

UNIVERSIDADE FEDERAL DO AMAZONAS
INSTITUTO DE SAÚDE E BIOTECNOLOGIA
BACHARELADO EM BIOTECNOLOGIA

ANGY YULIZA AYALA ELIZALDE

**Investigação do potencial biotecnológico da bactéria amazônica
Burkholderia gladioli Coa14 via *Genome Mining*.**

Coari
2023

ANGY YULIZA AYALA ELIZALDE

**Investigação do potencial biotecnológico da bactéria amazônica
Burkholderia gladioli Coa14 via *Genome Mining*.**

Trabalho de Conclusão de Curso IV (TCC IV),
apresentado ao Curso de Biotecnologia da
Universidade Federal do Amazonas (UFAM), como
requisito para obtenção do título de Bacharel em
biotecnologia.

Orientador: Prof. Dr. Eraldo Ferreira Lopes

Coari
2023

Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

E43i Elizalde, Angy Yuliza Ayala
Investigação do potencial biotecnológico da bactéria amazônica
Burkholderia gladioli Coa14 via Genome Mining / Angy Yuliza Ayala
Elizalde . 2023
26 f.: 31 cm.

Orientador: Eraldo Ferreira Lopes
TCC de Graduação (Biotecnologia) - Universidade Federal do
Amazonas.

1. AntiSMASH. 2. Data mining. 3. Genome mining. 4.
Burkholderia gladioli. I. Lopes, Eraldo Ferreira. II. Universidade
Federal do Amazonas III. Título

RESUMO

Introdução: Nos últimos anos vem ocorrendo um enorme crescimento e consolidação no campo de *data mining*. Este é um estudo sobre mineração de dados destacando suas funcionalidades, técnicas e aplicações utilizando o genoma de uma linhagem bacteriana amazônica de *Burkholderia gladioli*. Salientando principalmente o processo da *genome mining* como parte de um processo maior de pesquisa em busca de conhecimento na base de dados *antiSMASH*. Foi abordada e evidenciada a atividade da bactéria *Burkholderia gladioli* cuja finalidade e alcance foi a descoberta da bactéria amazônica e seu potencial biotecnológico. Dessa forma é comprovado como a natureza continua servindo de inspiração para preencher as lacunas de conhecimento e desenvolver novas estratégias de pesquisa. **Objetivo:** Investigar a presença dos *clusters* de genes relacionados a compostos bioativos no genoma da bactéria amazônica *Burkholderia gladioli*, depositados publicamente no NCBI, através do *antiSMASH*. **Resultados:** Por meio do *antiSMASH* que evidencia os *clusters* de genes biossintéticos silenciados ou pouco expressos, a *Burkholderia gladioli* apresenta uma ampla variedade de *clusters* benéficos, os mesmos foram separados por modelos e grupo de substâncias, de modo a facilitar e evidenciar a compreensão dos mesmos **Conclusão:** Há evidências significativas das vantagens da bactéria estudada, devido a ser uma bacterias promissora, assim como, buscou se contribuir para o aumento do aproveitamento biotecnológico das linhagens analisada com o presente trabalho.

Palavras-chave: *AntiSMASH*, *Data mining*, *Genome mining*, *Burkholderia gladioli*.

ABSTRACT

Introduction: In recent years there has been an enormous growth and consolidation in the field of data mining. This is a study on data mining highlighting its functionalities, techniques and applications using the genome of an Amazonian bacterial strain of *Burkholderia gladioli*. Mainly emphasizing the genome mining process as part of a larger research process in search of knowledge in the antiSMASH database. The activity of the bacterium *Burkholderia gladioli* was approached and evidenced, whose purpose and scope was the discovery of the Amazonian bacterium and its biotechnological potential. In this way, it is proven how nature continues to serve as an inspiration to fill knowledge gaps and develop new research strategies. **Objective:** To investigate the presence of gene clusters related to bioactive compounds in the genome of the Amazonian bacterium *Burkholderia gladioli*, publicly deposited at the NCBI, using antiSMASH. **Results:** Through the antiSMASH that shows clusters of silenced or poorly expressed biosynthetic genes, *Burkholderia gladioli* presents a wide variety of beneficial clusters, which were separated by models and group of substances, in order to facilitate and highlight their understanding. **Conclusion:** There is evidence of the advantages of the bacterium studied, due to it being a promising bacterium, as well as, it sought to contribute to the increase of the biotechnological use of the strains left with the present work.

Keywords: *AntiSMASH, Data mining, Genome mining, Burkholderia gladioli.*

LISTA DE TABELAS

Tabela 1 - <i>Clusters</i> identificados e divididos por grupo de substâncias.	18
---	----

SUMÁRIO

1. INTRODUÇÃO.....	7
2. FUNDAMENTAÇÃO TEÓRICA	7
2.1. <i>GENOME MINING</i>	7
2.2. BANCOS DE DADOS GENÔMICOS.....	8
2.3. <i>AntiSMASH</i>	9
2.4. AGRUPAMENTOS LOCALIZADOS PELO USO DO <i>AntiSMASH</i>	11
2.5. POTENCIAL BACTERIANO.....	13
3. OBJETIVO GERAL	16
4. OBJETIVO ESPECÍFICO	16
5. MATERIAL E MÉTODOS	17
5.1. METODOLOGIA	17
6. RESULTADOS.....	18
7. DISCUSSÃO.....	20
8. CONCLUSÃO.....	23
REFERÊNCIAS.....	24

1. INTRODUÇÃO

Os microrganismos representam uma fonte importante para novos produtos naturais, na forma de compostos químicos produzidos organicamente. Eles representam um rico reservatório de potencial biotecnológico e têm utilidade comprovada em várias áreas (HANNIGAN, 2019; AIGLE, 2014).

Os metabólitos microbianos especializados são uma importante fonte e inspiração para muitos produtos e processos de interesse biotecnológico e desempenham papéis fundamentais em processos ecológicos. Ainda assim, a maioria dos métodos de isolamento e identificação guiados por bioatividade, amplamente empregados em programas de descoberta de metabólitos, não exploram todo o potencial biossintético de um organismo (LEÃO, 2022).

É evidente o quanto muitos microrganismos são capazes de produzir muito mais compostos do que aqueles que são observados nas condições normais de laboratório. Muitos *clusters* de genes biossintéticos são pouco expressos ou silenciados na ausência de um estímulo particular como determinados nutrientes, compostos sinalizadores, sinais ambientais dentre outros muitas vezes desconhecidos (VARELLA, 2015).

Avanços na genômica evidenciaram que aproximadamente 90% da capacidade biossintética dos organismos ainda não foi explorada em termos biotecnológicos. O contínuo avanço das tecnologias da bioinformática, genômica microbiana e engenharia de vias metabólicas nos colocam na dianteira de uma nova era de descoberta de produtos naturais e desenvolvimento de medicamentos (TEIJARO, 2019).

A *Burkholderia gladioli* é uma bactéria móvel, Gram-negativa que pertence ao complexo *Burkholderia cepacia*. Detém atividades simbióticas sendo capaz de produzir doenças em plantas e em humanos imunossuprimidos. Possui uma ampla gama de aplicações biotecnológicas tais como controle biológico, biorremediação e bioestimulação (LOPES, 2018).

Inicialmente, trabalhos no campo da descoberta de produtos naturais mostraram que muitas moléculas com potencial biotecnológico são metabólitos secundários microbianos. Hoje em dia, eles são a base para a produção de antimicrobianos e antivirais, por exemplo. Sua síntese é orquestrada principalmente por sequências co-localizadas nos genomas, denominadas *clusters* (agrupamentos) de “Genes Biossintéticos” (HANNIGAN, 2019).

Estes *clusters* possuem padrões de sequências, que se repetem entre diferentes organismos. Com o advento do sequenciamento de DNA de nova geração e a criação de grandes bancos de dados de depósitos de sequências, surgiu o interesse em investigar, de maneira massiva, esta informação, na tentativa de estimar a presença dos *clusters* de genes, que possam ter escapado à outras análises.

O *data mining* (mineração de dados) envolve técnicas de diversas disciplinas, tais como tecnologia de bases de dados, armazenamento de dados, estatística, computação de alta performance, reconhecimento de padrões, colheita de informação e análise de dados. Quanto aplicado à Genômica o *data mining* têm seu nome adaptado para *Genome Mining*.

O *Genome Mining* (ou mineração do genoma) refere-se à aplicação de recursos computacionais (bioinformática) para análises de dados de sequenciamento de genoma, visando descobrir vias biológicas com potencial biotecnológico (YANG, 2012), principalmente relacionadas ao metabolismo secundário.

A mineração de genoma se dá início na identificação de genes biossintéticos conservados. Seus produtos gênicos são posteriormente analisados para obter informações sobre possíveis funções na biossíntese e, às vezes, sua especificidade de substrato (BLIN,2019).

As abordagens da mineração de genoma, que geralmente envolvem a identificação dos genes envolvidos na produção de metabólitos secundários, revelaram um potencial biossintético inédito em muitas espécies microbianas. Esses genes codificam as enzimas envolvidas na montagem de peptídeos, regulação,

resistência e síntese de um metabólito secundário, e são fisicamente agrupados em grupos chamados *clusters* (BELKNAP, 2020).

O objetivo deste trabalho é analisar o genoma bacteriano da bactéria amazônica *Burkholderia gladioli*, em quanto à presença de *clusters* de genes metabólicos secundários com capacidades biotecnológicas.

2. FUNDAMENTAÇÃO TEÓRICA

2.1. GENOME MINING

O termo genoma significa o conjunto completo de informações codificadas no DNA de uma unidade celular de um organismo, incluindo tanto os genes como as sequências não codificadoras (CASTRO, 2015).

A mineração de genoma é um método computacional para a detecção e anotação automática de *clusters* de genes biosintéticos (BGCs) a partir de dados genômicos. Esta abordagem tem sido cada vez mais utilizada na descoberta de produtos naturais devido à grande quantidade de dados de sequenciamento que agora estão disponíveis em bancos de dados públicos (RUSSELL, 2020).

A mineração de dados genômicos detecta a presença de vias biossintéticas que permitem que os organismos produzam moléculas de interesse (BLIN, 2019). Vários softwares estão disponíveis publicamente e têm como alvo o encontro de genes de interesse, como o Clutscan (STARCEVIC et. al, 2008) e o SBSPKS toolbox (ANAND et. al, 2010), por exemplo. Um dos principais softwares utilizados para este fim é o *antibiotics and secondary metabolite analysis shell - antiSMASH*.

O *antiSMASH* é capaz de localizar clusters de genes relacionados à produção de polinucleotídeos não ribossomais (NRP), poliquetídeos (PK), terpenos, bacteriocinas, indols, betalactâmicos, sideroforos, fosfoglicolipídios, aminoglicosídeos, lantibióticos, entre outros.

Através do *antiSMASH* utilizaram-se abordagens metabolômicas para descrever o potencial antitumoral do extrato bruto de *Streptomyces sp. BRB081*. Consistiu em uma importante etapa de pré-análises, revelando a biossíntese de

surugamidas e de alguns outros metabólitos citotóxicos, responsáveis pela elevada atividade antitumoral desse extrato (LEITE, 2022).

Pesquisas mais recentes com actinobacterias endofíticas como a *Streptomyces NRRL 30562*, foram identificadas as *munumbicinas* A, B, C, E-4 e E-5 que são promissores antibióticos, endófito de *Kennedia igriscans*, planta do território do norte australiano (VARELLA, 2015).

Uma evidência é a aplicação de co-culturas entre o fungo *liberatella* sp., isolado de a-proteobacteria marinha e de uma ascídia tendo como resultado da interação das células microbianas na co-culturas, levando à produção de diterpenos libertelenonas A - D, as mesmas demonstrando potente atividade citotóxica (VARELLA, 2015).

2.2. BANCOS DE DADOS GENÔMICOS

Atualmente existem três grandes repositórios de sequências de dados biológicos, o NCBI (National Center for Biotechnology Information), o EBI (European Bioinformatics Institute), e o DDBJ (Japan Data Bank).

Sendo talvez o mais importante, o NCBI torna as informações contidas no GenBank, um banco de dados de sequências anotadas, disponíveis sem custo através da Internet, por meio de servidores e uma ampla gama de serviços de recuperação e análise de bases.

É acessível através do NCBI, que é uma divisão da Biblioteca Nacional de Medicina dos Estados Unidos. O site do NCBI é responsável por integrar informações a partir das principais bases de dados de sequências de DNA e proteínas, juntamente com a taxonomia, genomas, mapeamentos, estruturas proteicas e informações de domínio, além da literatura biomédica da revista via PubMed. Através do BLAST que é uma ferramenta de comparação de sequenciamentos disponíveis no NCBI, são fornecidas pesquisas de similaridade de sequência do GenBank e outros bancos de dados de sequenciamento (CASTRO, 2015).

As sequências estão disponíveis para *download* gratuitamente, e podem ter como origem trabalhos de sequenciamento de genomas completos ou parciais

(*drafts*), sequências de RNAs, e aminoácidos. Para a obtenção, muitas vezes não se faz necessário sequer o cadastro na base de dados.

Entre os genomas que estão depositados publicamente no NCBI estão os *Thelebolus globosus culture CBS:113940* e *Arthrinium phaeospermum JCM 6018* cujo DNA genômico estão disponíveis na RIKEN BRC-JCM, são espécies de fungos que demonstraram potencial de biodegradar plástico, por serem fungos nativos da antártica possuem características extremas, os organismos que nesse ambiente vivem demonstram ser capazes de produzir enzimas com melhor capacidade de degradação, dessa forma, os polímeros sendo usados como fonte única de nutrientes e carbono (NOBRE, 2021).

Paenibacillus polymyxa MAS100 bactérias isoladas do solo de vida livre conhecidas por promover o crescimento de plantas e suprimir patógenos de plantas. As plantas tratadas com *Paenibacillus polymyxa* aumentaram a resistência a patógenos, com isso, uma menor incidência de doenças e redução máxima de doenças, seguido do aumentaram a resistência à seca. Este organismo demonstrou produzir uma série de compostos com atividade antifúngica (YANG, 2023).

2.3. AntiSMASH

Nos últimos anos, a mineração de genoma tem visto amplas aplicações na identificação e caracterização de novos compostos, bem como na engenharia metabólica. Desde 2011, o *antiSMASH* se tornou o padrão ouro para a análise de genomas em busca de compostos de interesse biotecnológico (BLIN, 2019).

O *antiSMASH* se estabeleceu como a ferramenta mais utilizada para identificar e analisar *clusters* de genes biossintéticos (BGCs) em sequências de genomas bacterianos e fúngicos devido a ser composto por um banco de dados que fornece acesso instantâneo a milhares de resultados de mineração de genomas pré-computados, de genomas disponíveis publicamente (BLIN, 2019).

Ele funciona comparando produtos de genes codificados em uma biblioteca de cluster com curadoria manual, que abrange uma variedade de genes biossintéticos de produtos naturais. Os BGCs são identificados de acordo com o método enzimas-

chave onde agrupamentos de genes são relacionados a grupos ortólogos, ou seja, genes de diferentes espécies que são similares entre si por serem derivados de um ancestral comum (RUSSELL, 2020).

Atualmente é usado por cientistas, acadêmicos e industriais em todo o mundo para identificar os chamados *clusters* (BGCs) de metabólitos secundários em seus genomas de interesse sendo a ferramenta de mineração de genoma mais usada, com mais de 670.000 trabalhos processados on-line no momento da redação (BLIN, 2019; RUSSELL, 2020).

A maioria das ferramentas de predição de produtos naturais utiliza a busca por domínios NRPS e PKS que são famílias de enzimas responsáveis pela síntese de metabólitos secundários com notável atividade biológica, atuando através do processamento de aminoácidos (pequenos blocos de construção) para a estruturação de metabólitos mais complexos, além de constituírem as principais classes de compostos de interesse biotecnológico com ênfase na indústria farmacêutica (CASTRO, 2015).

Os metabólitos secundários são derivados de quatro classes de enzimas: policetídeosintases (PKS), peptídeo sintases não ribossômicas (NRPS), terpeno sintases (TS) e dimetilalil difosfato triptofanosintases (DMATS). Cada classe utiliza metabólitos simples como substrato, os quais são rearranjados ou condensados em moléculas mais complexas, como policetídeos, peptídeos não-ribossômicos, terpenos e alcalóides. Os substratos para as PKS e NRPS são moléculas de acil coenzima A e aminoácidos e estas moléculas podem sofrer modificações como oxigenação, ciclização, isomerização e/ou condensação com outros compostos formando metabólitos altamente variáveis quanto à estrutura química e atividade biológica (FERREIRA, 2016).

As PKs (tipo I) são multifuncionais e codificadas por um único gene, podendo apresentar até oito domínios. As PKs (tipo II) de bactérias e plantas são complexos multienzimáticos onde cada domínio é representado por uma cadeia polipeptídica separada. Enzimas do tipo III (chalcona-sintases) são constituídas de uma cadeia polipeptídica, ocorrem em plantas e sintetizam flavonóides. Recentemente foram descritas em bactérias e fungos, ampliando ainda mais o potencial metabólico microbiano (FERREIRA, 2016).

O software *antiSMASH* é capaz de identificar loci biossintéticos cobrindo toda a gama de classes de compostos de metabólitos secundários conhecidos (poliketídeos, peptídeos não ribossomais, terpenos, aminoglicosídeos, aminocumarinas, indolocarbazoles, lantibióticos, bacteriocinas, nucleosídeos, beta-lactâmicos, butirolactonas, sideróforos, melaninas e outros). Os bancos de dados contêm todos os agrupamentos de genes conhecidos e integra ou cruza todos os métodos de análise de genes específicos de metabólitos secundários previamente disponíveis em uma visualização interativa (MEDEMA, 2011).

2.4. AGRUPAMENTOS LOCALIZADOS PELO USO DO *AntiSMASH*

Entre alguns dados obtidos referentes à *clusters* de genes já identificados e, por isto, presentes nos bancos de dados de buscadores como o *AntiSMASH*, e, aproveitados em biotecnologia, estão:

O *cluster spo* inclui todos os genes necessários para a biossíntese de enediína (KERSTEN, 2013). O *cluster cgc* (congocidina (também chamada de netropsina)) pertence à família dos antibióticos pirrolamidas caracterizados por uma porção 4-aminopirrol-2-carboxil (AIGLE, 2014).

Srm é referente ao cluster da espiramicina (sendo os mesmos antibióticos macrolídeos de anel de 16 membros usados na terapêutica humana como agentes antibacterianos e antiparasitários ativos contra *Toxoplasma* spp.)). Eles compreendem um anel central de macrolactona (platenólídeo I), ao qual estão ligados dois açúcares amino (micaminose, forosamina) e um açúcar neutro (micarose) (AIGLE, 2014).

Com o agrupamento de genes biossintéticos de lomaiviticina (*lom*), prevê-se que o locus *lom* inclua 59 ORFs, que incluem todos os genes biossintéticos que se pensa estarem envolvidos na construção do núcleo de diazofluoreno na cinamicina. (O cluster *lom* contém três genes PKS adicionais adjacentes aos genes PKS tipo II com uma unidade inicial de acetato na cinamicina. Produtos de genes biossintéticos putativos incluem duas glicosiltransferases e várias enzimas biossintéticas desoxiaçúcares que se correlacionam bem com o padrão de glicosilação das lomaiviticinas (KERSTEN, 2013).

Uma vez que apresenta imensa importância para a bioatividade a orfamida A pode desencadear eventos defensivos precoces, apresenta-se também como um metabólito inseticida contra *Myzus persicae*, e apresenta atividade de redução da tensão superficial (JANG, 2013).

Por meio da mineração do genoma e análises químicas evidenciou-se que as bactérias da rizosfera (*Paraburkholderia graminis*) produzem um novo tipo de sideróforo, a gramibactina, um lipodepsipeptídeo que se liga eficientemente ao ferro com um valor $\log\beta$ de 27,6. Os genes de biossíntese da gramibactina são conservados em numerosas bactérias associadas a plantas associadas ao arroz, trigo e milho, que podem utilizar o ferro do complexo (HERMENAU, 2018).

Encontram-se entre as mais complexas maquinarias protéicas conhecidas na natureza os poliquetideos, os quais são responsáveis pela biossíntese de inúmeros compostos utilizados. enzimas multifuncionais responsáveis pela biossíntese de inúmeros produtos naturais, muitos dos quais são atualmente usados como antibióticos, drogas antiparasitárias, agentes redutores de colesterol, imunossuppressores e quimioterapia para câncer (NIVINA, 2019).

Constituindo a maior classe de compostos naturais, os terpenos são extremamente valiosos do ponto de vista econômico devido às suas extensas propriedades físico-químicas e atividades biológicas. São comuns em todas as plantas (COUILLAUD, 2022).

O pyoverdina é um sideróforo de fluorescência solúvel em água com forte capacidade quelante de ferro do patógeno gram-negativo *Pseudomonas aeruginosa*. Comparado aos sideróforos comuns, o PVDI é um composto relativamente grande cuja síntese requer um grupo de enzimas com diferentes atividades catalíticas (YUAN Z, 2017).

Sendo um importante constituinte da parede celular de bactérias Gram-negativas, o lipopolissacarídeo (LPS) é vital para a integridade celular bacteriana, viabilidade e defesa contra o estresse ambiental. O lipopolissacarídeo é altamente conservado entre quase todas as bactérias Gram-negativas e é um potente indutor de respostas inflamatórias. Ele existe de forma ubíqua no ambiente e, portanto, o LPS

pode modular diretamente o sistema imunológico e a suscetibilidade a doenças (MAZGAEEN, 2020)

2.5. O POTENCIAL BACTERIANO

Produtos naturais (NPs) são substâncias frequentemente constituídas por estruturas químicas complexas e com uma orientação espacial bem definida, podendo atuar como modelos estruturais para síntese de novas substâncias ou protótipos ativos. Estes produtos naturais, também denominados metabólitos secundários ou compostos bioativos, são característicos e até mesmo únicos para determinados grupos específicos de organismos, e são sintetizados para interagir com eficácia com seus alvos biológicos (KINGSTON, 2011; UNESP, 2022).

Microrganismos e plantas produzem uma infinidade de produtos naturais (NPs) com uma variedade de bioatividades, incluindo antimicrobiana, anticancerígena, pesticida e imunossupressora. Como resultado, muitos desses compostos são altamente valiosos e amplamente utilizados na medicina, agricultura e indústria alimentícia (RUSSELL, 2020).

Os metabólitos secundários produzidos por bactérias e fungos são uma importante fonte de antimicrobianos e outros compostos bioativos. Os produtos naturais bacterianos e fúngicos constituem uma fonte chave para o desenvolvimento de antimicrobianos e outras drogas que age como mediador dessas interações ecológicas entre organismos de várias maneiras (BLIN, 2019).

De acordo com DA COSTA (2022) Foi identificado e caracterizado o *cluster* gênico biossintético para produção do *sideróforo petrobactin* com base no genoma completo encontrado no NCBI da bactéria isolada de sedimentos do rio Solimões (SOL 105), pertencente a *Bacillus anthracis* que é uma bactéria do gênero *Bacillus* cujos genes AsbABCDE codificam enzimas que participam do processo de transformação do aminoácido espermidina, ácido dihidroxibenzóico e citrato em petrobactina. CIRÍACO (2022), analisou com base no genoma completo encontrado no NCBI, revelando que a linhagem APUR 36.1 isolada de sedimentos do Rio Purus curso de

água do Amazonas, como um representante da espécie *Streptomyces sampsonii*. Estas bactérias já sequenciadas da localização do Amazonas são exemplos do quanto ainda falta por explorar, diante disso a *Burkholderia gladioli Coa14* foi escolhida por fazer parte do grupo de pesquisas dentro da vertente de bactérias amazônicas com um promissor potencial biotecnológico até então inexplorado.

Existe um grande número de *clusters* de genes de metabólitos secundários ainda desconhecidos em comparação com o número de metabólitos conhecidos por serem produzidos por uma única cepa geralmente é explicado pelo fato de alguns desses agrupamentos de genes serem críticos, ou seja, não expressos ou os produtos formados em um nível muito baixo para ser detectado em condições de crescimento de laboratório. Isso representa o principal gargalo para obter acesso a compostos potencialmente novos (AIGLE, 2014).

Burkholderia, um gênero bacteriano que compreende cerca de 120 espécies descritas, costuma habitar ambientes de solo e água. Estas bactérias Gram-negativas abrigam uma variedade de vias catabólicas. São extremamente versáteis e diversas, podendo tolerar e degradar uma variedade de compostos (aromáticos, monoaromáticos, aromáticos policíclicos e heterocíclicos) (MORYA, 2020).

As bactérias pertencentes ao gênero *Burkholderia* são considerados um hospedeiro cada vez mais promissor devido a estarem entre os microrganismos conhecidos por serem capazes de degradar compostos persistentes; suas cepas são usadas como modelos para estudar tal habilidade. O sequenciamento de alto rendimento dos mesmos permite que os pesquisadores alcancem um conhecimento mais amplo sobre a biodegradação por bactérias (CAUDURO, 2020; MORYA, 2020).

Muitas espécies dentro do gênero *Burkholderia* são capazes de colonizar plantas e animais, tanto de forma benéfica quanto patogênica. De fato, espécies capazes de promover o crescimento vegetal e isoladas de ambientes naturais foram recentemente agrupadas no gênero *Burkholderia*, possuem significativo potencial de controle biológico. Espécies de *Burkholderia* é amplamente reconhecida como uma bactéria que exhibe notável divergência de nichos ecológicos, ocorrendo comumente no solo, água, plantas, fungos, animais e humanos (SCOFFONE, 2021).

Uma peculiaridade das espécies de *Burkholderia* é seu grande genoma, tipicamente composto por dois ou três cromossomos, contendo mais de 5.000 genes (www.burkholderia.com), alto conteúdo de GC e altos níveis de resistência a antibióticos (SCOFFONE, 2021).

Seus complexos genomas multireplicons abrigam um número impressionante de genes de poliketido sintase (PKS) e peptídeo-sintetase não ribossômica (NRPS) que codificam a produção de metabólitos secundários antimicrobianos (SMs) (BACH, 2022).

Burkholderia gladioli Coa14 é uma bactéria isolada da água coletada do lago Coari (Amazonas, Brasil) que apresenta capacidade de sobrevivência em meio contendo apenas óleo como fonte de carbono (LOPES, 2018).

LOPES (2019) isolou a cepa *Burkholderia gladioli* Coa14 e verificou seu potencial de degradação de petróleo, além de realizar ensaios de biodegradação e resistência a antibióticos, obtendo resultados satisfatórios dentro da perspectiva de aplicar futuramente a biorremediação.

3. OBJETIVO GERAL

Investigar a presença de *clusters* de genes relacionados a compostos bioativos no genoma da bactéria amazônica *Burkholderia gladioli* depositados publicamente no NCBI, através do *antiSMASH*

4. OBJETIVO ESPECÍFICO

- Encontrar bioativos associados aos clusters obtidos;
- Identificar os grupos relacionados aos clusters terpenos, policetídeos não ribossomais (NRPS) e policetídeo (PKS);
- Analisar as sequências identificadas para observar e entender as classes pertencentes em maior número.

5. MATERIAL E MÉTODOS

5.1. METODOLOGIA

Para a análise, foi realizado o *download* do genoma da bactéria que se encontra depositado publicamente no NCBI (National Center for Biotechnology Information) com o número de referência ASM291790v1. Após o download, o mesmo foi analisado utilizando-se a plataforma *antiSMASH* v. 6.1.1 (<https://antismash.secondarymetabolites.org/#!/start>), na versão para a análise de genomas de bactérias. Para análise, além dos arquivos do genoma, também se utilizou os arquivos de anotação.

Toda a análise decorreu de maneira online, com as configurações do software na forma padrão, com a rigidez de detecção de maneira ampla. Para confirmação dos resultados, a sequência posteriormente foi submetida ao BLASTp, também de forma online, com *e-value* menor ou igual à 0.01 e percentual de identidade maior ou igual à 98%.

Após confirmação, os achados passaram a ser elencados em tabelas de acordo com a classe de metabólito secundário encontrado, sendo realizada, posteriormente, uma estatística descritiva em relação aos dados encontrados.

6. RESULTADOS

A tabela 1 mostra a classe de metabólitos secundários identificados através do *antiSMASH*, os mesmos foram analisados e retificados de acordo com os grupos de substâncias presentes.

Tabela 1 – *Clusters* identificados e divididos por grupo de substâncias.

MODELO	(SUBSTÂNCIAS) CLUSTERS MAIS SEMELHANTES	
NRPS	GRAMIBACTIN	NRP
NRPS, T1PKS	ORFAMIDE A/ ORFAMIDE C	NRP: CYCLIC DEPSIPEPTIDE
T1PKS, tipo transAT-PKS	ENACILOXIN LLA	NRP + POLYKETIDE
TERPENE	LASALOCID	POLYKETIDE
NRPS	RHIZOMIDE A/ RHIZOMIDE B/ RHIZOMIDE C	NRP
NRPS	PYOVERDIN	NRP
T1PKS	LIPOPOLYSACCHARIDE	SACCHARIDE: LIPOPOLYSACCHARIDE
TransAT-PKS	ISO-MIGRASTATIN/ MIGRASTATIN/ DORRIGOCIN A/ DORRIGOCIN B/ 13-EPI- DORRIGOCIN A	POLYKETIDE: MODULAR TYPE I + POLYKETIDE: transAT TYPE I
FOSFONATO	PHOSPHINOTHRICINTRIPEPTIDE	NRP
NRPS, TYPE RiPP	NAPSAMYCIN A/ NAPSAMYCIN B/ NAPSAMYCIN C/ NAPSAMYCIN D/ MUREIDOMYCIN A/ MUREIDOMYCIN B	NRP: URIDYLPEPTIDE + OTHER NUCLEOSIDE
REDOX-COFACTOR	LANKACIDIN C	NRP + POLYKETIDE

TERPENE	DESOTAMIDE	NRP
ECTOINE, NRPS	ICOSALIDE A/ ICOSALIDE B	NRP: LIPOPEPTIDE
RiPP-LIKE, BETALACTONE, TERPENE	BARBAMIDE	NRP + POLYKETIDE: MODULAR TYPE I

Fonte: Própria (2022). Sintetases de peptídeos não ribossômicos (NRPS); fragmentos similares a sintetase de peptídeo não ribossômico (NRPS-like); sintases de policetídeo tipo I (PKS).

7. DISCUSSÃO

O acesso ao sequenciamento parcial do genoma (*Draft*) da bactéria *Burkholderia gladioli*, permitiu identificar o abrangente potencial biotecnológico com mais de 50 *clusters* de genes de metabólitos secundários. Vários fatores e condições podem ser atribuídos à bactéria *Burkholderia gladioli* tal como a importância para a bioatividade desencadeando eventos defensivos precoces, e apresenta atividade de redução da tensão superficial e atividade inseticida sendo um componente de bactérias Gram-negativas.

Há bastante evidência de ser um combustível metabólico pois é apontada como um potente indutor de respostas inflamatórias, possui atividade antibiótica sendo uma atraente candidata para combater patógenos resistentes a antibióticos, outro fator importante é o fato de ser constituída pela maior classe de compostos naturais sendo os mesmos considerados extremamente valiosos do ponto de vista econômico, devido às suas extensas propriedades físico-químicas e atividades biológicas.

Ao observar detalhadamente é possível haver concordância perfeita entre os algoritmos individuais fazendo com que suas características dentro da biossíntese o tornam adequado para a evolução da biossíntese do metabolismo secundário evidenciando à pesquisa a importância da existência das bactérias amazônicas voltadas para a pesquisa na biotecnologia.

Como resultado, a tabela apresentada mostra evidências de atividade significativa de fator biotecnológico em torno da bactéria estudada, sendo um dos principais fatores que desempenham um importante papel na manutenção de bases sustentáveis para *clusters* promissores à biotecnologia.

LEE (2021) empregou a análise pangenômica, usou um total de 14 isolados de *B. gladioli* incluindo a *B. gladioli coa 14* e concluiu que todos exibem características únicas, o estudo genômico demonstrou que a adaptação bacteriana ao ambiente no hospedeiro está intimamente relacionada às capacidades biológicas, que são

alteradas pelo resultado do ganho/perda de genes ou redução/expansão do genoma, possivelmente contribuindo para uma adaptação bem-sucedida no ambiente. Enquanto isso, LOPES (2018) isolou a cepa *Burkholderia gladioli* Coa14 e realizou o sequenciamento do genoma que foi de 95,06%, dessa forma facilitando entender melhor os mecanismos utilizados para sobreviver em diferentes tipos de ambientes.

PKS e NRPS ambas classes utilizam metabólitos simples como substrato, e possuem uma facilidade de se condensar com outros compostos formando metabólitos altamente variáveis e muito diversificados, com uma gama extremamente ampla de atividades biológicas e propriedades biotecnológicas conforme a *Burkholderia gladioli*.

Policetídeos e Terpenos são moléculas complexas, em sua maioria são rearranjados de classes de enzimas como PKS, NRPS e terpeno sintases (TS) com metabólitos simples como substrato dando origem a maior classe de compostos naturais, pois estão entre as mais complexas maquinarias protéicas conhecidas na natureza, responsáveis pela biossíntese de inúmeros compostos utilizados. Enzimas multifuncionais responsáveis pela biossíntese de inúmeros produtos naturais, são extremamente valiosos do ponto de vista econômico devido às suas extensas propriedades físico-químicas e atividades biológicas. Presentes comuns em todas as plantas e na bactéria estudada.

LASALOCID é agente antibacteriano, a lasalocid é capaz de formar complexos neutros e transportá-los através da fase apolar assim como a membrana da bicamada lipídica, também tem a capacidade de transportar grandes cátions orgânicos como a dopamina.

SACCHARIDE e LIPOPOLYSACCHARIDE (LPS), são os padrões molecular comumente mais associado a patógenos maiormente estudado, o LPS pode modular diretamente o sistema imunológico e a suscetibilidade a doenças enquanto o SACCHARIDE é relevante na detecção de glicose em pacientes com diabetes. Sendo o principal combustível metabólico, produzido durante a fotossíntese. Ambos associados a detecção de patógenos.

Como representante de bioativos com potencial inseticida estão os *clusters* orfamida, orfamida A e orfamide C, apresentam-se como metabólito inseticida.

Apresentam uma potencial atividade antibiotica os *clusters* transAT-PKS, napsamycin D, mureidomycin A, uridylpeptideo, desotamide, icosalide A, lipopeptide e barbamide. No entantos *clusters* como o Migrastatin e lankacidin exibem uma importante atividade antitumoral.

8. CONCLUSÃO

O presente trabalho buscou evidenciar o promissor potencial da bactéria *Burkholderia gladioli* Coa14 através de uma análise das sequências obtidas e explorar a rica fonte de metabólitos bioativos da mesma. Através dos mais de 50 *clusters* encontrados e analisados foi possível identificar um grande potencial na atividade antibiótica diretamente relacionada aos *clusters* transAT-PKS, napsamycin D, mureidomycin A, uridylpeptideo, desotamide, icosalide A, lipopeptide e barbamide, da mesma forma que entre os policetídeos não ribossomais (NRPS) a plantaribaccitin é explorada por apresentar atividade antibiótica, já os policetídeos se apresentam com potencial antibiótico e antiparasitário, os *clusters* orfamide, orfamide A e orfamide C apresentam atividade inseticida. O *cluster* lipopolissacarídeo (LPS) apresenta-se como um potente indutor de resposta inflamatória e o *cluster* ectoine que promove a capacidade de proteger a bactéria de danos causados pelas condições ambientais seja calor extremo, frio, luz ultravioleta promovendo assim o fato da *Burkholderia gladioli* ser considerados um hospedeiro cada vez mais promissor devido a divergência de nichos ecológicos sendo capaz de colonizar plantas, fungos, animais, humanos, solo e água. Desta forma destaca-se a *Burkholderia gladioli* com um potencial biotecnológico abrangente.

REFERÊNCIAS

- AIGLE, Bertrand, et al. Genome mining of *Streptomyces ambofaciens*. **Journal of Industrial Microbiology and Biotechnology**, 41.2: 251-263, 2014.
- ANAND, Swadha, et al. SBSPKS: Structure based sequence analysis of polyketide synthases. **Nucleic Acids Research**, 38. 2: 487- 496, 2010.
- BACH, Evelise, et al. Burkholderia in the genomic era: from taxonomy to the discovery of new antimicrobial secondary metabolites. **Critical Reviews in Microbiology**, 48.2: 121-160, 2022.
- BELKNAP, Kaitlyn C., et al. Genome mining of biosynthetic and chemotherapeutic gene clusters in *Streptomyces* bacteria. **Scientific reports**, 10.1: 1-9, 2020.
- BLIN, Kai, et al. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. **Nucleic acids research**, 47.W1: W81-W87, 2019.
- CAUDURO, Guilherme Pinto, et al. Differential Expression and PAH Degradation: What *Burkholderia vietnamiensis* G4 Can Tell Us?. **International journal of microbiology**, 2020.
- CASTRO, Vanessa R. Avaliação De Softwares Para Predição De Clusters Gênicos: Uma Análise *In Silico* Com Cianobactérias Da Ordem Chroococcales, 2015.
- COUILLAUD, Julie; DUQUESNE, Katia; IACAZIO, Gilles. Extension of the terpene chemical space: the very first biosynthetic steps. **ChemBioChem**, 23.9: e202100642, 2022.
- DA COSTA, G. V., et al. Caracterização do cluster gênico biossintético para produção do sideróforo petrobactina por *Bacillus cereus* SOL 105 isolado de sedimentos do Rio Solimões. 2022.
- FERREIRA, Almir José. Análise e anotação do genoma de *Epicoccum nigrum* e metabolismo secundário. PhD Thesis. Universidade de São Paulo, 2016.
- HANNIGAN, Geoffrey D., et al. A deep learning genome-mining strategy for biosynthetic gene cluster prediction. **Nucleic acids research**, 47.18: e110-e110, 2019.

HERMENAU, Ron, et al. Gramibactin is a bacterial siderophore with a diazeniumdiolate ligand system. **Nature Chemical Biology**, 2018, 14.9: 841-843, 2018.

JANG, Ja Yeong, et al. Identification of orfamide A as an insecticidal metabolite produced by *Pseudomonas protegens* F6. **Journal of agricultural and food chemistry**, 61.28: 6786-6791, 2013.

KERSTEN, Roland D., et al. Bioactivity-guided genome mining reveals the lomaiviticin biosynthetic gene cluster in *Salinispora tropica*. **ChemBioChem**, 14.8: 955-962, 2013.

KINGSTON, David GI. Modern natural products drug discovery and its relevance to biodiversity conservation. **Journal of natural products**, 74.3: 496-511. 2011.

LEÃO, Tiago F., et al. NPOMix: A machine learning classifier to connect mass spectrometry fragmentation data to biosynthetic gene clusters. **PNAS Nexus**, 1.5: pgac257, 2022.

LEE, Hyun-Hee, et al. Pan-genome analysis reveals host-specific functional divergences in *Burkholderia gladioli*. **Microorganisms**, 9.6: 1123, 2021.

LEITE, Vida Marina Barreto. Caracterização da produção de moléculas antitumorais por *Streptomyces sp. BRB081*, guiada por mineração genômica. PhD Thesis. Universidade de São Paulo. 2022.

LOPES, Eraldo Ferreira, et al. Draft genome sequence of *Burkholderia gladioli* Coa14, a bacterium with petroleum bioremediation potential isolated from Coari Lake, Amazonas, Brazil. **Genome Announcements**, 6.16: e00301-18, 2018.

LOPES, Eraldo Ferreira, et al. Caracterização genômica e da capacidade de degradar componentes de petróleo de uma linhagem de *Burkholderia gladioli* isolada do Lago de Coari (Coari-Amazonas) - Universidade Federal do Amazonas, Manaus. 2019.

MAZGAEEN, Lalita; GURUNG, Prajwal. Recent advances in lipopolysaccharide recognition systems. **International journal of molecular sciences**, 21.2: 379, 2020.

MEDEMA, Marnix H., et al. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. **Nucleic acids research**, Vol 39. 2: W339-W346, 2011.

MORYA, Raj; SALVACHÚA, Davinia; THAKUR, Indu Shekhar. Burkholderia: an untapped but promising bacterial genus for the conversion of aromatic compounds. **Trends in Biotechnology**, 38.9: 963-975, 2020.

NIVINA, Aleksandra, et al. Evolution and diversity of assembly-line polyketide synthases: Focus review. **Chemical reviews**, 119.24: 12524-12547, 2019.

NOBRE, Steffany Virgolino Araujo, et al. Prospecção de sequências de nucleotídeos para enzimas relacionadas com a biodegradação de plástico em genomas de fungos endofíticos da Antártica. **Anais do Salão Internacional de Ensino, Pesquisa e Extensão**, 13.3, 2021.

RUSSELL, Alicia H.; TRUMAN, Andrew W. Genome mining strategies for ribosomally synthesised and post-translationally modified peptides. **Computational and Structural Biotechnology Journal**, 18: 1838-1851, 2020.

SCOFFONE, Viola Camilla, et al. Methodological tools to study species of the genus Burkholderia. **Applied Microbiology and Biotechnology**, 1-16, 2021.

STARCEVIC, Antonio, et al. ClustScan: an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures. **Nucleic acids research** 36.21: 6882-6892, 2008.

TEIJARO, Christiana N.; ADHIKARI, Ajeeth; SHEN, Ben. Challenges and opportunities for natural product discovery, production, and engineering in native producers versus heterologous hosts. **Journal of Industrial Microbiology and Biotechnology**, 46.3-4: 433-444, 2019.

UNESP, 2022. Disponível em http://unesp.br/prope/mostra_arg_multi.php?arquivo=6699. Acesso em 30 de Agosto de 2022.

VARELLA, Larissa. Estudo químico e estratégias para modular o metabolismo secundário de actinobactérias endofíticas. PhD Thesis. Universidade de São Paulo. 2015.

YANG, Shijia, et al. Screening of new Paenibacillus polymyxa S3 and its disease resistance of grass carp (*Ctenopharyngodon idellus*). *Journal of Fish Diseases*, 46.1: 17-29, 2023.

YANG, Yongliang; ADELSTEIN, S. James; KASSIS, Amin I. Target discovery from data mining approaches. **Drug discovery today**, 17: S16-S23, 2012.

YUAN Z, et al. Crystal structure of PvdO from *Pseudomonas aeruginosa*. **Biochem Biophys Res Commun**. PMID: 28109878, 2017.