

Universidade Federal do Maranhão
Centro de Ciências Biológicas e da Saúde
Programa de Pós-Graduação em Ciências da Saúde
Mestrado

**PROSPECÇÃO GENÔMICA DE METABÓLITOS BIOATIVOS
PROVENIENTES DE MICROALGAS EXTREMÓFILAS:
APLICAÇÕES NA ÁREA DA SAÚDE**

PEDRO FELIPE RODRIGUES SOUSA

São Luís
2021

PEDRO FELIPE RODRIGUES SOUSA

**PROSPECÇÃO DE METABÓLITOS BIOATIVOS
PROVENIENTES DE MICROALGAS EXTREMÓFILAS:
APLICAÇÕES NA ÁREA DA SAÚDE**

Dissertação apresentada ao Programa de Pós-Graduação em Ciências da Saúde da Universidade Federal do Maranhão, como requisito para avaliação de Defesa de Mestrado.

Orientadora: Prof. Dr. Hivana Patrícia Melo Barbosa Dall’Agnol

São Luís

2021

PEDRO FELIPE RODRIGUES SOUSA

**PROSPECÇÃO DE METABÓLITOS BIOATIVOS
PROVENIENTES DE MICROALGAS EXTREMÓFILAS:
APLICAÇÕES NA ÁREA DA SAÚDE**

Dissertação apresentada ao Programa de Pós- Graduação em Ciências da Saúde da Universidade Federal do Maranhão, como requisito para obtenção do título de Mestre em Ciências da Saúde.

Aprovada em 18 / 01 / 2021

BANCA EXAMINADORA

Prof^ª. Dr^ª. Hivana Patrícia Melo Barbosa Dall’Agnol (Orientadora)
Universidade Federal do Maranhão

Rafael Cardoso Carvalho (Titular)
Universidade Federal do Maranhão

Evonnildo Costa Gonçalves (Titular)
Universidade Federal do Pará

Cláudia Quintino Rocha (Titular)
Universidade Federal do Maranhão

Alex Ranieri Jerônimo Lima (Suplente)
Instituto Butantan

“Water off a duck's back!”
Jinkx Monsoon

In memoriam
À Carina Silva Sousa e a Ricardo Sousa Pinheiro:
que as pessoas aprendam a respeitar a vida do outro assim
como vocês respeitavam.
No entanto, não respeitaram as suas.

AGRADECIMENTOS

À Deusa e ao Deus como Forças Criadoras do Universo por terem me dado a chance de encarnar em tal época ao lado de pessoas queridas.

À Prof^ª. Dr^ª Hivana Patrícia Melo Barbosa Dall’Agnol pela orientação, paciência, exortações e oportunidades durante este curto período do mestrado: muito obrigado.

A todos os membros do Grupo de pesquisa em Biodiversidade, Bioprospecção e Biotecnologia - GB³ - em especial a Ana Carolina de Araújo Butarelli e a Igor Santana Ribeiro por terem me ajudado na parte experimental e discussão sobre as análises *in silico*. Agradeço também a Milene Viana Brito e Iolanda Karoline Barros dos Santos Rocha por terem cuidado das minhas amostras enquanto passava por um período de enfermidade.

Aos meus pais, Vânia Rodrigues Barata Sousa e a Felipe Dário Campos Sousa, por todo o apoio e amor que me deram ao longo destes anos.

À também a minha irmã, Ana Luíza Rodrigues Sousa, por estar comigo nos momentos mais marcantes da minha vida e seus conselhos muito oportunos “*a tua cabeça é o teu guia*”.

À minha tia, Leila Rodrigues Barata, que me socorreu nos momentos os quais fiquei enfermo.

A meus amigos da turma de Alimentos do CEFET-MA de 2009 sem eles eu literalmente não estaria aqui.

A todos os que me apoiaram até hoje: professores, colegas dos meus empregos passados e pessoas que passaram na minha vida as quais me ajudaram nos meus momentos difíceis.

Aos meus amores do passado, do presente e do futuro: pois sem vocês eu não aprenderia a ser eu.

À Fundação de Amparo à Pesquisa do Estado do Maranhão, pela bolsa de estudos.

Ao Laboratório de Genética e Biologia Molecular – LabGeM.

Ao Programa de Pós-Graduação em Ciências da Saúde –PPGCS.

RESUMO

Microalgas são microrganismos eucariotos capazes de utilizar em seu metabolismo vias autotróficas e heterotróficas. Podem ser encontradas nos mais diversos tipos de ambientes aquáticos. Determinados tipos de ambientes possuem condições extremas as quais organismos se adaptam para sobreviver. Essas adaptações induzem especializações e ativação de vias metabólicas secundárias capazes de produzir compostos que permitem a sobrevivência de organismos nas mais variadas condições. Diversos destes metabólitos secundários possuem atividades contra microrganismos devido a competição do meio, antibiofilme, atividade antioxidante dentre outras. Deste modo, este trabalho buscou avaliar o potencial biotecnológico em saúde, usando técnicas *in silico* de prospecção, de duas microalgas verdes *Chlamydomonas acidophila* e *Jaagichlorella hainangensis*. Ambas as microalgas foram isoladas de ambientes extremófilos localizados em área de mineração: *Chlamydomonas acidophila* de uma Drenagem Ácida de Mina de Cobre e *Jaagichlorella hainangensis* de um lago natural de Ferro. Após as etapas de obtenção dos genomas, controle de qualidade e montagem dos mesmos foi realizada a anotação destes utilizando a plataforma GenSAS. Para predição de agrupamentos gênicos ligados a metabólitos secundários foi utilizada a plataforma antiSMASH nas versões bactéria e fungo. O antiSMASH apontou diversos clusters gênicos ligados a vias biossintéticas de metabólitos secundários de interesse em ambas as microalgas tais como, terpenos, sideróforos e ectoínas. O resultado destes clusters preditos foram analisados através de gráficos de sentença com os agrupamentos depositados no MIBiG e as possíveis funções dos genes foram preditas através de BLAST no CDD e no banco de dados PFAM. O antiSMASH também apontou clusters classificados como peptídeos sintase não-ribossômicos (NRPS) e policetídeo sintases (PKS) no genoma das duas microalgas e clusters híbridos NRPS/PKS somente em *Jaagichlorella hainangensis*. Os genes relacionados a PKS tipo I foram os de maior prevalência em *Jaagichlorella hainangensis*, seguidos dos genes relacionados ao tipo III. No entanto, este perfil não é encontrado em *Chlamydomonas acidophila* na qual somente os genes relacionados a PKS tipo III foram identificados. Análises mais robustas com os clusters NRPS e PKS preditos nos genomas destas microalgas foram realizadas utilizando a plataforma NaPDoS. Somente em *Jaagichlorella hainangensis* 2 clusters foram vinculados com baixa similaridade a vias biossintéticas, sendo um destes um agrupamento PKS similar a sequência de epitolola, um metabólito já conhecido por sua atividade anticâncer. A falta de resultados em *Chlamydomonas acidophila* e nos clusters remanescentes em *Jaagichlorella hainangensis* apontam que estas sequências podem estar relacionadas com vias não categorizadas ou não inseridas no NaPDoS e nestes casos estão ligadas a compostos de caráter inédito. Obter estes dados preliminares nestas espécies, direcionam possíveis experimentos *in vitro* os quais possam ser realizados com objetivo de extração de metabólitos secundários bem como elucidar vias metabólicas, além de contribuir com bancos de dados genômicos com os agrupamentos gênicos preditos para possíveis relações sintéticas e expansão do conhecimento genômico de microalgas.

Palavras-chave: Mineração Genômica; Fitoplâncton; *Chlamydomonas*; *Jaagichlorella*; Bioprospecção; Bioinformática

ABSTRACT

Microalgae are eukaryotic microorganisms capable of using autotrophic and heterotrophic pathways in their metabolism. These can be found in the most diverse types of aquatic environments. Certain types of environments have extreme conditions to which organisms adapt to survive. These adaptations induce specializations and activation of secondary metabolic pathways capable of producing compounds that allow organisms to survive in the most varied conditions. Several of these secondary metabolites have activities against microorganisms due to media competition, antibiofilm, antioxidant activity, among others. Thus, this work sought to evaluate the biotechnological potential in health, using *in silico* prospecting techniques, of two green microalgae *Chlamydomonas acidophila* and *Jaagichlorella hainangensis*. Both microalgae were isolated from extremophile environments located in a mining area: *Chlamydomonas acidophila* from an acidic copper mine drainage and *Jaagichlorella hainangensis* from a natural iron lake. After the steps of obtaining the genomes, quality control and assembly, they were annotated using the GenSAS platform. To predict gene clusters linked to secondary metabolites, the antiSMASH platform was used in the bacterium and fungus versions. AntiSMASH pointed to several gene clusters linked to biosynthetic pathways of secondary metabolites of interest in both microalgae such as terpenes, siderophores and ectoines. The results of these predicted clusters were analyzed using synthesis plots with the groupings deposited in the MIBiG and the possible functions of the genes were predicted through BLAST in the CDD and in the PFAM database. AntiSMASH also identified clusters classified as non-ribosomal synthase peptides (NRPS) and polyketide synthases (PKS) in the genome of the two microalgae and NRPS / PKS hybrid clusters only in *Jaagichlorella hainangensis*. Genes related to PKS type I were the most prevalent in *Jaagichlorella hainangensis*, followed by genes related to type III. However, this profile is not found in *Chlamydomonas acidophila* in which only genes related to PKS type III have been identified. More robust analyzes with the NRPS and PKS clusters predicted in the genomes of these microalgae were performed using the NaPDoS platform. Only in *Jaagichlorella hainangensis* 2 clusters were linked with low similarity to biosynthetic pathways, one of which is a PKS cluster similar to the epothilone sequence, a metabolite already known for its anticancer activity. The lack of results in *Chlamydomonas acidophila* and in the remaining clusters in *Jaagichlorella hainangensis* indicate that these sequences may be related to pathways that are not categorized or not inserted in NaPDoS and in these cases are linked to compounds of unprecedented character. Obtaining these preliminary data in these species, direct possible *in vitro* experiments which can be carried out with the objective of extracting secondary metabolites as well as elucidating metabolic pathways, in addition to contributing to genomic databases with the predicted gene clusters for possible syntenic relationships and expansion of the genomic knowledge of microalgae.

Keywords: Genome Mining; Phytoplankton; *Chlamydomonas*; *Jaagichlorella*; Bioprospection; Bioinformatics

LISTA DE SIGLAS E ABREVIATURAS

antiSMASH	Antibiotics & Secondary Metabolite Analysis Shell
BLAST	Basic Local Alignment Search Tool
BGCs	Biosynthetic Gene Clusters
<i>C. reinhardtii</i>	<i>Chlamydomonas reinhardtii</i>
<i>C. acidophila</i>	<i>Chlamydomonas acidophila</i>
CDD	Conserved Domain Database
<i>Cluster</i>	Agrupamento
ITV01	Linhagem do Instituto Tecnológico Vale 01 (<i>Chlamydomonas acidophila</i>)
ITV09	Linhagem do Instituto Tecnológico Vale 09 (<i>Jaagichlorella hainangensis</i>)
<i>J. hainangensis</i>	<i>Jaagichlorella hainangensis</i>
MIBiG	Minimum Information about a Biosynthetic Gene cluster
NaPDoS	Natural Product Domain Seeker
NRPS	Peptídeos Sintetases Não-Ribossômicos
Pfam	Protein Family database
PKS	Policetídeos Sintases

LISTA DE FIGURAS

	Página
Figura 1	Resultados da busca por microalgas e aplicações biotecnológicas nos últimos 22 anos. 13
Figura 2	Ilustração de um BGC em uma região em particular de um cromossomo, os diferentes genes que o compõem (codificantes de produtos enzimáticos ou não), e o metabólito secundário (produto natural) derivado do mesmo. Adaptado de Chavali e Rhee (2017). 19
Figura 3	Microscopia das microalgas: A. <i>Chlamydomonas acidophila</i> ITV01. B. <i>Jaagichlorella hainangensis</i> ITV09 23
Figura 4	<i>Workflow</i> adaptado de Adamek <i>et al.</i> (2016) com o sugerido por O'Neill (2020) para as análises nos genomas das microalgas <i>Chlamydomonas acidophila</i> ITV01 e <i>Jaagichlorella hainangensis</i> ITV09. 26

SUMÁRIO

1. INTRODUÇÃO.....	11
2 Referencial Teórico	12
2.1 Microalgas e suas características.....	12
2.1.1 Gênero <i>Chlamydomonas</i>	14
2.1.1.1 <i>Chlamydomonas acidophila</i>	16
2.1.2 <i>Jaagichlorella hainangensis</i>	17
2.2 Bioprospecção: O metabolismo secundário como fonte de produtos naturais.....	17
2.2.1 Bioprospecção baseada em Genômica – <i>In silico</i>	18
2.2.2 <i>Softwares</i> de Bioprospecção <i>in silico</i>	21
3. Objetivos.....	22
3.1 Objetivo Geral	22
3.2 Objetivos Específicos	22
4. Materiais e Métodos	23
4.1 Linhagens Microbianas e Sequenciamento Genômico.....	23
4.2 Pré-processamento de sequência e montagem do genoma de <i>C acidophila</i>	24
4.3 Anotação dos genomas	24
4.4 Bioprospecção de BGCs relacionados a metabólitos secundários	24
5. Resultados (artigo).....	27
5.1 Prospecting of bioactive metabolites from extremophilic microalgae: an in silico approach. 27	
6. Conclusão	53
7. Referências Bibliográficas.....	54
Anexos	64

1. INTRODUÇÃO

As microalgas pertencem a um grupo extremamente diversificado de microrganismos que se adaptaram com êxito a inúmeros tipos de ambientes aquáticos sendo eles de água doce, salobra ou salgada. Devido a sua rica biodiversidade, este grupo possui diversas características funcionais, biológicas e papéis ecológicos. Estas possuem como características o tamanho microscópico, tipo celular eucariótico, pigmentação fotossintetizante, podendo assim utilizar vias autotróficas ou heterotróficas (BARRA *et al.*, 2014).

Pela perspectiva humana um ambiente em estado habitável possui condições normais de temperatura, pH neutro, pressão de 1atm dentre outros fatores, assim, ecossistemas que fogem a estas condições são considerados “ambientes extremos”, e os organismos que neles habitam “extremófilos”. A pesquisa por novas espécies que possuam a capacidade de sobreviver a estes ambientes vem sendo constantemente impulsionada, pois, a habilidade de sobrevivência aos mesmos está associada a qualidades especiais. Alguns exemplos são novas vias bioquímicas, enzimas e compostos com bioatividade com potencial uso nos diversos setores biotecnológicos da indústria, meio ambiente e saúde (RADDADI *et al.*, 2015 e SINGH *et al.*, 2019). Neste contexto, microalgas extremófilas seriam de grande interesse para bioprospecção.

Este estudo está vinculado a um projeto de origem no qual, em etapas anteriores, foram isoladas e identificadas duas linhagens de microalgas extremófilas do filo *Chlorophyta* (*Chlamydomonas acidophila* e *Jaagichlorella hainangensis*), isoladas de minas em funcionamento no norte do Brasil (SANTOS, 2017). Ambientes de mineração são extremos e geram elementos seletivos, não somente para características de resistência a metais pesados, como também a habilidade para produção de outros compostos de interesse que podem ser aplicados para o uso biotecnológico. Os estudos sobre esses organismos e o potencial biotecnológico dos mesmos ainda são escassos (BRIERLEY; BRIERLEY, 2013; SOUZA *et al.*, 2017).

Assim, este trabalho propõe-se a investigação do potencial biotecnológico destas duas espécies de microalgas extremófilas para a produção de metabólitos secundários (também chamados produtos naturais) com bioatividade na área da saúde. Utilizando metodologias de bioprospecção *in silico* baseadas na análise dos seus genomas com ferramentas de bioinformática, abordagem a qual auxilia na busca de metabólitos secundários e facilita processos biotecnológicos para a obtenção de novos compostos bioativos.

2 Referencial Teórico

2.1 Microalgas e suas características

As microalgas do filo *Chlorophyta*, apresentam coloração verde devido às clorofilas tipo A e B. Estão distribuídas tanto em ambientes marinhos quanto em água doce. Possuem ampla diversidade de características morfológicas: cloroplasto com 2 membranas, armazenamento de energia em forma de amido e alguns membros deste filo podem possuir flagelos. Atualmente o filo *Chlorophyta* está dividido em 11 classes com mais de 6.772 espécies: *Chlorodendrophyceae*, *Chloropicophyceae*, *Mamiellophyceae*, *Nephroselmidophyceae*, *Pedinophyceae*, *Picocystophyceae*, *Prasinophyceae*, *Pyramimonadophyceae*, *Ulvophyceae*, *Chlorophyceae* e *Trebouxiophyceae*. Sendo as 3 últimas classes consideradas o núcleo do filo (FUČÍKOVÁ et al., 2014; GUIRY, M.D. & GUIRY, 2020; PAL; CHOUDHURY, 2014).

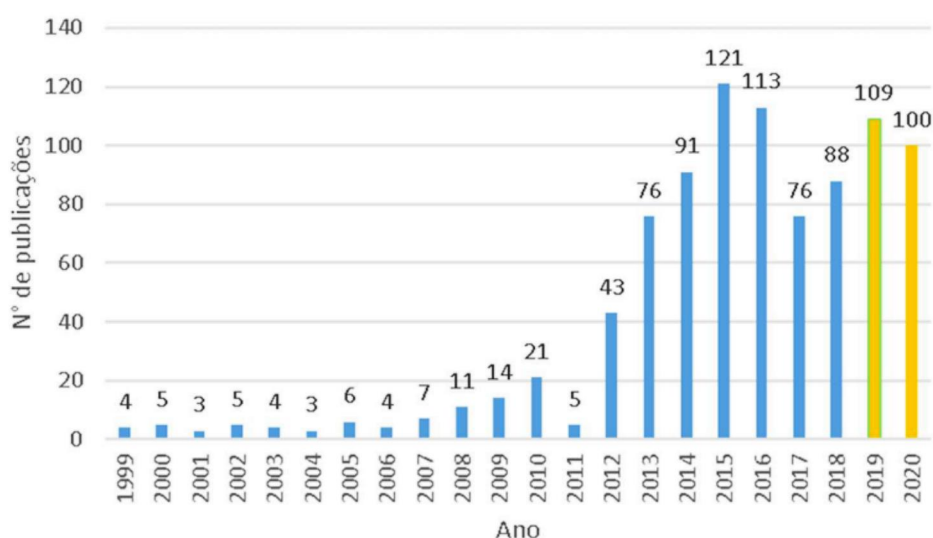
Como organismos eucariotos possuidores de plastídios, algumas microalgas possuem duas vias optativas, autotrófica e heterotrófica, de acordo com o habitat que ocupam. A via autotrófica fotossintetizante utiliza luz solar, nutrientes inorgânicos e dióxido de carbono (CO₂) para a síntese de proteínas, lipídios e carboidratos. Em condições de ausência de luz, as microalgas utilizam a via heterotrófica, realizando a fermentação de substratos orgânicos simples tais como: glicerol, glicose e acetato, como fontes de carbono para o crescimento celular (PEREZ-GARCIA et al., 2011).

As microalgas são produtoras de diversos compostos de interesse comercial com características químicas diversas. Carotenoides, vitaminas, proteínas, carboidratos, lipídeos e ácidos graxos que podem ser utilizados na produção de biodiesel e bioetanol. Além disso, as microalgas são produtoras de substâncias contendo propriedades antioxidantes, antitumorígenicas, imunomoduladoras dentre outras (DOLGANYUK et al., 2020). Outros compostos biologicamente ativos produzidos pelas microalgas possuem função de defesa contra diferentes organismos encontrados no meio em que habitam naturalmente, podendo ter efeitos antimicrobiano, antifúngico e antibiofilme. Estes pertencem a diversos grupos, dentre eles alcaloides, terpenos, lactonas, pirroles e esteroides. Tais metabólitos secundários possuem potencial para produção comercial em larga escala com a utilização de técnicas de maximização metabólica (QIAN; XU; FUSETANI, 2009).

Para o uso biotecnológico, as microalgas apresentam notáveis vantagens quando em comparação com os demais sistemas de produção de biocompostos de alto valor, ou mesmo de

proteínas recombinantes. Microalgas podem ser massivamente cultivadas sobre condições controladas em fotobiorreatores, com relativamente pequenas quantidades de micro e macronutrientes. As microalgas possuem diversas vantagens sobre o modelo de expressão bacteriano com as mais importantes sendo as modificações pós-traducionais que geram proteínas glicosiladas. Em comparação com os modelos de plantas geneticamente modificadas, as microalgas quando desenvolvidas em sistemas fechados, impedem o escape de transformantes para o ambiente. Microalgas que sintetizam compostos bioativos com potencial terapêutico podem ser administradas via oral ou pó liofilizado devido a diversas espécies pertencerem a categoria GRAS (Geralmente Reconhecido como Segura), do inglês *Generally Recognized As Safe* (ALVES, 2007; GONG et al., 2011; SCRANTON et al., 2015).

Vingiani et al. (2019) revisaram a literatura na base de dados PubMed, esta revisão demonstrou o crescente aumento em publicações na temática microalgas e aplicações biotecnológicas (Figura 1).



Pesquisa no bando de dados PubMed foi realizada com as palavras “microalgae” e “biotechnolog*”, com uso do asterístico (*) ampliando a possibilidade de termos (como “biotechnology”, “biotechnological”, e “biotechnologies”). Adaptado de Vingiani et al. (2019). Os anos de 2019 e até a data de 23 de novembro de 2020 foram acrescentados no gráfico adaptado neste trabalho.

Figura 1. Resultados da busca por microalgas e aplicações biotecnológicas nos últimos 22 anos.

Vingiani e colaboradores, 2019, observaram que as publicações dominavam quatro

principais setores: (1) Uso direto de microalgas para biorremediação e em suplementos alimentares; **(2) Extração de bioativos das microalgas para diferentes aplicações (Exemplo; em cosmecêuticos, nutracêuticos, e aplicações farmacêuticas,** e para produção de biocombustíveis); (3) Uso das microalgas como plataforma para expressão heteróloga, ou edição gênica, e superexpressão; **(4) Uso das microalgas como fonte de enzimas para aplicação industrial,** este último setor parece ser o menos estudado, e um dos limitantes seria a escassez de genomas de microalgas anotados disponíveis. A metodologia adotada na presente dissertação contempla dois dos citados setores de interesse biotecnológico com foco na área da saúde, em destaque **(2)** e **(4)**, avaliando os genomas de duas microalgas, de dois diferentes gêneros descritos a seguir.

2.1.1 Gênero *Chlamydomonas*

O gênero de microalgas *Chlamydomonas*, pertencente a classe *Chlorophyceae*, possui mais de 600 espécies das quais diversos membros apresentam rico potencial biotecnológico, no entanto apenas 12 genomas estão depositados nos bancos de dados de DNA GenBank (BENSON; LIPMAN; OSTELL, 1993), alguns exemplos estão dispostos na Tabela 1. Possuem como características morfológicas coloração verde, parede celular, 2 flagelos de igual tamanho geralmente próximos um do outro e um único cloroplasto provido de um microcompartimento singular denominado pirenoide. Apesar de diversas espécies isoladas, apenas algumas desempenharam papéis importantes como organismos de pesquisa. Dentre as quais se destaca como organismo modelo a *Chlorophyta Chlamydomonas reinhardtii* (*C. reinhardtii*), majoritariamente devido à sua capacidade de sintetizar acetato como única fonte de carbono de forma heterotrófica (HARRIS, 2001; PRÖSCHOLD et al., 2001).

No âmbito da investigação para uso clínico, o extrato metanólico de *C. reinhardtii* (Tabela 1) demonstrou inibir os fatores *Quorum Sensing*, hidrofobicidade da superfície celular, substâncias poliméricas extracelulares responsáveis pela formação de biofilme e piocianina ligadas a patogenicidade de *Pseudomonas aeruginosa* (NITHYA et al., 2014). O mesmo extrato demonstrou atividade antifúngica em *Aspergillus niger* e *Aspergillus fumigatus* (BAJPAI, 2016) e antibacteriana em *Staphylococcus aureus*, *Staphylococcus epidermidis*, *Bacillus subtilis*, *Escherichia coli* e *Salmonella typhi* (JAYSHREE; JAYASHREE; THANGARAJU, 2016). A microalga *C. reinhardtii* é a mais conhecida e a mais estudada do gênero *Chlamydomonas*. Possui diversos usos biotecnológicos como um organismo utilizado em modelos de expressão

heteróloga, e é utilizada como fornecedora de biomassa para produção de etanol (MAYFIELD et al., 2007; SCRANTON et al., 2015).

Tabela 1. Relevância biotecnológica já descrita na literatura no gênero *Chlamydomonas*.

Espécies com atividades descritas	Referências
<i>Chlamydomonas reinhardtii</i>	
Atividade antibacteriana em <i>Pseudomonas aeruginosa</i>	NITHYA et al., 2014
Atividade antifúngica em <i>Aspergillus spp.</i>	BAJPAI, 2016
Atividade antibacteriana em <i>Staphylococcus aureus</i> , <i>Staphylococcus epidermidis</i> , <i>Bacillus subtilis</i> , <i>Escherichia coli</i> e <i>Salmonella typhi</i> .	GHA SEMI et al., 2007; JAYSHREE; JAYASHREE; THANGARAJU, 2016
Atividade antioxidante	RENUKADEVI; SARAVANA; ANGAYARKANNI, 2011
Atividade anti α -sinucleína	CHOUDHARY; SAVE; VAVILALA, 2018
<i>Chlamydomonas debaryana</i>	DE LOS REYES et al., 2014
Atividade anti-inflamatória	ÁVILA-ROMÁN et al., 2016
<i>Chlamydomonas hedleyi</i>	
Atividade antioxidante	CHOI et al., 2015; SUH et al., 2014
Atividade anti-inflamatória	
<i>Chlamydomonas spp.</i>	
Atividade antibacteriana em <i>Staphylococcus aureus</i>	SENHORINHO; LANNÉR; SCOTT, 2019
Atividade antitumoral	
<i>Chlamydomonas acidophila</i>	
Atividade antioxidante	CUARESMA et al., 2011
Resistencia a metais pesados	OLSSON et al., 2015
<i>Chlamydomonas nivalis</i>	
Atividade antioxidante	ŘEZANKA et al., 2013
<i>Chlamydomonas moewusii</i>	
Biorremediação	OTTO et al., 2015

Outros integrantes do gênero também apresentaram compostos bioativos peculiares tal como *Chlamydomonas hedleyi* (Tabela 1) na qual foram encontrados aminoácidos do tipo micosporina. Estes são compostos que possuem ação antioxidante capazes de absorver e neutralizar a ação dos raios Ultravioleta que geram radicais livres. Além de demonstrar atividade anti-inflamatória em ensaios com células fibroblásticas linhagem *HaCat* e redução dos níveis de expressão do gene COX-2 que está relacionado com respostas inflamatórias (SUH et al., 2014).

Chlamydomonas debaryana, por sua vez, possui o mesmo potencial anti-inflamatório através da atividade de oxilipinas que atuam como inibidoras do TNF- α . Oxilipinas são metabólitos originados da oxidação de ácidos graxos poli-insaturados. Este grupo de compostos possui inúmeros membros e estão envolvidos na regulação de processos biológicos. Além disso, são sintetizados em resposta a estresses ambientais e a agentes patológicos (DE LOS REYES et al., 2014; MOSBLECH; FEUSSNER; HEILMANN, 2009).

2.1.1.1 *Chlamydomonas acidophila*

A microalga Chlorophyta *Chlamydomonas acidophila* (*C. acidophila*) é uma espécie encontrada em ambientes extremófilos, primeiramente descrita e isolada por Negoro (1944). Esta microalga é frequentemente a espécie dominante do fitoplâncton em lagos de ambientes de mineração e lagos vulcânicos, ambos ambientes ácidos e normalmente com altas concentrações de metais (DEAN et al., 2019; SANTOS, 2017). Análises transcriptômicas permitiram a identificação dos agrupamentos gênicos expressos quando uma linhagem de *C. acidophila* é exposta a altas concentrações de cobre. Estes atributos demonstram o seu potencial uso para ações de biorremediação em ambientes com contaminação, bem como outras potenciais aplicações biotecnológicas (OLSSON et al., 2015).

Quanto a prospecção de metabólitos, Cuaresma et al. (2011) isolaram uma linhagem de *C. acidophila* de um rio próximo a área de mina. Experimentos foram realizados envolvendo a utilização de fontes diferentes de carbono para explorar o potencial da síntese de carotenoides, compostos com conhecida atividade antioxidante. Os resultados destas análises demonstraram a relação da produção de Luteína, β -caroteno, Zeaxantina e Violaxantina com diferentes concentrações de glucose, amido, glicerol, glicina e ureia.

Experimentos de Senhorinho et al. (2019) com o extrato de um membro do gênero *Chlamydomonas spp.*, também isolado de ambiente de mineração, observaram atividade

antimicrobiana contra uma cepa de *Staphylococcus aureus* e atividade antitumoral em linhagens de células cancerosas sem afetar ou manifestar citotoxicidade em linhagens de células não cancerosas.

2.1.2 *Jaagichlorella hainangensis*

A microalga *Jaagichlorella hainangensis* (*J. hainangensis*), pertencente a classe *Trebouxiophyceae*, primeiramente descrita e isolada a partir da seringueira *Hevea brasiliensis*, no leste asiático por Zhang et al. (2008) como *Heveochlorella hainangensis* e posteriormente reclassificada no gênero *Jaagichlorella* por Darienko e Pröschold (2019). São descritas como organismos unicelulares, de formato esférico, que se reproduzem por autoesporulação e possuem cloroplastos providos de pirenoide. Um estudo sobre espécies deste gênero, isoladas do tronco da palmeira real *Roystonea regia* na China, indicaram que *J. hainangensis* necessita de carbono livre para um crescimento ideal (MA et al., 2013).

Atualmente esta microalga é descrita como de ampla distribuição, em regiões tropicais e subtropicais, podem ser encontradas de forma livre ou em endofitismo. Embora alguns estudos relatem características de cultivo, como as taxas de crescimento (ótimo em pH 7 e lento em pH abaixo de 3), fatores de distribuição, ecologia e metabolismo desta espécie, atualmente não existem trabalhos que explorem o potencial biotecnológico deste organismo (MA et al., 2013; SANDERS et al., 2016).

2.2 Bioprospecção: O metabolismo secundário como fonte de produtos naturais

A relação entre os seres humanos e os produtos de origem natural é milenar. Diversos medicamentos e produtos biologicamente ativos foram desenvolvidos a partir de plantas, animais e microrganismos. Mesmo com diversos fármacos sintéticos (que possuem base produtos de origem biológica), é notório que a natureza permanece como uma rica fonte de novos compostos com estruturas singulares e novas classes químicas (NEWMAN; CRAGG, 2020). Deste modo, a exploração da natureza e sua biodiversidade com o intuito de encontrar compostos com atividades de interesse é denominada Bioprospecção. Historicamente, os resultados da Bioprospecção são utilizados em indústrias de cunho farmacêutico, cosmético e agricultura. Contemporaneamente, as abordagens relacionadas a aquicultura, biorremediação, engenharias biossintéticas e miméticas, nanotecnologia e biomineração de genomas receberam destaque devido aos avanços relacionados a estas áreas (BEATTIE et al., 2011).

Deste modo, a bioprospecção revelou a importância de diversos organismos, dos mais variados habitats, como fontes de biocompostos, genes e vias metabólicas para a indústria e economia global. Dentre os quais se destacam os microrganismos produtores de metabólitos secundários. Estes metabólitos são um grupo de moléculas das mais variadas classes químicas. Diferentemente dos metabólitos primários que são necessários para a sobrevivência do organismo, estes metabólitos não são essenciais para o mesmo, entretanto, possuem funções singulares como defesa, sinalização intra e entre espécies, simbiose dentre outras (BEATTIE et al., 2011; THIRUMURUGAN et al., 2018).

Uma das metodologias de bioprospecção *in vitro* que mais contribuiu para a descoberta de novos antibióticos é nomeada como plataforma Waksman, desenvolvida pelo pesquisador Selman Waksman, na qual se baseia no cultivo de microrganismos provenientes do solo cultivados em variados meios. Após o cultivo, diversos testes de atividade são realizados para identificação de metabólitos com capacidade antimicrobiana. Esta padronização da prospecção de produtos naturais revolucionou a forma de como buscar compostos de interesse, desde então as pesquisas *in vitro* baseiam-se em variações desta metodologia. Este modelo rendeu a Waksman um Nobel em Medicina através da descoberta da estreptomicina, revisado em LEWIS, (2013).

No entanto, mesmo com a enorme contribuição que as metodologias *in vitro* demonstraram, as mesmas possuem desvantagens na prospecção dos metabólitos com atividade de interesse: diversos desses compostos são encontrados em baixas concentrações e com frequência os microrganismos produtores deixam de sintetizar estas moléculas devido as diversas condições laboratoriais. Sendo assim, diversas destas metodologias acabam sendo laboriosas e de elevado custo (WRIGHT, 2014).

2.2.1 Bioprospecção baseada em Genômica – *In silico*

A análise de genomas microbianos tem revelado que muitos metabólitos secundários, tem a maquinaria biossintética responsável pela sua produção codificada por um grupo de genes localizados fisicamente próximos uns aos outros no cromossomo, sendo assim conhecidos como conjuntos ou agrupamentos gênicos biossintéticos (do inglês *Biosynthetic gene clusters*, BGCs) (Figura 2).

O fenômeno de “*clusters*” (agrupamentos) de genes é comum em bactérias, onde até 50% de todos os genes costumam estar organizados em estruturas de operons sobre o controle de um

mesmo promotor, por outro lado, em eucariotos, genes organizados em *clusters* estariam similarmente próximos fisicamente, mas com promotores independentes. Assim, alguns benefícios deste fenômeno em eucariotos seriam; regulações co-transcricionais ou pós-transcricionais, como, por exemplo, relacionadas à síntese, compartimentalização e exportação de proteínas (CHAVALI; RHEE, 2018).

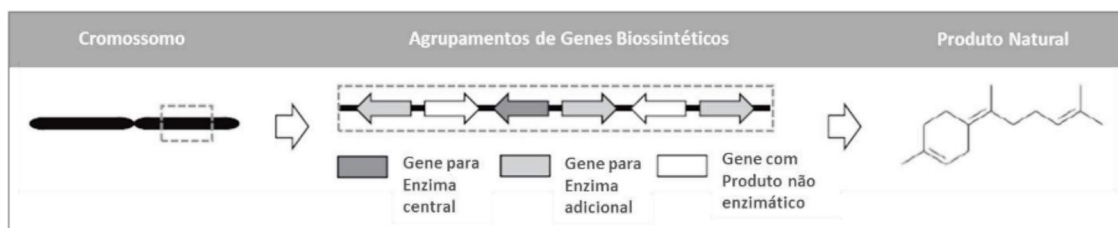


Figura 2. Ilustração de um BGC em uma região em particular de um cromossomo, os diferentes genes que o compõem (codificantes de produtos enzimáticos ou não), e o metabólito secundário (produto natural) derivado do mesmo. Adaptado de Chavali e Rhee (2017).

O uso de ferramentas computacionais, o avanço das ciências ômicas alinhados a queda do custo do sequenciamento de DNA permitiram a mineração do genoma de várias espécies de microalgas, o que revelou uma série de novos conjuntos ou agrupamentos de BGCs. Demonstrando um potencial destes microrganismos para a produção de compostos com aplicabilidade tanto para áreas industriais quanto para as biomédicas. As recentes aplicações da genômica comparativa possibilitam o exame de variações de genes e funcionalidades através das comparações de BGCs intra e entre espécies, fornecendo assim uma visão geral da capacidade dos mesmos (GUARNIERI; PIENKOS, 2015).

Dentre alguns exemplos de metabólitos secundários os quais possuem BGCs já descritos na literatura, com relevância na área da saúde, destacam-se:

Terpenos, ou isoprenóides, são metabólitos secundários que possuem como fórmula geral $(C_5H_8)_n$. Esta classe de compostos é a maior família de metabólitos secundários conhecida. Carotenóides são pigmentos com propriedades antioxidantes formados por 8 unidades básicas de isoprenóides. Estes por sua vez são divididos em dois grupos de acordo com a presença ou não de oxigênio em sua composição: xantofilas, os quais possuem oxigênio, e carotenos, que não possuem oxigênio. Exemplos de carotenos estão o β -caroteno e licopeno. Dentre as xantofilas destacam-se a zeaxantina e a astaxantina, esta última destaca-se por ser um pigmento avermelhado descrito em literatura como um dos mais potentes antioxidantes relatados. Quando comparado com outros

antioxidantes a astaxantina é 10 vezes mais potente que o β -caroteno e zeaxantina e 65 vezes mais potente que a vitamina C (NOVOVESKÁ et al., 2019; SHAH et al., 2016).

O ferro é necessário para diversos processos essenciais no metabolismo de diversos seres vivos. Mesmo sendo abundante no ambiente, o mesmo não se encontra biodisponível para o uso, este fato gerou uma pressão evolutiva a qual diversos organismos desenvolveram mecanismos para a captação de ferro. Sideróforos são os metabólitos secundários sintetizados por diversos organismos para coletar ferro do ambiente e permitir o seu uso (HIDER; KONG, 2010). A comunidade científica tem grandes expectativas para o uso terapêutico de sideróforos, com diversos usos descritos em literatura tais como o uso em entrega seletiva de drogas na estratégia cavalo-de-Tróia (GÓRSKA; SLODERBACH; MARSZAŁŁ, 2014).

Ectoína é um metabólito secundário de organismos extremófilos pertencente a um grupo denominado solutos compatíveis. Esta classe de compostos tem como principal função o controle do estresse osmótico o qual estes organismos são expostos por seu habitat e também está relacionada a proteção de proteínas impedindo a sua desnaturação (CZECH et al., 2018). Devido as suas atividades biológicas, ectoínas tem demonstrado um potencial uso em diversas áreas relacionadas a patologias por defeitos no enovelamento de proteínas e outras áreas da saúde em geral. A atividade deste metabólito na estabilização de proteínas e resultados positivos nos modelos de doenças neurodegenerativas é promissora (JORGE et al., 2016). Estudo realizado por Werkhäuser e colaboradores (2014) exemplifica outra aplicação deste metabólito, este grupo demonstrou que a adição de ectoína em spray nasal e a um colírio diminuíram os sintomas nasais e oculares, respectivamente, de pacientes com rinite alérgica sem apresentar efeitos colaterais.

Dentre os diversos BGCs, 2 tipos de grupos gênicos destacam-se como indicativos de produção de moléculas de interesse: Peptídeos Sintetases Não-Ribossômicos (NRPS) e Policetídeos Sintases (PKS). Essas enzimas podem ser essenciais para a formação metabólitos secundários tais como antibióticos, antifúngicos e antitumorais (WANG et al., 2014). Os PKS formam cadeias de policetídeos através da oligomerização de pequenos ácidos carboxílicos e os NRPS utilizam aminoácidos para formar ligação amida ou éster (WANG et al., 2015). Um exemplo de produto mediado por enzima do tipo PKS é a eritromicina um conhecido antibiótico já utilizado em ampla escala no uso clínico (SHELEST et al., 2015).

Chlamydomonas reinhardtii como microalga modelo teve seu genoma nuclear finalizado

em 2007, desde então diversas técnicas pós-genômicas e genéticas foram utilizadas nesta microalga. As demais espécies de microalgas que também já tiveram genomas sequenciados foram escolhidas graças a sua distribuição filogenética, papéis ecológicos e potencial nocivo. Os dados gerados por estes sequenciamentos fornecem referências para análises pós-genômicas, tais como estudos transcriptômicos, proteômicos, metabolômicos, bem como para a bioprospecção *in silico* (CADORET; GARNIER; SAINT-JEAN, 2012).

2.2.2 Softwares de Bioprospecção *in silico*

O avanço no sequenciamento do genoma de inúmeras espécies microbianas e a disponibilidade destas em bancos de dados possibilitou a busca de agrupamentos gênicos e a mineração de genomas *in silico*. Uma das ferramentas para a busca de metabólitos secundários mais popularizada é conhecida pela sigla “antiSMASH” (*Antibiotics & Secondary Metabolite Analysis Shell*) que permite a busca de produtos naturais provenientes de diversos organismos. Esta plataforma de mineração de genomas, microbianos e de plantas, utiliza uma segmentação de instruções (*pipeline*), baseada no modelo oculto de Markov, para identificar 45 tipos de vias biossintéticas já descritas na literatura que geram metabólitos secundários (BLIN et al., 2019a; MEDEMA et al., 2011). Diversos estudos envolvendo a plataforma antiSMASH foram utilizados para evidenciar o potencial de organismos e a mineração genômica de metabólitos secundários. Com o auxílio desta ferramenta Shelest et al. (2015) prospectaram agrupamentos gênicos do tipo PKS em 15 diferentes espécies de microalgas clorofitas.

O modelo oculto de Markov é um modelo estatístico de probabilidade de eventos finitos acontecerem de acordo com uma base de dados pré-fornecidos. Este modelo probabilístico é utilizado em diversas plataformas de bioinformática com diversos objetivos. Destacam-se o alinhamento de sequências para avaliação de homologia, desenho de mapas gênicos com distâncias entre os diferentes *loci* e análises filogenéticas através de sequências biológicas (GOLLERY, 2008).

A plataforma online NaPDoS (Natural Product Domain Seeker) realiza análises de sequências de genoma baseadas em filogenia e no modelo oculto de Markov para busca de produtos naturais. Esta plataforma tem como foco a busca de domínios PKS e NRPS. O NaPDoS é otimizado para a utilização em procariotos, no entanto devido a história evolutiva, existe a possibilidade da identificação e classificação de genes em eucariotos (ZIEMERT et al., 2012).

3. Objetivos

3.1 Objetivo Geral:

Avaliar o potencial biotecnológico para a área da saúde das linhagens das microalgas extremófilas *Chlamydomonas acidophila* ITV01 e *Jaagichlorella hainangensis* ITV09 analisando seus aspectos genômicos.

3.2 Objetivos Específicos:

- I. Aplicar filtros de qualidade e montar o genoma de *Chlamydomonas acidophila* ITV01
- II. Anotar funcionalmente os genomas das microalgas *Chlamydomonas acidophila* e *Jaagichlorella hainangensis*
- III. Avaliar os agrupamentos gênicos das microalgas envolvidos na síntese de compostos bioativos de interesse biotecnológico para a área da saúde.

4. Materiais e Métodos

4.1 Linhagens Microbianas e Sequenciamento Genômico

Foram utilizadas as linhagens ITV01 (*Chlamydomonas acidophila*) e ITV09 (*Jaagichlorella hainangensis*) isoladas a partir de amostra de Drenagem Ácida de Mina de cobre (Coordenadas: 6°26'42.3"S 50°04'36.9"W), e de um lago natural de uma mina de ferro (Coordenadas: 6°24'02.5"S 50°21'06.7"W), respectivamente (SANTOS, 2017). Ambientes os quais apresentam elevada concentração de metais e são ácidos (pH < 3); ambos com atividade de mineração estão situados na região de Carajás, estado do Pará, Brasil. Estas linhagens foram cedidas da Coleção de Microalgas e Cianobactérias do Instituto Tecnológico Vale (ITV), localizado na capital Belém – Pará.

Ambas as espécies de microalgas foram identificadas morfológicamente e molecularmente, através de amplificação e sequenciamento de genes código de barra, como *rbcl*, *16S rRNA* e *18S rRNA*.

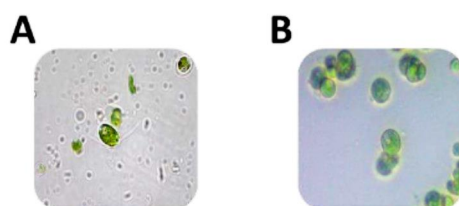


Figura 3. Microscopia das microalgas: A. *Chlamydomonas acidophila* ITV01. B. *Jaagichlorella hainangensis* ITV09

A sequência de DNA genômico da linhagem de *Chlamydomonas acidophila* (ITV01) e *Jaagichlorella hainangensis* (ITV09) foram obtidas por sequenciamento de DNA de nova geração, realizado pela equipe do Instituto Tecnológico Vale Desenvolvimento Sustentável (ITV-DS), no qual foi utilizada a plataforma Illumina/MiSeq (Illumina, Inc., EUA) de acordo com os protocolos descritos pelos fabricantes.

Os genomas das microalgas foram cedidos em arquivos contendo sequências de DNA para as análises bioinformáticas. Para a microalga *Jaagichlorella hainangensis* (ITV09), o genoma foi previamente montado por Santos (2017), sendo cedida a sequência do mesmo em formato FASTA, para anotação e mineração no presente trabalho. Para a microalga *Chlamydomonas acidophila* (ITV01), o dado bruto proveniente do sequenciador de DNA Illumina, arquivo em formato

FASTAQ, foi cedido para montagem, anotação e mineração para clusters de produtos naturais neste trabalho.

4.2 Pré-processamento de sequência e montagem do genoma de *C acidophila*

Antes de iniciar a montagem do genoma de ITV01, a qualidade do sequenciamento foi primeiramente visualizada através da ferramenta FASTQC (ANDREWS *et al.*, 2010), para remoção das leituras de baixa qualidade seguindo a escala Phred (EWING; GREEN, 1998) (*phread* 20) e das bases das extremidades dos reads, permanência mínima de 80% das reads com a qualidade selecionada e *reads* com tamanho de 420pb - foi utilizado o pacote FASTX-Toolkit (GORDON e HANNON, 2017). A montagem do genoma foi realizada utilizando o *software* SPAdes, com os parâmetros padrão do programa (BANKEVICH *et al.*, 2012).

4.3 Anotação dos genomas

Para a anotação automática foi utilizada a plataforma online GenSAS v.6.0 (HUMANN *et al.*, 2019) que reúne diversos programas utilizados no pipeline de anotação de genomas eucariotos. Para a predição de genes foi utilizado o programa AUGUSTUS (STANKE *et al.*, 2006) com os parâmetros disponibilizado nesta plataforma da microalga modelo *Chlamydomonas reinhardtii* CC-503 MT (Bioproject: PRJNA12260). Para o alinhamento das proteínas preditas pelo webAugustus e anotação foi utilizado o DIAMOND (BUCHFINK; XIE; HUSON, 2015) tendo como referencia a base de dados do SwissProt versão 9.22 (BATEMAN, 2019). O N50 das sequencias submetidas no GenSAS foram calculadas utilizando a ferramenta online Quast (GUREVICH *et al.*, 2013).

4.4 Bioprospecção de BGCs relacionados a metabólitos secundários

Para a prospecção dos BGCs ligados a metabólitos secundários de interesse, foi utilizada a metodologia sugerida para microalgas eucariotas de O'Neill (2020) com adaptações baseadas na mineração de genomas descrita por Adamek e colaboradores (2017). O *workflow* utilizado neste trabalho está descrito na Figura 04.

Os possíveis clusters gênicos de metabólitos secundários foram identificados através da ferramenta de predição antiSMASH (antibiotics & Secondary Metabolite Analysis Shell), versão 5.0 (BLIN *et al.*, 2019b) utilizando a versão para bactérias e a versão para fungos. Primeiramente

tendo como *input* somente os arquivos formato FASTA dos genomas e uma segunda submissão tendo como *input* os arquivos FASTA e os arquivos formato GFF3 obtidos da anotação realizada no GenSAS. O parâmetro utilizado para o rigor da detecção dos agrupamentos foi o *relaxed* com todos os algoritmos extras (*KnownClusterBlast*, *ActiveSiteFinder*, *ClusterBlast*, *Cluster Pfam analysis*, *SubClusterBlast* e *Pfam-based GO term annotation*) disponibilizados pelo antiSMASH ligados.

Todos as sequencias dos agrupamentos preditos pelo antiSMASH com maior similaridade apontadas pela plataforma foram submetidas na ferramenta onlince BLAST (MADDEN, 2013) do NCBI. Este procedimento foi realizado para identificar se os agrupamentos eram pertencentes as microalgas ou se eram provenientes de uma possível contaminação destes genomas.

Para filtragem e classificação dos *clusters* mais conhecidos a partir do predito pelo antiSMASH foram utilizados bancos de dados específicos. O MIBiG (Minimum Information about a Biosynthetic Gene cluster) versão 2.0 (KAUTSAR et al., 2020) é um repositório de BGCs o qual possui curadoria manual. O banco de dados de domínios conservados (CDD) do NCBI (MARCHLER-BAUER et al., 2017) e o Pfam (EL-GEBALI et al., 2019) foram utilizados para análise dos genes ligados a síntese de produtos naturais e inferência das possíveis funções.

A partir dos clusters preditos pelo antiSMASH, foi realizada uma análise comparativa dos genes de cada BGC apontados nos genomas das microalgas com os genes dos clusters depositados no banco de dados MIBiG através de gráficos de sintenia plotados no programa Easyfig 2.2.5 (SULLIVAN; PETTY; BEATSON, 2011). Os arquivos utilizados foram as saídas em formato genbank do antiSMASH e os arquivos em formato genbank dos BGCs contidos no MIBiG. Para a construção dos gráficos de sintenia foram utilizados os parâmetros tBLASTx com um e-value ≤ 0.001 , tamanho de alinhamento ≥ 50 e alinhamento de identidade ≥ 35 . Para uma investigação da identidade dos genes, sejam eles do tipo cerne ou adicional, foram realizadas buscas por BLAST (Basic Local Alignment Search Tool) no banco de dados do NCBI, inferência da função através do alinhamento de proteínas BLASTP no CDD e PFAM.

Os *contigs* contendo os agrupamentos preditos como NRPS e PKS foram submetidos no servidor NaPDoS (Natural Product Domain Seeker) para a busca de domínios “KS” e “C” (ZIEMERT et al., 2012). Primeiramente foi realizada uma busca com os parâmetros padrões e posteriormente a sensibilidade dos parâmetros foi diminuída para a busca destes domínios.

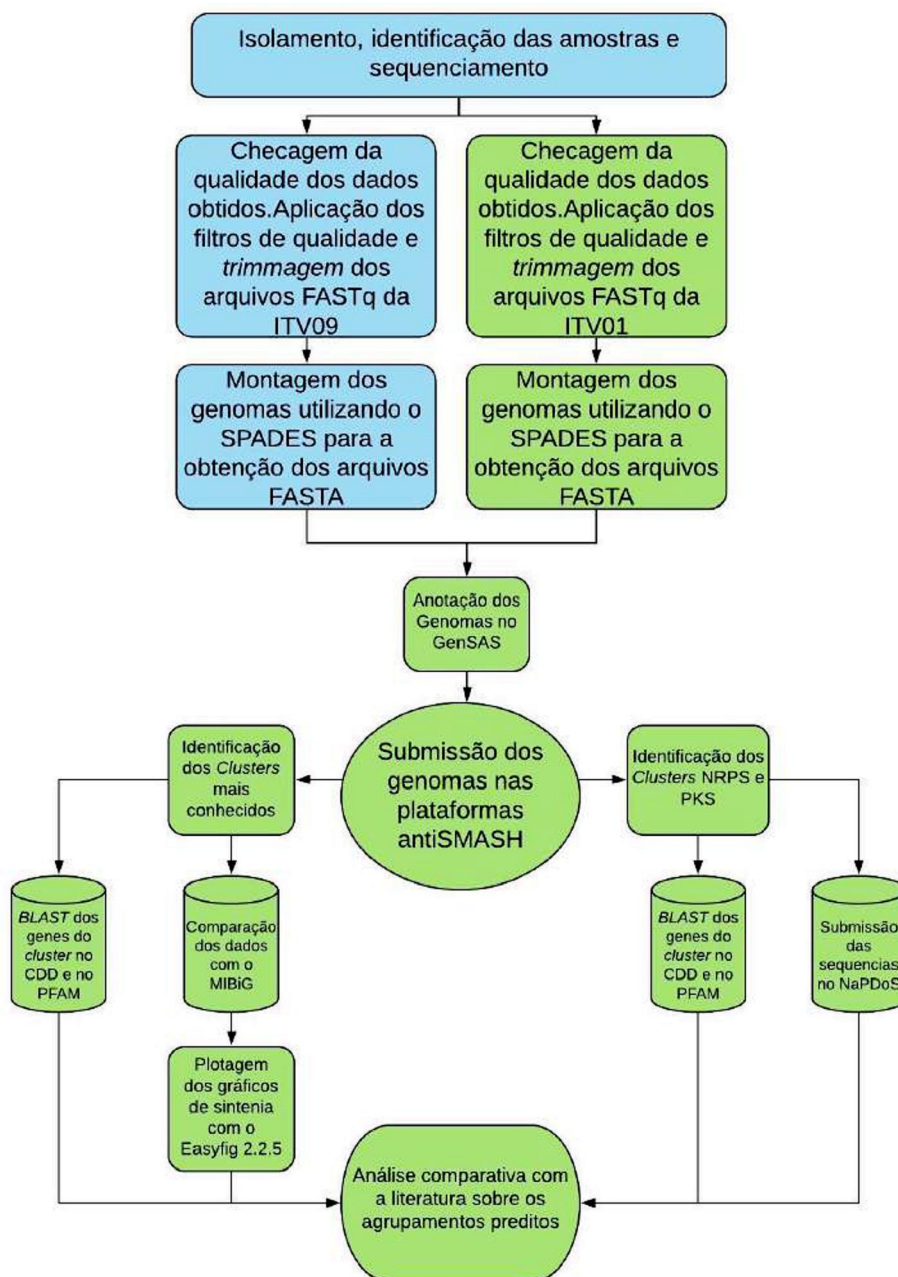


Figura 4: Workflow adaptado de Adamek *et al.* (2016) com o sugerido por O'Neill (2020) para as análises nos genomas das microalgas *Chlamydomonas acidophila* ITV01 e *Jaagichlorella hainangensis* ITV09. As cores indicam quais processos foram realizados em trabalho anterior (azul) e neste estudo (verde).

5. Resultados (artigo)

5.1 Prospecting of bioactive metabolites from extremophilic microalgae: an *in silico* approach

Pedro Felipe Rodrigues Sousa^{1,2}, Bruno Garcia Simões dos Santos³, José Tasso Felix Guimarães³, Guilherme Oliveira³, Ana Carolina de Araújo Butarelli^{1,4}, Leonardo Teixeira Dall’Agnol^{1,2}, Alex Ranieri Jerônimo Lima⁵, Hivana Patricia Melo Barbosa Dall’Agnol^{1,2},

¹ Research Group in Biodiversity, Bioprospection and Biotechnology, Center for 10 Biological and Health Sciences, Federal University of Maranhão, São Luís, Brazil

² Federal University of Maranhão - UFMA

³ Vale Institute of Technology - ITV

⁴ São Paulo University – USP

⁵ Butantan Institute

Corresponding author:

Hivana Patricia Melo Barbosa Dall’Agnol¹

Abstract

Background: Microalgae are eukaryotic microorganisms capable of using autotrophic and heterotrophic pathways in their metabolism. Certain types of environments have extreme conditions to which organisms adapt in order to survive. These adaptations induce specializations and activation of secondary metabolic pathways capable of producing compounds that allow organisms to survive in the most varied conditions. Several of these secondary metabolites have activities against microorganisms due to the competition of the medium, antibiofilm, antioxidant, protective activity and others. Thus, this work sought to evaluate the biotechnological potential in health, using *in silico* prospection techniques, of 2 green microalgae *Chlamydomonas acidophila* and *Jaagichlorella hainangensis*. Both microalgae were isolated from extremophile environments: *Chlamydomonas acidophila* from an acidic copper mine drainage and *Jaagichlorella hainangensis* from a iron natural lake, also in a mining area.

Results: The results after submission in the web platform antiSMASH pointed several gene clusters linked to biosynthetic pathways of secondary metabolites of interest in both microalgae such as terpenes, siderophores and ectoines. The results of these predicted clusters were analyzed using synteny plots with the agroupments deposited in the MIBiG and the possible functions of the genes were predicted with BLAST in CDD and PFAM databases. AntiSMASH also pointed out clusters of NRPS and PKS in the genome of the 2 microalgae and a hybrid NRPS / PKS clusters only in *Jaagichlorella hainangensis*. Genes related to PKS type I were the most prevalent in *Jaagichlorella hainangensis*. This profile is not found in *Chlamydomonas acidophila*. Further analyzes were performed using the NaPDoS platform. Only in *Jaagichlorella hainangensis*, 2 clusters were linked with low similarity to biosynthetic pathways, one of them a PKS cluster similar to the sequence of a metabolite already known to have anticancer activity. The low or the non-similarity of the clusters may indicate compounds of unprecedented character.

Conclusions: Obtaining these preliminary data in these species, direct possible *in vitro* experiments which can be carried out with the objective of extracting secondary metabolites as well as elucidating metabolic pathways. In addition to contributing with genomic databases with predicted gene clusters for possible synthetic relationships and expansion of microalgae genomic knowledge.

Keywords

Chlamydomonas, *Jaagichlorella*, Genome Mining, Biosynthetic Gene Clusters.

Background

Microalgae belong to a diverse group of microorganisms that successfully adapted to numerous types of aquatic environments as well as fresh, brackish or salt water. These have as characteristics microscopic size, eukaryotic cell type, photosynthetic pigmentation, being able to use autotrophic or heterotrophic pathways. Due to its rich biodiversity, this group has several functional, biological and ecological roles, thus showing very promising uses for biotechnological applications (BARRA et al., 2014).

The search for new species that have the ability to survive in hostile environments (extreme temperatures, high variation of pH, variants of 1atm pressure and others) it has been constantly

promoted because of their ability to survive is associated with special qualities. In this context, extremophilic microalgae would be of great interest for bioprospection. Some examples of bioprospection are new biochemical pathways, enzymes and compounds that have bioactivity and applicability in various biotechnological industry, environment and health (RADDADI et al., 2015; SINGH et al., 2019).

In order to prospecting bioactive metabolites, the use of computational tools, the advancement of omic's sciences, aligned with the decrease in the cost of DNA sequencing allowed microbial genome mining. These genome's data *in silico* analysis showed their abilities to produce compounds with applicability for both industrial and biomedical areas (GUARNIERI; PIENKOS, 2015).

The analysis of microbial genomes has revealed that many secondary metabolites have the biosynthetic machinery responsible for their production encoded by a group of genes located physically close to each other on the chromosome, thus being known as biosynthetic gene clusters (BGCs). The phenomenon of gene clusters is common in bacteria, where up to 50% of all genes are organized in structures of operons under the control of the same promoter. On the other hand, in eukaryotes, the genes organized in clusters would be similar physically close, but with independent promoters (CHAVALI; RHEE, 2018). In microalgae, recently, O'Neill (2020) demonstrated that programs used to search BGC's signatures in bacterial and fungal genomes may be applied on eukaryotic microalgae genomes.

Our study performed the analyses of two extremophilic microalgae genomes from the phylum Chlorophyta (*Chlamydomonas acidophila* and *Jaagichlorella hainangensis*), isolated and identified from operating copper and iron mines, respectively, in northern Brazil. Mining environments are extreme, as they generate selective elements for several types of microorganisms, selecting not only characteristics of resistance to heavy metals, but also the ability to produce other compounds of interest that may be applied for biotechnological use (BRIERLEY; BRIERLEY, 2013; SOUZA et al., 2017).

The microalgae genus *Chlamydomonas* has more than 600 species of which several members have a wide biotechnological application already known. These microalgae belong to the class

Chlorophyceae, their morphological characteristics are green colour, cell wall, 2 flagella of equal size usually close to each other and a single chloroplast provided with a unique microcompartment called pyrenoid. This genus stands out for owning the biological model organism *Chlamydomonas reinhardtii*, a microalga known by its historical studies about cycle cell and photosynthesis mainly (HARRIS, 2001; SALOMÉ; MERCHANT, 2019). However, only 12 genomes of this genus are available at the DNA data banks, including the model *C. reinhardtii*, and so far there is no genome submitted for *C. acidophila* (BENSON; LIPMAN; OSTELL, 1993).

The microalgae *Chlamydomonas acidophila* is a species found in extremophile environments, first described and isolated by Negoro in 1944 (NEGORO, 1944). This microalgae is often the dominant species of phytoplankton in lakes of mining sites and volcanic lakes, both of them are acidic environments and usually have high concentrations of metals (DEAN et al., 2019). Transcriptomic analyses allowed the identification of gene clusters expressed when a strain of *C. acidophila* is exposed to high concentrations of copper. These attributes demonstrate its potential use for bioremediation actions in environments with metal contamination, as well as other potential biotechnological applications (OLSSON et al., 2015).

The microalgae *Jaagichlorella hainangensis* (*J. hainangensis*) belongs to the class Trebouxiophyceae, first described and isolated from the rubber tree *Hevea brasiliensis*, in East Asia as *Heveochlorella hainangensis* (ZHANG et al., 2008) and recently reclassified into the genus *Jaagichlorella* (DARIENKO; PRÖSCHOLD, 2019). The literature describe as single-celled organisms, spherical in shape, which reproduction is due to by self-sporulation and have chloroplasts provided with pyrenoids (MA et al., 2013).

Currently, this genus is described as widely distributed, in tropical and subtropical regions and can be found living as endophytes or not. Although some studies report cultivation characteristics, such as growth rates (excellent at pH 7 and slow at pH below 3), distribution factors, ecology and metabolism of this species (SANDERS et al., 2016; ZHANG et al., 2008). Until the publication of this article only the mitochondrial genome of *Heveochlorella*.

Results and Discussion

Microalgae's genomes features

The search for gene clusters of secondary metabolites was carried out for the genomes of the microalgae *C. acidophila* ITV01 *J. hainangensis* ITV09 and, which are in their Draft versions. The genome of *J. hainangensis* containing about 23,000 contigs and *C. acidophila* containing about 94,000 contigs. For the annotation of the genomes, the Augustus software was used which identified 18,861 genes in *C. acidophila* and 16,630 genes in *J. hainangensis* (Table 1).

Table 1 Draft Genome data and predicted features of *C. acidophila* ITV01 and *J. hainangensis* ITV09.

Genome	<i>C. acidophila</i> ITV01	<i>J. hainangensis</i> ITV09
Number of contigs	94.008	23.782
N50	4165	106246
Length of largest contigs	39.215	1.627.548
GC content (%)	58%	42%
Genes	18.861	16.630
mRNA	18.861	16.630
rRNA	21	42
Proteins	16.956	16.238
tRNA	258	605

Genome mining for natural products synthesis

The online platform antiSMASH 5.0 (BLIN et al., 2019b) identified a total of 58 BGCs in *J. hainangensis* and 24 BGCs in *C. acidophila* (Fig. 3). In *J. hainangensis* clusters were found clusters related to Terpenes' synthesis pathways (10), PKS (9), Hybrids –PKS / NRPS (4), NRPS (10), Ectoines (1), Siderophores (4) and others (20). In *C. acidophila* were found clusters related to Terpenes' synthesis pathways (9), PKS (4), NRPS (2), Ectoine (1), Siderophores (2) and others (6) (Fig. 1).

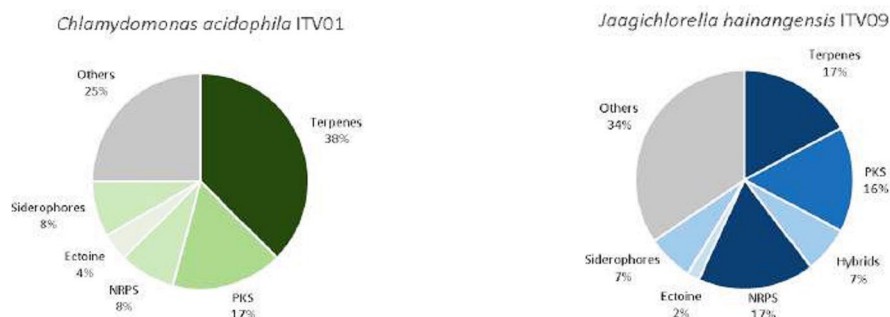


Fig. 1 Distribution of BGCs predicted by antiSMASH in the genomes of *Chlamydomonas acidophila* ITV01 and *Jaagichlorella hainangensis* ITV09

Synthesis of carotenoids

Carotenoids, which are tetraterpenes, are pigments with antioxidant properties formed by 8 basic units of isoprenoids. The review study by Novoveská et al. (2019) highlighted the growth of the market of carotenoids synthesised by microalgae (e.g. Astaxanthin produced by *Haemotococcus pluvialis*) due to the growing demand for natural products in the food (for humans and animals), cosmetic and nutraceutical industries (NOVOVESKÁ et al., 2019).

The homology with BGCs of other microorganisms supports the inference of the metabolic pathways of carotenoid synthesis in microalgae. Majority the genes, enzymes and metabolic pathways specific to microalgae are still unknown, the ones related were found through homology of the known genes of cyanobacterial and land plants (TAKAICHI, 2020). Nonetheless, current databases have made it possible to identify the possible synthesis of relevant carotenoids through analysis of the conserved domains.

Terpenes, the basic units that forms carotenoids, were the majority of the clusters pointed out by antiSMASH in both microalgae. The analyses carried out through BLASTx at the NCBI CDD (MARCHLER-BAUER et al., 2015) and PFAM database (EL-GEBALI et al., 2019) in the amino acid and nucleotide sequences of the core and additional genes, allowed to infer their possible bioactivities. Among the BGCs predicted for the synthesis of terpenes in *J. hainangensis*, the region 3.1 stands out. The analysis by BLASTx identity (%) identified the core genes related to the enzymes Phytoene synthase *crtB* (76%), phytoene dehydrogenase *crtI* (87%) and lycopene beta-

cyclase *crtY* (82%) which are linked to the synthesis of Zeaxanthin in *Xanthobacter autotrophicus* Py2 (MIBiG BGC0000656) (Fig. 2a).

C. acidophila shows the 1583.1 region pointed out by the antiSMASH database as the most similar to carotenoid synthesis. It is due to the core phytoene / squalene synthase family protein *crtB* (100%), phytoene desaturase (phytoene dehydrogenase) *crtI* (100%) and lycopene beta-cyclase *crtY* (99%). These three core enzymes are linked to Zeaxanthin synthesis. Nevertheless, the BGC predicted in *C. acidophila* showed two other genes linked to the carotenoid metabolic pathways. Those are Zeaxanthin glucosyltransferase *crtX* (98.77%) and beta-carotene hydroxylase *crtZ* (99.4%) which are associated with changes in the synthesis of carotenoids and it shows similarity to the MIBiG BGC BGC0000642 of *Enterobacteriaceae* bacterium DC413 (See figure 2.b).

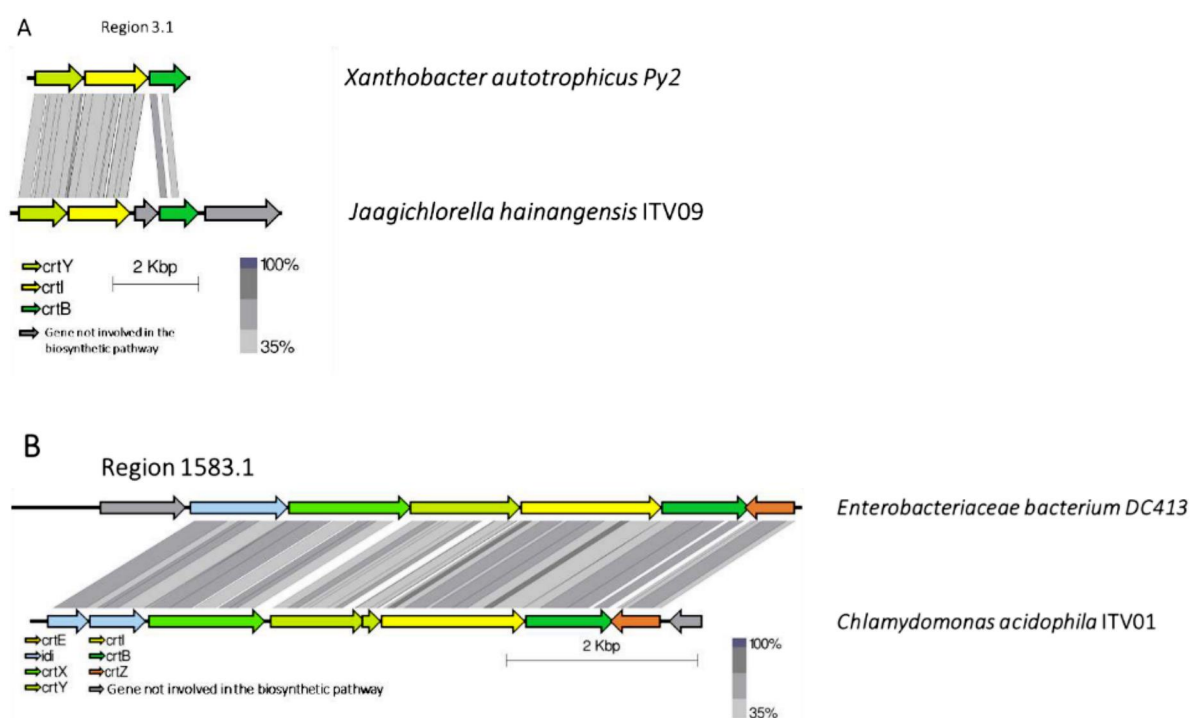


Fig. 2 Easyfig tBLASTx synteny graphs of the MIBiG clusters with the BGCs predicted by antiSMASH in the microalgae genomes related to: **a** and **b** as carotenoids.

Although MIBiG database only stores bacteria and fungi BGCs, the literature shows that the majority of the genes found in *C. acidophila* ITV01 and *J. hainangensis* ITV09, related to

carotenoid synthesis, they are described in eukaryotic microalgae already. This correlation data can be visualized in table 2.

Table 2. Enzymes coded by genes found in the *Chlamydomonas acidophila* ITV01 and *Jaagichlorella hainangensis* ITV09 and correlation with an organism previously described in literature containing these genes.

Enzyme	<i>C.</i> <i>acidophila</i> ITV01	<i>J.</i> <i>hainangensis</i> ITV09	Organism	Reference
Phytoene synthase <i>crtB</i>	X	X	<i>Chlamydomonas reinhardtii</i>	(STEINBRENNER; SANDMANN, 2006)
Lycopene beta-cyclase <i>crtY</i>	X	X	<i>Chlamydomonas reinhardtii</i>	(COUSO et al., 2012)
Phytoene desaturase/dehydrogenase <i>crtI</i>	X	X	<i>Dunaliella salina</i>	(SRINIVASAN; BABU; GOTHANDAM, 2017)
Zeaxanthin Glucosyltransferase <i>crtX</i>	X		<i>Pantoea agglomerans</i>	(HUNDLE et al., 1992)
Beta-carotene hydroxylase	X		<i>Haematococcus pluvialis</i>	NIES- (LINDEN, 1999) 144

Siderophores

Siderophores are the secondary metabolites synthesized by various organisms in order to uptake iron from the environment and allow its use. The scientific community has high expectations for the therapeutic use of siderophores, with several uses described in the literature, such as iron overload, malaria treatment and the use of selective delivery of antibacterial drugs in the Trojan

horse strategy. This strategy uses the iron uptake system of bacteria in order to actively deliver the drug into its cells (GÓRSKA; SLODERBACH; MARSZAŁŁ, 2014). Eukaryotic microalgae can also use the siderophore-iron complexes in order to acquire the iron needed (SUTAK; CAMADRO; LESUISSE, 2020).

Three clusters related to the synthesis of siderophores were identified in *J. hainangensis*. In one of these clusters, the core gene of the region 7.1 *dfoC* (97.89%) and the additional lysine / ornithine N-monooxygenase *dfoA* gene (95.91%) are similar to the core gene and to the one of the additional genes in the MIBiG cluster related to the synthesis of deferoxamine E of *Pantoea agglomerans* (MIBiG BGC0001572). Further analysis in the region 7.1 revealed another three genes: an additional acetyltransferase *AT* gene (98.51%), a regulatory gene of the Helix-turn-helix *htx* type (98.75%) and a multi-drug efflux transporter gene type MFS *MFS_t* (94.19 %) which activity is similar to the BGC0001572 *dfoS* gene (Fig 3a).

ITV09 presented two gene clusters with 85% similarity with the cluster related to ochrobactin synthesis of *Ochrobactrum anthropi* deposited in the MIBiG database (BGC0001713). This cluster consists of two core *IucA/Iuc* genes and two additional genes. The analysis showed that the two predicted clusters have two core genes and two additional genes with high identity rates and they are related to the genes found in the MIBiG database (Fig. 3b and 3c). The core genes of one of the predicted clusters in ITV09 demonstrated more than 96% and 82% identity with the family of *IucA/IucC* enzymes related to siderophore biosynthesis (CHEUNG et al., 2009).

The antiSMASH identified In *C. acidophila* a cluster with high similarity to the one of deferoxamine E synthesis in *Streptomyces sp. ID38640* (MIBiG BGC0001478). The predicted cluster is similar to the core gene *IucA/IucC* siderophore biosynthesis family of proteins (99.79%) and to the lysine/ornithine N-monooxygenase additional gene (98.16%). The analysis of *C. acidophila* showed that it has the MFS transporter *MFS_t* gene (99.48%) in the predicted cluster and the MIBiG cluster has pyridoxal-dependent decarboxylase *DesD* (Fig. 3d).

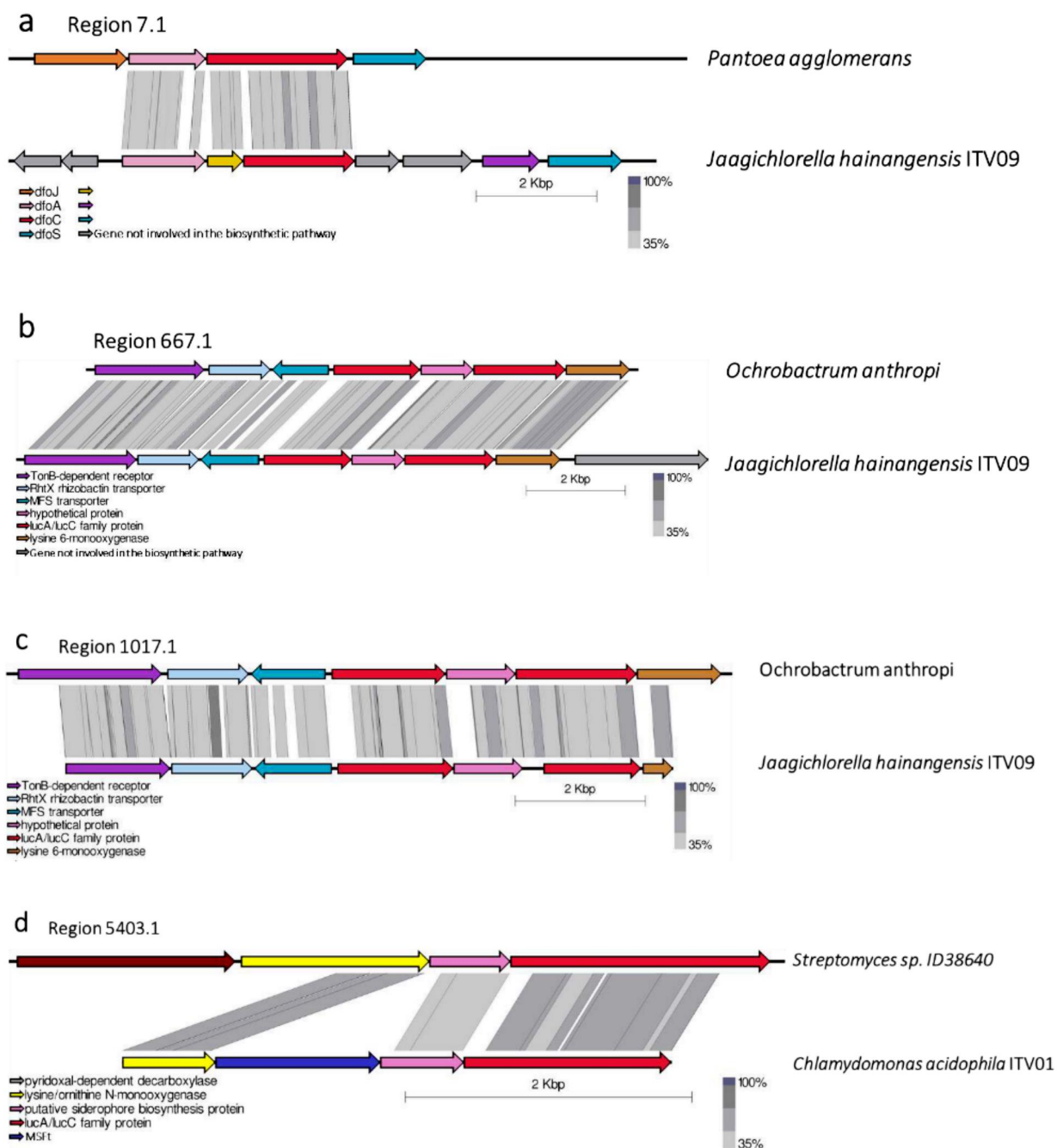


Fig. 3 Easyfig tBLASTx synteny graphs of the MIBiG clusters with the BGCs predicted by antiSMASH in the microalgae genomes related to: **a**, **b**, **c** and **d** as siderophores.

There are reports of siderophore synthesis in red and green microalgae described in literature such as *Rhodomonas ovalis* (NAITO et al., 2001), *Closterium aciculare* (NAITO et al., 2004) and *Scenedesmus incrassatulus* (BENDERLIEV; IVANOVA, 1994). An *in silico* study of Palenik et al (PALENIK et al., 2007) in the genome of a green microalga from the genus *Ostreococcus* demonstrated that this organism has genes related to siderophore-iron uptake with similarities with prokaryotic pathways and they suggest siderophore synthesis by this microalga. Thus, it is possible to infer due to the same pattern is found in *C. acidophila* ITV01 and *J. hainangensis* ITV09 and may be possible these microalgae are able synthesize siderophores as well.

Ectoines

Ectoine is a secondary metabolite of extremophile organisms that belongs to a group called compatible solutes. This class of compounds has as its main function the control of osmotic stress, which these organisms are exposed to in their habitat and is also related to the protection of proteins preventing their denaturation. It was known to only be produced by bacteria, however an experimental study of Fenizia *et al.* indicated that eukaryotic microalgae synthesize these types of metabolite (CZECH et al., 2018; FENIZIA et al., 2020).

Due to their biological activities, ectoines have demonstrated a potential use in several areas related to pathologies due to defects in protein folding and other health areas in general. The activity of this metabolite in protein stabilization and positive results in models of neurodegenerative diseases is promising (JORGE et al., 2016). A study by Werkhäuser et al. (2014) exemplifies another application of this metabolite, their study demonstrated that the addition of bacterial ectoine in nasal spray and eye drops decreased nasal and ocular symptoms, respectively, in patients with allergic rhinitis without showing side effects (WERKHÄUSER; BILSTEIN; SONNEMANN, 2014).

The same BGC related to ectoine synthesis was found in the genome of both microalgae. In *C. acidophila* the cluster indicated by antiSMASH contained a core ectoine synthase *ectC* gene (98.50%) and an additional diaminobutyrate - 2-oxoglutarate transaminase *ectB* gene (100%). In

the synteny plot, (Fig. 4a) it was visualized that other two genes were similar to the ectoine cluster of *Methylomicrobium alcaliphilum* 20Z (MIBiG BGC0000854).

J. hainangensis exhibited 1 core ectoine synthase gene *ectC* (100%) and 1 additional gene diaminobutyrate - 2-oxoglutarate transaminase *ectB* (99.54%) which are present in the MIBiG cluster (BGC0000858) of the ectoine synthesis in *Methylobacter marinus*. The cluster predicted by antiSMASH also pointed the regulatory gene *MarR* of the transcriptional regulator family (99.45%) in *J. hainangensis* (Fig. 4b).

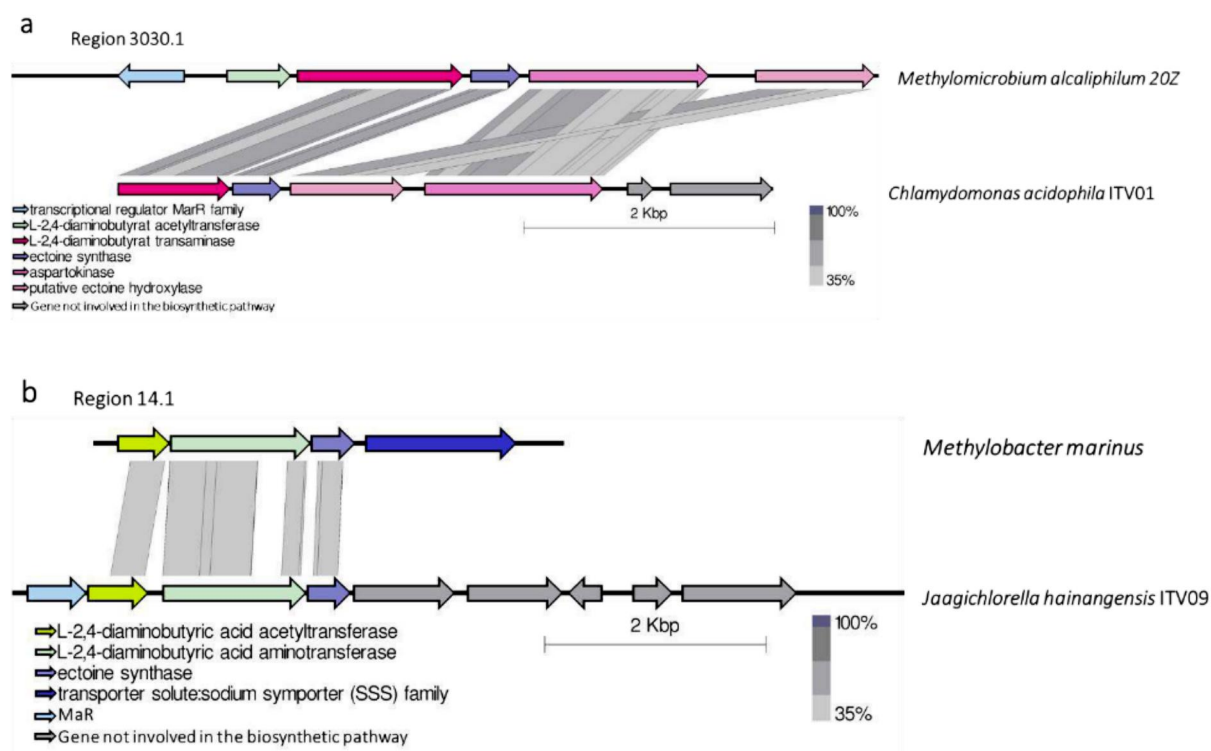


Fig. 4 Easyfig tBLASTx synteny graphs of the MIBiG clusters with the BGCs predicted by antiSMASH in the microalgae genomes related to: **a** and **b** as ectoines.

NRPS and PKS gene clusters

Among the various BGCs, two types of gene groups stand out as indicative of the production of molecules of interest: Nonribosomal Peptide Synthetases (NRPS) and Polyketide Synthases (PKS). These secondary metabolites may be essential for the formation of antibiotics, antifungals and

antitumor agents (WANG et al., 2014). The polyketides are produced through oligomerization of carboxylic acids mediated by PKS and NRPS use amino acids to form an amide or ester bond (WANG et al., 2014). An example of a PKS-type enzyme-mediated product is erythromycin, a well-known antibiotic already widely used in clinical use (SHELEST et al., 2015)

The antiSMASH identified clusters classified as PKS, and NRPS in the genome of the two microalgae and hybrid NRPS / PKS clusters only in *J. hainangensis*. All predicted clusters were submitted to manual curation, according to the criteria related to these types of clusters: for PKS the presence of the β -ketoacyl synthase (KS), acyltransferase (AT), acyl carrier protein (ACP) modules; for NRPS the presence of the domains adenylation (A), thiolation (T) and condensation (C) (SINGH; CHAUDHARY; SAREEN, 2017).

Shelest and co-workers (2015) (SHELEST et al., 2015) identified that type I PKS clusters are the most frequent in the microalgae genomes of the phylum Chlorophyta, and one or no NRPS clusters were present in the analysed species. Both microalgae showed the same profile: the most prevalent clusters were from the PKS type. Genes related to PKS type I were the most prevalent in *J. hainangensis*, followed by genes related to type III. On the other hand this profile was different in *C. acidophila* which only genes related to PKS type III have been identified.

AntiSMASH identified 23 clusters related to NRPS and PKS pathways in *J. hainangensis*. All clusters were submitted to the NaPDoS platform (ZIEMERT et al., 2012) to search for similarity in the KS and C domains. Nevertheless, only two clusters of this microalgae were related to metabolic pathways. In the analysis of the platform, the sequences of the domains related to the functions described in the literature were identified and these domains were confirmed using the BLAST tool at NCBI (Table 2).

None of the 6 clusters identified by antiSMASH, submitted to the NaPDoS platform of microalgae *C. acidophila* were related to a biosynthetic pathway, even with a reduction in the alignment restrictions (HMM e-value, minimum amino acid alignment size and e-value of the BLASTP). One of the clusters in *J. hainangensis* was related by NaPDoS to have similarity with the epothilone metabolic pathway, a secondary metabolite with anticancer activity from mixobacteria (CHENG; HUANG; HUANG, 2018). According to the developers of the NaPDoS platform, its database is

continuously updated, however not all metabolic pathways are deposited. In the case of the percentage of identity below 85% here is a probability that the related synthetic routes has not been inserted or categorized and in these cases it may be possible that the sequence is related to a compound of an unprecedented character (ZIEMERT et al., 2012). All the clusters submitted to NaPDoS showed percentage of identity below 85%.

Table 2. NaPDoS's results of the clusters identified in *J. hainangensis* ITV09, the metabolic pathway indicated by NaPDoS and their BLAST identity percentage.

Region * with the predicted cluster in antiSMASH	Metabolic pathway of NaPDoS (percentage of identity) <i>organism</i> (reference paper)	BLAST at NCBI with identity percentage
6.1	Epothilone (58%) <i>Sorangium cellulosum</i> (MOLNÁR et al., 2000)	Polyketide synthase type I (99.06%)
10.1	Fatty acid synthesis (60%) <i>Streptomyces coelicolor</i> (BENTLEY et al., 2002)	Beta-ketoacyl-ACP synthase II (98.69%)

* The number of the region that identifies the contig of origin in the *J. hainangensis* ITV09 genome.

Conclusions

This study used an *in silico* approach in order to bioprospection of gene clusters related to secondary metabolites with an emphasis on red biotechnology in two new microalgae genomes, extremophilic microalgae *C. acidophila* and *J. hainangensis*. Through genomic mining platforms, the relationship of conserved domains and similarity between BGCs deposited in databases, it was possible to infer the possible functions of the genes contained in the clusters of these microalgae.

The analyses of the *C. acidophila* and *J. hainangensis* clusters suggest the synthesis of carotenoids, ectoines and siderophores. The results for the biosynthetic pathways NRPS and PKS of *C. acidophila* does not exclude the potential for secondary metabolites of these pathways, since new pathways may be inserted in the databases in the future. On the other hand *J. hainangensis* presented a PKS cluster similar to a metabolite already known to have anticancer activity.

These preliminary data in these species guides possible *in vitro* experiments such as the ones with the objective of extracting secondary metabolites as well as elucidating metabolic pathways. In addition to contributing with genomic databases with predicted gene clusters for possible synthetic relationships and expansion of the knowledge of microalgae genomes.

Methods

Samples

The strains ITV01 (*Chlamydomonas acidophila*) and ITV09 (*Jaagichlorella hainangensis*), respectively, were isolated from a copper mine acid drainage sample (Coordinates: 6 ° 26'42.3 "S 50 ° 04'36.9" W), and of an iron mine natural lake (Coordinates: 6 ° 24'02.5 "S 50 ° 21'06.7" W) (SANTOS, 2017). These environments have a high concentration of metals and are acidic; both with mining activity are located in the Carajás region, state of Pará, Brazil. These strains were obtained from the Collection of Microalgae and Cyanobacteria of the Instituto Tecnológico Vale (ITV), located in the capital Belém - Pará.

Both microalgae species were identified morphologically and molecularly, through amplification and sequencing of barcode genes, such as *rbcl*, 16S rRNA and 18S rRNA.

Sequencing, quality control and assembly of the genomes

The genomic DNA sequence of the *Chlamydomonas acidophila* lineage (ITV01) and *Jaagichlorella hainangensis* (ITV09) were obtained by sequencing a new generation DNA already carried out by the team from the Instituto Tecnológico Vale Desenvolvimento Sustentável (ITV-DS) it was used the platform Illumina / MiSeq (Illumina, Inc., USA) according to the protocols described by the manufacturers. The DNA sequences of the microalgae genomes were provided in FASTA format files for ITV09 and FASTQ for ITV01.

In order to assembly and obtain the ITV01 FASTA file, the quality of the sequencing was first analysed using the FASTQC tool (ANDREWS; OTHERS, 2010). The FASTX-Toolkit package was used (GORDON; HANNON, 2017) to remove the low quality readings and the bases of the ends of the reads - phread 20 (EWING; GREEN, 1998), minimum stay of 80% of the reads with

the selected quality and reads with a size of 420 bp -. The genome assembly was performed using the SPAdes software in standard setting (BANKEVICH et al., 2012).

Genome annotation

For automatic annotation of the nuclear genomes of *C. acidophila* and *J. hainangensis* , the online platform GenSAS v.6.0 (HUMANN et al., 2019) was used, which gathers several programs used in the eukaryotic genome annotation pipeline. For the prediction of genes, the AUGUSTUS program (STANKE et al., 2006) trained with the microalgae model *Chlamydomonas reinhardtii* was used. For the alignment of the predicted proteins, DIAMOND (BUCHFINK; XIE; HUSON, 2015) was used, with the SwissProt database version 9.22 as reference (BATEMAN, 2019).

Bioprospection of BGCs related to secondary metabolites

For the prospection of BGCs linked to secondary metabolites of interest, the suggested methodology for eukaryotic microalgae from O'Neill (O'NEILL, 2020) was used with adaptations based on the genome mining described by Adamek *et al.* (ADAMEK et al., 2017). The workflow used in this work is described in the additional file 01.

The possible gene clusters of secondary metabolites were identified using the antiSMASH (antibiotics & Secondary Metabolite Analysis Shell) version 5.0 (BLIN et al., 2019b) using the both bacteria and fungi versions. Firstly, using only the FASTA format files of the genomes and a second submission using the FASTA files and the GFF3 format files obtained from the annotation performed in GenSAS. The parameter used for the accuracy of detecting clusters was relaxed with all the extra algorithms provided by antiSMASH (KnownClusterBlast, ActiveSiteFinder, ClusterBlast, Cluster Pfam analysis, SubClusterBlast and Pfam-based GO term annotation) turned on.

To filter and classify the most well-known clusters based on the predicted BGCs by antiSMASH, specific databases were used. The MIBiG (Minimum Information about a Biosynthetic Gene cluster) version 2.0 (KAUTSAR et al., 2020) is a repository of BGCs which has manual curation. The NCBI's database of conserved domains (CDD) (MARCHLER-BAUER et al., 2017) and Pfam (EL-GEBALI et al., 2019) were used for analysis of genes linked to the synthesis of natural products and inference of possible functions.

From the clusters predicted by antiSMASH, a comparative analysis of the predicted genes of each BGC in the genomes of the microalgae was carried out with the genes of the clusters deposited in the MIBiG database through synteny plots plotted with Easyfig 2.2.5 (SULLIVAN; PETTY; BEATSON, 2011). For input were used the files of the outputs in genbank format of antiSMASH and the files in genbank format of the MIBiG's BGCs. The tBLASTx was performed with an e-value of ≤ 0.001 , alignment size ≥ 50 and identity alignment ≥ 35 . For an investigation of the identity of the genes, whether core or additional, searches were performed using BLAST (Basic Local Alignment Search Tool) in the NCBI database. The inference of the functions was carried out through the alignment of BLASTP proteins in the CDD and PFAM.

The contigs containing the predicted clusters such as NRPS and PKS were submitted on the NaPDoS (Natural Product Domain Seeker) server in order to search for "KS" and "C" domains (ZIEMERT et al., 2012). Initially a search was conducted using the standard parameters and later with less stringent to search for these domains.

References

ADAMEK, M. et al. Mining bacterial genomes for secondary metabolite gene clusters. In: **Antibiotics**. [s.l.] Springer, 2017. p. 23–47.

ALVES, G. S. a Biotecnologia Dos Transgênicos: Precaução É a Palavra De Ordem. **Holos**, v. 2, 2007.

ANDREWS, S.; OTHERS. **FastQC: a quality control tool for high throughput sequence data** Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom, , 2010.

ÁVILA-ROMÁN, J. et al. Anti-inflammatory effects of an oxylipin-containing lyophilised biomass from a microalga in a murine recurrent colitis model. **British Journal of Nutrition**, v. 116, n. 12, p. 2044–2052, 2016.

BAJPAI, V. K. Antimicrobial bioactive compounds from marine algae: A mini review. **Indian Journal of Geo-Marine Sciences**, v. 45, n. 9, p. 1076–1085, 2016.

BANKEVICH, A. et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. **Journal of Computational Biology**, v. 19, n. 5, p. 455–477, 2012.

- BARRA, L. et al. **The challenge of ecophysiological biodiversity for biotechnological applications of marine microalgae** *Marine Drugs*, 2014.
- BATEMAN, A. UniProt: A worldwide hub of protein knowledge. **Nucleic Acids Research**, v. 47, n. D1, p. D506–D515, 2019.
- BEATTIE, A. J. et al. Ecology and bioprospecting. **Austral Ecology**, v. 36, n. 3, p. 341–356, 2011.
- BENDERLIEV, K. M.; IVANOVA, N. I. High-affinity siderophore-mediated iron-transport system in the green alga *Scenedesmus incrassatulus*. **Planta**, v. 193, n. 2, p. 163–166, 1994.
- BENSON, D.; LIPMAN, D. J.; OSTELL, J. GenBank. v. 21, n. 13, p. 2963–2965, 1993.
- BENTLEY, S. D. et al. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). **Nature**, v. 417, n. 6885, p. 141–147, 2002.
- BLIN, K. et al. The antiSMASH database version 2: A comprehensive resource on secondary metabolite biosynthetic gene clusters. **Nucleic Acids Research**, v. 47, n. D1, p. D625–D630, 2019a.
- BLIN, K. et al. AntiSMASH 5.0: Updates to the secondary metabolite genome mining pipeline. **Nucleic Acids Research**, v. 47, n. W1, p. W81–W87, 2019b.
- BRIERLEY, C. L.; BRIERLEY, J. A. Progress in bioleaching: Part B: Applications of microbial processes by the minerals industries. **Applied Microbiology and Biotechnology**, v. 97, n. 17, p. 7543–7552, 2013.
- BUCHFINK, B.; XIE, C.; HUSON, D. H. Fast and sensitive protein alignment using DIAMOND. **Nature methods**, v. 12, n. 1, p. 59–60, 2015.
- CADORET, J. P.; GARNIER, M.; SAINT-JEAN, B. **Microalgae, Functional Genomics and Biotechnology**. [s.l.] Elsevier, 2012. v. 64
- CHAVALI, A. K.; RHEE, S. Y. Bioinformatics tools for the identification of gene clusters that biosynthesize specialized metabolites. **Briefings in bioinformatics**, v. 19, n. 5, p. 1022–1034, 2018.

CHENG, H.; HUANG, H.; HUANG, G. Synthesis and antitumor activity of epothilone B. **European Journal of Medicinal Chemistry**, v. 157, p. 925–934, 2018.

CHEUNG, J. et al. Molecular characterization of staphyloferrin B biosynthesis in *Staphylococcus aureus*. **Molecular Microbiology**, v. 74, n. 3, p. 594–608, 2009.

CHOI, Y. H. et al. Mycosporine-like amino acids promote wound healing through focal adhesion kinase (FAK) and mitogen-activated protein kinases (MAP kinases) signaling pathway in keratinocytes. **Marine Drugs**, v. 13, n. 12, p. 7055–7066, 2015.

CHOU DHARY, S.; SAVE, S. N.; VAVILALA, S. L. Unravelling the inhibitory activity of *Chlamydomonas reinhardtii* sulfated polysaccharides against α -Synuclein fibrillation. **Scientific Reports**, v. 8, n. 1, p. 1–12, 2018.

COUSO, I. et al. Synthesis of carotenoids and regulation of the carotenoid biosynthesis pathway in response to high light stress in the unicellular microalga *Chlamydomonas reinhardtii*. **European Journal of Phycology**, v. 47, n. 3, p. 223–232, 2012.

CUARESMA, M. et al. Productivity and selective accumulation of carotenoids of the novel extremophile microalga *Chlamydomonas acidophila* grown with different carbon sources in batch systems. **Journal of Industrial Microbiology and Biotechnology**, v. 38, n. 1, p. 167–177, 2011.

CZECH, L. et al. Role of the extremolytes ectoine and hydroxyectoine as stress protectants and nutrients: Genetics, phylogenomics, biochemistry, and structural analysis. **Genes**, v. 9, n. 4, p. 1–58, 2018.

DARIENKO, T.; PRÖSCHOLD, T. The genus *Jaagichlorella* Reisi gl (Trebouxiophyceae, Chlorophyta) and its close relatives: An evolutionary puzzle. **Phytotaxa**, v. 388, n. 1, p. 47–68, 2019.

DE LOS REYES, C. et al. Oxylipins from the microalgae *Chlamydomonas debaryana* and *Nannochloropsis gaditana* and their activity as TNF- α inhibitors. **Phytochemistry**, v. 102, p. 152–161, 2014.

DEAN, A. P. et al. Metabolic adaptation of a *Chlamydomonas acidophila* strain isolated from

acid mine drainage ponds with low eukaryotic diversity. **Science of the Total Environment**, v. 647, p. 75–87, 2019.

DOLGANYUK, V. et al. Microalgae: A Promising Source of Valuable Bioproducts. **Biomolecules**, v. 10, n. 8, p. 1–24, 2020.

EL-GEBALI, S. et al. The Pfam protein families database in 2019. **Nucleic Acids Research**, v. 47, n. D1, p. D427–D432, 2019.

EWING, B.; GREEN, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. **Genome Research**, v. 8, n. 3, p. 186–194, 1998.

FENIZIA, S. et al. Ectoine from Bacterial and Algal Origin Is a Compatible Solute in Microalgae. p. 1–13, 2020.

FUČÍKOVÁ, K. et al. New phylogenetic hypotheses for the core Chlorophyta based on chloroplast sequence data. **Frontiers in Ecology and Evolution**, v. 2, n. OCT, p. 1–12, 2014.

GHASEMI, Y. et al. Antifungal and antibacterial activity of the microalgae collected from paddy fields of Iran: characterization of antimicrobial activity of *Chroococcus dispersus*. 2007.

GOLLERY, M. Handbook of Hidden Markov models in bioinformatics. **Handbook of Hidden Markov Models in Bioinformatics**, p. 1–157, 2008.

GONG, Y. et al. Microalgae as platforms for production of recombinant proteins and valuable compounds: Progress and prospects. **Journal of Industrial Microbiology and Biotechnology**, v. 38, n. 12, p. 1879–1890, 2011.

GORDON, A.; HANNON, G. Fastx-toolkit. FASTQ/A short-reads pre-processing tools. 2010. **Unpublished available online at: http://hannonlab.cshl.edu/fastx_toolkit**, 2017.

GÓRSKA, A.; SLODERBACH, A.; MARSZAŁŁ, M. P. Siderophore-drug complexes: Potential medicinal applications of the “Trojan horse” strategy. **Trends in Pharmacological Sciences**, v. 35, n. 9, p. 442–449, 2014.

GUARNIERI, M. T.; PIENKOS, P. T. Algal omics: Unlocking bioproduct diversity in algae cell factories. **Photosynthesis Research**, v. 123, n. 3, p. 255–263, 2015.

- GUIRY, M.D. & GUIRY, G. M. **AlgaeBase. World-wide electronic publication, National University of Ireland, Galway.** Disponível em: <<https://www.algaebase.org>>. Acesso em: 11 ago. 2020.
- GUREVICH, A. et al. QUASt: Quality assessment tool for genome assemblies. **Bioinformatics**, v. 29, n. 8, p. 1072–1075, 2013.
- HARRIS, E. H. Chlamydomonas as a model organism. **Annual Review of Plant Biology**, v. 52, n. 1, p. 363–406, 2001.
- HIDER, R. C.; KONG, X. Chemistry and biology of siderophores. **Natural Product Reports**, v. 27, n. 5, p. 637–657, 2010.
- HUMANN, J. L. et al. Structural and functional annotation of eukaryotic genomes with GenSAS. In: **Gene Prediction**. [s.l.] Springer, 2019. p. 29–51.
- HUNDLE, B. S. et al. Functional expression of zeaxanthin glucosyltransferase from *Erwinia herbicola* and a proposed uridine diphosphate binding site. **Proceedings of the National Academy of Sciences of the United States of America**, v. 89, n. 19, p. 9321–9325, 1992.
- JAYSHREE, A.; JAYASHREE, S.; THANGARAJU, N. *Chlorella vulgaris* and *Chlamydomonas reinhardtii*: Effective antioxidant, antibacterial and anticancer mediators. **Indian Journal of Pharmaceutical Sciences**, v. 78, n. 5, p. 575–581, 2016.
- JORGE, C. D. et al. Potential applications of stress solutes from extremophiles in protein folding diseases and healthcare. **Extremophiles**, v. 20, n. 3, p. 251–259, 2016.
- KAUTSAR, S. A. et al. MIBiG 2.0: A repository for biosynthetic gene clusters of known function. **Nucleic Acids Research**, v. 48, n. D1, p. D454–D458, 2020.
- LEWIS, K. Platforms for antibiotic discovery. **Nature Reviews Drug Discovery**, v. 12, n. 5, p. 371–387, 2013.
- LINDEN, H. Carotenoid hydroxylase from *Haematococcus pluvialis*: cDNA sequence, regulation and functional complementation. **Biochimica et Biophysica Acta - Gene Structure and Expression**, v. 1446, n. 3, p. 203–212, 1999.

- MA, S. et al. A novel species in the genus *Heveochlorella* (Trebouxiophyceae, Chlorophyta) witnesses the evolution from an epiphytic into an endophytic lifestyle in tree-dwelling green algae. **European Journal of Phycology**, v. 48, n. 2, p. 200–209, 2013.
- MADDEN, T. The BLAST sequence analysis tool. **The BLAST Sequence Analysis Tool**, n. Md, p. 1–17, 2013.
- MARCHLER-BAUER, A. et al. CDD: NCBI's conserved domain database. **Nucleic Acids Research**, v. 43, n. D1, p. D222–D226, 2015.
- MARCHLER-BAUER, A. et al. CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. **Nucleic Acids Research**, v. 45, n. D1, p. D200–D203, 2017.
- MAYFIELD, S. P. et al. *Chlamydomonas reinhardtii* chloroplasts as protein factories. **Current Opinion in Biotechnology**, v. 18, n. 2, p. 126–133, 2007.
- MEDEMA, M. H. et al. AntiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. **Nucleic Acids Research**, v. 39, n. SUPPL. 2, p. 339–346, 2011.
- MOLNÁR, I. et al. The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from *Sorangium cellulosum* So ce90. **Chemistry and Biology**, v. 7, n. 2, p. 97–109, 2000.
- MOSBLECH, A.; FEUSSNER, I.; HEILMANN, I. Oxylipins: Structurally diverse metabolites from fatty acid oxidation. **Plant Physiology and Biochemistry**, v. 47, n. 6, p. 511–517, 2009.
- NAITO, K. et al. The pursuit of siderophore secreted by marine phytoplankton *Rhodomonas ovalis*. **Analytical Sciences**, v. 17, n. Table 1, p. 817–819, 2001.
- NAITO, K. et al. Secretion of iron-complexing ligands from *Closterium aciculare* (Charophyceae, Chlorophyta) under iron-deficient conditions. **Phycologia**, v. 43, n. 5, p. 632–634, 2004.
- NEGORO, K. Untersuchungen über die Vegetation der mineralogenazidotrophen Gewässer Japans. **Science Reports of the Tokyo Bunrika Daigaku, Sect. B.**, v. 6, p. 231–374, 1944.

- NEWMAN, D. J.; CRAGG, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. **Journal of Natural Products**, v. 83, n. 3, p. 770–803, 2020.
- NITHYA, C. et al. Biofilm inhibitory potential of *Chlamydomonas* sp. extract against *Pseudomonas aeruginosa*. **J. Algal Biomass Utiln**, v. 5, n. 4, p. 74–81, 2014.
- NOVOVESKÁ, L. et al. Microalgal carotenoids: A review of production, current markets, regulations, and future direction. **Marine Drugs**, v. 17, n. 11, p. 1–21, 2019.
- O'NEILL, E. Mining natural product biosynthesis in eukaryotic algae. **Marine Drugs**, v. 18, n. 2, 2020.
- OLSSON, S. et al. Transcriptional response to copper excess and identification of genes involved in heavy metal tolerance in the extremophilic microalga *Chlamydomonas acidophila*. **Extremophiles**, v. 19, n. 3, p. 657–672, 2015.
- OTTO, B. et al. Laccase-like enzyme activities from chlorophycean green algae with potential for bioconversion of phenolic pollutants. **FEMS Microbiology Letters**, v. 362, n. 11, p. 1–8, 2015.
- PAL, R.; CHOUDHURY, A. K. **An introduction to phytoplanktons: Diversity and ecology**. [s.l: s.n.].
- PALENIK, B. et al. The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. **Proceedings of the National Academy of Sciences of the United States of America**, v. 104, n. 18, p. 7705–7710, 2007.
- PEREZ-GARCIA, O. et al. Heterotrophic cultures of microalgae: Metabolism and potential products. **Water Research**, v. 45, n. 1, p. 11–36, 2011.
- PRÖSCHOLD, T. et al. Molecular phylogeny and taxonomic revision of *Chlamydomonas* (Chlorophyta). I. Emendation of *Chlamydomonas ehrenberg* and *Chloromonas gobi*, and description of *Oogamochlamys* gen. nov. and *Lobochlamys* gen. nov. **Protist**, v. 152, n. 4, p. 265–300, 2001.
- QIAN, P. Y.; XU, Y.; FUSETANI, N. Natural products as antifouling compounds: recent

progress and future perspectives. **Biofouling**, v. 26, n. 2, p. 223–234, 2009.

RADDADI, N. et al. Biotechnological applications of extremophiles, extremozymes and extremolytes. **Applied Microbiology and Biotechnology**, v. 99, n. 19, p. 7907–7913, 2015.

RENUKADEVI, K. P.; SARAVANA, P. S.; ANGAYARKANNI, J. ANTIMICROBIAL AND ANTIOXIDANT ACTIVITY OF CHLAMYDOMONAS REINHARDTII SP. **International journal of Pharmaceutical Sciences and Research**, v. 2, n. 6, p. 1467, 2011.

ŘEZANKA, T. et al. LC-MS/APCI identification of glucoside esters and diesters of astaxanthin from the snow alga *Chlamydomonas nivalis* including their optical stereoisomers. **Phytochemistry**, v. 88, p. 34–42, 2013.

SALOMÉ, P. A.; MERCHANT, S. S. A series of fortunate events: Introducing *chlamydomonas* as a reference organism. **Plant Cell**, v. 31, n. 8, p. 1682–1707, 2019.

SANDERS, W. B. et al. *Heveochlorella* (Trebouxiophyceae): a little-known genus of unicellular green algae outside the Trebouxiales emerges unexpectedly as a major clade of lichen photobionts in foliicolous communities. **Journal of Phycology**, v. 52, n. 5, p. 840–853, 2016.

SANTOS, B. G. S. DOS. Prospecção de Microalgas para uma Mineração Sustentável: Caracterização de um Isolado Fitoplantônico da Floresta Nacional de Carajás – PA. **Dissertação (Mestrado Profissional em Uso Sustentável de Recursos Tropicais) – Instituto Tecnológico Vale. Pará**, 2017.

SCRANTON, M. A. et al. *Chlamydomonas* as a model for biofuels and bio-products production. **Plant Journal**, v. 82, n. 3, p. 523–531, 2015.

SENHORINHO, G. N. A.; LANNÉR, C.; SCOTT, J. A. Effect of green microalgal extracts exhibiting antibacterial activity on viability of human malignant and non-malignant cells. **Phycological Research**, v. 67, n. 2, p. 145–151, 2019.

SHAH, M. M. R. et al. Astaxanthin-producing green microalga *Haematococcus pluvialis*: From single cell to high value commercial products. **Frontiers in Plant Science**, v. 7, n. APR2016, 2016.

SHELEST, E. et al. Multimodular type I polyketide synthases in algae evolve by module duplications and displacement of AT domains in trans. **BMC Genomics**, v. 16, n. 1, p. 1–15, 2015.

SINGH, M.; CHAUDHARY, S.; SAREEN, D. Non-ribosomal peptide synthetases: Identifying the cryptic gene clusters and decoding the natural product. **Journal of Biosciences**, v. 42, n. 1, p. 175–187, 2017.

SINGH, P. et al. **Microbial Community Dynamics of Extremophiles/Extreme Environment**. [s.l.] Elsevier Inc., 2019.

SOUZA, L. S. et al. Morphological and ultrastructural characterization of the acidophilic and lipid-producer strain *Chlamydomonas acidophila* LAFIC-004 (Chlorophyta) under different culture conditions. **Protoplasma**, v. 254, n. 3, p. 1385–1398, 2017.

SRINIVASAN, R.; BABU, S.; GOTHANDAM, K. M. Accumulation of phytoene, a colorless carotenoid by inhibition of phytoene desaturase (PDS) gene in *Dunaliella salina* V-101. **Bioresource Technology**, v. 242, p. 311–318, 2017.

STANKE, M. et al. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. **BMC bioinformatics**, v. 7, n. 1, p. 62, 2006.

STEINBRENNER, J.; SANDMANN, G. Transformation of the green alga *Haematococcus pluvialis* with a phytoene desaturase for accelerated astaxanthin biosynthesis. **Applied and Environmental Microbiology**, v. 72, n. 12, p. 7477–7484, 2006.

SUH, S. S. et al. Anti-inflammation activities of mycosporine-like amino acids (MAAs) in response to UV radiation suggest potential anti-skin aging activity. **Marine Drugs**, v. 12, n. 10, p. 5174–5187, 2014.

SULLIVAN, M. J.; PETTY, N. K.; BEATSON, S. A. Easyfig: A genome comparison visualizer. **Bioinformatics**, v. 27, n. 7, p. 1009–1010, 2011.

SUTAK, R.; CAMADRO, J. M.; LESUISSE, E. Iron Uptake Mechanisms in Marine Phytoplankton. **Frontiers in Microbiology**, v. 11, n. November, p. 1–14, 2020.

- TAKAICHI, S. Carotenoids in Phototrophic Microalgae: Distributions and Biosynthesis. In: **Pigments from Microalgae Handbook**. [s.l.] Springer, 2020. p. 19–41.
- THIRUMURUGAN, D. et al. An introductory chapter: secondary metabolites. **Second metab—sources Appl**, p. 1–21, 2018.
- VINGIANI, G. M. et al. Microalgal enzymes with biotechnological applications. **Marine Drugs**, v. 17, n. 8, p. 1–20, 2019.
- WANG, H. et al. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. **Proceedings of the National Academy of Sciences of the United States of America**, v. 111, n. 25, p. 9259–9264, 2014.
- WANG, L. et al. Identification of genetic variations associated with epsilon-poly-lysine biosynthesis in *Streptomyces albulus* ZPM by genome sequencing. **Scientific reports**, v. 5, p. 9201, 2015.
- WERKHÄUSER, N.; BILSTEIN, A.; SONNEMANN, U. Treatment of Allergic Rhinitis with Ectoine Containing Nasal Spray and Eye Drops in Comparison with Azelastine Containing Nasal Spray and Eye Drops or with Cromoglycic Acid Containing Nasal Spray. **Journal of Allergy**, v. 2014, p. 1–13, 2014.
- WRIGHT, G. D. Something old, something new: Revisiting natural products in Antibiotic drug discovery. **Canadian Journal of Microbiology**, v. 60, n. 3, p. 147–154, 2014.
- ZHANG, J. et al. Morphology and phylogenetic position of a trebouxiophycean green alga (Chlorophyta) growing on the rubber tree, *Hevea brasiliensis*, with the description of a new genus and species. **European Journal of Phycology**, v. 43, n. 2, p. 185–193, 2008.
- ZIEMERT, N. et al. The natural product domain seeker NaPDoS: A phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. **PLoS ONE**, v. 7, n. 3, p. 1–9, 2012.

6. Conclusão

As análises dos clusters de *C. acidophila* e *J. hainangensis* sugerem a síntese de carotenoides, ectoínas e sideróforos. A falta de retorno na busca das vias biossintéticas NRPS e PKS de *C. acidophila* não exclui o potencial de metabólitos secundários destas vias, visto que novas vias podem ser inseridas futuramente nos bancos de dados.

Obter estes dados preliminares nestas espécies, direcionam possíveis experimentos *in vitro* os quais possam ser realizados com objetivo de extração de metabólitos secundários bem como elucidar vias metabólicas. Além de contribuir com bancos de dados genômicos com os agrupamentos gênicos preditos para possíveis relações sintenicas e expansão do conhecimento genômico de microalgas.

7. Referências Bibliográficas

- ADAMEK, M. et al. Mining bacterial genomes for secondary metabolite gene clusters. In: **Antibiotics**. [s.l.] Springer, 2017. p. 23–47.
- ALVES, G. S. a Biotecnologia Dos Transgênicos: Precaução É a Palavra De Ordem. **Holos**, v. 2, 2007.
- ANDREWS, S.; OTHERS. **FastQC: a quality control tool for high throughput sequence data** Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom, , 2010.
- ÁVILA-ROMÁN, J. et al. Anti-inflammatory effects of an oxylipin-containing lyophilised biomass from a microalga in a murine recurrent colitis model. **British Journal of Nutrition**, v. 116, n. 12, p. 2044–2052, 2016.
- BAJPAI, V. K. Antimicrobial bioactive compounds from marine algae: A mini review. **Indian Journal of Geo-Marine Sciences**, v. 45, n. 9, p. 1076–1085, 2016.
- BANKEVICH, A. et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. **Journal of Computational Biology**, v. 19, n. 5, p. 455–477, 2012.
- BARRA, L. et al. **The challenge of ecophysiological biodiversity for biotechnological applications of marine microalgae** **Marine Drugs**, 2014.
- BATEMAN, A. UniProt: A worldwide hub of protein knowledge. **Nucleic Acids Research**, v. 47, n. D1, p. D506–D515, 2019.
- BEATTIE, A. J. et al. Ecology and bioprospecting. **Austral Ecology**, v. 36, n. 3, p. 341–356, 2011.
- BENDERLIEV, K. M.; IVANOVA, N. I. High-affinity siderophore-mediated iron-transport system in the green alga *Scenedesmus incrassatulus*. **Planta**, v. 193, n. 2, p. 163–166, 1994.
- BENSON, D.; LIPMAN, D. J.; OSTELL, J. GenBank. v. 21, n. 13, p. 2963–2965, 1993.
- BENTLEY, S. D. et al. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). **Nature**, v. 417, n. 6885, p. 141–147, 2002.
- BLIN, K. et al. The antiSMASH database version 2: A comprehensive resource on secondary metabolite biosynthetic gene clusters. **Nucleic Acids Research**, v. 47, n. D1, p. D625–D630, 2019a.
- BLIN, K. et al. AntiSMASH 5.0: Updates to the secondary metabolite genome mining pipeline. **Nucleic Acids Research**, v. 47, n. W1, p. W81–W87, 2019b.

- BRIERLEY, C. L.; BRIERLEY, J. A. Progress in bioleaching: Part B: Applications of microbial processes by the minerals industries. **Applied Microbiology and Biotechnology**, v. 97, n. 17, p. 7543–7552, 2013.
- BUCHFINK, B.; XIE, C.; HUSON, D. H. Fast and sensitive protein alignment using DIAMOND. **Nature methods**, v. 12, n. 1, p. 59–60, 2015.
- CADORET, J. P.; GARNIER, M.; SAINT-JEAN, B. **Microalgae, Functional Genomics and Biotechnology**. [s.l.] Elsevier, 2012. v. 64
- CHAVALI, A. K.; RHEE, S. Y. Bioinformatics tools for the identification of gene clusters that biosynthesize specialized metabolites. **Briefings in bioinformatics**, v. 19, n. 5, p. 1022–1034, 2018.
- CHENG, H.; HUANG, H.; HUANG, G. Synthesis and antitumor activity of epothilone B. **European Journal of Medicinal Chemistry**, v. 157, p. 925–934, 2018.
- CHEUNG, J. et al. Molecular characterization of staphyloferrin B biosynthesis in *Staphylococcus aureus*. **Molecular Microbiology**, v. 74, n. 3, p. 594–608, 2009.
- CHOI, Y. H. et al. Mycosporine-like amino acids promote wound healing through focal adhesion kinase (FAK) and mitogen-activated protein kinases (MAP kinases) signaling pathway in keratinocytes. **Marine Drugs**, v. 13, n. 12, p. 7055–7066, 2015.
- CHOUHDARY, S.; SAVE, S. N.; VAVILALA, S. L. Unravelling the inhibitory activity of *Chlamydomonas reinhardtii* sulfated polysaccharides against α -Synuclein fibrillation. **Scientific Reports**, v. 8, n. 1, p. 1–12, 2018.
- COUSO, I. et al. Synthesis of carotenoids and regulation of the carotenoid biosynthesis pathway in response to high light stress in the unicellular microalga *Chlamydomonas reinhardtii*. **European Journal of Phycology**, v. 47, n. 3, p. 223–232, 2012.
- CUARESMA, M. et al. Productivity and selective accumulation of carotenoids of the novel extremophile microalga *Chlamydomonas acidophila* grown with different carbon sources in batch systems. **Journal of Industrial Microbiology and Biotechnology**, v. 38, n. 1, p. 167–177, 2011.
- CZECH, L. et al. Role of the extremolytes ectoine and hydroxyectoine as stress protectants and nutrients: Genetics, phylogenomics, biochemistry, and structural analysis. **Genes**, v. 9, n. 4, p. 1–58, 2018.
- DARIENKO, T.; PRÖSCHOLD, T. The genus *Jaagichlorella* Reisingl (Trebouxiophyceae,

- Chlorophyta) and its close relatives: An evolutionary puzzle. **Phytotaxa**, v. 388, n. 1, p. 47–68, 2019.
- DE LOS REYES, C. et al. Oxylipins from the microalgae *Chlamydomonas debaryana* and *Nannochloropsis gaditana* and their activity as TNF- α inhibitors. **Phytochemistry**, v. 102, p. 152–161, 2014.
- DEAN, A. P. et al. Metabolic adaptation of a *Chlamydomonas acidophila* strain isolated from acid mine drainage ponds with low eukaryotic diversity. **Science of the Total Environment**, v. 647, p. 75–87, 2019.
- DOLGANYUK, V. et al. Microalgae: A Promising Source of Valuable Bioproducts. **Biomolecules**, v. 10, n. 8, p. 1–24, 2020.
- EL-GEBALI, S. et al. The Pfam protein families database in 2019. **Nucleic Acids Research**, v. 47, n. D1, p. D427–D432, 2019.
- EWING, B.; GREEN, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. **Genome Research**, v. 8, n. 3, p. 186–194, 1998.
- FENIZIA, S. et al. Ectoine from Bacterial and Algal Origin Is a Compatible Solute in Microalgae. p. 1–13, 2020.
- FUČÍKOVÁ, K. et al. New phylogenetic hypotheses for the core Chlorophyta based on chloroplast sequence data. **Frontiers in Ecology and Evolution**, v. 2, n. OCT, p. 1–12, 2014.
- GHASEMI, Y. et al. Antifungal and antibacterial activity of the microalgae collected from paddy fields of Iran: characterization of antimicrobial activity of *Chroococcus dispersus*. 2007.
- GOLLERY, M. Handbook of Hidden Markov models in bioinformatics. **Handbook of Hidden Markov Models in Bioinformatics**, p. 1–157, 2008.
- GONG, Y. et al. Microalgae as platforms for production of recombinant proteins and valuable compounds: Progress and prospects. **Journal of Industrial Microbiology and Biotechnology**, v. 38, n. 12, p. 1879–1890, 2011.
- GORDON, A.; HANNON, G. Fastx-toolkit. FASTQ/A short-reads pre-processing tools. 2010. **Unpublished available online at: http://hannonlab.cshl.edu/fastx_toolkit**, 2017.
- GÓRSKA, A.; SLODERBACH, A.; MARSZAŁŁ, M. P. Siderophore-drug complexes: Potential medicinal applications of the “Trojan horse” strategy. **Trends in Pharmacological Sciences**, v. 35, n. 9, p. 442–449, 2014.

- GUARNIERI, M. T.; PIENKOS, P. T. Algal omics: Unlocking bioproduct diversity in algae cell factories. **Photosynthesis Research**, v. 123, n. 3, p. 255–263, 2015.
- GUIRY, M.D. & GUIRY, G. M. **AlgaeBase. World-wide electronic publication, National University of Ireland, Galway**. Disponível em: <<https://www.algaebase.org>>. Acesso em: 11 ago. 2020.
- GUREVICH, A. et al. QUASt: Quality assessment tool for genome assemblies. **Bioinformatics**, v. 29, n. 8, p. 1072–1075, 2013.
- HARRIS, E. H. Chlamydomonas as a model organism. **Annual Review of Plant Biology**, v. 52, n. 1, p. 363–406, 2001.
- HIDER, R. C.; KONG, X. Chemistry and biology of siderophores. **Natural Product Reports**, v. 27, n. 5, p. 637–657, 2010.
- HUMANN, J. L. et al. Structural and functional annotation of eukaryotic genomes with GenSAS. In: **Gene Prediction**. [s.l.] Springer, 2019. p. 29–51.
- HUNDLE, B. S. et al. Functional expression of zeaxanthin glucosyltransferase from *Erwinia herbicola* and a proposed uridine diphosphate binding site. **Proceedings of the National Academy of Sciences of the United States of America**, v. 89, n. 19, p. 9321–9325, 1992.
- JAYSHREE, A.; JAYASHREE, S.; THANGARAJU, N. *Chlorella vulgaris* and *Chlamydomonas reinhardtii*: Effective antioxidant, antibacterial and anticancer mediators. **Indian Journal of Pharmaceutical Sciences**, v. 78, n. 5, p. 575–581, 2016.
- JORGE, C. D. et al. Potential applications of stress solutes from extremophiles in protein folding diseases and healthcare. **Extremophiles**, v. 20, n. 3, p. 251–259, 2016.
- KAUTSAR, S. A. et al. MIBiG 2.0: A repository for biosynthetic gene clusters of known function. **Nucleic Acids Research**, v. 48, n. D1, p. D454–D458, 2020.
- LEWIS, K. Platforms for antibiotic discovery. **Nature Reviews Drug Discovery**, v. 12, n. 5, p. 371–387, 2013.
- LINDEN, H. Carotenoid hydroxylase from *Haematococcus pluvialis*: cDNA sequence, regulation and functional complementation. **Biochimica et Biophysica Acta - Gene Structure and Expression**, v. 1446, n. 3, p. 203–212, 1999.
- MA, S. et al. A novel species in the genus *Heveochlorella* (Trebouxiophyceae, Chlorophyta) witnesses the evolution from an epiphytic into an endophytic lifestyle in tree-dwelling green

- algae. **European Journal of Phycology**, v. 48, n. 2, p. 200–209, 2013.
- MADDEN, T. The BLAST sequence analysis tool. **The BLAST Sequence Analysis Tool**, n. Md, p. 1–17, 2013.
- MARCHLER-BAUER, A. et al. CDD: NCBI's conserved domain database. **Nucleic Acids Research**, v. 43, n. D1, p. D222–D226, 2015.
- MARCHLER-BAUER, A. et al. CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. **Nucleic Acids Research**, v. 45, n. D1, p. D200–D203, 2017.
- MAYFIELD, S. P. et al. *Chlamydomonas reinhardtii* chloroplasts as protein factories. **Current Opinion in Biotechnology**, v. 18, n. 2, p. 126–133, 2007.
- MEDEMA, M. H. et al. AntiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. **Nucleic Acids Research**, v. 39, n. SUPPL. 2, p. 339–346, 2011.
- MOLNÁR, I. et al. The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from *Sorangium cellulosum* So ce90. **Chemistry and Biology**, v. 7, n. 2, p. 97–109, 2000.
- MOSBLECH, A.; FEUSSNER, I.; HEILMANN, I. Oxylipins: Structurally diverse metabolites from fatty acid oxidation. **Plant Physiology and Biochemistry**, v. 47, n. 6, p. 511–517, 2009.
- NAITO, K. et al. The pursuit of siderophore secreted by marine phytoplankton *Rhodomonas ovalis*. **Analytical Sciences**, v. 17, n. Table 1, p. 817–819, 2001.
- NAITO, K. et al. Secretion of iron-complexing ligands from *Closterium aciculare* (Charophyceae, Chlorophyta) under iron-deficient conditions. **Phycologia**, v. 43, n. 5, p. 632–634, 2004.
- NEGORO, K. Untersuchungen uber die Vegetation der mineralogenazidotrophen Gewasser Japans. **Science Reports of the Tokyo Bunrika Daigaku, Sect. B.**, v. 6, p. 231–374, 1944.
- NEWMAN, D. J.; CRAGG, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. **Journal of Natural Products**, v. 83, n. 3, p. 770–803, 2020.
- NITHYA, C. et al. Biofilm inhibitory potential of *Chlamydomonas* sp. extract against *Pseudomonas aeruginosa*. **J. Algal Biomass Utiln**, v. 5, n. 4, p. 74–81, 2014.
- NOVOVESKÁ, L. et al. Microalgal carotenoids: A review of production, current markets,

- regulations, and future direction. **Marine Drugs**, v. 17, n. 11, p. 1–21, 2019.
- O'NEILL, E. Mining natural product biosynthesis in eukaryotic algae. **Marine Drugs**, v. 18, n. 2, 2020.
- OLSSON, S. et al. Transcriptional response to copper excess and identification of genes involved in heavy metal tolerance in the extremophilic microalga *Chlamydomonas acidophila*. **Extremophiles**, v. 19, n. 3, p. 657–672, 2015.
- OTTO, B. et al. Laccase-like enzyme activities from chlorophycean green algae with potential for bioconversion of phenolic pollutants. **FEMS Microbiology Letters**, v. 362, n. 11, p. 1–8, 2015.
- PAL, R.; CHOUDHURY, A. K. **An introduction to phytoplanktons: Diversity and ecology**. [s.l: s.n.].
- PALENIK, B. et al. The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. **Proceedings of the National Academy of Sciences of the United States of America**, v. 104, n. 18, p. 7705–7710, 2007.
- PEREZ-GARCIA, O. et al. Heterotrophic cultures of microalgae: Metabolism and potential products. **Water Research**, v. 45, n. 1, p. 11–36, 2011.
- PRÖSCHOLD, T. et al. Molecular phylogeny and taxonomic revision of *Chlamydomonas* (Chlorophyta). I. Emendation of *Chlamydomonas ehrenberg* and *Chloromonas gobi*, and description of *Oogamochlamys* gen. nov. and *Lobochlamys* gen. nov. **Protist**, v. 152, n. 4, p. 265–300, 2001.
- QIAN, P. Y.; XU, Y.; FUSETANI, N. Natural products as antifouling compounds: recent progress and future perspectives. **Biofouling**, v. 26, n. 2, p. 223–234, 2009.
- RADDADI, N. et al. Biotechnological applications of extremophiles, extremozymes and extremolytes. **Applied Microbiology and Biotechnology**, v. 99, n. 19, p. 7907–7913, 2015.
- RENUKADEVI, K. P.; SARAVANA, P. S.; ANGAYARKANNI, J. ANTIMICROBIAL AND ANTIOXIDANT ACTIVITY OF *CHLAMYDOMONAS REINHARDTII* SP. **International journal of Pharmaceutical Sciences and Research**, v. 2, n. 6, p. 1467, 2011.
- ŘEZANKA, T. et al. LC-MS/APCI identification of glucoside esters and diesters of astaxanthin from the snow alga *Chlamydomonas nivalis* including their optical stereoisomers. **Phytochemistry**, v. 88, p. 34–42, 2013.
- SALOMÉ, P. A.; MERCHANT, S. S. A series of fortunate events: Introducing *chlamydomonas*

- as a reference organism. **Plant Cell**, v. 31, n. 8, p. 1682–1707, 2019.
- SANDERS, W. B. et al. Heveochlorella (Trebouxiophyceae): a little-known genus of unicellular green algae outside the Trebouxiales emerges unexpectedly as a major clade of lichen photobionts in foliicolous communities. **Journal of Phycology**, v. 52, n. 5, p. 840–853, 2016.
- SANTOS, B. G. S. DOS. Prospecção de Microalgas para uma Mineração Sustentável: Caracterização de um Isolado Fitoplanctônico da Floresta Nacional de Carajás – PA. **Dissertação (Mestrado Profissional em Uso Sustentável de Recursos Tropicais) – Instituto Tecnológico Vale. Pará**, 2017.
- SCRANTON, M. A. et al. Chlamydomonas as a model for biofuels and bio-products production. **Plant Journal**, v. 82, n. 3, p. 523–531, 2015.
- SENHORINHO, G. N. A.; LANNÉR, C.; SCOTT, J. A. Effect of green microalgal extracts exhibiting antibacterial activity on viability of human malignant and non-malignant cells. **Phycological Research**, v. 67, n. 2, p. 145–151, 2019.
- SHAH, M. M. R. et al. Astaxanthin-producing green microalga Haematococcus pluvialis: From single cell to high value commercial products. **Frontiers in Plant Science**, v. 7, n. APR2016, 2016.
- SHELEST, E. et al. Multimodular type I polyketide synthases in algae evolve by module duplications and displacement of AT domains in trans. **BMC Genomics**, v. 16, n. 1, p. 1–15, 2015.
- SINGH, M.; CHAUDHARY, S.; SAREEN, D. Non-ribosomal peptide synthetases: Identifying the cryptic gene clusters and decoding the natural product. **Journal of Biosciences**, v. 42, n. 1, p. 175–187, 2017.
- SINGH, P. et al. **Microbial Community Dynamics of Extremophiles/Extreme Environment**. [s.l.] Elsevier Inc., 2019.
- SOUZA, L. S. et al. Morphological and ultrastructural characterization of the acidophilic and lipid-producer strain Chlamydomonas acidophila LAFIC-004 (Chlorophyta) under different culture conditions. **Protoplasma**, v. 254, n. 3, p. 1385–1398, 2017.
- SRINIVASAN, R.; BABU, S.; GOTHANDAM, K. M. Accumulation of phytoene, a colorless carotenoid by inhibition of phytoene desaturase (PDS) gene in Dunaliella salina V-101. **Bioresource Technology**, v. 242, p. 311–318, 2017.

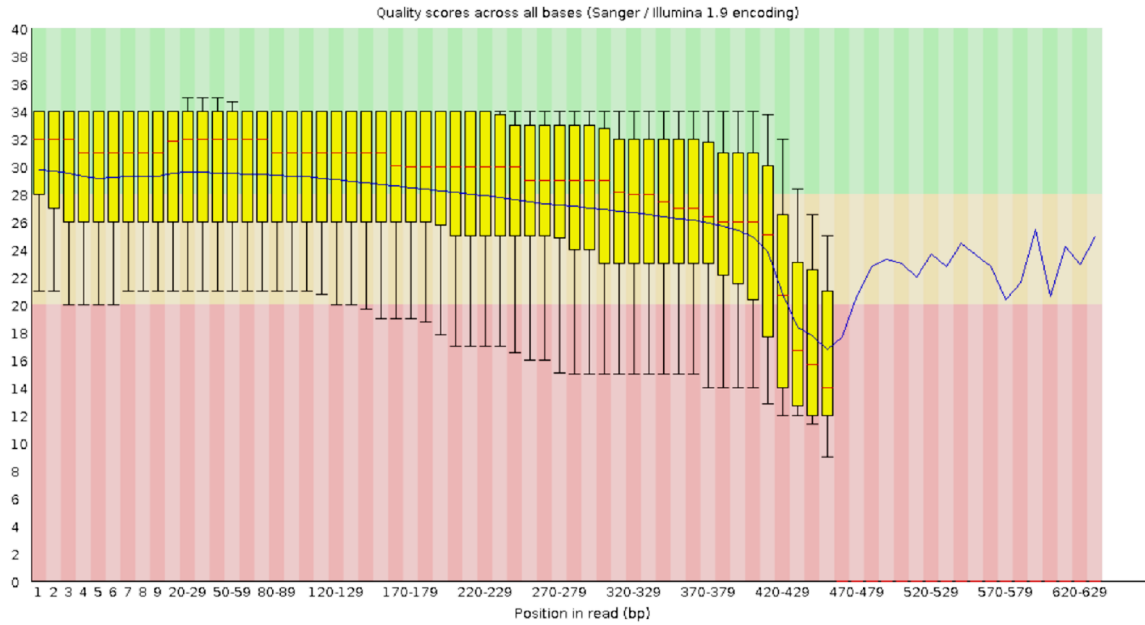
- STANKE, M. et al. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. **BMC bioinformatics**, v. 7, n. 1, p. 62, 2006.
- STEINBRENNER, J.; SANDMANN, G. Transformation of the green alga *Haematococcus pluvialis* with a phytoene desaturase for accelerated astaxanthin biosynthesis. **Applied and Environmental Microbiology**, v. 72, n. 12, p. 7477–7484, 2006.
- SUH, S. S. et al. Anti-inflammation activities of mycosporine-like amino acids (MAAs) in response to UV radiation suggest potential anti-skin aging activity. **Marine Drugs**, v. 12, n. 10, p. 5174–5187, 2014.
- SULLIVAN, M. J.; PETTY, N. K.; BEATSON, S. A. Easyfig: A genome comparison visualizer. **Bioinformatics**, v. 27, n. 7, p. 1009–1010, 2011.
- SUTAK, R.; CAMADRO, J. M.; LESUISSE, E. Iron Uptake Mechanisms in Marine Phytoplankton. **Frontiers in Microbiology**, v. 11, n. November, p. 1–14, 2020.
- TAKAICHI, S. Carotenoids in Phototrophic Microalgae: Distributions and Biosynthesis. In: **Pigments from Microalgae Handbook**. [s.l.] Springer, 2020. p. 19–41.
- THIRUMURUGAN, D. et al. An introductory chapter: secondary metabolites. **Second metab—sources Appl**, p. 1–21, 2018.
- VINGIANI, G. M. et al. Microalgal enzymes with biotechnological applications. **Marine Drugs**, v. 17, n. 8, p. 1–20, 2019.
- WANG, H. et al. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. **Proceedings of the National Academy of Sciences of the United States of America**, v. 111, n. 25, p. 9259–9264, 2014.
- WANG, L. et al. Identification of genetic variations associated with epsilon-poly-lysine biosynthesis in *Streptomyces albulus* ZPM by genome sequencing. **Scientific reports**, v. 5, p. 9201, 2015.
- WERKHÄUSER, N.; BILSTEIN, A.; SONNEMANN, U. Treatment of Allergic Rhinitis with Ectoine Containing Nasal Spray and Eye Drops in Comparison with Azelastine Containing Nasal Spray and Eye Drops or with Cromoglycic Acid Containing Nasal Spray. **Journal of Allergy**, v. 2014, p. 1–13, 2014.
- WRIGHT, G. D. Something old, something new: Revisiting natural products in Antibiotic drug discovery. **Canadian Journal of Microbiology**, v. 60, n. 3, p. 147–154, 2014.

ZHANG, J. et al. Morphology and phylogenetic position of a trebouxiophycean green alga (Chlorophyta) growing on the rubber tree, *Hevea brasiliensis*, with the description of a new genus and species. **European Journal of Phycology**, v. 43, n. 2, p. 185–193, 2008.

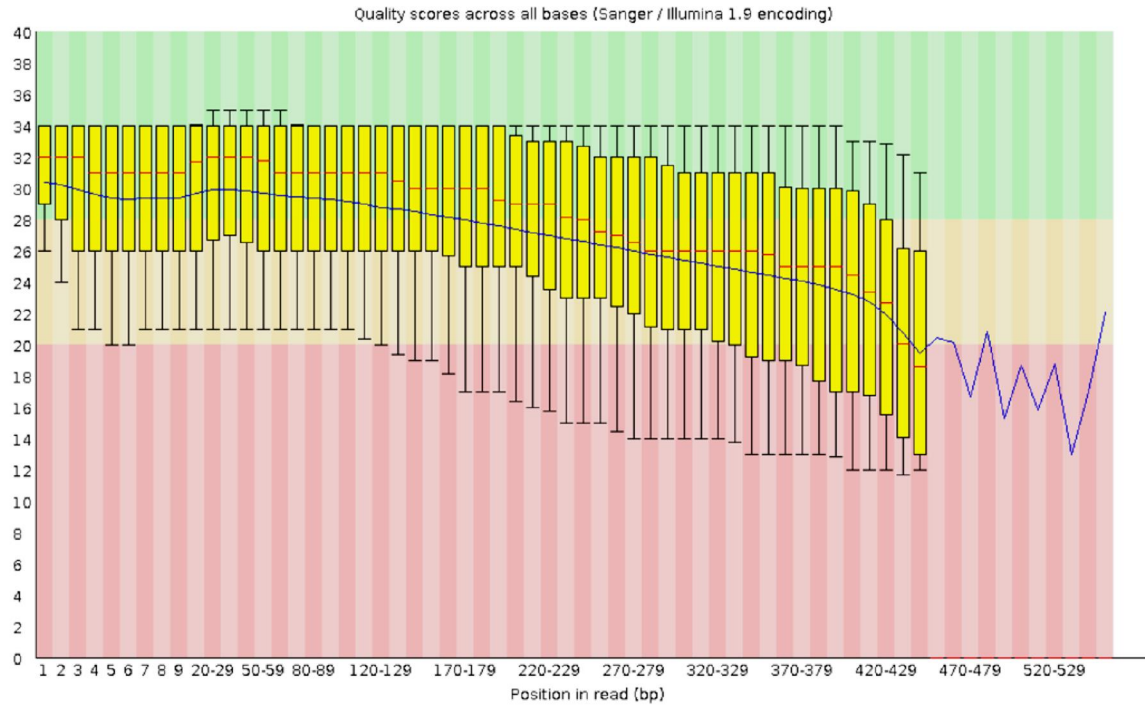
ZIEMERT, N. et al. The natural product domain seeker NaPDoS: A phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. **PLoS ONE**, v. 7, n. 3, p. 1–9, 2012.

Anexo 1, Gráficos de qualidade dos sequenciamentos obtidos pelo programa FASTQC em 1. *Chlamydomonas acidophila* e em 2. *Jaagichlorella hainangensis*.

1. *Chlamydomonas acidophila*



2. *Jaagichlorella hainangensis*



Anexo 2. Table of all clusters identified by antiSMASH on the fungal and bacterial versions on the genome contigs of *Chlamydomonas acidophila* ITV01

Region_*	Type	Most known similar cluster	Bacteria			Fungi			Status of the identified cluster
			with annotation	without annotation	with annotation	without annotation	with annotation	without annotation	
Region 19.1	Terpene	none	X			X		yes	
Region 152.1	Terpene	none	X	X		X	X	yes	
Region 204.1	T3PKS	none	X	X		X		yes	
Region 569.1	NAGGN	none	X	X		X		yes	
Region 865.1	NRPS	none	X	X		X		yes	
Region 1480.1	Terpene	carotenoid	X	X		X	X	yes	
Region 1583.1	Terpene	carotenoid	X	X		X	X	yes	
Region 2676.1	Terpene	none	X	X		X		yes	
Region 3030.1	ectoine			X				yes	
Region 3272.1	arylpolyene	arylpolyene	X	X		X	X	yes	
Region 3470.1	arylpolyene	apevf	X	X		X		yes	
Region 4163.1	Terpene	none						yes	
Region 4348.1	NRPS	none	X			X	X	yes	
Region 5403.1	siderophore	desferrioxamine E		X			X	yes	
Region 5786.1	Homoserina Lactona						X	yes	
Region 6984.1	T3PKS	none	X			X		yes	
Region 7075.1	Terpene						X	yes	
Region 8629.1	Terpene			X				yes	
Region 9013.1	siderophore			X				yes	
Region 10003.1	T3PKS						X	yes	
Region 11672.1	Betalactone			X				yes	
Region 12599.1	Terpene		X	X		X	X	yes	
Region 15493.1	Homoserina Lactona		X	X		X	X	yes	
Region 14177.2	T3PKS			X			X	yes	

* The number of the region that identifies the contig of origin in the microalga genome

Anexo 3. Table of BLAST analysis performed in the identified genes of each Biosynthetic Gene Cluster in *Chlamydomonas acidophila* ITV01

Region*	Type	Most known similar cluster (similarity in percentage)	identification of the gene in the genome	BLASTP Identity (identity in %)
Region 1480. 1	Terpene	Carotenoid (42%)	ctg1480_2	phytoene/squalene synthase family protein 70.50%
			ctg1480_3	phytoene desaturase 73.62%
			ctg1480_4	RNA polymerase sigma factor 88.24%
			ctg1480_5	iron complex outermembrane receptor protein 53.04%
			ctg1480_6	MFS transporter 60.36%
			ctg1583_3	phytoene/squalene synthase family protein 100%
Region 1583. 1	Terpene	Carotenoid (85%)	ctg1583_4	phytoene desaturase 100%
			ctg1583_6	lycopene beta-cyclase CrTY 99.64%
			ctg1583_7	glycosyltransferase family I protein 100%
			ctg3030_5	ectoine synthase 98.50%
Region 3030. 1	Ectoine	Ectoine (66%)	ctg3030_6	diaminobutyrate--2-oxoglutarate transaminase 100%
			ctg3030_4	ectoine hydroxylase 99.34%
			ctg3030_3	aspartate kinase 100%
			ctg3470_1	beta-ketoacyl-[acyl-carrier-protein] synthase family protein 100%
Region 3470. 1	Arylpolyene	APE Vf (15%)	ctg3470_3	Select seq ref WP_019405671.1 3-oxoacyl-ACP reductase FabG (100%)
			ctg3470_4	beta-ketoacyl-ACP synthase (93.04%)
			ctg3470_7	LysR family transcriptional regulator (100%)
			ctg5403_1	lucA/lucC family siderophore biosynthesis protein (99.79%)
Region 5403. 1	Siderophore	Deferoxamina E (75%)	ctg5403_2	acetyltransferase (99.49%)
			ctg5403_3	MFS transporter (99.48%)
			ctg5403_4	SidA/lucD/PvdA family monooxygenase (100%)

* The number of the region that identifies the contig of origin in the microalga genome

Anexo 4. Table of the NRPS and PKS domains found in each BGC predicted by antiSMASH in *Chlamydomonas acidophila* ITV01

Region (Type)	Domains (CDD/PFAM)	key:
204 (T3PKS)	3KS, CSS	NRPS Non-Ribosomal Peptide Synthetase
865.1 (NRPS)	A, PP-BIND	PKS polyketide synthase
4348.1 (NRPS)	A, PP-BIND	T3PKS polyketide synthase type III
6984.1 (T3PKS)	3KS	3KS 3-ketoacyl-CoA synthase
10003.1 (T3PKS)	3KS	CSS Chalcone and stilbene synthase (C-terminal domain)
15493.1 (T3PKS)	CSS, CSSt	A adenylation domain
		PP-BIND Phosphopantetheine attachment site
		CSSt Chalcone and stilbene synthase (T-terminal domain)

Anexo 5. Table of all clusters identified by antiSMASH on the fungal and bacterial versions on the genome contigs of *Jaagichlorella hainangensis* ITV09

Region_*	Type	Most known similar cluster	bacteria		fungi		Status of the identified cluster	
			with anotation	without anotation	with anotation	without anotation	contig edge	cluster
Region 2.1	Terpeno		x	x	x	x		no
Region 3.1	Terpeno	carotenoid	x	x	x	x		no
Region 6.1	T1PKS		x	x	x	x		no
Region 6.2	Terpeno		x	x	x	x		no
Region 7.1	Siderórfó	Deferoxamina E	x	x	x	x		no
Region 10.1	NRPS		x	x	x			no
Region 10.2	NRPS		x		x			no
Region 11.1	Siderórfó			x		x		no
Region 13.1	Resorcinol		x	x	x			no
Region 14.1	Ectoína	Ectoíne	x	x	x	x		no
Region 14.2	NRPS/Aril Polieno	APE Ec	x	x	x	x		no
Region 27.1	Bacteriocina		x	x	x	x		no
Region 31.1	Lassopectideo			x				no
Region 43.1	Betalactona		x	x	x			no
Region 51.1	T3PKS		x	x	x	x		no
Region 52.1	Homoserina Lactona/Lassopectideo	Subtserin	x	x	x			no
Region 67.1	NRPS/Aril Polieno	APE Vf	x	x	x	x		no
Region 71.1	NRPS		x		x	x		no
Region 73.1	Betalactona	Fengicina	x	x	x	x		no
Region 91.1	Betalactona		x	x	x			no
Region 97.1	Terpeno		x		x	x		no
Region 100.1	Bacteriocina		x	x	x	x		no
Region 150.1	T3PKS		x		x	x		no
Region 152.1	Fosfonato	Salecan	x	x	x	x		no
Region 168.1	NRPS	Puwainaphycin	x	x	x			no
Region 188.1	T1PKS		x		x			no

Cont.

Region_*	Type	Most known similar cluster	bacteria		fungi		Status of the indentified cluster	
			with annotation	without annotation	with annotation	without annotation	contig edge	cluster
Region 257.1	NRPS	Fuscachelin	x	x	x	x	no	no
Region 291.1	TIPKS		x		x	x	no	no
Region 306.1	T3PKS		x	x	x	x	no	no
Region 314.1	Resorcinol		x	x	x	x	yes	yes
Region 320.1	Fosfonato	Echocide	x	x	x	x	no	no
Region 405.1	Homosserina Lactona			x			no	no
Region 574.1	Bacteriocina		x	x	x	x	no	no
Region 632.1	NRPS		x	x	x	x	no	no
Region 636.1	Betalactona		x	x	x	x	yes	yes
Region 667.1	Siderórfo	Ochrobactin	x	x	x	x	no	no
Region 702.1	Aril Polieno		x		x	x	no	no
Region 718.1	Betalactona		x	x	x	x	yes	yes
Region 924.1	Resorcinol	Chejuenolide	x	x	x	x	yes	yes
Region 935.1	Aril Polieno			x		x	yes	yes
Region 957.1	T3PKS		x	x	x	x	yes	yes
Region 1017.1	Siderórfo		x	x	x	x	yes	yes
Region 1423.1	Homosserina Lactona			x			yes	yes
Region 1496.1	Terpeno			x			yes	yes
Region 1543.1	Terpeno			x			yes	yes
Region 1631.1	Bacteriocina		x	x	x	x	yes	yes
Region 1901.1	TIPKS	Pyrronazol B		x			yes	yes
Region 2019.1	NRPS		x	x	x	x	yes	yes
Region 2151.1	TIPKS			x			yes	yes
Region 2259.1	Homosserina Lactona		x		x		yes	yes
Region 2293.1	Terpeno			x		x	yes	yes
Region 2614.1	Terpeno			x			yes	yes
Region 2875.1	TIPKS			x		x	yes	yes

Cont.

Region*	Type	Most known similar cluster	bacteria			fungi		Status of the identified cluster
			with anotation	without anotation	with anotation	without anotation	contig edge	
Region 3043.1	Terpeno		x	x	x		yes	
Region 3424.1	Terpeno		x	x	x	x	yes	
Region 3461.1	NRPS	Icosalide A/Icosalide B		x			yes	
Region 3573.1	NRPS	Anabenozeptin		x			yes	
Region 3843.1	T1PKS			x		x	yes	

* The number of the region that identifies the contig of origin in the microalga genome

Anexo 6. Table of BLAST analysis performed in the identified genes of each Biosynthetic Gene Cluster in *Jaagichlorella hainangensis* ITV09

Most similar known cluster, MIBiG Access and Reference organism	identification of the gene in the genome	BLASTP Identity (identity in %)	
Carotenoid (Região 3.1): BGC0001086 <i>Sphingomonas</i> spp.	ctg3_288	Fitoeno sintase (76.75%)	
	ctg3_290	fitoeno desidrogenase (87.60%)	
	ctg3_291	licopeno beta-ciclase CrtY (82.43%)	
	ctg3_283	Peptidase da família S9 (70.83%)	
	ctg3_285	alfa / beta hidrolase (83.01%)	
	ctg3_292	proteína hipotética (90.69%)	
	ctg3_297	proteína hipotética (77.44%)	
	ctg3_302	poli-hidroxitirato despolimerase (87.75%)	
	ctg7_277	LucA_LucC (97.89%)	
	ctg7_275:	lisina / ornitina N-monooxigenase (95.91%)	
Deferoxamine E (Região 7.1) BGC0001572 <i>Pantoea agglomerans</i>	ctg7_276:	acetiltransferase (98.51%)	
	ctg7_281	transportador de efluxo de múltiplas drogas tipo MFS (94.19%)	
	ctg14_45	ectoína sintetase (100%)	
	ctg14_44	diaminobutirato - 2-oxoglutarato transaminase (99.54%)	
	ctg52_42	asparagina sintase (55.03%)	
	ctg52_41	proteína B2 da biossíntese de lassopeptídeo (100%)	
	ctg52_43	Isopeptidase do lassopeptídeo da família Atxe2 (61.68%)	
	ctg257_35	subunidade F da enterobactina sintase (99%)	
	ctg257_38	Proteína de ligação a AMP (99%)	
	ctg257_16	alfa / beta hidrolase (98.53%)	
Fuscachelin (Região 257.1): BGC0000359 <i>Thermobifida fusca</i> YX	ctg257_34	2,3-di-hidro-2,3-di-hidroxi-benzoato desidrogenase (98.41%)	
	ctg257_36	proteína de ligação à fosfopantetina (98.82%)	
	ctg257_37	proteína da família isochorismatase (98.57%)	
	ctg257_39	isochorismate sintase (98.74%)	
	ctg257_42	Oxidoreductase da classe LLM da família TIGR03571	
	ctg257_46	proteína da família flavina reductase (99.48%)	
	ctg257_57	desidrogenase de ligação ao zinco (99.38%)	
	Ectoine BGC0000858 <i>Methylobacter marinus</i>	ctg14_44	diaminobutirato - 2-oxoglutarato transaminase (99.54%)
		ctg52_42	asparagina sintase (55.03%)
		ctg52_41	proteína B2 da biossíntese de lassopeptídeo (100%)
ctg52_43		Isopeptidase do lassopeptídeo da família Atxe2 (61.68%)	
ctg257_35		subunidade F da enterobactina sintase (99%)	
ctg257_38		Proteína de ligação a AMP (99%)	
ctg257_16		alfa / beta hidrolase (98.53%)	
ctg257_34		2,3-di-hidro-2,3-di-hidroxi-benzoato desidrogenase (98.41%)	
ctg257_36		proteína de ligação à fosfopantetina (98.82%)	
ctg257_37		proteína da família isochorismatase (98.57%)	
ctg257_39	isochorismate sintase (98.74%)		
Subterisin (Região 52.1): BGC0001786 <i>Novosphingobium subterraneum</i>	ctg257_42	Oxidoreductase da classe LLM da família TIGR03571	
	ctg257_46	proteína da família flavina reductase (99.48%)	
	ctg257_57	desidrogenase de ligação ao zinco (99.38%)	
	ctg14_44	diaminobutirato - 2-oxoglutarato transaminase (99.54%)	
	ctg52_42	asparagina sintase (55.03%)	
	ctg52_41	proteína B2 da biossíntese de lassopeptídeo (100%)	
	ctg52_43	Isopeptidase do lassopeptídeo da família Atxe2 (61.68%)	
	ctg257_35	subunidade F da enterobactina sintase (99%)	
	ctg257_38	Proteína de ligação a AMP (99%)	
	ctg257_16	alfa / beta hidrolase (98.53%)	

Cont.

Most similar known cluster, MIBiG Access and Reference organism	identification of the gene in the genome	BLASTP Identity (identity in %)
Echosides (Região 320.1) BGC0000340 <i>Streptomyces sp. LZ35</i>	ctg320_9	fosfoenolpiruvato fosfomutase (100%)
	ctg320_10	aminotransferase dependente de fosfato piridoxal (99.74%)
	ctg320_19	proteína hipotética [Mesorhizobium sp. Pch-S] (98.92%)
	ctg320_21	proteína da família fosfocolina citidililtransferase (100%)
	ctg320_22	enzima aminotransferase classe I / II dependente de fosfato de piridoxal (99.01%)
Ochrobactin (Região 667.1) BGC0001713 <i>Ochrobactrum anthropi</i>	ctg667_23	fosfoenolpiruvato mutase (100%)
	ctg667_7	Proteína da família lucA / lucC (96.89%)
	ctg667_8	N-acetiltransferase da Família GNAT(97.76%)
	ctg667_9	Proteína da família lucA / lucC (99.67%)
	ctg667_10	Monooxigenase da família SidA / lucD / PvdA (99.77%)
Ochrobactin (Região 1017.1) BGC0001713 <i>Ochrobactrum anthropi</i>	ctg1017_1	Monooxigenase da família SidA / lucD / PvdA (79.14%)
	ctg1017_4	Proteína da família lucA / lucC (82.91%)
	ctg1017_3	N-acetiltransferase da Família GNAT(88.52%)
	ctg1017_2	Proteína da família lucA / lucC (92.20%)

Anexo 7. Table of the NRPS and PKS domains found in each BGC predicted by antiSMASH in *Jaagichlorella hainangensis* ITV09

Region (Type)	Domains (CDD/PFAM)	Key:
6.1 (hybrid)	ACP, A, A, MT, PCP, KS, AT, DH, ER, KR, PCP, NAD, AmT	NRPS Non-Ribosomal Peptide Synthetase
10.1 (NRPS)	AmT, (2) SDR, A, ACP	PKS polyketide synthase
10.2 (NRPS)	A, ACP	KS ketosynthase
51.1 (T3PKS)	3KS, CSS,	AT acyltransferase
71.1 (NRPS)	A, NAD(P)-bind, Acyl-CoA, MT	DH dehydratase
150.1 (T3PKS)	3KS, ACP	ER enoylreductase
168.1 (NRPS)	A, C, ACP	KR ketoreductase
188.1 (T1PKS)	KS, AT, ER, DH	ACP acyl carrier protein
257.1 (NRPS)	C, A, PCP, TE, KR, A, ER	MT methyltransferase
291.1 (T1PKS)	KS, AT, DH, ER, KR, PP-BIND	TE thioesterase
306.1 (T3PKS)	3KS, ACP	C condensation domain
632.1 (NRPS)	A, PCP	A adenylation domain
957.1 (T3PKS)	3KS, ACP	PCP peptidyl carrier protein
1901.1 (T1PKS)	KS, PCP	NAD(P)-bind NAD(P)-binding domain
2019.1 (NRPS)	C, PCP	AmT Aminotransferase
2875.1 (T1PKS)	KS	SDR Short-chain dehydrogenase/reductase
		Acyl-CoA Acyl-CoA dehydrogenase
		3KS 3-ketoacyl-CoA synthase
		CSS Chalcone and stilbene synthase (C-terminal domain)
		PP-BIND Phosphopantetheine attachment site

Anexo 8. Table of PKS identified by NaPDoS in *Jaagichlorella hainangensis* ITV09

Contig	NaPDoS database	Product	Organism
6	EpoD_Q9L8C7_4mod	epothilone	<i>Sorangium cellulosum</i>
10	FabF_Ecoli_FAS	fatty acid synthesis	<i>Streptomyces coelicolor</i>