvania geminata	KM462888
Mesostigma viride	NC_002186
Micromonas sp. RCC 299	NC_012575
Microthamnion kuetzingianum	KM462876
Monomastix sp. OKE-1	NC_012101
Myrmecia israelensis	KM462861
Neocystis brevis	KM462873
Nephroselmis astigmática	KJ746600
Nephroselmis olivacea	NC_000927
Oedogonium cardiacum	NC_011031
Oltmannsiellopsis viridis	NC_008099
Oocystis solitaria	FJ968739
Ostreococcus tauri	NC_008289
Pabia signiensis	KM462866
Parachlorella kessleri	NC_012978
Paradoxia multiseta	NC_025540
Parietochloris pseudoalveolaris	KM462869
Pedinomonas minor	NC_016733
Pedinomonas tuberculata	NC_025530.1
Picocystis salinarum	KJ746599
Planctonema lauterbornii	KM462880
Prasinococcus sp. CCMP 1194	KJ746597
Prasinoderma coloniale	KJ746598
Prasinophyceae sp. CCMP1205	KJ746601
Prasinophyceae sp. MBIC 106222	KJ746602
Prasiola crispa	KR017748, KR017749, KR017750
Prasiolopsis sp. SAG 84.81	KM462862
Prototheca wickerhamii	KJ001761.1
Pseudendoclonium akinetum	NC_008114
Pseudochloris wilhelmii	KM462886

Pycnococcus provasolii	NC_012097
Pyramimonas parkeae	NC_012099
Schizomeris leibleinii	NC_015645
Stichococcus bacillaris	KM462864
Stigeoclonium helveticum	NC_008372
Trebouxia aggregata	EU123962, EU124002
Trebouxiophyceae sp. MX-AZ01	NC_018569
Volvox carteri f. nagariensis	GU084820
Watanabea reniformis	KM462863
Xylochloris irregularis	KM462872

Table S2. GenBank accession numbers for sequences used in the mitochondrial phylogenomic analyses.

Species	GenBank access number
Andalucia godoyi	NC_021124.1
Arabidopsis thaliana	NC_001284
Auxenochlorella protothecoides	NC_026009
Chaetosphaeridium globosum	NC_004118
Chara vulgaris	NC_005255
Chlorella sorokiniana	KM241869.1
Chlorella sp. ArM0029B	KF554428.1
Chlorella variabilis	NC_025413
Chlorokybus atmophyticus	EF463011
Chondrus crispus	NC_001677
Coccomyxa subellipsoidea	NC_015316
Colpomenia peregrina	NC_025302.1
Cyanidioschyzon merolae	NC_000887
Helicosporidium sp.	NC_017841
Marchantia polymorpha	NC_001660
Micromonas sp.RCC 299	FJ859351.1
Neochloris aquatica	NC_024761
Nephroselmis olivacea	NC_008239
Oltmannsiellopsis viridis	NC_008256
Physcomitrella patens	NC_007945
Phytophthora andina	NC_015619.1
Phytophthora ipomoeae	NC_015622.1
Porphyra purpurea	NC_002007
Prasiola crispa	KR017746, KR017747
Prototheca wickerhamii	NC_001613
Reclinomonas americana	NC_001823.1
Trebouxiophyceae sp. MX-AZ01	NC_018568

9. ANEXO II

O Anexo II é o material suplementar do Artigo 2 intitulado *"De novo assembly and annotation of the Antartic alga Prasiola crispa transcriptome"*.

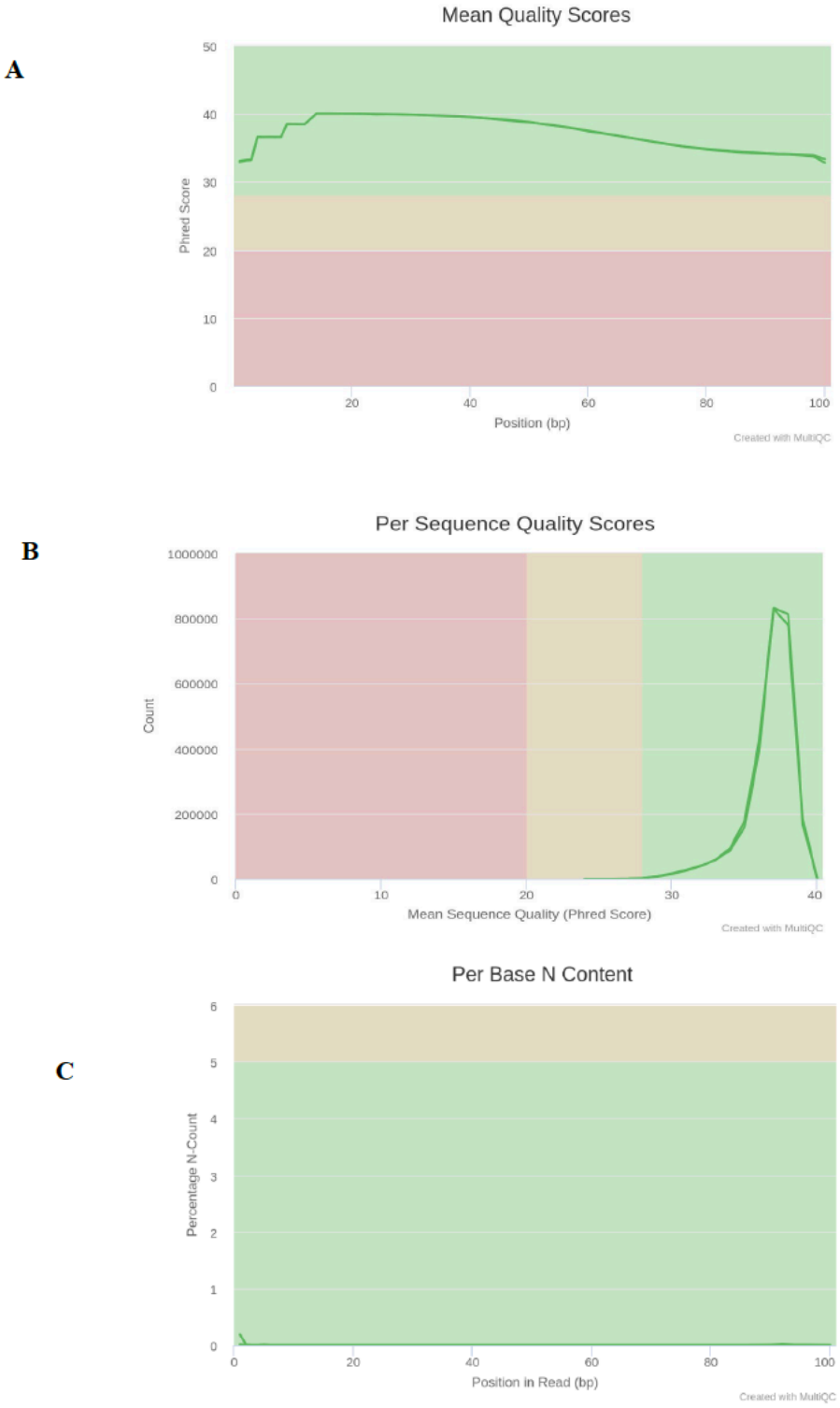
A *Table S1* apresenta a comparação do transcriptoma de *Prasiola crispa* e outras algas da classe Trebouxiophyceae e a *Figure S1* contém os gráficos sobre a qualidade do sequenciamento.

Table S1. Comparison between *Prasiola crispa* and organisms from the Trebouxiophyceae class with transcriptome sequenced.

Attributes	*Prasiola crispa*	*Chlorella minutissima*	*Trebouxia gelatinosa*	*Coccomyxa subellipsoidea*	*Chlorella sorokiniana*	*Botryococcus braunii*
Total raw reads	42,978,976	69,011,712	243,763,578	-	244,291,069	-
Total processed reads	5,233,428	67,559,338	237,404,631	46,000,000	229,228,757	-
Number of contigs	17,201	14,905	19,601	9,409	63,811	61,220
Mean length	763.1	2,998.04	1,605	-	1,022	-
Sequences with at least one blast hit (%)	52.19	53.50	53.60	-	36.80	-
Largest contig (pb)	12,802	-	31,749	-	15,932	-
N50	1,036	-	3,594	-	2,502	-
Assembler	Trinity	Trinity	Trinity	Reference genome	Trinity	Trinity

(-) Data not reported.

Figure S1. Graphs about sequencing read quality generated with FastQC. (A) Per base quality phred. (B) Per sequence quality. (C) Per base N content. The colored areas separate the metrics into High (green), Medium (yellow) and Low (red) quality.



10. ANEXO III

O artigo intitulado *"HTT-DB: new features and updates"* de autoria de Dotto, B.R., **Carvalho, E.L.**, Da Silva, A.F., Dezordi, F.Z., Pinto, P.M., Campos, T.L., Rezende, A.M., Wallau, G.L. foi publicado no periódico *Database* (ISSN: 1758-0463), no ano de 2018.

Neste artigo, trazemos uma atualização do HTT-DB: *Horizontally transferred transposable elements database*, um banco de dados de elementos transponíves transferidos horizontalmente criado pelo nosso grupo em 2015 (Dotto et al., 2015). Novos recursos e atualizações como por exemplo, a transferência horizontal de vírus (HVT) foram adicionados.

Minha participação neste trabalho consistiu na obtenção de dados de HVTs através de revisões bibliográficas.

DATABASE
The Journal of Biological Databases and Curation

Database, 2018, 1–5
doi: 10.1093/database/bax102
Database update

Database update

HTT-DB: new features and updates

Bruno Reis Dotto[1], Evelise Leis Carvalho[2], Alexandre Freitas da Silva[3], Filipe Zimmer Dezordi[1,3], Paulo Marcos Pinto[1,2], Tulio de Lima Campos[4], Antonio Mauro Rezende[5] and Gabriel da Luz Wallau[6,*]

[1]Campus São Gabriel, Universidade Federal do Pampa, Avenida Antonio Trilha, 1847, São Gabriel, Rio Grande do Sul, 97300-000, [2]Pos Graduation in Biological Sciences, Universidade Federal do Pampa, Avenida Antonio Trilha, 1847, São Gabriel, Rio Grande do Sul, 97300-000, [3]Pos Graduation in Biosciences and Health Biotechnology, Aggeu Magalhães Institute (IAM), Avenida Professor Moraes Rego, s/n, Recife, Pernambuco, 50740-465, [4]Bioinformatic Core, Aggeu Magalhães Institute (IAM), Avenida Professor Moraes Rego, s/n, Recife, Pernambuco, 50740-465, [5]Department of Microbiology, Aggeu Magalhães Institute (IAM), Avenida Professor Moraes Rego, s/n, Recife, Pernambuco, 50740-465 and [6]Department of Entomology, Aggeu Magalhães Institute (IAM), Avenida Professor Moraes Rego, s/n, Recife, Pernambuco, 50740-465

*Corresponding author: Tel/Fax: +55 55 21012549; Email: gabriel.wallau@cpqam.fiocruz.br

Citation details: Dotto,B.R., Carvalho,E.L., da Silva,A.F. *et al.* HTT-DB: new features and updates. *Database* (2017) Vol. 2017: article ID bax102; doi:10.1093/database/bax102

Received 2 August 2017; Revised 2 November 2017; Accepted 12 December 2017

Abstract

Horizontal Transfer (HT) of genetic material between species is a common phenomenon among Bacteria and Archaea species and several databases are available for information retrieval and data mining. However, little attention has been given to this phenomenon among eukaryotic species mainly due to the lower proportion of these events. In the last years, a vertiginous amount of new HT events involving eukaryotic species was reported in the literature, highlighting the need of a common repository to keep the scientific community up to date and describe overall trends. Recently, we published the first HT database focused on HT of transposable elements among eukaryotes: the Horizontal Transposon Transfer DataBase: Database URL: (http://lpa.saogabriel.unipampa.edu.br:8080/httdatabase/). Here, we present new features and updates of this unique database: (i) its expansion to include virus-host exchange of genetic material, which we called Horizontal Virus Transfer (HVT) and (ii) the availability of a web server for HT detection, where we implemented the online version of vertical and horizontal inheritance consistence analysis (VHICA), an R package developed for HT detection. These improvements will help researchers to navigate through known HVT cases, take data-informed decision and export figures based on keywords searches. Moreover, the availability of the VHICA

as an online tool will make this software easily reachable even for researchers with no or little computation knowledge as well as foster our capability to detect new HT events in a wide variety of taxa.

Database URL: http://lpa.saogabriel.unipampa.edu.br: 8080/httdatabase/

Introduction

Genetic inheritance is the main mode of genetic material transmission from ancestral to descendent individuals or species. Such process is widely known as vertical transfer of genetic material. However, there is another phenomenon which allows the transfer of genetic material between biological entities known as Horizontal Transfer (HT). Such HT events are very frequent in prokaryotic species, having a great impact on the exchange of different classes of genes such as anti-biotic resistance genes. However, in recent years, investigation of HT events frequency and their impact in multi-cellular eukaryotic genomes have been underscored (1).

Historically HTs are named based on the genetic entity which is transferred between species as Horizontal Gene Transfer (HGT) and Horizontal Transposon Transfer (HTT). HGT is a rare phenomenon among eukaryotes and few examples are known, taking place in particular conditions such as host-parasite relationships (2). HTT is a much more frequent phenomenon among eukaryotes in which transposable elements (TEs), DNA fragments that are capable of moving themselves between different genomic positions, transfer from one host species to another by means other than sexual reproduction (3).

Most of the knowledge about HT events is dispersed on original research articles, but some effort has been made to compile such information, at least for HGT events among Bacteria and Archaea (4, 5). However, no database was available for HT events among eukaryotic species until 2015. Based on this gap, we developed the Horizontally Transferred Transposable Element Database (HTT-DB) which allows researchers to have access to all known HTT cases among eukaryotes and perform searches on these data using TEs [Repbase—(6) and host species (https://www.ncbi.nlm.nih.gov/taxonomy)] keywords (7). To date, our database was cited in 7 other publications and had >1000 page views since its publication. Moreover, 2276 new HTT cases were added to the database since its publication covering a wide variety of taxa (Cumulative number of HTT in the last 10 years panel–http://lpa.saogabriel.unipampa.edu.br:8080/httdatabase/).

Here, we present HTT-DB's new features and updates which make it richer in details/information regarding HT of genetic material between viruses and host (Host-Virus Transfer–HVT) as well as the availability of a web server implementing a recently published method for HT detection (8).

Database interface for horizontal virus transfer

HT of genetic material between viruses and their hosts is another phenomenon not fully appreciated so far. Nevertheless, in the last years, accumulating evidence has shown that this exchange is more common than previously recognized (9–11). Such events are also known as endogenization, and the viral remains, found integrated in the host genomes, are commonly called endogenous viral elements (EVEs) (12–14). Due to the absence of a database for such events which can have huge impact on understanding of genome biology of host species (15, 16), we decided to expand the HTT database to include such endogenization events in the same fashion as HTT events. Here, we call such events Horizontal Virus Transfer (HVT) in order to emphasize the genetic entities that are being transferred: the viral sequences. The main difference of HVT compared to HTT is the database searching that can be performed with viral classification scheme established by the International Committee on Taxonomy of Virus (ICTV–https://talk.ictvonline.org/) instead of the TEs Repbase classification. Moreover, database navigation and image generation depending on the user selected keywords follow the same rationale already implemented in HTT-DB. Such new database layer will be updated annually with new published information obtained from manuscripts recovered from PubMed-NCBI and Google Scholar searchers with the following terms 'endogenous virus,' 'EVEs,' 'endogenization' and 'viral derived elements.' Furthermore, any user can contribute sending new data to the database through the 'Add new data–HVT database (Virus)' menu directly through the system or downloading an example table. The user also can submit nucleotide sequences of TEs and endogenous virus which will be available along with its corresponding metadata to download.

Based on our curation of HVT events reported in the literature and added to the HTT-DB we found a total of 1563 HVT events with the majority of cases reported in Metazoan genomes (1279 endogenization events), and a high proportion of cases found in the Phylum Arthropoda (721) and

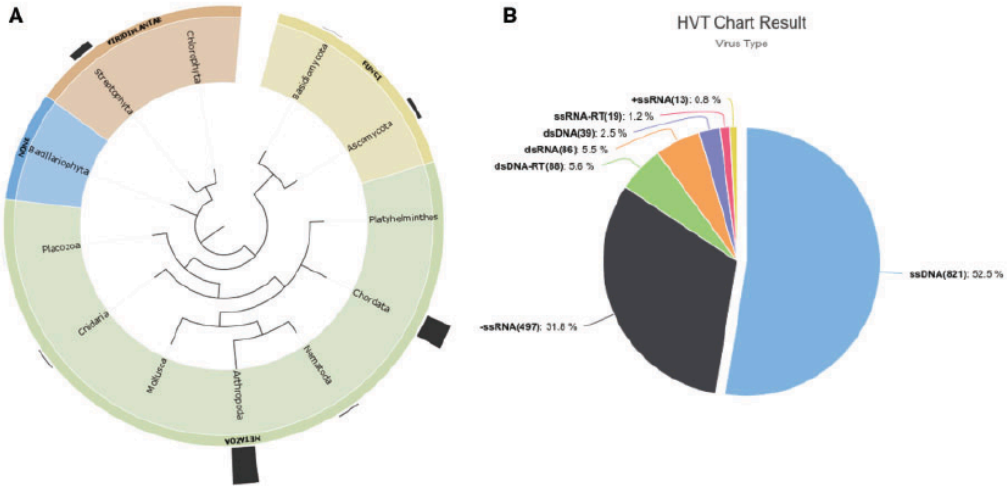


Figure 1. **(A)** Distribution of HVT events (endogenization) across different Phylum, outermost bars represent the number of HVT cases in the taxa shown. **(B)** HVT chart result from database search including all cases reported in the database, ssDNAs >50% of all known HVT events.

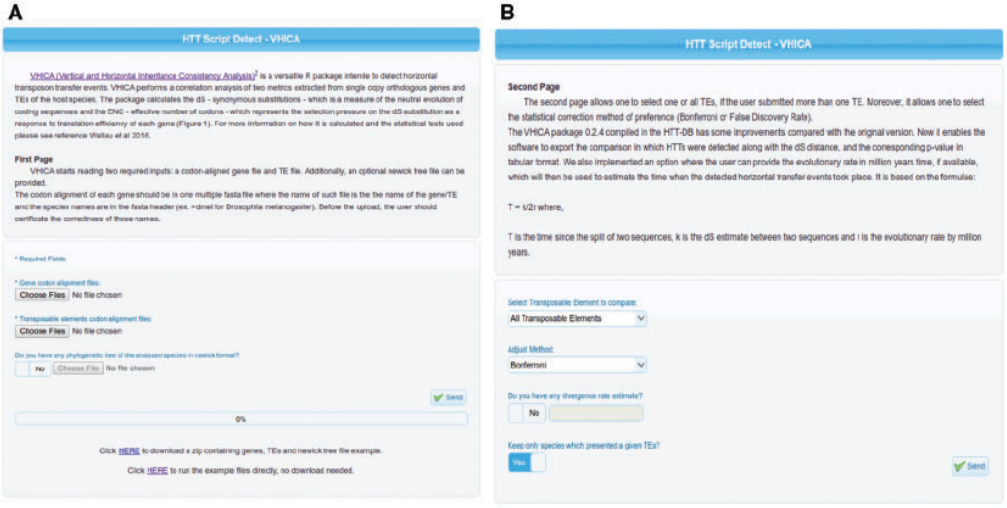


Figure 2. **(A)** First panel of VHICA interface available for HTT detection through the HTT-DB database. Two optional (Gene Files and TE Files) and one optional box (Phylotree File) for files upload are available as well as an example of all those files used in the original VHICA paper (Wallau *et al.* 2016). **(B)** Second panel of VHICA interface. Here, the user can select one, a set of or all TEs for run VHICA. Since VHICA performs multiple comparisons there are two correction methods available for the user choice: Bonferroni and False Discovery Rate. TE divergence rate can be added, and then VHICA will estimate the HT time in Mya. User can also choose if they want to plot the results only for hosts' species which presented a given TE or maintain all the other species in the plotting step.

Chordata (516) (Figure 1A). Exploring the viral classification scheme from the ICTVs, we can observe that the largest amount of endogenization events is accounted by the single strand DNA viruses (ssDNA—821) followed by negative single strand RNA viruses (-ssRNA—497) (Figure 1B). Deeper exploration at lower taxonomic levels can be performed by the users both at the host and viral taxa following the drop down menus of the 'Database search button.'

Interface for vertical and horizontal inheritance consistence analysis

Vertical and Horizontal Inheritance Consistence Analysis (VHICA) is an R package developed to detect HT events (8). This package implements several command line R functions capable of reading codon aligned host and TEs sequences, extracting codon bias and synonymous substitutions and performing statistical analysis on these extracted parameters in

A

Dromar6

	seq	dS	sp1	sp2	Time(Mya)	pvalor
9955	Dromar6.a.b	0.158112700	dere.a	drho.b	4.941022000	0.006163747
9959	Dromar6.b.a	0.136416900	drho.b	dyak.a	4.263028000	0.006675676
9941	Dromar6.a	0.067462300	dana.a	dere.a	2.108197000	0.000117252
9944	Dromar6.a	0.057300010	dana.a	dyak.a	1.790625000	0.000057234
9940	Dromar6.a	0.246577300	dana.a	dele.a	7.705539000	0.002872525
9943	Dromar6.a.b	0.159505300	dana.a	drho.b	4.984541000	0.001636277
9942	Dromar6.a	0.195430400	dana.a	dfic.a	6.107199000	0.000548159
9935	Dromar6.b.a	0.217867700	dana.b	dere.a	6.808366000	0.000432800
9938	Dromar6.b.a	0.134232400	dana.b	dyak.a	4.194762000	0.000070364
9937	Dromar6.b	0.183232000	dana.b	drho.b	5.725999000	0.000982421
9936	Dromar6.b.a	0.263341800	dana.b	dfic.a	8.229431000	0.000521882
9946	Dromar6.a	0.083717890	dbip.a	dere.a	2.616184000	0.000178717
9949	Dromar6.a	0.044465800	dbip.a	dyak.a	1.389556000	0.000444730
9945	Dromar6.a	0.254473600	dbip.a	dele.a	7.952301000	0.002543994
9948	Dromar6.a.b	0.186048400	dbip.a	drho.b	5.814011000	0.001587570
9947	Dromar6.a	0.140651800	dbip.a	dfic.a	4.395367000	0.000338208

Statistic	N	Mean	St. Dev.	Min	Max
dS	16	0.158052300	0.069636130	0.044465800	0.263341800
Time(Mya)	16	4.939133000	2.176129000	1.389556000	8.229431000
pvalor	16	0.001573222	0.002081161	0.000057234	0.006675676

B

==========

No HTT detected

Figure 3. Three possible VHICA outputs. **(A)** Summary of all statistically supported HTT cases and associated data. **(B)** Resulting output when no significant HTT signal is detected. **(C)** Graphical output showing all host species and the significant pairwise comparison following *P*-value the legend colour.

order to evaluate the presence of HT signal. Although bioinformatics curriculum is being implemented in a growing number of under-graduate courses in biological sciences, several biologists and scientists from associated areas still have difficulties facing command line software (17–19). In order to overcome such issue and make VHICA available to a wide variety of researchers, we implemented it as a web server in the HTT-DB. As can be seen in first panel of VHICA, there are three fields where the user can upload the input data (Figure 2A). Besides, in the same fields, one can have access to example files. Example files are the full dataset used by Wallau *et al*. 2016 which the user can apply as input in order to see the VHICA output or check the alignments and gene/TE names formats needed to run VHICA properly. In addition, we also added detailed information regarding file formats and VHICA analysis in the 'About' Section 'VHICA Online–Vertical and Horizontal Inheritance Consistency Analysis (VHICA package)' sub-section of HTT-DB. In the second panel, the user can choose a specific TE or all TEs to be used and tested in VHICA, the multiple-testing correction method, and an optional field: the divergence rate. The last option is a new VHICA feature which allows an estimation of the time when the HT event took place using the equation $T = k/2r$ (20) (Figure 2B). T represents the divergence time between TEs, k is the synonymous divergence (dS) between TEs and r is the TE specific evolutionary rate.

VHICA can output two types of results: (i) if all TEs were selected for analysis, it will report a summary of all statistically significant pairwise comparison as well as the dS, the pair of species involved, the estimated HT time in Mya and the associated *P*-value or a text file with 'No HTT detected message' (Figure 3A and B). (ii) A plot in .pdf format containing the *P*-value matrix for all TEs evaluated (Figure 3C). All these results become available as a compressed .zip file that will remain stored in the server and available for download for 10 days.

Conclusion

In summary, we present here the expansion of the only available database about HT among eukaryotic species. Currently, HTT-DB expansion includes HVT (endogenization of viral sequences into the host genomes) and we made available a R package (VHICA) developed to detect HT events through the HTT-DB interface.

Conflict of interest. None declared.

References

1. Syvanen,M. (2012) Evolutionary implications of horizontal gene transfer. *Annu. Rev. Genet.*, **46**, 341–358.
2. Keeling,P.J. and Palmer,J.D. (2008) Horizontal gene transfer in eukaryotic evolution. *Nat. Rev. Genet.*, **9**, 605–618.
3. Wallau,G.L., Ortiz,M.F. and Loreto,E.L.S. (2012) Horizontal transposon transfer in eukarya: detection, bias, and perspectives. *Genome Biol. Evol.*, **4**, 801–811.
4. Jeong,H., Sung,S., Kwon,T. *et al.* (2016) HGTree: database of horizontally transferred genes determined by tree reconciliation. *Nucleic Acids Res.*, **44**, D610–D619.

5. Podell,S., Gaasterland,T. and Allen,E.E. (2008) A database of phylogenetically atypical genes in archaeal and bacterial genomes, identified using the DarkHorse algorithm. *BMC Bioinformatics*, **9**, 419.
6. Bao,W., Kojima,K.K. and Kohany,O. (2015) Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA*, **6**, 11.
7. Dotto,B.R., Carvalho,E.L., Silva,A.F. *et al.* (2015) HTT-DB: horizontally transferred transposable elements database. *Bioinformatics*, **31**, 2915–2917.
8. Wallau,G.L., Capy,P., Loreto,E. *et al.* (2016) VHICA, a new method to discriminate between vertical and horizontal transposon transfer: application to the mariner family within drosophila. *Mol. Biol. Evol.*, **33**, 1094–1109.
9. Katzourakis,A. and Gifford,R.J. (2010) Endogenous viral elements in animal genomes. *PLoS Genet.*, **6**, e1001191.
10. Thézé,J., Leclercq,S., Moumen,B. *et al.* (2014) Remarkable diversity of endogenous viruses in a crustacean genome. *Genome Biol. Evol.*, **6**, 2129–2140.
11. Li,C.-X., Shi,M., Tian,J.-H. *et al.* (2015) Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *eLife*, **4**.
12. Holmes,E.C. (2011) The evolution of endogenous viral elements. *Cell Host Microbe*, **10**, 368–377.
13. Feschotte,C. and Gilbert,C. (2012) Endogenous viruses: insights into viral evolution and impact on host biology. *Nat. Rev. Genet.*, **13**, 283–296.
14. Patel,M.R., Emerman,M. and Malik,H.S. (2011) Paleovirology—ghosts and gifts of viruses past. *Curr. Opin. Virol.*, **1**, 304–309.
15. Aswad,A. and Katzourakis,A. (2012) Paleovirology and virally derived immunity. *Trends Ecol. Evol.*, **27**, 627–636.
16. Malfavon-Borja,R. and Feschotte,C. (2015) Fighting fire with fire: endogenous retrovirus envelopes as restriction factors. *J. Virol.*, **89**, 4047–4050.
17. Welch,L., Lewitter,F., Schwartz,R. *et al.* (2014) Bioinformatics curriculum guidelines: toward a definition of core competencies. *PLoS Comput. Biol.*, **10**, e1003496.
18. Brazas,M.D., Lewitter,F., Schneider,M.V. *et al.* (2014) A quick guide to genomics and bioinformatics training for clinical and public audiences. *PLoS Comput. Biol.*, **10**, e1003510.
19. Loman,N. and Watson,M. (2013) So you want to be a computational biologist? *Nat. Biotechnol.*, **31**, 996–998.
20. Graur,D. and Li,W.-H. (1999) *Fundamentals of Molecular Evolution*, 2nd edn. Sinauer Associates, Inc, Sunderland, Massachusetts, USA, pp. 1–481.

11. ANEXO IV

O artigo intitulado *"Are the bacteria and their metabolites contributing for gut inflammation on GSD-Ia patients?"* de autoria de Colonetti, K., **Carvalho, E.L.**, Rangel, D.L., Pinto, P.M., Roesch, L.F.W., Pinheiro, F.C., Schwartz, I.V.D. foi publicado no periódico *Metabolites* (ISSN: 2218-1989), no ano de 2022.

A doença de armazenamento do glicogênio (GSD) é uma doença de origem genética, onde a GSD I é a forma mais frequente e grave, causando disbiose intestinal, baixo pH fecal e um desequilíbrio dos ácidos graxos de cadeia curta (SCFAs). SCFAs são produtos da fermentação microbiana no intestino grosso e exercem um papel biológico sobre o hospedeiro, sendo capazes de modular a comunidade bacteriana. A espécie bacteriana, assim como seus metabólitos e o sistema imunológico do hospedeiro, podem influenciar a homeostase celular, iniciando ou não um processo inflamatório.

Sendo assim, o objetivo desse estudo foi quantificar os SCFAs (ácido fórmico, ácido lático, ácido propiônico e ácido succínico de amostras de pacientes com GSD, por cromatografia líquida de alta eficiência (HPLC).

Os níveis de ácido succínico foram mais elevados nos pacientes com GSD, em comparação com o controle, consistente com os casos de disbiose. Os resultados indicam uma associação de entre a doença inflamatória intestinal e o aumento de SCFAs.

Minha participação neste trabalho consistiu na parte metodológica: preparação das amostras e análise por HPLC.

metabolites

MDPI

Brief Report

Are the Bacteria and Their Metabolites Contributing for Gut Inflammation on GSD-Ia Patients?

Karina Colonetti [1], Evelise Leis de Carvalho [2], Darlene Lopes Rangel [2], Paulo Marcos Pinto [2], Luiz Fernando Wurdig Roesch [3], Franciele Cabral Pinheiro [1,4,5] and Ida Vanessa Doederlein Schwartz [1,5,6,*]

[1] Laboratory of Basic Research and Advanced Investigations in Neurosciences (BRAIN), Porto Alegre 90035-903, RS, Brazil
[2] Laboratório de Proteômica Aplicada (LPA), Universidade Federal do Pampa, São Gabriel 97300-000, RS, Brazil
[3] Department of Microbiology and Cell Science, Institute of Food and Agricultural Sciences, University of Florida, Gainesville 32611, FL, USA
[4] Laboratório de Biologia, Universidade Federal do Pampa, Itaqui 97650-000, RS, Brazil
[5] Post-Graduate Program in Genetics and Molecular Biology, Universidade Federal do Rio Grande do Sul, Porto Alegre 91501-970, RS, Brazil
[6] Medical Genetics Service, Hospital de Clínicas de Porto Alegre, Porto Alegre 90035-903, RS, Brazil
* Correspondence: ischwartz@hcpa.edu.br

Abstract: Recently, patients with glycogen storage disease (GSD) have been described as having gut dysbiosis, lower fecal pH, and an imbalance in SCFAs due to an increase in acetate and propionate levels. Here, we report the fecal measurement of bacterial-related metabolites formic, acetic, lactic, propionic, and succinic acid, a key metabolite of both host and microbiota, on a previously described cohort of 24 patients (GSD Ia = 15, GSD Ib = 5, 1 GSD III = 1 and GSD IX = 3) and 16 healthy controls, with similar sex and age, using the high-performance liquid chromatography technique. The succinic acid levels were higher in the GSD patients than in the controls (patients = 38.02; controls = 27.53; $p = 0.045$), without differences between the groups for other metabolites. Fecal pH present inverse correlation with lactic acid ($R = -0.54$; $p = 0.0085$), while OTUs were inversely correlated with both lactic ($R = -0.46$; $p = 0.026$) and formic ($R = -0.54$; $p = 0.026$) acids. Using two distinct metrics of diversity, borderline significance was obtained for propionic acid, affecting the microbial structure on Euclidean basis in 8% ($r^2 = 0.081$; $p = 0.079$), and for lactic acid, affecting 6% of microbial structure using Bray–Curtis distance ($r^2 = 0.065$; $p = 0.060$). No correlation was found between SCFAs and total carbohydrate consumption among the participants or uncooked cornstarch consumption among the patients.

Keywords: glycogen storage disease; inflammation; short-chain fatty acids; gut microbiota; dysbiosis; fecal pH

Citation: Colonetti, K.; de Carvalho, E.L.; Rangel, D.L.; Pinto, P.M.; Roesch, L.F.W.; Pinheiro, F.C.; Schwartz, I.V.D. Are the Bacteria and Their Metabolites Contributing for Gut Inflammation on GSD-Ia Patients? *Metabolites* **2022**, *12*, 873. https://doi.org/10.3390/metabo12090873

Academic Editor: Tiemin Liu

Received: 4 August 2022
Accepted: 12 September 2022
Published: 16 September 2022

1. Introduction

Hepatic glycogen storage diseases (GSD) are genetic diseases treated mainly by dietetic management, with simple sugar restriction and an overload of uncooked cornstarch (UCCS) [1,2]. There are several types of GSD, but GSD type I is the most frequent and severe in terms of UCCS dosage, and patients frequently present as overweight. GSD type I is classified as GSD Ia (OMIM#232200) or GSD Ib (OMIM#232220), and both are the result of the impairment of the glucose-6-phosphatase complex (G6PC). While GSD Ia affects one catalytic subunit of the complex (G6PC1 gene), GSD Ib affects the transporter protein on the endoplasmic reticle membrane (SLC37A4 gene). Once the transporter is affected, the whole complex will be affected due to the restriction of substrate availability [3]. Clinically, these patients also will be different since GSD Ib patients have a classical immunological impairment and inflammatory bowel disease (IBD) [4,5]. Despite not being a classic feature, IBD has also been described in GSD Ia patients, and the alteration of the gut microbiome

because of the massive intake of UCCS may play a role in that [6,7]. Host intestinal health depends on the intestinal homeostasis between the innate/adaptive immune system and the microbiome. Numerous studies suggest that gut microbiota is constantly monitored by the host mucosal immune system, and any slight disturbance in the microbial communities may contribute to intestinal immune disruption and increased susceptibility to IBD, a chronic relapsing inflammatory condition of the gastrointestinal tract [8].

Fecal short-chain fatty acids (SCFA) and small molecules such as succinate are products of microbial fermentation in the large bowel from food components unabsorbed/undigested in the small intestine; they are characterized by containing fewer than six carbons (C), existing in straight, and branched-chain conformation. Acetic acid (C2), propionic acid (C3), and butyric acid (C4) are the most abundant, representing 90–95% of the SCFA present in the colon [9]. Those metabolites exert a biological role over the host and are also capable of interfering with bacterial community fitness. Their production is largely influenced by pH and the available substrates [10,11]. On the other hand, the types of SCFAs produced influence the luminal pH [12]. In addition to that, the luminal pH in the colon is modulated by mucosal bicarbonate, lactate production, the bacterial fermentation of carbohydrates, and the mucosal absorption of SCFA [13]. The association between these factors and lower pH, SCFA and gut inflammation has been reported by several studies [14,15].

The bacterial species, their metabolic products, and the host immune system may influence whether host cellular homeostasis is maintained or inflammatory mechanisms are triggered. Our previous study [16] demonstrated that GSD patients presented low fecal pH and high calprotectin levels, in addition to intestinal dysbiosis, with increased Proteobacteria, recently confirmed by Ceccarani and colleagues [17]. Therefore, the aim of this study was to analyze how these aspects influence dysbiosis in patients with GSD, which was previously described. Thus, we report the quantification of fecal formic, acetic, lactic, propionic, and succinic acid of a previous well-characterized cohort of patients described in our aforementioned study and their association with fecal pH, microbial diversity, diet, and calprotectin levels found in the patients and controls.

2. Materials and Methods

This was an observational, cross-sectional, controlled study with convenience sampling. Twenty-four GSD patients (GSD Ia = 15, GSD Ib = 5, 1 GSD III = 1, and GSD IX = 3) receiving UCCS treatment were recruited from the outpatient clinics of the Medical Genetics Service at Hospital de Clínicas de Porto Alegre (MGS-HCPA), Brazil from January 2016 to May 2017, as previously described [16]. Briefly, as inclusion criteria, the subjects (patients and controls) should be ≥3 years old, have no signs of infection, have not been vaccinated for at least 15 days prior to sample collection, not be taking antibiotics, nor should they have been designated to receive/received an organ transplant. The healthy controls (n = 16) were recruited by invitation as they came to routine appointments at Santa Cecília Basic Health Unit, Porto Alegre, Brazil. The minimum–maximum age of the included patients and controls were 10–19.75 and 10–23.25 years, respectively. All of the subjects received a kit and printed instructions for their own stool collection, as well as storage and transport. Upon returning to the clinic, the fecal samples were frozen at −80 °C until use.

2.1. Sample Preparation and Fecal Short Chain Fatty Acids Measurement

To determine the SCFAs in the feces, 150 mg of each of the frozen fecal samples from the patients and controls was aliquoted at room temperature (20 °C) and homogenized in 1.5 mL H_2SO_4 0.15 mM for 2 min. Then, the samples were centrifuged for 5 min at 10,000× g rpm/4 °C. The supernatant was removed, filtered on a 0.22 µm pore nitrocellulose membrane, and frozen at −20 °C until analysis on a high-performance liquid chromatography (HPLC) system.

Commercial standards of acetic (AA), formic (FA), lactic (LA), succinic (SA), and propionic (PA) acid were used to obtain the calibration curve and thus quantify the SCFAs

using the HPLC system (Shimadzu Prominence®UFLC, LC-6AD pump, SPD-20AV UV detector; Shimadzu, Kyoto, Japan).

Separation was achieved on a Nucleosil 100-5 C18 EC (d.i. 250 mm × 4.6 mm, 100 Å e 5.0 µm; Macherey-Nagel, Dueren, Germany) HPLC column attached to a pre-column Security Guard Cartridge System (Phenomenex, Torrance, USA) using monosodium phosphate solution (20 mM; pH 2.2) (A) and acetonitrile (B) and an isocratic mode 95:5 with a flow rate of 1.25 mL/min. The injection volume was 20 µL. The chromatograms were monitored at 210 nm. The LC Solution software (Shimadzu) was used to obtain the retention time and the chromatograms and to perform the quantification of the peak area. Protocol standardization was performed using triplicate samples.

The quantification of SCFAs in the fecal sample was performed according to De Baere et al. (2013) [18], with some modifications. The samples were compared with a mix of standard SCFAs and succinic acid, composed of FA, LA, AA, SA, and PA, with degrees of purity ranging from 85 to 99.9%. The working solution (5 M) was prepared, and serial dilutions (200–0.048 mM) were performed to obtain the retention time of the fatty acids (AA = 1.739 min; FA = 2.718 min; LA = 3.247 min; PA = 4.302 min; SA = 4,462 min) and the calibration curves [AA y = 26,960x − 31,217 (r = 0.9846); FA y = 23,042x + 25,601 (r = 0.9982); LA y = 27,648x − 60,305 (r = 0.9714); PA y = 35,066x + 86,765 (r = 0.9911); SA y = 12,766x − 289,128 (r = 0.9631)].

The limit of detection (LoD) and limit of quantification (LoQ) for each metabolite were performed: FA LoD = 0.024 mM, LoQ = 0.048 mM; LA LoD = 0.048 mM, LoQ = 0.195 mM; AA LoD = 0.195 mM, LoQ = 0.390 mM; SA LoD = 0.098 mM, LoQ = 0.390 mM; PA LoD = 0.098 mM; LoQ = 0.195 mM.

2.2. Statistical Analysis

Statistical analysis among the groups was performed using PASW Statistics for Windows software (Vs18.0, 2009, SPSS Inc., Chicago, IL, USA) and R Studio software Version 1.3.959, using the phyloseq [19] and vegan [20] packages after dataset rarefaction to the minimum library size [20,21]. The numerical variables were compared using the Mann–Whitney U test. The categorical variables were compared using X^2, Fisher's exact test, or Continuity Correction, when necessary (statistically significant results were determined by the threshold $p \leq 0.05$). The graphs were constructed in R Studio software. To analyze the relationship between calprotectin and SCFAs, the valid reads of these last ones were used as the limit of the points on the graph. The valid reads of SCFAs were plotted on the Y-axis.

3. Results

3.1. Cohort Description, Metabolite Quantification and Differences on Microbial Community between Patients and Controls

The clinical characteristics, food intake, and gut microbial profile of the 24 patients and 16 controls that comprised this study were previously described in detail, as well as the features of the gut microbiota [16].

Using the set of data with patients and controls, formic (FA), acetic (AA), lactic (LA), propionic (PA), and succinic (SA) acids were compared between the groups (Table 1). Among all of the metabolites, only SA was present on different levels, and was higher in patients (patients = 38.02, controls = 27.53; $p = 0.045$). However, taking the characteristics of the microbial community into account, none of the metabolites isolated were directly related to alterations to the microbial structure at the significance level threshold ($p = 0.05$), with LA presenting borderline significance ($p = 0.06$) and the effect size over the microbial structure of 6.5%. Considering the fact that SA was different between the groups, the borderline significance of some metabolites, and the correlation among AA and SA (r = 0.52; $p = 0.0055$); AA and PA (r = 0.64; $p = 0.00095$), FA and LA (r = 0.85; $p = 0.0016$); and PA and SA (r = 0.47; $p = 0.023$) (Figure 1), we also inquire about the combined effect of these metabolites on the microbial structure found in the participants. SA seems to be an

important adjuvant for both LA and PA effects on the microbial structure, increasing the effect size to 33% and 11%, respectively, both with statistical significance ($p < 0.05$) (Table 1).

Table 1. Metabolite quantification between patient and control group.

Metabolite [1]	N [2] (Patient/Control)	Patient (mM) [3]	Control (mM) [3]	*p*-Value [4]	Microbial Community Difference between Groups (r^2; *p*-Value)	
					Euclidean	Bray-Curtis
Formic acid	09/08	71.74 (42.28–106.88)	41.82 (20.08–86.38)	0.149	0.057; 0.443	0.058; 0.523
Acetic acid	18/15	91.28 (44.96–143.20)	85.27 (48.75–112.19)	0.942	0.0276; 0.521	0.017; 0.989
Lactic acid	13/10	28.86 (13.91–56.14)	17.87 (13.08–25.78)	0.321	0.0562; 0.256	0.065; 0.060
Propionic acid	11/14	32.68 (11.37–10.51)	42.06 (22.06–52.98)	>0.999	0.081; 0.079	0.027; 0.905
Succinic acid	18/16	38.02 (24.60–60.33)	27.53 (9.50–39.99)	**0.045**	0.045; 0.131	0.034; 0.234
Formic + Succinic	7/8	61.28 (30.52–100.04)	37.89 (15.28–68.06)	0.224	0.054; 0.557	0.062; 0.627
Acetic + Succinic	13/15	101.59 (42.78–189.65)	89.31 (46.93–129.04)	0.440	0.067; 0.069	0.030; 0.701
Lactic + Succinic	10/10	49.18 (27.66–78.17)	35.28 (11.93–52.49)	0.061	0.073; 0.201	**0.332; 0.019**
Propionic + Succinic	10/14	45.17 (12.02–78.36)	50.91 (12.02–87.41)	0.564	**0.111; 0.030**	0.035; 0.688
Formic + Acetic	5/7	105.97 (51.83–188.98)	85.27 (41.10–151,36)	0.696	0.070; 0.465	0.079; 0.440
Propionic + Acetic	10/14	94.92 (14.56–189.54)	107.77 (64.65–134.86)	0.540	0.078; 0.087	0.223; 0.237
Formic + Lactic	6/5	46.03 (0–90.23)	27.49 (18.85–50.43)	0.714	0.099; 0.468	0.310; 0.054

[1] Common name = IUPAC name: Formic acid = methanoic acid; Acetic acid = ethanoic acid; Lactic acid = 2-hydroxypropanoic acid; Propionic acid = Propanoic acid; Succinic acid = Butanedioic acid. [2] The number of patients analyzed changed between analyses; [3] Numeric variables were reported as medians (Q1–Q3); [4] Due to the not-normal distribution, numeric variables were subjected to the Mann–Whitney test. *p*-values < 0.05 were considered significant and are shown in bold.

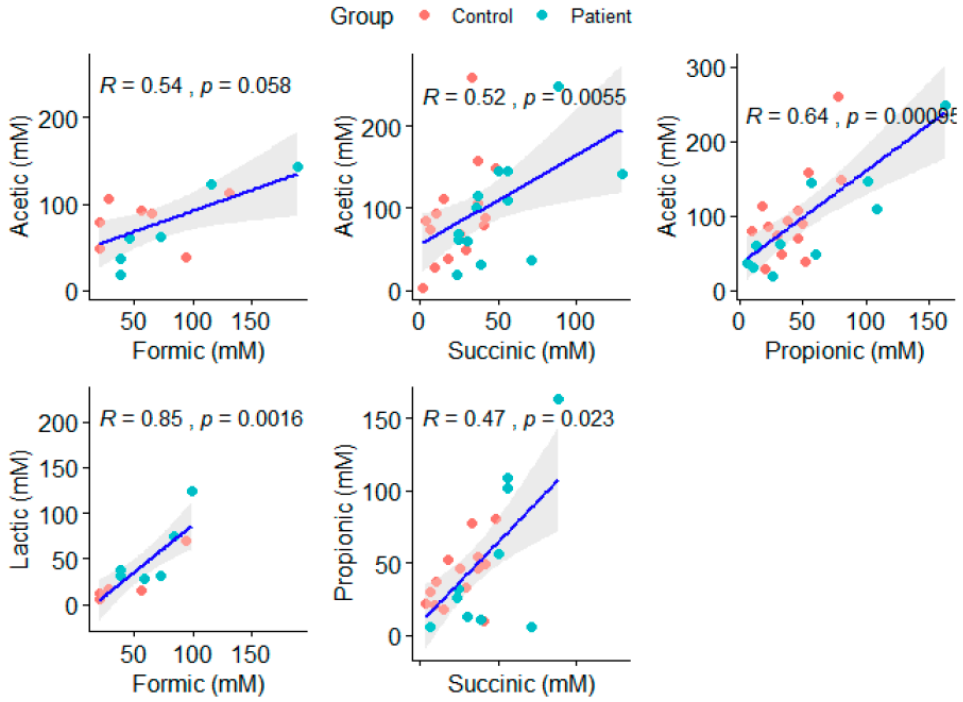


Figure 1. Significative correlations ($p > 0.05$) between acetic, propionic, succinic, lactic and formic acids. As the acetic and formic acids has a related metabolic pathway, this correlation is also shown above although it is not significant.

3.2. Metabolites, Fecal pH and Observed Operational Taxonomic Units

As second step, tests were performed to determine if there were correlations between each metabolite and the fecal pH (Figure 2A) and also between the metabolites and the

number of operational taxonomic units observed on subjects (Figure 2B). Microbial diversity was found decreased when FA (r = −0.54; $p = 0.026$) and LA (r = −0.46; $p = 0.026$) are high, while low fecal pH is associated with higher levels of LA (r = −0.54; $p = 0.0085$). Lactic and formic acid were found to be strongly correlated between them (r = 0.85; $p = 0.0016$).

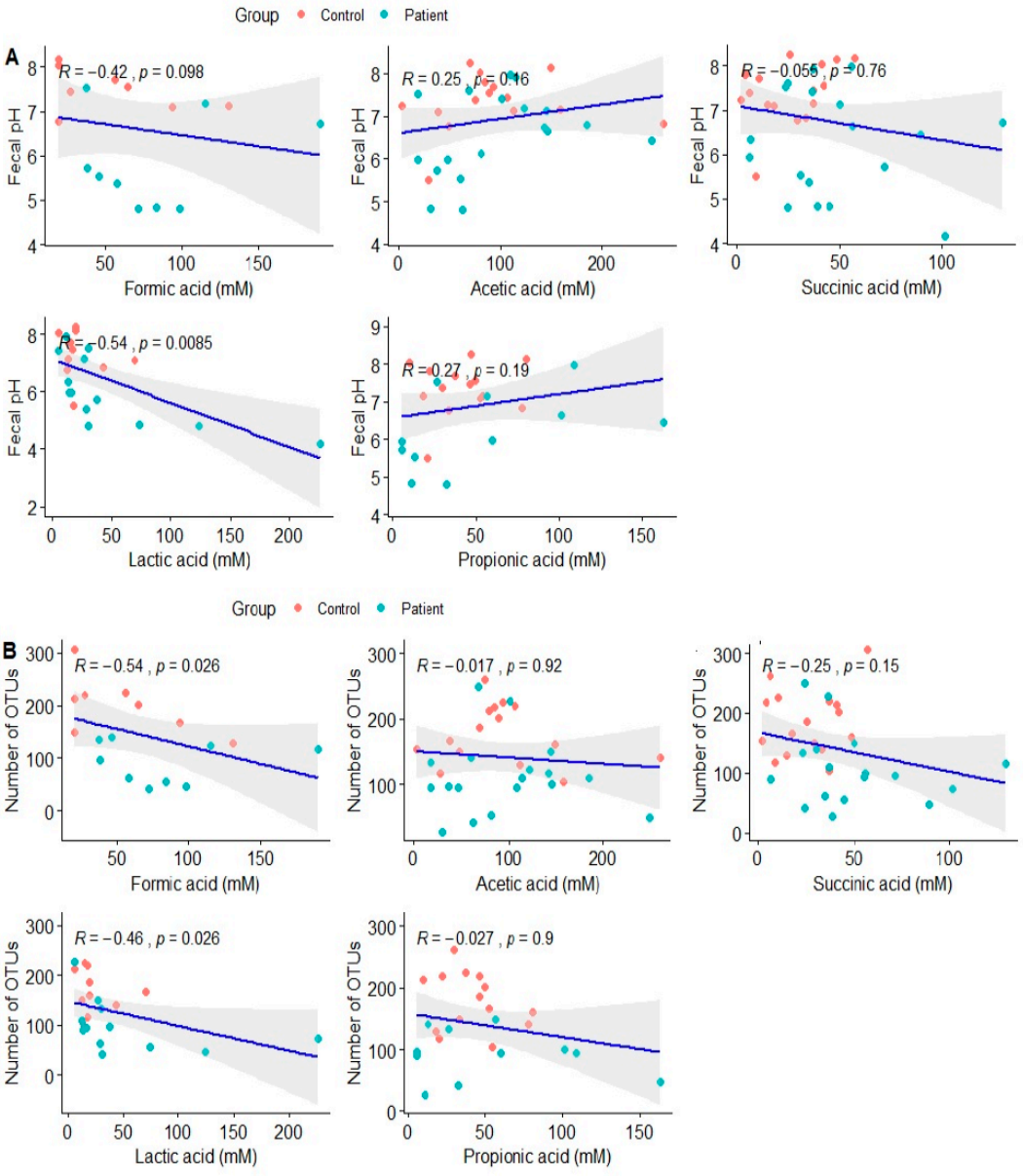


Figure 2. Panel show correlations of acetic, propionic, succinic, lactic and formic acids with (**A**) fecal pH; and (**B**) Observed OTUs between OTUs in fecal samples of patients and controls.

No variation was found among the metabolites when GSD Ia and GSD Ib patients were compared. The number of valid observations did not allow a comparison with GSD type III and IX (Table 2). Metabolites were not directly related to the calprotectin levels found on patients (Figure 3). Additionally, the relationship between BMI and SCFAs was analyzed (data not shown), but no statistically significant difference was observed.

Table 2. Description of clinical parameters and measured metabolites among GSD types.

Variable	Valid Observations GSD Ia/Ib [1]	GSD Ia (n = 15) [2]	GSD Ib (n = 5) [3]	p-Value [4]	GSD III (n = 1)	GSD IX [3] (n = 3)
Formic acid (methanoic acid)	06/02	52.00 (38.34–74.76)	98.50–189.58	NA	NA	115.26
Acetic acid (ethanoic acid)	10/04	71.68 (32.09–145.19)	95.15 (35.27–221.79)	0.777	109.25	68.28–122.67
Lactic acid (2-hydroxypropanoic acid)	10/02	29.78 (14.22–47.08)	16.71–124.19	NA	NA	5.34
Propionic acid (Propanoic acid)	07/03	26.45 (5.86–56.85)	11.37–162.97	0.305	109.10	-
Succinic acid (Butanedioic acid)	12/03	35.95 (23.75–54.95)	38.97–129.04	0.083	55.67	24.65–36.39

[1] The number of patients analyzed changed between analyses; [2] Numeric variables were reported as medians (Q1–Q3); or [3] Min–Max for less than four observations; [4] Due to the not-normal distribution, numeric variables were subjected to the Mann–Whitney test. p-values < 0.05 were considered significant. NA = not accessed due to small number of valid observations. All GSD IX patients had valid measurements to ALT, AST, and acetic acid; lactic and formic acid obtained for one patient, and propionic acid had no reads.

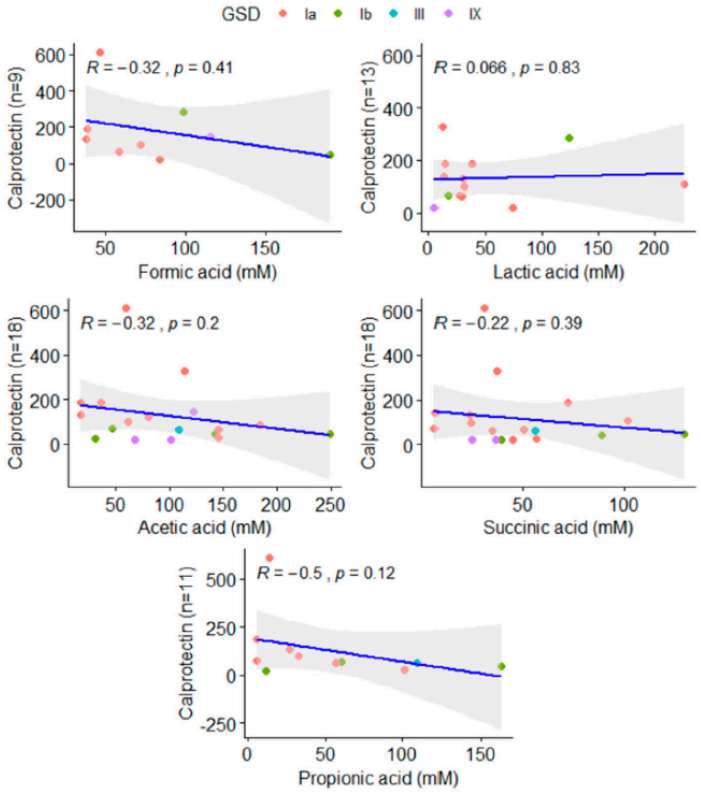


Figure 3. Correlation of acetic, propionic, succinic, lactic, and formic acids and calprotectin levels found in patients. The number of valid reads available in each comparison was added to Y-axis.

4. Discussion

The pathophysiology and gut microbial alteration during inflammatory bowel disease in GSD patients are still not completely understood. In our previous study, we described the biochemical alterations in the gut environment of GSD patients, such as acidification and inflammation. Recently, propionate and acetate have been found to be increased in a group of nine GSD patients (Ia = 4; Ib = 5) and twelve healthy controls [17]. Instead, our study did not find propionate or acetate accumulation but an increase in succinic acid and a tendency

towards formate accumulation in patients. Both studies are concordant regarding dysbiosis in GSD patients and the overrepresentation of Phylum Proteobacteria in patients, but they are not comparable due to factors previously mentioned by Ceccarani and colleagues, such as sequencing technologies and the reference database used for taxonomic classification. Additionally, fecal pH was not characterized by Ceccarani and colleagues (2020), as well as the relationships between metabolites and the structure of the microbial community, which are detailed in this study. This, among other factors already considered by Ceccarani and colleagues, might help to explain the different results in the microbial divergences at lower taxa levels found in the studies.

The profile of SCFAs is influenced by lifestyle, including physical activity and body composition [22]. However, it is known that diet is a major driver of the gut microbiome [23], and as previously published, this was found to explain the majority of differences between the patients and the controls [16]. SCFA formation represents the major flow of carbon from the diet through the microbiome to the host [15]. The principal SCFAs that results from both carbohydrate and amino acid fermentation are acetate, propionate, and butyrate, although formate and other metabolites are also produced in lesser amounts [24]. Other fermentation products such as lactate, ethanol, and succinate, which are intermediates in the global fermentation process in the microbiota, are to varying extents metabolized to SCFA by cross-feeding species in the ecosystem, and they do not usually accumulate in the bowel [25]. Some studies suggest that butyrogenic capacity can be related to lactate availability and that different gut microbial communities can metabolize lactate in different ways [26]. Furthermore, pH is a critical factor in fermentation reactions interfering with metabolite production [27], and thus, its role in metabolite production cannot be underestimated.

In this study, the metabolites did not interfere with the microbial structure in a statistically significant way when they were analyzed alone, but the number of samples was small, the gut is a complex environment, and a synergistic effect is more likely than an isolated effect. Table 1 shows that although some SCFAs can be statistically different between patients/controls, they are not able to affect the gut microbiome alone. The combined effect can be much more interesting when studied together. However, our technique seems not to be the most efficient one to study the SCFAs, given the low number of reads for some SCFAs, e.g., Formic and Propionic. Our number of patients/limited valid reads did not allow us to explore with confidence some trends in statistical patterns that arose from the data. Worthy of note, there was a difference between the patients and the controls for just one metabolite, SA. Although there was no difference in bacterial metabolite levels between the patients and controls for LA and PA, these metabolites alone reached borderline significance and had an interesting effect size on the microbial structure (6.5% and 8%, respectively). This effect was increased and became statistically significant when considering the sum of these metabolites and the amounts of SA, which were higher in patients than in controls (Table 1), indicating that SA has an important role in the gut environment.

SA was found to be increased in patients but had no impact on the microbial structure when considered apart. SA is an important metabolite in both host and microbial processes, playing a role in the activation of immune cells via SUCNR1, a G-protein coupled receptor, enhancing inflammation. Succinate accumulation is infrequently reported for human fecal samples, even in the presence of overweight [27]. However, SA is increased in dysbiosis, patients with IBD, and animal models of intestinal inflammation [28,29]. Previously, we have reported that calprotectin levels, a gut inflammation marker, were abnormal among 70% of GSD Ia patients. There was no correlation between calprotectin and SCFAs in this study (Figure 3).

Taking the biochemical aspects into account, SA is the precursor for propionate formation, but when sufficient carbohydrate is present, there is a reduced need to decarboxylate succinate, and this metabolite accumulates instead of propionate [30]. The UCCS used in the treatment of GSD patients is a mix of amylose and amylopectin, with15% of ileal effluents that leads to microbial fermentation in the colon [31] and might be related to the succinate levels found in the patients. Importantly, none of the metabolites alone or

the tested sums correlated with the total carbohydrate intake among the participants or uncooked cornstarch consumption among the patients (data not shown).

Relatively little is known about the role of formate in the gut. Formic acid in the intestinal lumen has been reported as a product of microbial activity. It has been linked to methanogenesis and appears to be elevated in inflammatory conditions [32,33]. In vivo, formate served as an electron donor in conjunction with oxygen as the terminal electron acceptor. Formate oxidation and oxygen respiration have been described as metabolic signatures for inflammation-associated dysbiosis [34], thus consistent with reduced observed numbers of OUTs and dysbiosis and also with the depletion of methanogen previously reported in patients.

In this study, the levels of butyrate were not accessed. In addition, our technique was different from that used by Ceccarani and collaborators (HPLC vs. CG), and we are unable to verify if the pH conditions in the gut environment were similar among the subjects of this and the aforementioned study.

Here, we described an altered microbial metabolite profile on an SA basis, which are consistent with dysbiosis. Considering the limitations of this study, such as the limited data on gut inflammation and the absence of physical activity data in patients and controls, our results indicate that the association between IBD and the increased penetration of SCFAs [14] can be extended to patients with GSD. However, the mechanism underlying the immune activation of the gut environment in those patients remains unclear.

Author Contributions: Conceptualization, K.C., I.V.D.S. and L.F.W.R.; methodology, E.L.d.C., D.L.R. and P.M.P.; software, K.C., E.L.d.C., D.L.R. and P.M.P.; validation, E.L.d.C., D.L.R., P.M.P. and L.F.W.R.; formal analysis, K.C. and L.F.W.R..; investigation, K.C., E.L.d.C., D.L.R., P.M.P., L.F.W.R., F.C.P. and I.V.D.S.; resources, P.M.P. and I.V.D.S.; data curation, P.M.P. and L.F.W.R.; writing—original draft preparation, K.C., E.L.d.C., D.L.R. and L.F.W.R.; writing—review and editing, K.C., E.L.d.C., D.L.R., P.M.P., L.F.W.R., F.C.P. and I.V.D.S.; visualization, K.C.; supervision, I.V.D.S. and L.F.W.R.; project administration, I.V.D.S.; funding acquisition, I.V.D.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS)-EDITAL PRONEX FAPERGS/CNPq 12/2014, processo 16/2551-0000492-7; Fundo de Incentivo à Pesquisa e Eventos do Hospital de Clínicas de Porto Alegre (FIPE-HCPA n. 20150218; n. 20190457) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Institutional Review Board Statement: The study protocol was approved by the Ethics Committee of Hospital de Clínicas de Porto Alegre (IRB numbers 20150218 and 20190457).

Informed Consent Statement: All participants and/or legal guardians signed an informed consent form. Written informed consent was obtained from the patient(s) to publish this paper.

Data Availability Statement: The data presented in this study are available in the main article.

Acknowledgments: The authors would like to thank FIPE-HCPA, CAPES, CNPq, FAPERGS and the UFRGS Post-Graduate Program in Genetics and Molecular Biology for their support.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study, in the collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

References

1. Weinstein, D.A.; Steuerwald, U.; De Souza, C.F.; Derks, T. Inborn Errors of Metabolism with Hypoglycemia: Glycogen Storage Diseases and Inherited Disorders of Gluconeogenesis. *Pediatr. Clin. N. Am.* **2018**, *65*, 247–265. [CrossRef] [PubMed]
2. Kishnani, P.S.; Austin, S.L.; Abdenur, J.E.; Arn, P.; Bali, D.S.; Boney, A.; Chung, W.K.; Dagli, A.I.; Dale, D.; Koeberl, D.; et al. Diagnosis and management of glycogen storage disease type I: A practice guideline of the American College of Medical Genetics and Genomics. *Genet. Med.* **2014**, *16*, e1–e29. [CrossRef] [PubMed]

3. Chou, J.Y.; Mansfield, B.C. Molecular Genetics of Type 1 Glycogen Storage Diseases. *Trends Endocrinol. Metab.* **1999**, *10*, 104–113. [CrossRef]
4. Labrune, P.; Ullrich, K.; Smit, P.G.; Rake, J.; Visser, G.; Leonard, J.V. Guidelines for management of glycogen storage disease type I—European Study on Glycogen Storage Disease Type I (ESGSD I). *Eur. J. Pediatr.* **2002**, *161*, S112–S119. [CrossRef] [PubMed]
5. Visser, G.; Rake, J.-P.; Fernandes, J.; Labrune, P.; Leonard, J.V.; Moses, S.; Ullrich, K.; Smit, G.A. Neutropenia, neutrophil dysfunction, and inflammatory bowel disease in glycogen storage disease type Ib: Results of the European Study on Glycogen Storage Disease Type I. *J. Pediatr.* **2000**, *137*, 187–191. [CrossRef]
6. Lawrence, N.T.; Chengsupanimit, T.; Brown, L.M.; Derks, T.; Smit, G.P.A.; Weinstein, D.A. Inflammatory Bowel Disease in Glycogen Storage Disease Type Ia. *J. Pediatr. Gastroenterol. Nutr.* **2017**, *64*, e52–e54. [CrossRef]
7. Lawrence, N.T.; Chengsupanimit, T.; Brown, L.M.; Weinstein, D.A.; Zschocke, J. *High Incidence of Serologic Markers of Inflammatory Bowel Disease in Asymptomatic Patients with Glycogen Storage Disease Type Ia*; Springer: Berlin/Heidelberg, Germany, 2015; Volume 24, pp. 123–128. [CrossRef]
8. Yue, B.; Luo, X.; Yu, Z.; Mani, S.; Wang, Z.; Dou, W. Inflammatory Bowel Disease: A Potential Result from the Collusion between Gut Microbiota and Mucosal Immune System. *Microorganisms* **2019**, *7*, 440. [CrossRef]
9. Ríos-Covián, D.; Ruas-Madiedo, P.; Margolles, A.; Gueimonde, M.; De Los Reyes-Gavilán, C.G.; Salazar, N. Intestinal Short Chain Fatty Acids and their Link with Diet and Human Health. *Front. Microbiol.* **2016**, *7*, 185. [CrossRef]
10. Belenguer, A.; Duncan, S.H.; Holtrop, G.; Anderson, S.E.; Lobley, G.E.; Flint, H.J. Impact of pH on Lactate Formation and Utilization by Human Fecal Microbial Communities. *Appl. Environ. Microbiol.* **2007**, *73*, 6526–6533. [CrossRef]
11. Walker, A.W.; Duncan, S.H.; McWilliam Leitch, E.C.; Child, M.W.; Flint, H.J. pH and Peptide Supply Can Radically Alter Bacterial Populations and Short-Chain Fatty Acid Ratios within Microbial Communities from the Human Colon. *Appl. Environ. Microbiol.* **2005**, *71*, 3692–3700. [CrossRef]
12. McLoughlin, R.F.; Berthon, B.S.; Jensen, M.E.; Baines, K.J.; Wood, L.G. Short-chain fatty acids, prebiotics, synbiotics, and systemic inflammation: A systematic review and meta-analysis. *Am. J. Clin. Nutr.* **2017**, *106*, 930–945. [CrossRef] [PubMed]
13. Nugent, S.G.; Kumar, D.; Rampton, D.S.; Evans, D.F. Intestinal luminal pH in inflammatory bowel disease: Possible determinants and implications for therapy with aminosalicylates and other drugs. *Gut* **2001**, *48*, 571–577. [CrossRef] [PubMed]
14. Jaworska, K.; Konop, M.; Bielinska, K.; Hutsch, T.; Dziekiewicz, M.; Banaszkiewicz, A.; Ufnal, M. Inflammatory bowel disease is associated with increased gut-to-blood penetration of short-chain fatty acids: A new, non-invasive marker of a functional intestinal lesion. *Exp. Physiol.* **2019**, *104*, 1226–1236. [CrossRef] [PubMed]
15. Morrison, D.J.; Preston, T. Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. *Gut Microbes* **2016**, *7*, 189–200. [CrossRef] [PubMed]
16. Colonetti, K.; Dos Santos, B.B.; Nalin, T.; De Souza, C.F.M.; Triplett, E.W.; Dobbler, P.C.T.; Schwartz, I.V.D.; Roesch, L.F.W. Hepatic glycogen storage diseases are associated to microbial dysbiosis. *PLoS ONE* **2019**, *14*, e0214582. [CrossRef]
17. Ceccarani, C.; Bassanini, G.; Montanari, C.; Casiraghi, M.C.; Ottaviano, E.; Morace, G.; Biasucci, G.; Paci, S.; Borghi, E.; Verduci, E. Proteobacteria Overgrowth and Butyrate-Producing Taxa Depletion in the Gut Microbiota of Glycogen Storage Disease Type 1 Patients. *Metabolites* **2020**, *10*, 133. [CrossRef]
18. De Baere, S.; Eeckhaut, V.; Steppe, M.; De Maesschalck, C.; De Backer, P.; Van Immerseel, F.; Croubels, S. Development of a HPLC–UV method for the quantitative determination of four short-chain fatty acids and lactic acid produced by intestinal bacteria during in vitro fermentation. *J. Pharm. Biomed. Anal.* **2013**, *80*, 107–115. [CrossRef]
19. McMurdie, P.J.; Holmes, S. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **2013**, *8*, e61217. [CrossRef]
20. Oksanen, J.; Simpson, G.L.; Blanchet, F.G.; Kindt, R.; Legendre, P.; Minchin, P.R.; O'Hara, R.B.; Solymos, P.; Stevens, M.H.H.; Szoecs, E.; et al. Vegan: Community Ecology Package Version. 2013. Available online: https://github.com/vegandevs/vegan (accessed on 25 March 2022).
21. Lemos, L.N.; Fulthorpe, R.R.; Triplett, E.W.; Roesch, L.F. Rethinking microbial diversity analysis in the high throughput sequencing era. *J. Microbiol. Methods* **2011**, *86*, 42–51. [CrossRef]
22. Della Guardia, L.; Codella, R. Exercise tolls the bell for key mediators of low-grade inflammation in dysmetabolic conditions. *Cytokine Growth Factor Rev.* **2021**, *62*, 83–93. [CrossRef]
23. Hasan, N.; Yang, H. Factors affecting the composition of the gut microbiota, and its modulation. *PeerJ* **2019**, *7*, e7502. [CrossRef] [PubMed]
24. Macfarlane, S.; Macfarlane, G.T. Proteolysis and Amino Acid Fermentation. In *Human Colonic Bacteria: Role in Nutrition, Physiology, and Pathology*; Gibson, G.R., Macfarlane, G.T., Eds.; CRC Press: Boca Raton, FL, USA, 1995; p. 75.
25. Bernalier, A.; Dore, J.; Durand, M. Biochemistry of Fermentation. In *Colonic Microbiota, Nutrition and Health*; Springer: Dordrecht, The Netherlands, 1999; pp. 37–53. [CrossRef]
26. Bourriaud, C.; Robins, R.; Martin, L.; Kozlowski, F.; Tenailleau, E.; Cherbut, C.; Michel, C. Lactate is mainly fermented to butyrate by human intestinal microfloras but inter-individual variation is evident. *J. Appl. Microbiol.* **2005**, *99*, 201–212. [CrossRef] [PubMed]
27. Ilhan, Z.E.; Marcus, A.K.; Kang, D.-W.; Rittmann, B.E.; Krajmalnik-Brown, R. pH-Mediated Microbial and Metabolic Interactions in Fecal Enrichment Cultures. *mSphere* **2017**, *2*, e00047-17. [CrossRef] [PubMed]

28. Gilissen, J.; Jouret, F.; Pirotte, B.; Hanson, J. Insight into SUCNR1 (GPR91) structure and function. *Pharmacol. Ther.* **2016**, *159*, 56–65. [CrossRef]
29. Lei, W.; Ren, W.; Ohmoto, M.; Urban, J.F.; Matsumoto, I.; Margolskee, R.F.; Jiang, P. Activation of intestinal tuft cell-expressed Sucnr1 triggers type 2 immunity in the mouse small intestine. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 5552–5557. [CrossRef]
30. Macy, J.M.; Ljungdahl, L.G.; Gottschalk, G. Pathway of Succinate and Propionate Formation in *Bacteroides fragilis*. *J. Bacteriol.* **1978**, *134*, 84–91. [CrossRef]
31. Nalin, T.; Venema, K.; Weinstein, D.A.; De Souza, C.F.M.; Perry, I.D.S.; Van Wandelen, M.T.R.; Van Rijn, M.; Smit, G.P.A.; Schwartz, I.V.D.; Derks, T.G.J. In vitro digestion of starches in a dynamic gastrointestinal model: An innovative study to optimize dietary management of patients with hepatic glycogen storage diseases. *J. Inherit. Metab. Dis.* **2014**, *38*, 529–536. [CrossRef]
32. Vanderhaeghen, S.; Lacroix, C.; Schwab, C. Methanogen communities in stools of humans of different age and health status and co-occurrence with bacteria. *FEMS Microbiol. Lett.* **2015**, *362*, fnv092. [CrossRef]
33. Bereswill, S.; Fischer, A.; Plickert, R.; Haag, L.-M.; Otto, B.; Kühl, A.A.; Dashti, J.I.; Zautner, A.E.; Muñoz, M.; Loddenkemper, C.; et al. Novel Murine Infection Models Provide Deep Insights into the "Ménage à Trois" of Campylobacter jejuni, Microbiota and Host Innate Immunity. *PLoS ONE* **2011**, *6*, e20953. [CrossRef]
34. Hughes, E.R.; Winter, M.G.; Duerkop, B.A.; Spiga, L.; De Carvalho, T.F.; Zhu, W.; Gillis, C.C.; Büttner, L.; Smoot, M.P.; Behrendt, C.L.; et al. Microbial Respiration and Formate Oxidation as Metabolic Signatures of Inflammation-Associated Dysbiosis. *Cell Host Microbe* **2017**, *21*, 208–219. [CrossRef]

12. ANEXO V

O artigo intitulado *"Venom characterization of the Brazilian Pampa snake Bothrops pubescens by top-down and bottom-up proteomics"* de autoria de Rangel, D.L., Melani, R.D., **Carvalho, E.L.**, Boldo, J.T., Dos Santos, T.G., Kelleher, N.L., Pinto, P.M. foi publicado no periódico *Toxicon* (ISSN: 0041-0101), no ano de 2022.

Bothrops pubescens é uma cobra peçonhenta endêmica do Bioma Pampa, no Sul do Brasil. O envenenamento causado por cobras do gênero *Bothrops* são caracterizados por efeitos locais, como hemorragia, necrose, entre outros; e efeitos sistêmicos, como insuficiência renal e coagulopatias. As principais famílias de toxinas encontradas no veneno desse gênero são as metaloproteinases, serino-proteases e fosfolipase A2.

A caracterização dos venenos é um importante fator para entender o processo de envenenamento, apoiar estudos de filogenia e na descoberta de novos produtos farmacêuticos e biotecnológicos.

Através de análises proteômicas tipo *bottom-up* e *top-down* do veneno de 5 espécimes de *B. pubescens*, conseguimos identificar 89 grupos de proteínas pertencentes a 13 famílias de toxinas e 40 proteoformas uniúnicas pertencentes a 6 famílias de toxinas. Também identificamos um complexo multi-proteofórmico de uma L-aminoácido oxidase.

Minha participação neste trabalho foi no projeto e execução dos experimentos, análise dos dados proteômicos e participação na redação do manuscrito.

Toxicon 220 (2022) 106937

Contents lists available at ScienceDirect

Toxicon

journal homepage: www.elsevier.com/locate/toxicon

Short communication

Venom characterization of the Brazilian Pampa snake *Bothrops pubescens* by top-down and bottom-up proteomics

Darlene Lopes Rangel [a,b,1], Rafael D. Melani [c,1], Evelise Leis Carvalho [a,b], Juliano Tomazzoni Boldo [b], Tiago Gomes dos Santos [d], Neil L. Kelleher [c], Paulo Marcos Pinto [a,b,*]

[a] Applied Proteomics Laboratory, Federal University of Pampa, São Gabriel, Brazil
[b] Pos Graduation in Biological Sciences, Federal University of Pampa, São Gabriel, Brazil
[c] Departments of Molecular Biosciences, Chemistry, and the Feinberg School of Medicine, Northwestern University, Evanston, IL, 60208, United States
[d] Pampa Biodiversity Studies Laboratory (LEBIP), Federal University of Pampa, São Gabriel, Brazil

ARTICLE INFO

Handling Editor: Ray Norton

Keywords:
Venomics
Bothrops
Envenomation
L-amino acid Oxidase

ABSTRACT

The envenomation from the *Bothrops* genus is characterized by systemic and local effects caused by the main toxin families in the venom. In *Bothrops pubescens* venom we were able to identify 89 protein groups belonging to 13 toxin families with the bottom-up proteomics approach and 40 unique proteoforms belonging to 6 toxin families with the top-down proteomics approach. We also identified multi-proteoform complexes of dimeric L-amino acid oxidase using native top-down mass spectrometry.

Main

The pit viper genus *Bothrops* is responsible for ~90% of snakebite envenoming in Brazil. The envenomation by *Bothrops* genus is generally characterized by local (hemorrhage, necrosis, vessel damage, and edema) and systemic effects (renal failure and coagulopathies) (Mamede et al., 2020; Gutierrez et al., 2010). These effects are caused by the main toxin families found in the venom, such as snake venom metalloproteinase (SVMP), snake venom serine protease (SVSP), and phospholipase A_2 (PLA2). SVMPs have fibrin(ogen)olytic activity, prothrombin activator action and are responsible for proteolytic degradation of the capillary basement membrane (Markland and Swenson, 2013). SVSPs are responsible for the cleavage of proteins involved in the clotting system and homeostasis leading to activation/inhibition of proteins involved in the coagulation system and homeostasis. (Sajevic et al., 2011). Additionally, PLA2 is accountable for neurotoxicity, myotoxicity, cardiotoxicity, and platelet aggregation/inhibition (Hiu and Yap, 2020)

Using snake venomics (Calvete et al., 2007) and other proteomics-based methods many *Bothrops* species have had their venom characterized in the past decade (Nery et al., 2016; Jorge et al., 2015; Mora-Obando et al., 2014, 2020; Mora-Obando et al., 2014, 2020; Gay et al., 2015; Amorim et al., 2018; Sanz et al., 2020a, 2020b; Goncalves-Machado et al., 2016; Ohler et al., 2010). These studies are essential for understanding the envenomation process, discovering potential pharmaceutical/biotechnological products (Waheed et al., 2017), and supporting phylogeny studies (Waheed et al., 2017; Segura et al., 2013). *Bothrops pubescens* is a venomous snake endemic to the Pampa biome in south Brazil (Machado et al., 2014). Although there are few phylogenetic studies about *B. pubescens* (Machado et al., 2014), the species is neglected regarding its venom composition. In this study, we present for the first time the venom composition of *B. pubescens* that was analyzed by top-down proteomics (TDP), native top-down mass spectrometry, and bottom-up proteomics (BUP).

B. pubescens venom was milked from 5 different adult specimens (2 males and 3 females) collected in the State of Rio Grande do Sul - Brazil. After extraction, the venom was pooled, lyophilized, and kept at −80 °C until further use. For the size exclusion chromatography (SEC), crude venom (500 μg) was diluted in 100 mM ammonium acetate and fractionated on a Yarra 3 μm SEC-3000 LC column 300 × 4.6 mm (Phenomenex) using an Agilent 1100 Series liquid chromatography system at a flow rate of 0.2 mL/min. Protein separation was monitored at 280 nm using Agilent OpenLab software. The fractions were subjected to SDS-PAGE, according to Laemmli (1970), and the gel was silver stained.

* Corresponding author. Applied Proteomics Laboratory, Federal University of Pampa, São Gabriel, Brazil.
E-mail address: paulopinto@unipampa.edu.br (P.M. Pinto).
[1] Equal contribution.

https://doi.org/10.1016/j.toxicon.2022.106937
Received 17 August 2022; Received in revised form 29 September 2022; Accepted 29 September 2022
Available online 10 October 2022

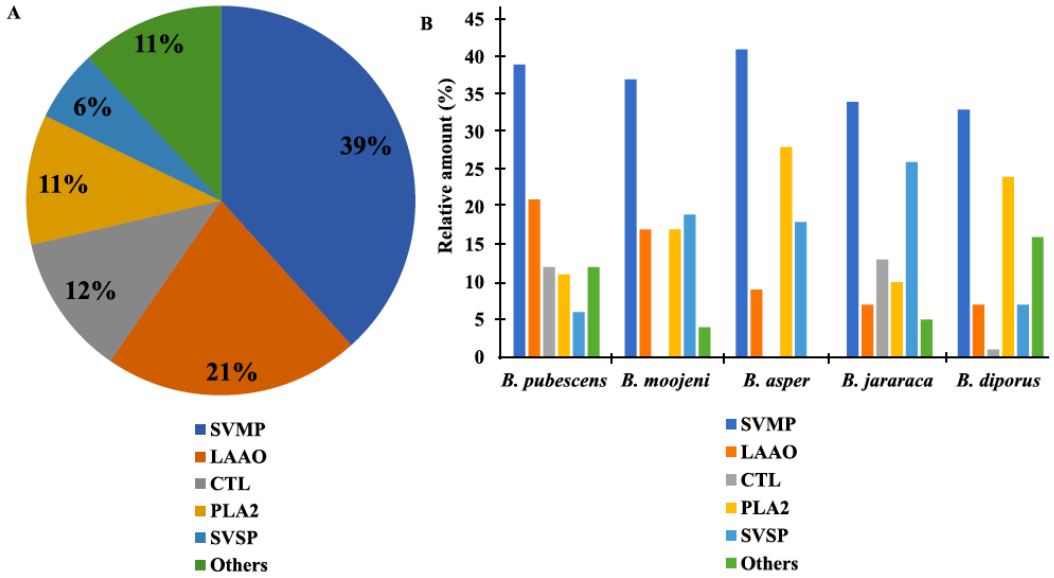


Fig. 1. Relative protein composition of *B. pubescens* venom. The venom was analyzed by LC-MS/MS. Peptide search was performed on ProLuCID v1.3, and peptides were filtered by Search Engine Processor (SEPro). (A) Relative protein composition (%) of *B. pubescens* venom was estimated by label-free protein quantitation performed according to the normalized spectral abundance factor (NSAF). (B) Comparison between *B. pubescens*, *B. moojeni* (Melani et al., 2016), *B. asper* (Laemmli), *B. jararaca* (Mora-Obando et al., 2020), and *B. diporus* (Amorim et al., 2018) venom composition. SVMP - snake venom metalloproteinase, LAAO - L-aminoacidic oxidase, CTL - C-type lectin, PLA_2 - phospholipase A_2, SVSP - snake venom serine protease. Toxin families with less than 2% in the venom (others) include VEGF – vascular endothelial growth factor, BPP - bradykinin potentiating peptides, CRISP - cysteine-rich secretory protein, NGF - nerve growth factor, PDE - phosphodiesterase, GPC - glutamine-cyclotransferase, PLB - phospholipase B, NT - nucleotidases.

All the proteomics analysis was conducted as described in Melani et al. (2016). In brief, for BUP, 300 µg of *B. pubescens* crude venom was resuspended in 7 M urea, 2 M thiourea, and reduced using 30 mM DTT for 1 h at 37 °C. Cysteine residues were carboxamide methylated with 30 mM iodoacetamide for 1 h in the dark at room temperature. The sample was diluted to 1 M urea with Tris 100 mM pH 8.2. MS grade trypsin (Promega) was added (1:25, protease:substrate), and the sample was incubated overnight at 37 °C. Peptides were fractionated using an Ultimate 3000 nanoLC (Thermo Fisher Scientific) system on an in-house packed 2 cm × 150 µm i.d. trap-column, and 25 cm × 75 µm i.d. column (Jupiter C18, 3 µm particle size, 300 Å pore size, Phenomenex) coupled to a Q-Exactive HF (Thermo Fisher Scientific). Chromatography was performed at 300 nL/min flow rate with 95% water, 5% ACN, and 0.2% formic acid (FA) as mobile phase A and 95% ACN, 5% water, and 0.2% FA as phase B on a 120 min Gradient (5% B for 8 min, 5–10% B over 4 min, followed by 10–45% B over 88 min, 45%–94% B over 4 min, 95% for 4 min, 90–5% B for 2 min, and 5% for 10 min). Mass spectra were acquired by Xcalibur software operating in data-dependent acquisition mode, switching between full scan MS1 (60,000 resolution, 50 ms maximum injection time, AGC $3e^6$ charges, spectrum range from 300 to 1800 m/z) and MS2 (30,000 resolution, 100 ms maximum injection time, AGC $1e^5$ charges, spectrum range from 200 to 2,000 m/z). MS2 spectra were obtained by high-energy collision dissociation (HCD) fragmentation using 28% normalized collision energy. The peptides were analyzed in three technical replicates. Peptide search was performed using ProLuCID v1.3 search engine against sequences in the Serpente database, downloaded from UniprotKB on February 15, 2021. Carboxamidomethylation of cysteines and oxidation of methionine was set as fixed and variable modifications, respectively. Resultant peptides were processed and evaluated by Search Engine Processor (SEPro) (Reid et al., 2018) with the following parameters: 10 ppm deviation from theoretical peptide precursor, peptides longer than six amino acid residues, and a 1% estimated protein-level false discovery rate (FDR). Label-free protein quantitation was performed according to the normalized spectral abundance factor (NSAF).

For the TDP approach, the pooled venom was resuspended in mobile phase A and then directly analyzed by LC-MS/MS. NanoRPLC analysis was performed using an Ultimate 3000 nanoLC (Thermo Fisher Scientific) system on an in-house packed polymeric reverse phase resin (PLRP-S, 5 µm particle size, 1,000 Å pore size, Agilent Technologies) for the trap-column (2 cm × 150 µm i.d.) and analytical column (20 cm × 75 µm i.d.). The sample was analyzed using a 120 min Gradient (5% B for 10 min, 5–15% B over 2 min, followed by 15–50% B over 88 min, 50%–95% B over 2 min, 95% for 5 min, followed by 90–5% B over 3 min, and 5% for 10 min). Mass spectra were acquired by Xcalibur. MS1 spectra were acquired at 120,000 resolving power (AGC $1e^6$ charges, 50 ms maximum injection time, spectrum range 400–2,000 m/z) using 4 µscans. Data-dependent MS2 were acquired at 60,000 resolving power (AGC 1 e^6 charges, maximum injection time 800 ms, spectrum range 200–2,000 m/z) using 4 µscans on a Q Exactive HF. MS2 spectra were obtained by HCD fragmentation using an isolation window of 4 m/z and 21%, 22%, and 25% stepped normalized collision energy. The raw data was searched using TDPortal against the Serpente database downloaded from UniProtKB on March 8, 2020, using 2.2 Da precursor window tolerance, 10 ppm fragment tolerance, and 1% FDR.

For the native top-down characterization, the fractions from SEC were individually analyzed into a Q-Exactive UHMR (Thermo Fisher Scientific) mass spectrometer using a Nanospray Flex Ion Source (Thermo Fisher Scientific), spray voltages between 1,500 and 2,000 V, and the ion transfer tube was set to 310 °C. The mass spectrometer was run in positive mode, data collected from 700 to 15,000 m/z with 20 µscans, resolution of 17,500 (at 200 m/z), and a target AGC of $1e^6$ charges. The S-lens RF level was set to 200% and extended trapping set to 100 V. MS2 spectra were obtained by HCD fragmentation using 21% normalized collision energy. Raw spectra from each experiment were summed across scans, and mass deconvolution was performed using

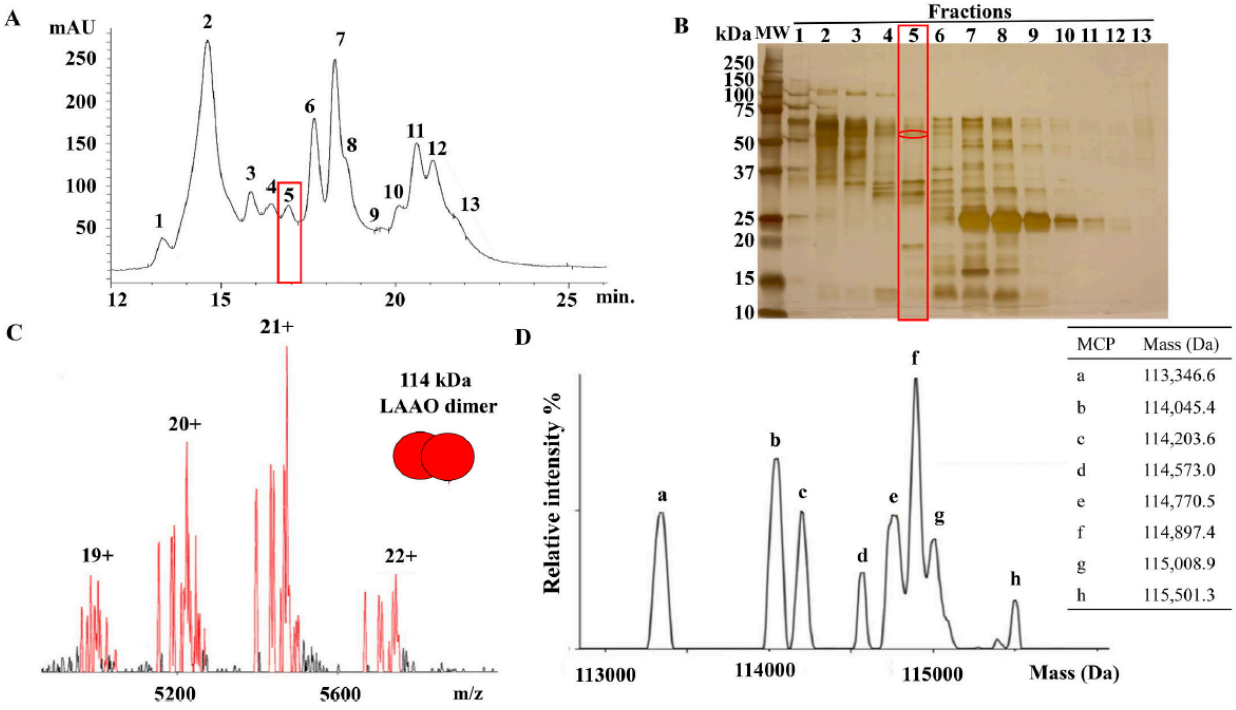


Fig. 2. Identification of dimeric L-amino acid oxidase (LAAO) in *B. pubescens* venom. (A) Chromatogram trace at 280 nm of the venom proteins highlighting the 13 fractions collected by size-exclusion chromatography (SEC). (B) Silver-stained SDS-PAGE analysis of the 13 SEC fractions. (C) Spectrum obtained using native top-down proteomics showing in red the charge state distribution of different LAAO multiproteoform complexes (MPC). (D) Deconvoluted MS1 spectrum and the intact mass of 8 different LAAO MPCs (a–h). The red squares indicate the fraction containing LAAO, and the red circle the LAAO bands region in the SDS-PAGE gel. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

UniDec (Melani et al., 2016). All data are available via ProteomeXchange with identifier PXD027650.

We were able to identify 89 proteins (Supplementary Table 1) by BUP that were assigned to 13 toxin families. We estimated the relative abundance of the venom composition using label-free quantification, and SVMP was the most abundant toxin family corresponding to 39% of the total (Fig. 1A). The venom composition of *B. pubescens* showed to be similar to other *Bothrops* species, presenting around 80–100 proteins belonging to characteristic toxin families (Fig. 1B) (Gay et al., 2015; Amorim et al., 2018; Alape-Girón et al., 2008; Galizio et al., 2018). When comparing the amount of L-amino acid oxidase (LAAO) among different *Bothrops* venoms, *B. pubescens* (21%) has a slightly higher level than *B. moojeni* (17.7%), and superior amounts than venoms from B. *asper* (9.2%), *B. jararaca* (7.9%), and *B. diporus* (7.4%). This could indicate that *B. pubescens* has a stronger myonecrosis, edema-forming, hemorrhage-promoting, platelet-aggregating and/or -inhibiting effect than other *Bothrops* venoms (Tan et al., 2018; Ribeiro et al., 2016). Furthermore, *B. pubescens* venom showed a lower content of SVSP (6%) compared to *B. moojeni* (19.6%), *B. asper* (18.2%), and *B. jararaca* (26.65%). Such a low amount is comparable to *B. diporus* (7.2%), which belongs to a phylogenetic sister group of *B. pubescens* (Machado et al., 2014).

With the TDP approach, we identified 40 proteoforms belonging to 6 toxin families (Supplementary Table 2). From the total, 12 proteoforms mapped to SVMP protein family, 13 to PLA2, 7 to bradykinin potentiating peptide (BPP), 6 to LAAO, 1 to SVSP, and 1 to SVSP inhibitor. All proteoforms and the SVSP inhibitor family were only identified by TDP, even without performing multiple fractionation steps before LC-MS/MS and not having genomic data on *B. pubescens*. Further, we were able to identify 5 alpha-amino acetylated residues at the N-terminus of SVSPs, SVMPs, PLA2, and SVSP inhibitor families. TDP applied to snake venoms is a growing field that allows the analysis and characterization of intact toxin proteoforms compared to identifying peptides by BUP (Melani et al., 2017). However, few snake venoms have been characterized by this technique (Carvalho et al., 2012; Gocmen et al., 2015; Petras et al., 2015, 2016; Calderon-Celis et al., 2016; Zhou et al., 2020; Ainsworth et al., 2018; Kazandjian et al., 2021; Calvete et al., 2021), and the method still has limitations in identifying proteins larger than 30 kDa. In summary, TDP we were able to locate PTMs, identify unique proteoforms and a toxin family not identified by de BUP approach. This indicates the potential of the technique since a lot of information is lost in the BUP approach.

On the other hand, native top-down analysis maintains the non-covalent associations of protein complexes, while this information is inferred in BUP and denaturing TDP (Zhou et al., 2020). For native TDP, we collected 13 fractions using SEC (Fig. 2A), and they were submitted to SDS-PAGE (Fig. 2B) to check their molecular composition. In fraction 5, we observed a cluster of peaks in the charge states 19–22+ (Fig. 2C), which presented intact masses from 113.3 to 115.5 kDa after deconvolution (Fig. 2D). Subsequently, a whole charge state was fragmented using HCD, and an amino acid sequence was partially obtained. The sequence tag "PYQFQHFSEALTA" was identified by BLASTP against the protein sequences characterized in the BUP approach as LAAO with 100% sequence identity to B5AR80, Q6TGQ9, P0CC17, and P56742. The observed intact masses correspond to multi-proteoform complexes (MPC) of two monomers of LAAO, a protein complex known to occur in snake venom (Gay et al., 2015; Amorim et al., 2018; Ohler et al., 2010; Alape-Girón et al., 2008; Galizio et al., 2018).

In summary, this work shows the first characterization of the *B. pubescens* venom, a neglected species from the southern cone, using multiple proteomics approaches. BUP and TDP combined were able to identify 14 toxin families and demonstrate that *B. pubescens* venom is a

classical *Bothrops* venom, although it presents some peculiarities. Combining complementary BUP and TDP approaches allowed us to understand the *B. pubescens* venom composition, including toxin families, toxin proteoforms, and posttranslational modifications (Laemmli, 1970). Also, with the application of native top-down mass spectrometry, we have a better view of the isoforms we have in the venom composition (Gocmen et al., 2015). Furthermore, TDP was used to analyze a bothropic venom for the first time.

Associated data (for reviewers only)

Data are available via ProteomeXchange with identifier PXD027650.
Username: reviewer_pxd027650@ebi.ac.uk.
Password: CXtXYyFe

Ethical statement

Paulo Marcos Pinto (corresponding author), consciously assure that for the manuscript entitled "Venom Characterization of the Brazilian Pampa Snake *Bothrops pubescens* by Top-Down and Bottom-Up Proteomics", the following is fulfilled:

1) This material has not been published in whole or in part elsewhere;
2) The manuscript is not currently being considered for publication in another journal;
3) All authors have been personally and actively involved in substantive work leading to the manuscript and will hold themselves jointly and individually responsible for its content.

Credit author statement

DLR, RDM, ELC and PMP conceived and designed the experiments. DLR, RDM, ELC and TGS performed the experiments. DLR, RDM, ELC, JTB, NLK and PMP analyzed the proteomics data. DLR, RDM, ELC and PMP analyzed the results and wrote the original draft manuscript. All authors participated in the writing, review, and editing of the final manuscript.

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

This work was supported by the National Council for Scientific and Technological Development (CNPq-Brazil), the Coordination for the Improvement of Higher Education Personnel (CAPES-Brazil), the Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS-Brazil), the National Institute of Health in a visiting scholar program supported by a grant from the National Institute of General Medical Sciences P41 GM108569 (N.L.K.), and the NIH Office of Director award S10 OD025194. D.L.R. received a research fellowship from CAPES-Brazil, and E.L.C. received a research fellowship from FAPERGS-Brazil. This is for Alice.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.toxicon.2022.106937.

References

Ainsworth, S., Petras, D., Engmark, M., Sussmuth, R.D., Whiteley, G., Albulescu, L.O., Calvete, J.J., 2018. The medical threat of mamba envenoming in sub-Saharan Africa revealed by genus-wide analysis of venom composition, toxicity and antivenomics profiling of available antivenoms. J. Proteomics 172, 173–189. https://doi.org/10.1016/j.jprot.2017.08.016.

Alape-Girón, A., Sanz, L., Escolano, J., Flores-Diaz, M., Madrigal, M., Sasa, M., Calvete, J. J., 2008. Snake venomics of the lancehead pitviper Bothrops asper: geographic, individual and ontogenetic variations. J. Proteome Res. 7, 3556–3571. https://doi.org/10.1021/pr800332p.

Amorim, F.G., Costa, T.R., Baiwir, D., De Pauw, E., Quinton, L., Sampaio, S.V., 2018. Proteopeptidomic, functional and immunoreactivity characterization of Bothrops moojeni snake venom: influence of snake gender on venom composition. Toxins 10 (5). https://doi.org/10.3390/toxins10050177.

Calderon-Celis, F., Diez-Fernandez, S., Costa-Fernandez, J.M., Encinar, J.R., Calvete, J.J., Sanz-Medel, A., 2016. Elemental mass spectrometry for absolute intact protein quantification without protein-specific standards: application to snake venomics. Anal. Chem. 88 (19), 9699–9706. https://doi.org/10.1021/acs.analchem.6b02585.

Calvete, J.J., Juarez, P., Sanz, L., 2007. Snake venomics. Strategy and applications. J. Mass Spectrom. 42 (11), 1405–1414. https://doi.org/10.1002/jms.1242.

Calvete, J.J., Pla, D., Els, J., Carranza, S., Damm, M., Hempel, B.F., Encinar, J.R., 2021. Combined molecular and elemental mass spectrometry approaches for absolute quantification of proteomes: application to the venomics characterization of the two species of desert black cobras, Walterinnesia aegyptia and Walterinnesia morgani. J. Proteome Res. 20 (11), 5064–5078. https://doi.org/10.1021/acs.jproteome.1c00608.

Carvalho, P.C., Fischer, J.S., Xu, T., Cociorva, D., Balbuena, T.S., Valente, R.H., Barbosa, V.C., 2012. Search engine processor: filtering and organizing peptide spectrum matches. Proteomics 12 (7), 944–949. https://doi.org/10.1002/pmic.201100529.

Galizio, N.D.C., Serino-Silva, C., Stuginski, D.R., Abreu, P.A.E., Sant'Anna, S.S., Grego, K. F., Morais-Zani, K., 2018. Compositional and functional investigation of individual and pooled venoms from long-term captive and recently wild-caught Bothrops jararaca snakes. J Proteomics 186, 56–70. https://doi.org/10.1016/j.jprot.2018.07.007.

Gay, C., Sanz, L., Calvete, J.J., Pla, D., 2015. Snake venomics and antivenomics of Bothrops diporus, a medically important pitviper in northeastern Argentina. Toxins 8 (1). https://doi.org/10.3390/toxins8010009.

Gocmen, B., Heiss, P., Petras, D., Nalbantsoy, A., Sussmuth, R.D., 2015. Mass spectrometry guided venom profiling and bioactivity screening of the Anatolian Meadow Viper, Vipera anatolica. Toxicon 107 (Pt B), 163–174. https://doi.org/10.1016/j.toxicon.2015.09.013.

Goncalves-Machado, L., Pla, D., Sanz, L., Jorge, R.J.B., Leitao-De-Araujo, M., Alves, M.L. M., Calvete, J.J., 2016. Combined venomics, venom gland transcriptomics, bioactivities, and antivenomics of two Bothrops jararaca populations from geographic isolated regions within the Brazilian Atlantic rainforest. J. Proteomics 135, 73–89. https://doi.org/10.1016/j.jprot.2015.04.029.

Gutierrez, J.M., Rucavado, A., Escalante, T., Lomonte, B., Angulo, Y., Fox, J.W., 2010. Tissue pathology induced by snake venoms: how to understand a complex pattern of alterations from a systems biology perspective? Toxicon 55 (1), 166–170. https://doi.org/10.1016/j.toxicon.2009.08.006.

Hiu, J.J., Yap, M.K.K., 2020. Cytotoxicity of snake venom enzymatic toxins: phospholipase A2 and l-amino acid oxidase. Biochem. Soc. Trans. 48 (2), 719–731. https://doi.org/10.1042/BST20200110.

Jorge, R.J., Monteiro, H.S., Goncalves-Machado, L., Guarnieri, M.C., Ximenes, R.M., Borges-Nojosa, D.M., Pla, D., 2015. Venomics and antivenomics of Bothrops erythromelas from five geographic populations within the Caatinga ecoregion of northeastern Brazil. J. Proteomics 114, 93–114. https://doi.org/10.1016/j.jprot.2014.11.011.

Kazandjian, T.D., Petras, D., Robinson, S.D., van Thiel, J., Greene, H.W., Arbuckle, K., Casewell, N.R., 2021. Convergent evolution of pain-inducing defensive venom components in spitting cobras. Science 371 (6527), 386–390. https://doi.org/10.1126/science.abb9303.

Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680–685.

Machado, T., Silva, V.X., Silva, M.J., 2014. Phylogenetic relationships within Bothrops neuwiedi group (Serpentes, Squamata): geographically highly-structured lineages, evidence of introgressive hybridization and Neogene/Quaternary diversification. Mol. Phylogenet. Evol. 71, 1–14. https://doi.org/10.1016/j.ympev.2013.10.003.

Mamede, C.C.N., de Sousa Simamoto, B.B., da Cunha Pereira, D.F., de Oliveira Costa, J., Ribeiro, M.S.M., de Oliveira, F., 2020. Edema, hyperalgesia and myonecrosis induced by Brazilian bothropic venoms: overview of the last decade. Toxicon 187, 10–18. https://doi.org/10.1016/j.toxicon.2020.08.016.

Markland Jr., F.S., Swenson, S., 2013. Snake venom metalloproteinases. Toxicon 62, 3–18. https://doi.org/10.1016/j.toxicon.2012.09.004.

Melani, R.D., Skinner, O.S., Fornelli, L., Domont, G.B., Compton, P.D., Kelleher, N.L., 2016. Mapping proteoforms and protein complexes from king cobra venom using both denaturing and native top-down proteomics. Mol. Cell Proteomics 15 (7), 2423–2434. https://doi.org/10.1074/mcp.M115.056523.

Melani, R.D., Nogueira, F.C.S., Domont, G.B., 2017. It is time for top-down venomics. J. Venom. Anim. Toxins Incl. Trop. Dis. 23, 44. https://doi.org/10.1186/s40409-017-0135-6.

Mora-Obando, D., Guerrero-Vargas, J.A., Prieto-Sanchez, R., Beltran, J., Rucavado, A., Sasa, M., Lomonte, B., 2014. Proteomic and functional profiling of the venom of Bothrops ayerbei from Cauca, Colombia, reveals striking interspecific variation with

Bothrops asper venom. J Proteomics 96, 159–172. https://doi.org/10.1016/j.jprot.2013.11.005.
Mora-Obando, D., Salazar-Valenzuela, D., Pla, D., Lomonte, B., Guerrero-Vargas, J.A., Ayerbe, S., Calvete, J.J., 2020. Venom variation in Bothrops asper lineages from North-Western south America. J Proteomics 229, 103945. https://doi.org/10.1016/j.jprot.2020.103945.
Nery, N.M., Luna, K.P., Fernandes, C.F., Zuliani, J.P., 2016. An overview of Bothrops erythromelas venom. Rev. Soc. Bras. Med. Trop. 49 (6), 680–686. https://doi.org/10.1590/0037-8682-0195-2016.
Ohler, M., Georgieva, D., Seifert, J., Von Bergen, M., Arni, R.K., Genov, N., Betzel, C., 2010. The venomics of Bothrops alternatus is a pool of acidic proteins with predominant hemorrhagic and coagulopathic activities. J. Proteome Res. 9, 2422–2437. https://doi.org/10.1021/pr901128x.
Petras, D., Heiss, P., Sussmuth, R.D., Calvete, J.J., 2015. Venom proteomics of Indonesian king cobra, ophiophagus hannah: integrating top-down and bottom-up approaches. J. Proteome Res. 14 (6), 2539–2556. https://doi.org/10.1021/acs.jproteome.5b00305.
Petras, D., Heiss, P., Harrison, R.A., Sussmuth, R.D., Calvete, J.J., 2016. Top-down venomics of the East African green mamba, Dendroaspis angusticeps, and the black mamba, Dendroaspis polylepis, highlight the complexity of their toxin arsenals. J. Proteom. 146, 148–164. https://doi.org/10.1016/j.jprot.2016.06.018.
Reid, D.J., Diesing, J.M., Miller, M.A., Perry, S.M., Wales, J.A., Montfort, W.R., Marty, M. T., 2018. MetaUniDec: high-throughput deconvolution of native mass spectra. J. Am. Soc. Mass Spectrom. 30 (1), 118–127. https://doi.org/10.1007/s13361-018-1951-9.
Ribeiro, P.H., Zuliani, J.P., Fernandes, C.F., Calderon, L.A., Stabeli, R.G., Nomizo, A., Soares, A.M., 2016. Mechanism of the cytotoxic effect of l-amino acid oxidase isolated from Bothrops alternatus snake venom. Int. J. Biol. Macromol. 92, 329–337. https://doi.org/10.1016/j.ijbiomac.2016.07.022.
Sajevic, T., Leonardi, A., Krizaj, I., 2011. Haemostatically active proteins in snake venoms. Toxicon 57 (5), 627–645. https://doi.org/10.1016/j.toxicon.2011.01.006.
Sanz, L., Quesada-Bernat, S., Perez, A., De Morais-Zani, K., Sant'Anna, S.S., Hatakeyama, D.M., Caldeira, C.A.S., 2020a. Danger in the canopy. Comparative proteomics and bioactivities of the venoms of the south American palm pit viper Bothrops bilineatus subspecies bilineatus and smaragdinus and antivenomics of B. B. bilineatus (Rondonia) venom against the Brazilian pentabothropic antivenom. J. Proteome Res. 19 (8), 3518–3532. https://doi.org/10.1021/acs.jproteome.0c00337.
Sanz, L., Perez, A., Quesada-Bernat, S., Diniz-Sousa, R., Calderon, L.A., Soares, A.M., Caldeira, C.A.S., 2020b. Venomics and antivenomics of the poorly studied Brazil's lancehead, Bothrops brazili (Hoge, 1954), from the Brazilian State of Para. J. Venom. Anim. Toxins Incl. Trop. Dis. 26, e20190103 https://doi.org/10.1590/1678-9199-JVATITD-2019-0103.
Segura, A., Herrera, M., Villalta, M., Vargas, M., Gutierrez, J.M., Leon, G., 2013. Assessment of snake antivenom purity by comparing physicochemical and immunochemical methods. Biologicals 41 (2), 93–97. https://doi.org/10.1016/j.biologicals.2012.11.001.
Tan, K.K., Bay, B.H., Gopalakrishnakone, P., 2018. L-amino acid oxidase from snake venom and its anticancer potential. Toxicon 144, 7–13. https://doi.org/10.1016/j.toxicon.2018.01.015.
Waheed, H., Moin, S.F., Choudhary, M.I., 2017. Snake venom: from deadly toxins to life-saving therapeutics. Curr. Med. Chem. 24 (17), 1874–1891. https://doi.org/10.2174/0929867324666170605091546.
Zhou, M., Lantz, C., Brown, K.A., Ge, Y., Pasa-Tolic, L., Loo, J.A., Lermyte, F., 2020. Higher-order structural characterisation of native proteins and complexes by top-down mass spectrometry. Chem. Sci. 11 (48), 12918–12936. https://doi.org/10.1039/d0sc04392c.