

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO
CURSO DE CIÊNCIAS BIOMÉDICAS - TURMA V
DEPARTAMENTO DE BIOLOGIA CELULAR E MOLECULAR E
BIOAGENTES PATOGÊNICOS



Análise *in silico* de metabólitos secundários de isolados bacterianos

ISABELA AKEMI MARUYAMA

Ribeirão Preto
2022

ISABELA AKEMI MARUYAMA

**Análise *in silico* de metabólitos secundários de isolados
bacterianos**

Trabalho de conclusão de curso apresentado como parte das exigências para a conclusão da ênfase em Biotecnologia e para a obtenção do grau de bacharel em Ciências Biomédicas.

Orientador: Prof. Dr. Rafael Silva Rocha
Co-orientador: Edson Alexandre do Nascimento Silva

Ribeirão Preto
Julho de 2022

RESUMO

Os estudos de análise genômica estão focados em diversos microrganismos em busca de vias biossintéticas que produzam moléculas bioativas com aplicações biotecnológicas, como por exemplo: na agricultura visando os bioagentes e microrganismos multifuncionais, e na saúde com a busca constante de novos possíveis antibióticos. Neste trabalho, foram realizadas a identificação *in silico* e a predição de vias biossintéticas de metabólitos secundários do genoma de duas amostras de bactérias isoladas do solo de Sertãozinho: a *Priestia megaterium* (MAPA1) e a *Arthrobacter sp.* (M4). Para isso utilizou-se as ferramentas de bioinformática antiSMASH, deepBCG e o KEGG. Com a ferramenta antiSMASH foram identificados na *Priestia megaterium* e *Arthrobacter sp.* 9 e 5 regiões com cluster gênico de metabólitos secundários, respectivamente. Já com a ferramenta deepBCG foram encontrados 46 clusters na *Priestia megaterium* sendo 2% relacionados com função citotóxica, 59% antibacteriana e 39% não determinada e 62 clusters na *Arthrobacter sp.* sendo 1% com função citotóxica, 2% inibitória, 72% antibacteriana e 25% não determinada. Por fim, com a ferramenta KEGG foram analisadas as rotas metabólicas e foram encontradas 25 rotas na *Priestia megaterium* e 49 rotas na *Arthrobacter sp.*, porém as amostras apresentaram apenas 1-2 genes presentes em cada uma das rotas. Então, embora as ferramentas apresentem os resultados citados, não são todas as regiões encontradas que possuem clusters com alta similaridade com o banco de dados, as rotas identificadas possuem poucos genes presentes, e as ferramentas possuem resultados divergentes entre si. Portanto, a análise *in silico* auxilia quanto ao gerenciamento do tempo e diminuição dos custos de pesquisas *in vitro* para detectar produtos naturais, pois com as predições das análises de bioinformática é possível fazer uma melhor escolha da amostra. Com isso, após as análises temos um melhor parâmetro para escolher quais espécies passarão por uma análise manual com experimentos *in vitro* para a validação dos resultados encontrados.

Palavras-chave: análise genômica; metabólitos secundários; bioinformática; *in silico*

ABSTRACT

Genomic analysis studies are focused on various microorganisms in search of biosynthetic pathways that produce bioactive molecules with biotechnological applications, such as: in agriculture targeting bioagents and multifunctional microorganisms, and in health with the constant search for new possible antibiotics. In this work, the *in silico* identification and prediction of biosynthetic pathways of secondary metabolites from the genome of two samples of bacteria isolated from the soil of Sertãozinho were carried out: *Priestia megaterium* (MAPA1) and *Arthrobacter sp.* (M4). For this, the bioinformatics tools antiSMASH, deepBCG and KEGG were used. With the antiSMASH tool, *Priestia megaterium* and *Arthrobacter sp.* 9 and 5 regions with secondary metabolite gene clusters, respectively. With the deepBCG tool, 46 clusters were found in *Priestia megaterium*, being 2% related to cytotoxic function, 59% antibacterial and 39% undetermined, and 62 clusters in *Arthrobacter sp.* 1% with cytotoxic function, 2% inhibitory, 72% antibacterial and 25% undetermined. Finally, with the KEGG tool, the metabolic routes were analyzed and 25 routes were found in *Priestia megaterium* and 49 routes in *Arthrobacter sp.*, however the samples showed only 1-2 genes present in each of the routes. So, although the tools present the cited results, not all regions found have clusters with high similarity with the database, the identified routes have few genes present, and the tools have divergent results. Therefore, *in silico* analysis helps to manage time and reduce costs of *in vitro* research to detect natural products, because with the predictions of bioinformatics analysis it is possible to make a better choice of sample. Thus, after the analyzes we have a better parameter to choose which species will undergo a manual analysis with *in vitro* experiments to validate the results found.

Key-words: genomic analysis; secondary metabolites; bioinformatic; *in silico*

Lista de Figuras

Figura 1: Foto do isolado MAPA1 em placa de Petri tirada com celular.....	19
Figura 2: Foto do isolado M4 em placa de Petri tirada no <i>ChemiDoc</i>	19
Figura 3: Imagem retirada do resultado da análise de qualidade do MiGA Online do isolado MAPA1.....	21
Figura 4: Imagem do <i>Full report</i> da Figura 3.....	21
Figura 5: Imagem retirada do resultado da análise de qualidade do MiGA Online do isolado M4.....	22
Figura 6: Imagem do <i>Full report</i> da Figura 5.....	22
Figura 7: Imagem retirada do resultado da análise filogenética do TYGS do isolado MAPA1 - GBDP <i>tree</i> (16S rDNA <i>gene sequence-based</i>).....	24
Figura 8: Imagem retirada do resultado da análise filogenética do TYGS do isolado MAPA1 - GBDP <i>tree</i> (<i>whole-genome sequence-based</i>).....	24
Figura 9: Imagem retirada do resultado da análise filogenética do TYGS do isolado MAPA1 - GBDP <i>tree</i> (<i>whole-proteome-based</i>).....	25
Figura 10: Imagem retirada do resultado da análise filogenética do TYGS do isolado M4 - GBDP <i>tree</i> (16S rDNA <i>gene sequence-based</i>).....	26
Figura 11: Imagem retirada do resultado da análise filogenética do TYGS do isolado M4 - GBDP <i>tree</i> (<i>whole-genome sequence-based</i>).....	27
Figura 12: Imagem retirada do resultado da análise de metabólitos secundário e BGCs do antiSMASH da <i>Priestia megaterium</i>	28
Figura 13: Imagem retirada do resultado da análise de BGCS e metabólitos secundário do antiSMASH da <i>Arthrobacter sp</i>	29

Lista de tabelas

Tabela 1: Informações finais sobre a montagem do genoma MAPA1.....	20
Tabela 2: Informações finais sobre a montagem do genoma M4.....	20
Tabela 3: Resultado da análise filogenética do TYGS do isolado MAPA1 - valor de dDDH.....	23
Tabela 4: Resultado da análise filogenética do TYGS do isolado M4 - valor de dDDH.....	26
Tabela 5: Resultado da análise de rotas metabólicas da <i>Priestia megaterium</i> pelo KAAS.....	30
Tabela 6: Resultado da análise de rotas metabólicas da <i>Arthrobacter sp.</i> pelo KAAS.....	32
Tabela 7: Resultado da análise de função de BGCs da <i>Priestia megaterium</i> pelo deepBGC.....	36
Tabela 8: Resultado da análise de classe de BGCs da <i>Priestia megaterium</i> pelo deepBGC.....	36
Tabela 9: Resultado da análise de função de BGCs da <i>Arthrobacter sp.</i> pelo deepBGC.....	37
Tabela 10: Resultado da análise de classe de BGCs da <i>Arthrobacter sp.</i> pelo deepBGC.....	37

SUMÁRIO

1. INTRODUÇÃO.....	9
1.1. METABÓLITOS SECUNDÁRIOS.....	9
1.1.1. Sintetases de Peptídeos não Ribossomais (NRPSs).....	10
1.1.2. Sintases de Policetídeos (PKSs).....	10
1.1.3. Terpenóides.....	11
1.1.4. Sideróforos.....	11
1.1.5. Peptídeos Ribossomais Modificados Pós-Tradução (RiPPs).....	12
1.2. BIOPROSPECÇÃO E BIOTECNOLOGIA.....	12
2. OBJETIVOS.....	13
3. MATERIAIS E MÉTODOS.....	13
3.1. ISOLADOS BACTERIANOS.....	13
3.2. MONTAGEM DOS GENOMAS.....	14
3.3. ANÁLISE DA QUALIDADE DOS GENOMAS.....	15
3.3.1. <i>Microbial Genomes Atlas Online (MiGA Online)</i>	15
3.4. ANÁLISE DA FILOGENIA DOS GENOMAS.....	15
3.4.1. <i>Type (Strain) Genome Server (TYGS)</i>	16
3.5. ANÁLISE DE METABÓLITOS SECUNDÁRIOS E ROTAS METABÓLICAS.....	16
3.5.1. antiSMASH.....	16
3.5.2. <i>Kyoto Encyclopedia of Genes and Genomes (KEGG)</i> <i>Automatic Annotation Server (KAAS)</i>	17
3.5.3. deepBGC.....	18
4. RESULTADOS E DISCUSSÃO.....	18
4.1. ISOLADOS BACTERIANOS.....	19
4.2. MONTAGEM DOS GENOMAS.....	19
4.3. ANÁLISE DA QUALIDADE DOS GENOMAS.....	20
4.3.1. Análise MAPA1	20
4.3.2. Análise M4	21
4.4. ANÁLISE DA FILOGENIA DOS GENOMAS.....	22
4.4.1. Análise MAPA1	23
4.4.2. Análise M4	25

4.5.	ANÁLISE DE METABÓLITOS SECUNDÁRIOS E ROTAS METABÓLICAS.....	27
4.5.1.	antiSMASH	27
4.5.1.1.	Análise da <i>Priestia megaterium</i>	27
4.5.1.2.	Análise da <i>Arthrobacter sp</i>	28
4.5.2.	KEGG Automatic Annotation Server (KAAS)	29
4.5.2.1.	Análise da <i>Priestia megaterium</i>	29
4.5.2.2.	Análise da <i>Arthrobacter sp</i>	31
4.5.3.	deepBGC	35
4.5.3.1.	Análise da <i>Priestia megaterium</i>	35
4.5.3.2.	Análise da <i>Arthrobacter sp</i>	36
5.	CONCLUSÃO	38
6.	OBJETIVOS FUTUROS	38
7.	REFERÊNCIAS	39

1. INTRODUÇÃO

De acordo com a Organização Mundial da Saúde (OMS), a resistência aos antibióticos é uma das maiores ameaças à saúde global. A crise de resistência aos antibióticos é alertada há muito tempo pela comunidade científica e o Centros de Controle e Prevenção de Doenças (CDC) classificou várias bactérias como apresentando ameaças urgentes, sérias e preocupantes (VENTOLA, 2015).

A resistência aos antibióticos ocorre naturalmente, porém o uso indevido e excessivo de antibióticos em humanos e animais, a prescrição incorreta e a falta de investimento em pesquisa e desenvolvimento de novos fármacos contra bactérias resistentes, está acelerando o processo (VENTOLA, 2015).

Um número crescente de infecções como pneumonia, tuberculose e gonorreia está se tornando mais difícil de tratar à medida que os antibióticos usados se tornam menos eficazes e isso leva a internações hospitalares mais longas, custos médicos mais altos e aumento da mortalidade (VENTOLA, 2015). Portanto, encontrar novas moléculas com potencial antimicrobiano é crucial, a fim de prevenir o rápido desenvolvimento e disseminação da resistência aos antibióticos (KONGKHAM et al., 2020).

Uma das abordagens para combater a resistência a antibióticos é o uso de metabólitos secundários em combinação com antibióticos convencionais. Os metabólitos secundários são eles próprios antimicrobianos ou potencializam a atividade do antibiótico através de uma série de mecanismos. No entanto, seu uso está atrasado devido ao pouco conhecimento do modo de ação, relações de estrutura e atividade e farmacocinética KONGKHAM et al., 2020; NODWELL, 2007).

1.1. METABÓLITOS SECUNDÁRIOS

Os metabólitos secundários ou produtos naturais são um grupo de moléculas heterogêneas que derivam do metabolismo primário, porém ao contrário destes não são essenciais ao crescimento, desenvolvimento e reprodução do organismo que o produz (THIRUMURUGAN et al., 2018). Entretanto, contribuem para a sobrevivência da espécie e sua adaptação, realizando funções de simbiose, defesa contra predadores, transporte de metais, sinalização, entre outras funções. O estudo e

mapeamento dos metabólitos secundários é de grande importância devido suas inúmeras aplicações para os seres humanos tanto no campo da saúde com os antibióticos e antifúngicos, como na agricultura com os pesticidas e ferormônios (THIRUMURUGAN et al., 2018).

Os metabólitos secundários são produzidos a partir dos clusters biossintéticos (BGCs), ou seja, um agrupamento de genes responsáveis pela transcrição de enzimas que levam à formação dos produtos naturais e cada organismo produz seus metabólitos particulares. Portanto, existe uma diversidade de BGCs descritos e dentre eles os principais são: os sideróforos, terpenos, policetídeos sintases (PKS), sintase de peptídeos não ribossomais (NRPS) e peptídeos ribossomais modificados pós-tradução (RiPPs) (THIRUMURUGAN et al., 2018).

1.1.1. Sintetases de Peptídeos não Ribossomais (NRPSs)

As sintetases de peptídeos não ribossomais (NRPSs) são enzimas que podem utilizar 500 substratos diferentes de acil para a produção de metabólitos secundários que ainda podem ser modificados durante ou após a síntese, permitindo uma diversidade de topologias e funções, como: antifúngica (bacilomicina), antibacteriana (daptomicina), antiviral (luzopeptina), antitumoral (actinomicina D), e imunossupressora (ciclosporina) (BLOUDOFF, SCHMEING, 2017; WALSH, 2001). Para a síntese do metabólito secundário, as NRPSs utilizam uma linha de montagem a partir de um complexo enzimático onde cada enzima é responsável por adicionar um aminoácido específico ao alongamento de um peptídeo (BLOUDOFF, SCHMEING, 2017; WALSH, 2001).

1.1.2. Sintases de Policetídeos (PKSs)

As sintases de policetídeos (PKSs) podem ser divididas em três categorias de acordo com o mecanismo de ação e com a estrutura: PKS do tipo I, tipo II e tipo III.

As PKSs do tipo I são proteínas multifuncionais que se organizam de acordo com a síntese do policetídeo, de modo iterativo e não-iterativo, é a categoria composta pelas principais moléculas com atividade farmacológica. Já as PKSs do tipo II são heterodímeras e monofuncionais e produzem policetídeos aromáticos. Enquanto que as PKSs do tipo III são homodímeras e constroem o policetídeo em

um único sítio ativo de modo repetitivo (RISDIAN, MOZEF, WINK, 2019; GRININGER, 2020).

O mecanismo de formação dos policetídeos é semelhante ao da sintase de ácido graxo (FAS) porém no processo das PKSs a catalização pela ceto redutase (KR), da desidratase (DH) e da enoil redutase (ER) é opcional na cascata de reação das PKSs. Portanto, isso permite variações de estruturas finais de policetídeos com grupos ceto, hidroxí e/ou ligações duplas em diferentes locais da molécula (RISDIAN, MOZEF, WINK, 2019; GRININGER, 2020).

1.1.3. Terpenóides

Os terpenóides são os mais diversificados dos metabólitos secundários com mais de 50.000 terpenos estruturalmente caracterizados (KUZUYAMA, 2017). Eles podem ser produzidos a partir de duas vias distintas: a via do fosfato de metileritritol (MEP) e a via do mevalonato (MVA) (REYES et al., 2018). Essas vias são responsáveis pela produção de dois blocos de construção de cinco carbonos: o isopentenil difosfato (IPP) e o dimetilalil difosfato (DMAPP) que são os precursores da biossíntese dos terpenóides (REYES et al., 2018). Essas moléculas passam por uma série de reações enzimáticas de modificação como ciclizações, hidroxilações e glicosilações gerando a diversidade e complexidade estrutural dos terpenóides que podem ser diferenciados pelo seu esqueleto básico e grupos funcionais (ISAH et al., 2018). Essa diversidade estrutural possibilita uma gama de aplicações e funções aos terpenos, desde agentes antitumorais e anti maláricos até sua utilização em produtos alimentícios, combustível de aviação, fragrâncias, atrativos ou repelentes de insetos (BIAN et al., 2018; PEMBERTON et al., 2017). Além disso, contribuem também para o sabor, aromas e cor das folhas, flores e frutos das plantas (BIAN et al., 2018; PEMBERTON et al., 2017).

1.1.4. Sideróforos

Os sideróforos são quelantes de ferro produzidos por microrganismos como forma de tentar suprir a baixa aderência e captação de ferro em organismos hospedeiros ou do solo. Após a liberação do sideróforo, o mesmo forma um complexo ferro-sideróforo que é reabsorvido para o interior da célula bacteriana. Os

sideróforos podem ser biossintetizados pela rota do NRPS ou pela rota independente de NRPS (NIS) a partir do ácido dicarboxílico e blocos de diamina ou aminoálcool (CHALLIS, 2005).

1.1.5. Peptídeos Ribossomais Modificados Pós-tradução (RiPPs)

Os peptídeos ribossomais modificados pós-tradução (RiPPs) são sintetizados por um gene inicial que codifica um peptídeo precursor. Este gene inicial é co-localizado com os genes responsáveis por codificar as enzimas que o transformam no metabólito secundário final. Dentre as suas funções, eles podem ser citotóxicos, pois agem em processos celulares essenciais para o organismo como: replicação, transcrição e tradução do DNA (ARNISON et al., 2013). Os RiPPs também interagem extensivamente com o envelope celular de bactérias, tanto Gram-negativas quanto as Gram-positivas (CAO et al., 2021). As modificações pós-traducionais permitem uma grande diversidade estrutural que não é encontrada nos peptídeos ribossomais naturais (ARNISON et al., 2013).

1.2. BIOPROSPECÇÃO E BIOTECNOLOGIA

Bioprospecção é a busca por metabólitos em microrganismos, plantas e animais que sejam úteis para a humanidade (FILHO, SILVA, BIGI, 2014), por exemplo na agricultura, permitindo a redução de fertilizantes e pesticidas sintéticos (REZENDE et al., 2021) ou na saúde, com a busca de novos fármacos e antimicrobianos (FILHO, SILVA, BIGI, 2014). Portanto, esta busca pode ser realizada em diversos ambientes, como desertos, fontes termais, florestas, águas ou solos, entre outros (FILHO, SILVA, BIGI, 2014).

Embora exista uma gama de diferentes metabólitos secundários e uma constante bioprospecção, apenas poucos organismos são cultivados em laboratório e dentre eles apenas alguns são induzidos a produzirem moléculas de interesse (KALLIO, 2008). Entretanto, o avanço do sequenciamento genético, o melhoramento de ferramentas de bioinformáticas e a disposição dos genes em clusters, permitem a detecção e isolamento completo de uma rota biossintética para análise, comparação com banco de dados, mapeamento de função e manipulação (COWAN et al., 2005; NETT, IKEDA, MOORE, 2009).

Portanto, a utilização da metagenômica juntamente com a bioprospecção, permitem a análise de uma diversidade de organismos independente de seu cultivo em laboratório, possibilitando a busca de novos genes de interesse (COWAN et al., 2005; NETT, IKEDA, MOORE, 2009).

2. OBJETIVOS

Em um trabalho prévio realizado em conjunto com o laboratório de metagenômica funcional da Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto (FFCLRP - USP) foi feita a coleta de amostras de solo da área de proteção da USP e do solo agrícola de cana-de-açúcar de Sertãozinho - SP. A partir desta coleta foram extraídos, em escala massiva, os genomas presentes nas amostras de solos.

Para este trabalho de conclusão de curso foram selecionados dois dentre os vários microrganismos isolados com o objetivo de: analisar a qualidade do sequenciamento e montagem do genoma; identificar os isolados MAPA1 e M4 através da plataforma TYGS; identificar através das ferramentas antiSMASH, KEGG e deepBGC, possíveis rotas metabólicas e cluster gênicos envolvidos na biossíntese de metabólitos secundários e antimicrobianos, para a sua predição, comparação e identificação.

3. MATERIAIS E MÉTODOS

Os genomas dos isolados bacterianos MAPA1 e M4 foram sequenciados com base na plataforma Illumina Novaseq6000 com cobertura de 300X. Posteriormente, a partir dos resultados do sequenciamento foi feita a montagem do genoma. Em seguida, o programa MiGA online foi usado para analisar a qualidade da sequência, o programa *Type (Strain) Genome Server* (TYGS) foi utilizado para a identificação do isolado e construção de árvores filogenéticas e os genomas foram submetidos ao *National Center for Biotechnology Information* (NCBI) (SMITH, 2008). Por fim, os programas antiSMASH, *Kyoto Encyclopedia of Genes and Genomes* (KEGG) e deepBGC foram utilizados para a identificação de BGCs, predição dos metabólitos secundários e vias metabólicas.

3.1. ISOLADOS BACTERIANOS

Os dois microrganismos MAPA1 e M4 foram isolados a partir de amostras de solo da área de proteção da USP e do solo agrícola de cana-de-açúcar de Sertãozinho - SP, respectivamente. Para a coleta, foi retirada a camada de serrapilheira e, aproximadamente 50 gramas de solo foi coletado na profundidade de até 15 cm, utilizando instrumentos esterilizados. Os solos foram acondicionados em sacos plásticos e levados ao laboratório para processamento imediato e estocagem adequada.

Para o cultivo dos microrganismos foi utilizado meio mínimo M9 (Na_2HPO_4 , KH_2PO_4 , NH_4Cl , NaCl , MgSO_4 , CaCl_2) com glicerol como fonte de carbono.

O processamento das amostras de solo se deu pelo método direto que consiste em pesar 1 grama do solo diluído em 50 mL de água destilada e esterilizada. A amostra foi agitada em vortex por 10 minutos e, após, uma alíquota do sobrenadante foi diluída até 10^{-6} . 200 μL de cada diluição e da alíquota não diluída foram plaqueados nos diferentes meios de cultura. As placas foram incubadas em estufa a 30°C por até 7 dias, sendo que o crescimento microbiano foi monitorado diariamente.

O DNA genômico de cada isolado foi extraído utilizando o kit *Wizard® Genomic DNA Purification* (Promega, EUA) de acordo com as recomendações do fabricante. Os DNAs obtidos foram enviados à Novogene (Beijing, China) onde foram realizados os sequenciamentos dos genomas dos isolados previamente selecionados. Para isso, as amostras de DNA foram quantificadas fluorometricamente (Qubit® 3.0, kit Qubit® dsDNA Broad Range Assay, Life Technologies, EUA) e cada DNA purificado foi preparado para sequenciamento usando o kit Nextera XT DNA Library Prep (Illumina, EUA).

3.2. MONTAGEM DOS GENOMAS

Para a montagem do genoma foram realizados os seguintes passos: análise da qualidade do sequenciamento com a ferramenta FastQC (ANDREW, 2019); trimagem das sequências com a remoção de *reads* com baixa qualidade e de possíveis adaptadores dos kits de sequenciamento com o programa Trimmomatic (BOLGER, LOHSE, USADEL, 2014). Após estes passos foi feita uma nova medição de qualidade.

Na tentativa de otimizar as montagens em contiguidade, os *reads* de extremidade emparelhadas foram combinados com a ferramenta FLASH que estende o comprimento dos pequenos *reads* e sobrepõem os *reads* de extremidade emparelhadas na tentativa de prover comprimentos maiores de N50 (MAGOČ, SALZBERG, 2011).

Para a montagem do genoma foram utilizados o montador SPAdes (BANKEVICH et al., 2012) e o MEGAHIT (LI et al., 2015) como montador secundário otimizado para metagenomas e a ferramenta MAC (TANG et al., 2020), um recém lançado reconciliador de montagem.

Por fim, o FGAP (PIRO et al., 2014), uma ferramenta que preenche as bases indefinidas de montagens, foi utilizado com as bases de extremidades pareadas extendidas. A anotação foi realizada usando o Prokka (SEEMANN, 2014).

3.3. ANÁLISE DA QUALIDADE DOS GENOMAS

3.3.1. *Microbial Genomes Atlas Online (MiGA Online)*

O *Microbial Genomes Atlas Online* (MiGA Online) é um sistema de gerenciamento e processamento de dados para genomas e metagenomas microbianos. O MiGA Online fornece dados relacionados à qualidade da sequência genômica de consulta (RODRIGUEZ et al., 2018).

Os conjuntos de dados são aparados, retirados de sequências adaptadoras normalmente usadas e verificados quanto à qualidade usando SolexaQA++ (COX et al., 2010), Scythe com parâmetros padrão (BUFFALO, 2014) e FastQC (ANDREW, 2019), respectivamente.

O relatório de qualidade da sequência é baseado nos dados de completude e contaminação. A completude é medida pela presença de 111 genes que são observados em quase todos os genomas procarióticos, enquanto a contaminação é medida pela frequência com que esses genes estão presentes em mais de uma cópia, ou seja, duplicados. As pontuações de qualidade são então calculadas como porcentagem de integridade menos cinco vezes a porcentagem de contaminação (RODRIGUEZ et al., 2018).

3.4. ANÁLISE DA FILOGENIA DOS GENOMAS

3.4.1. *Type (Strain) Genome Server (TYGS)*

Os isolados MAPA1 e M4 foram identificados com a ferramenta TYGS que utiliza métodos de filogenia e classificação baseados no distanciamento genômico. Ao adicionar o isolado no programa, a ferramenta armazena a sequência em seu banco de dados de genomas de espécies com nomes validamente publicados (MEIER-KOLTHOFF e GÖKER, 2019).

Em seguida, passa pela seguinte sequência de passos: a ferramenta extrai as sequências gênicas de 16S rRNA utilizando o RNAmmer, essas sequências são comparadas com o banco de dados utilizando NCBI BLAST+ (CAMACHO et al., 2009) e as 50 espécies com as maiores pontuações são selecionadas para o cálculo da distância GBDP (*Genome Blast Distance Phylogeny*) entre elas. O cálculo da GBDP seleciona 10 estirpes com as menores distâncias entre a sequência do isolado e os genomas do banco de dados.

Os resultados são utilizados para a predição dos valores de hibridização DNA:DNA digital (dDDH) com o dDDH *calculator* e para a inferência das árvores filogenéticas do genoma completo e do 16S com suporte de ramificação com bootstraps de 100. Para a construção da árvore filogenética é utilizado o programa FastME versão 2.1.4 (LEFORT; DESPER; GASCUEL, 2015).

3.5. ANÁLISE DE METABÓLITOS SECUNDÁRIOS E ROTAS METABÓLICAS

3.5.1. *antiSMASH*

A análise dos BGCs foi realizada no programa antiSMASH versão 6.0 (BLIN et al., 2021). Na primeira etapa da análise, todos os genes da sequência analisada são pesquisados em um banco de dados de perfis de Modelos Ocultos de Markov (HMM) de enzimas altamente conservadas que são indicativos de uma classe específica de BGC.

Em seguida, regras de cluster predefinidas são empregadas para definir “clusters” individuais codificados na região. Sempre que o programa encontra um gene que codifica para uma proteína com determinado domínio, uma nova

característica é gerada dentro da região. Esta característica compreende o(s) produto(s) do gene central que desencadeiam a regra e se estendem à esquerda e à direita dos genes centrais em 20 kb conforme definido na regra. Os valores para os diferentes tipos de cluster são determinados empiricamente e geralmente tendem a superestimar, ou seja, incluem também genes adjacentes.

Depois que as características são atribuídas eles são verificados quanto a sobreposição e são agrupados em vários tipos de “*cluster* candidato” pois muitos BGCs compreendem várias classes de maquinaria biossintética. Os *clusters* candidatos são classificados em 4 tipos: I - híbrido químico, contém *clusters* que compartilham sequências codificadoras; II - intercalado, contém *clusters* que não compartilham sequências codificadoras mas seus locais se sobrepõem; III - vizinhança, contém *clusters* que se sobrepõem transitivamente em suas vizinhanças. Cada *cluster* do tipo vizinhança também pertencerá ao tipo híbrido químico ou intercalado; IV - único, apenas um *cluster* está contido naquela região.

Os *clusters* identificados são pesquisados em banco de dados como o MIBiG, uma coleção de dados com curadoria manual sobre clusters de genes (KAUTSAR et al., 2020). Esta informação é então usada para fornecer uma suposta função dos produtos do gene.

3.5.2. *Kyoto Encyclopedia of Genes and Genomes (KEGG) Automatic Annotation Server (KAAS)*

O KAAS (*KEGG Automatic Annotation Server*) fornece anotações funcionais de genes a partir da comparação das sequências de aminoácidos com banco de dados de grupos ortólogos curados manualmente em KEGG GENES. O KEGG GENES é um banco de dados de genes e proteínas de genomas completos gerados a partir de bancos publicamente disponíveis, como o NCBI RefSeq (PRUITT et al., 2007) e GenBank (BENSON et al., 2013). O resultado contém atribuições KEGG *Orthology* (KO), caminhos KEGG e classificações funcionais. Os métodos de atribuição de KO podem ser executados utilizando o melhor *hit* bi-direcional (BBH, padrão) ou no melhor *hit* unidirecional (SBH). Se as sequências forem de um genoma completo ou parcial, o método baseado em BBH será mais preciso do que o baseado em SBH. Entretanto, se o número da sequência for muito grande, como os de metagenomas, o método baseado em SBH é suficiente (MORIYA et al., 2007).

No primeiro passo, a sequência é comparada com cada genoma do banco de dados KEGG GENES. Realiza a busca de genes enzimáticos homólogos nas direções direta e reversa, tomando cada gene no genoma A como uma consulta de similaridade de sequências e posicionamento correlacional dos genes do genoma B, e vice-versa. Em seguida, os homólogos com pontuações de *bits* inferiores a 60 são removidos e os superiores são selecionados como candidatos a ortólogos e divididos em grupos KO. Finalmente, a pontuação de atribuição é calculada com base na probabilidade para cada grupo KO (KANEHISA et al., 2016).

Os dados gerados dão origem aos mapas metabólicos que representam diversas redes moleculares, como redes de interação/reação para metabolismo, processamento de informações genéticas, processamento de informações ambientais e outros processos celulares (KANEHISA et al., 2016).

3.5.3. deepBGC

A ferramenta deepBGC detecta BGCs através da abordagem de aprendizado profundo (*deep learning*) usando redes neurais recorrentes (RNNs) de memória bidirecional de longo e curto prazo (BiLSTM) e representações vetoriais de domínios da família de proteínas (Pfam) que juntos são capazes de detectar semelhanças e diferenças de curto e longo prazo entre genomas distantes ou adjacentes (HANNIGAN et al., 2019). As sequências genômicas brutas são usadas para predição de genes pelo Prodigal (HYATT et al., 2010). Os domínios Pfam são atribuídos a cada fase de leitura aberta (ORF) usando HMMER (EDDY, 1998). O BiLSTM gera a pontuação de classificação para cada domínio que são atribuídas entre os genes. Os possíveis genes consecutivos de um BGC candidatos são montados em BGCs putativos (HANNIGAN et al., 2019). Uma etapa de pós-processamento opcional permite a fusão de BGCs vizinhos com base na presença de uma via biossintética conhecida, comprimento mínimo do cluster e lacunas entre BGCs adjacentes. Os BGCs são classificados com base na classe do composto e na atividade molecular (HANNIGAN et al., 2019). Para este trabalho a ferramenta foi executada em modo *default*, onde a pontuação mínima para a detecção de clusters biossintéticos foi de 0,5.

4. RESULTADOS E DISCUSSÃO

4.1. ISOLADOS BACTERIANOS



Figura 1: Foto do isolado MAPA1 em placa de Petri tirada com celular.

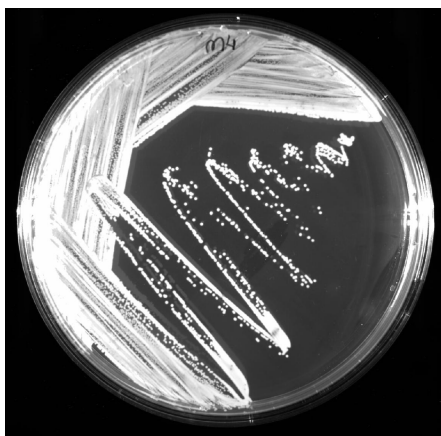


Figura 2: Foto do isolado M4 em placa de Petri tirada no *ChemiDoc*.

4.2. MONTAGEM DOS GENOMAS

Após a finalização da montagem, o genoma do isolado MAPA1 apresenta 5.997.981 pb, 45 *contigs* e 6.178 regiões codificadoras (CDSs) (Tabela 1). Enquanto o genoma do isolado M4 apresenta 5.062.732 pb, 60 *contigs* e 4.694 CDSs (Tabela 2).

Tabela 1: Informações finais sobre a montagem do genoma MAPA1

Tamanho do genoma (MAPA1)	5.997.981 pb
Quantidade de <i>contigs</i>	45
CDSs (proteínas)	6.178
rRNAs	5, 6, 2 (5S, 16S, 23S)

Tabela 2: Informações finais sobre a montagem do genoma M4

Tamanho do genoma (M4)	5.062.732 pb
Quantidade de <i>contigs</i>	60
CDSs (proteínas)	4.694
rRNAs	1, 1, 2 (5S, 16S, 23S)

4.3. ANÁLISE DA QUALIDADE DOS GENOMAS

Os genomas dos isolados bacterianos MAPA1 e M4 foram analisados pelo programa MiGA Online com o objetivo de determinar a qualidade dos genomas.

4.3.1. Análise MAPA1

O genoma do isolado MAPA1 contém 105 dos 106 genes de cópia única encontrados, resultando em uma completude de 99,1% (muito alta). O isolado apresenta taxa de contaminação de 7,5% (baixa), pois contém 7 genes que apresentam entre duas ou três cópias e também apresenta a ausência de 1 dos genes de cópia única. Portanto, resulta na qualidade de 61,59% (alta qualidade) (Figura 3 e 4).

A partir deste resultado de uma completude muito alta e uma baixa taxa de contaminação é possível utilizar o genoma do isolado MAPA1 nas próximas etapas de análise dos metabólitos secundários, BGCs e rotas metabólicas.

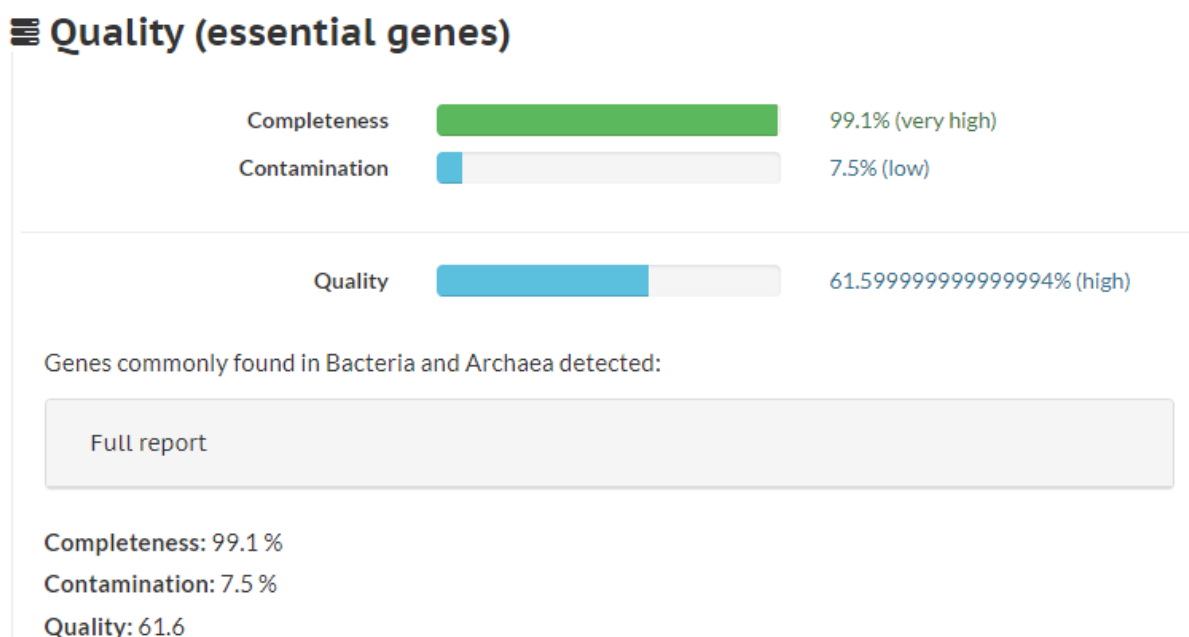


Figura 3: Imagem retirada do resultado da análise de qualidade do MiGA Online do isolado MAPA1.

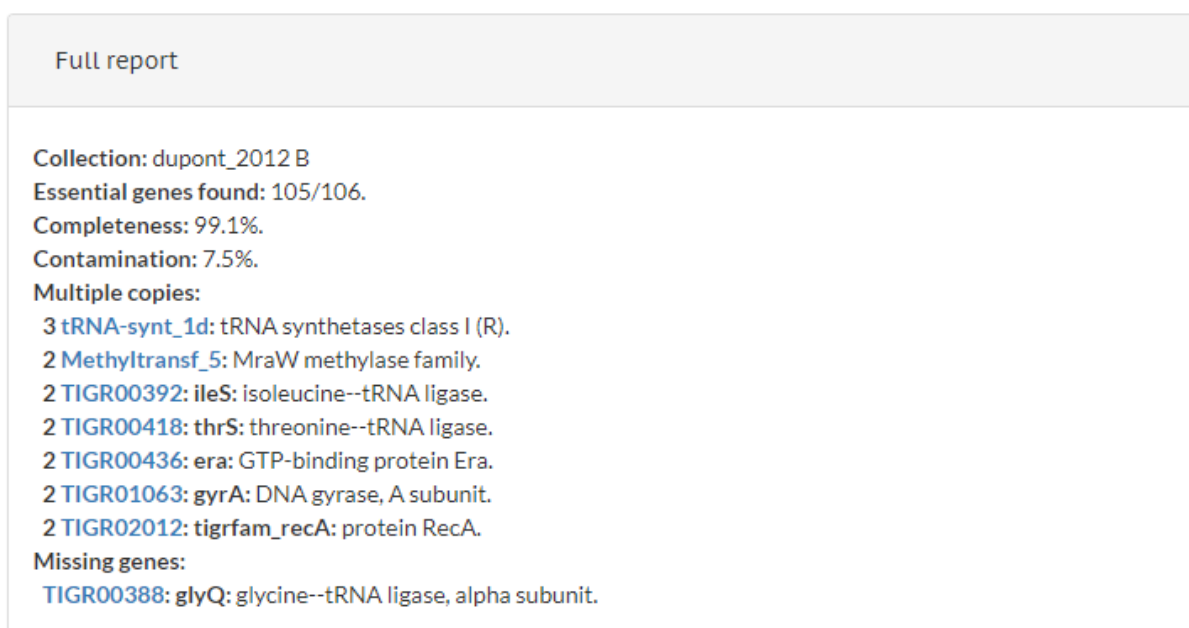


Figura 4: Imagem do *Full report* da Figura 3.

4.3.2. Análise M4

O genoma do isolado M4 contém 104 dos 106 genes de cópia única encontrados, resultando em uma completude de 98,1% (muito alta). O isolado apresenta taxa de contaminação de 0,9% (muito baixa), pois contém 1 gene que

apresenta duas cópias e também apresenta a ausência de 2 dos genes de cópia única. Portanto, resulta na qualidade de 93,6% (excelente) (Figura 5 e 6).

A partir deste resultado de uma completude muito alta e uma taxa de contaminação muito baixa é possível utilizar o genoma do isolado M4 nas próximas etapas de análise dos metabólitos secundários, BGCs e rotas metabólicas.

☰ Quality (essential genes)

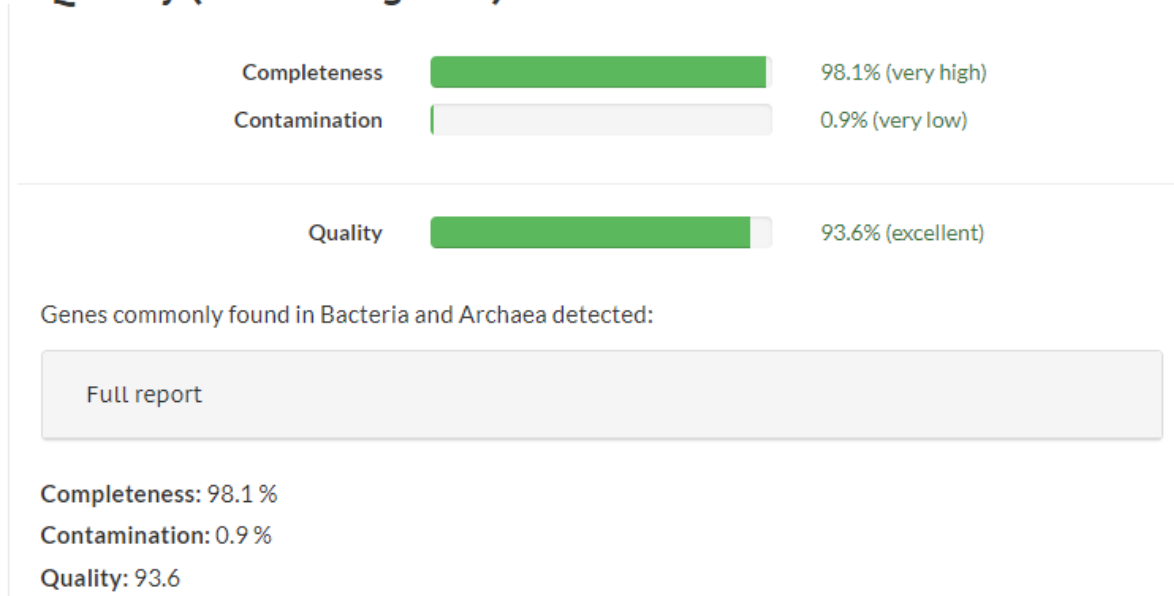


Figura 5: Imagem retirada do resultado da análise de qualidade do MiGA Online do isolado M4.

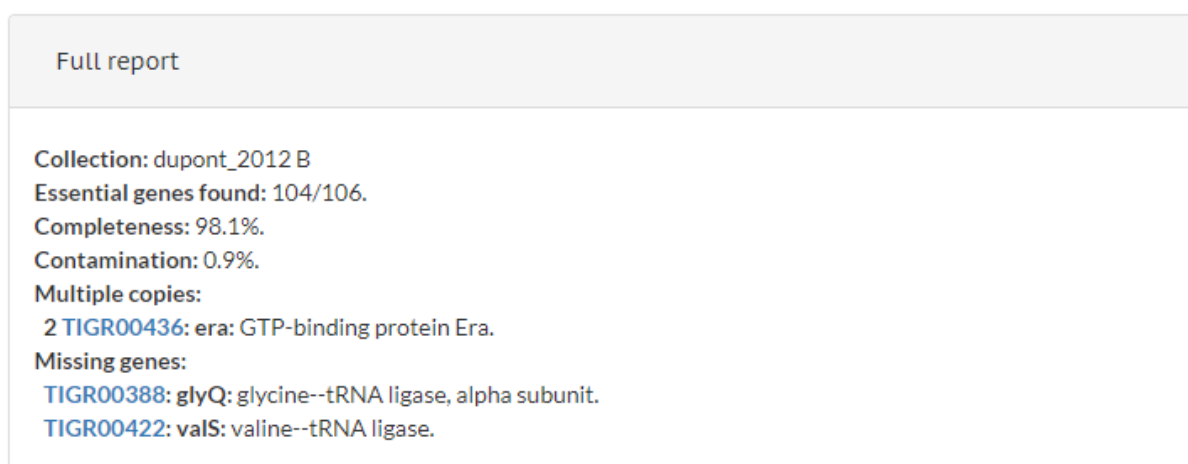


Figura 6: Imagem do *Full report* da Figura 5.

4.4. ANÁLISE DA FILOGENIA DOS GENOMAS

Os genomas dos isolados bacterianos MAPA1 e M4 foram analisados pelo programa TYGS para a identificação das espécies e reconstrução de árvores filogenéticas. Caso o genoma da linhagem do usuário comparado com um genoma do banco de dados obtivesse um limiar igual ou acima de 70%, estes seriam considerados como mesma espécie, em contrapartida, aquelas abaixo do limiar seriam consideradas distintas, ou seja, possíveis novas espécies (AUCH et al., 2010). E quanto menor o valor σ , maior a semelhança entre as espécies relacionadas (HOLLAND et al., 2002).

4.4.1. Análise MAPA1

A análise filogenética do isolado MAPA1 resultou no valor de dDDH de 72% com a espécie *Priestia megaterium* a partir do valor da fórmula d_4 , pois sua utilização independe do comprimento do genoma e, portanto, é robusta contra o uso de genomas incompletos. Além disso, a diferença percentual do conteúdo de G+C do isolado e da *Priestia megaterium* é de 0,28. (Tabela 3)

Outro dado apresentado foi a inferência de três árvores filogenéticas baseadas no 16S (Figura 7), no genoma completo (Figura 8) e no proteoma (Figura 9). Entre elas, as árvores do 16S e do proteoma apresentam um bom suporte de bootstrap nos ramos.

Tabela 3: Resultado da análise filogenética do TYGS do isolado MAPA1 - valor de dDDH.

Genoma analisado	Espécie comparada	dDDH (%)	Diferença percentual do conteúdo G+C
MAPA1	<i>Priestia megaterium</i> ATCC 14581	72,0	0,28

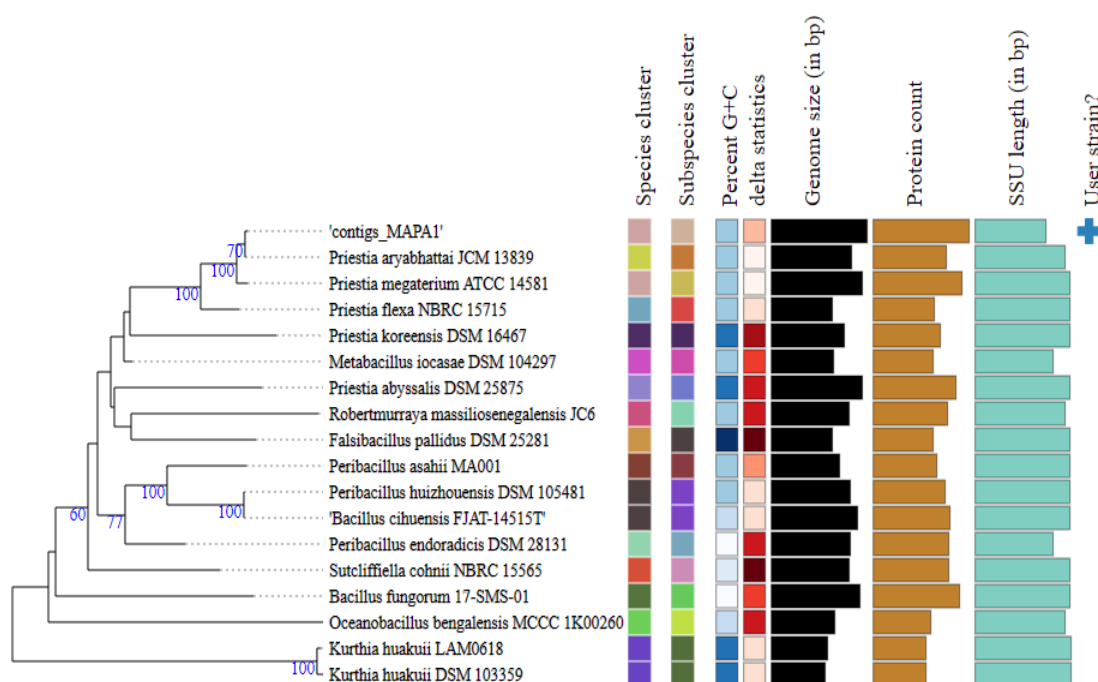


Figura 7: Imagem retirada do resultado da análise filogenética do TYGS do isolado MAPA1 - GBDP *tree* (16S rDNA *gene sequence-based*).

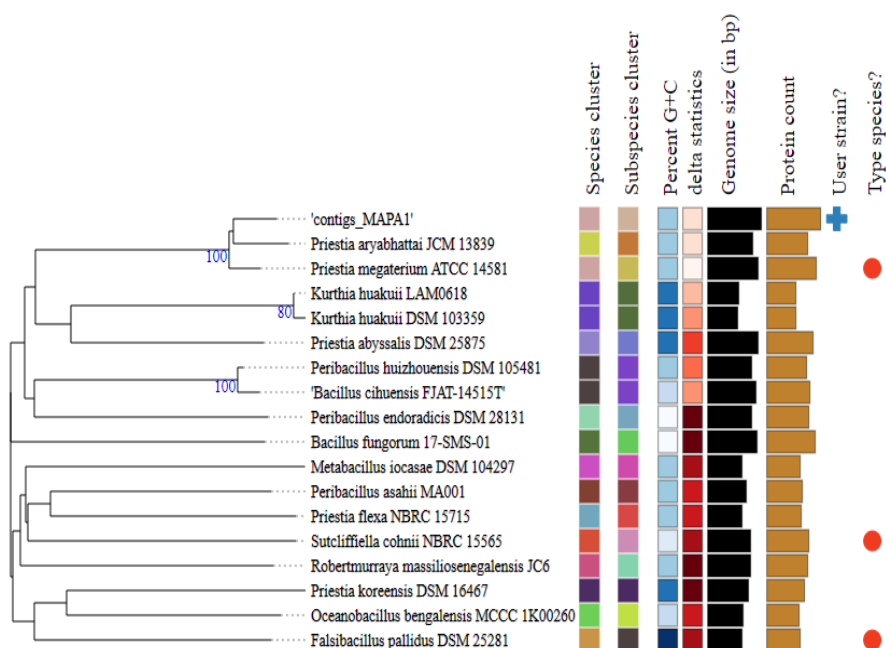


Figura 8: Imagem retirada do resultado da análise filogenética do TYGS do isolado MAPA1 - GBDP tree (whole-genome sequence-based).

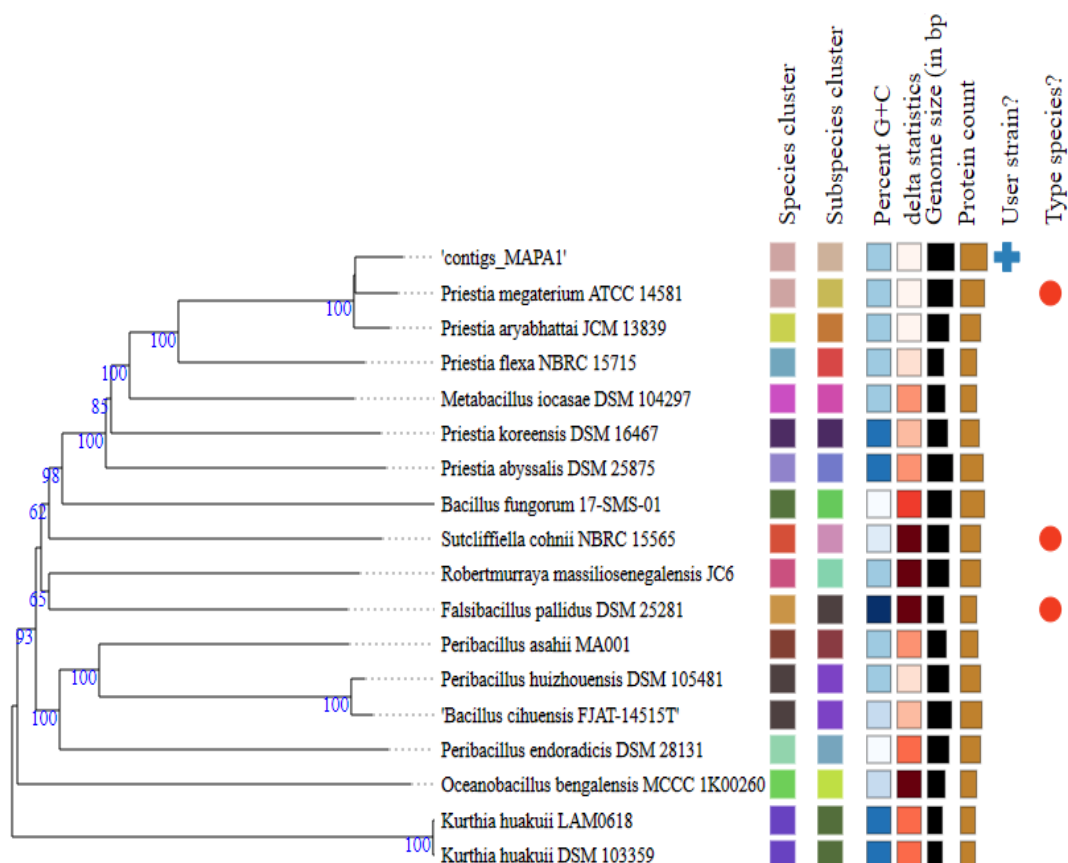


Figura 9: Imagem retirada do resultado da análise filogenética do TYGS do isolado MAPA1 - GBDP tree (whole-proteome-based).

Estes resultados indicam que o isolado MAPA1 é uma cepa da espécie *Priestia megaterium*. *Priestia* é um gênero de bactérias aeróbicas, principalmente Gram-positiva em forma de bastonete, capazes de formar endósporos (GUPTA et al., 2020). A maioria das espécies são móveis e podem ser isoladas de diversas fontes, incluindo fezes, solo e sedimentos marinhos. A faixa de temperatura para o crescimento é de 5 a 48 °C, com o crescimento ideal ocorrendo na faixa de 28 a 37 °C (GUPTA et al., 2020).

4.4.2. Análise M4

A análise filogenética do isolado M4 resultou em valores de dDDH entre 20 e 21,5% com espécies de *Arthrobacter*. Além disso, as diferenças percentuais do conteúdo de G+C do isolado e da *Arthrobacter* são maiores que 3 (Tabela 4).

Outro dado apresentado foi a inferência de duas árvores filogenéticas baseadas no 16S (Figura 10) e no genoma completo (Figura 11), ambas com um

bom suporte de *bootstrap* nos ramos das árvores.

Tabela 4: Resultado da análise filogenética do TYGS do isolado M4 - valor de dDDH.

Genoma analisado	Espécie comparada	dDDH (%)	Diferença percentual do conteúdo G+C
M4	<i>Arthrobacter celericrescens</i> NEAU-SA2	21,5	3,91
M4	<i>Arthrobacter silvisoli</i> NEAU-SA1	21,4	3,6
M4	<i>Arthrobacter cupressi</i> DSM 24664	21,3	3,77
M4	<i>Arthrobacter cupressi</i> CGMCC 1.10783	21,2	3,83

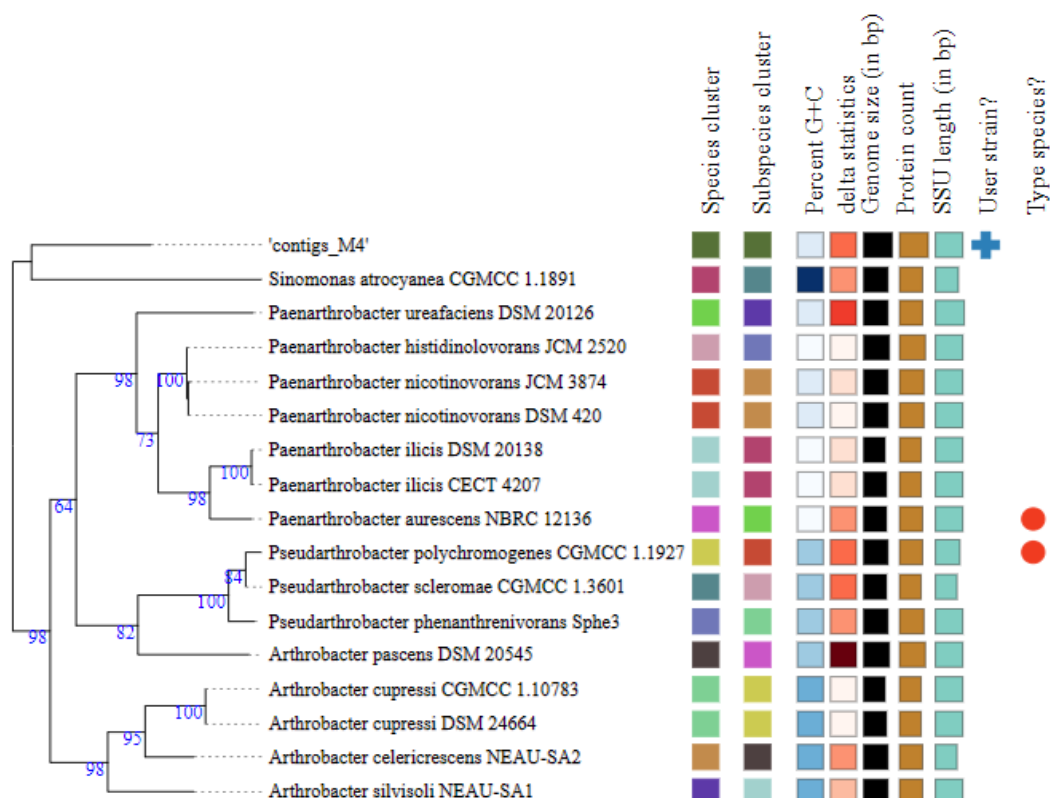


Figura 10: Imagem retirada do resultado da análise filogenética do TYGS do isolado M4 - GBDP tree (16S rDNA gene sequence-based).

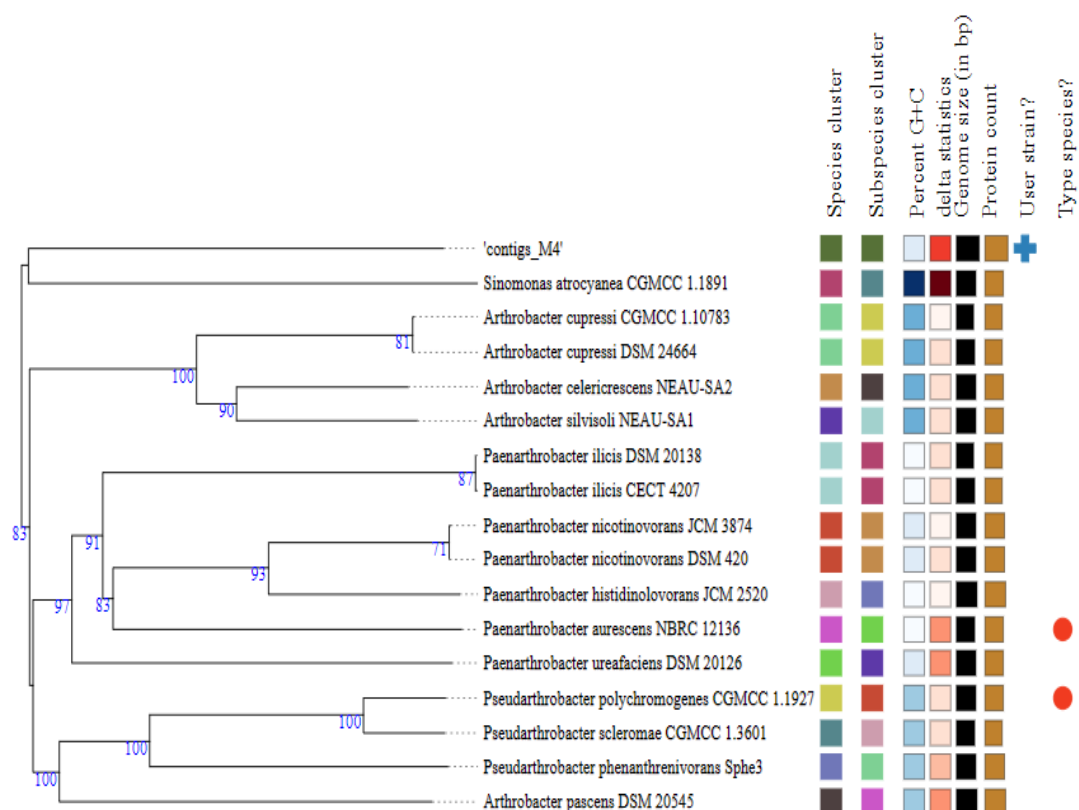


Figura 11: Imagem retirada do resultado da análise filogenética do TYGS do isolado M4 - GBDP tree (whole-genome sequence-based).

Estes resultados indicam que o isolado M4 é uma nova espécie de *Arthrobacter*, um gênero de bactérias principalmente do solo cuja principal característica distintiva é um ciclo de crescimento bastonete-coco (GOBBETTI e RIZZELLO, 2014). As células cocoides podem assumir formas grandes quando em condições de estresse nutricional severo. Ambas as formas bastonetes e cocoides são gram-positivas. As células não formam endósporos e são aeróbios obrigatórios e catalase positiva (GOBBETTI e RIZZELLO, 2014).

4.5. ANÁLISE DE METABÓLITOS SECUNDÁRIOS E ROTAS METABÓLICAS

4.5.1. antiSMASH

4.5.1.1. Análise da *Priestia megaterium*

Com base no genoma da *Priestia megaterium* foram identificadas 9 regiões de *clusters* biossintéticos, contendo 1 *cluster* para produção de sideróforo, 3 para produção de terpenos, 1 para a produção de fosfonato, 1 para NRP, 1 para PKS do tipo III e 2 para RiPP (Figura 12).

Dentre estes 9 *clusters*, 3 deles apresentaram uma baixa similaridade com BGCs com funções já determinadas depositados nos banco de dados. A região 1.2 apresenta similaridade de 13% com o *cluster* responsável pela produção de surfactina, responsável por reduzir a tensão superficial (BARROS et al., 2007); a região 3.1 apresenta similaridade de 50% com o *cluster* responsável pela produção de carotenóide, um grupo de pigmento utilizado como corantes alimentícios (UENOJO et al., 2007); e a região 3.3 apresenta similaridade de 60% com o *cluster* responsável pela produção de paeninodina, um lassopeptídeo sem uma função muito bem caracterizada, porém estudos apontam que ela pode servir como uma molécula sinalizadora ou participar da autoimunidade no organismo produtor (ZHU et al., 2016) (Figura 12). As outras 6 regiões não apresentaram similaridade alguma com o banco de dados.

Estes resultados podem indicar que esses BGCs possivelmente estão relacionados a vias de biossíntese de novas moléculas ou que pertencem a vias não caracterizadas de moléculas já conhecidas. Além disso, esses dados revelam que a *Priestia megaterium* possui um inexplorado arsenal metabólico.

Region	Type	From	To	Most similar known cluster		Similarity
Region 1.1	siderophore	1	15,841			
Region 1.2	terpene	182,898	201,403	surfactin	NRP:Lipopeptide	13%
Region 1.3	terpene	1,708,316	1,730,184			
Region 3.1	terpene	20,462	41,310	carotenoid	Terpene	50%
Region 3.2	phosphonate	149,041	190,558			
Region 3.3	lassopeptide	471,836	495,754	paeninodin	RiPP	60%
Region 4.1	T3PKS	21,339	62,424			
Region 11.1	lanthipeptide-class-i	7,882	34,392			
Region 26.1	RiPP-like	6,404	17,840			

Figura 12: Imagem retirada do resultado da análise de metabólitos secundário e BGCs do antiSMASH da *Priestia megaterium*.

4.5.1.2. Análise *Arthrobacter sp.*

Com base no genoma da *Arthrobacter sp.* foram identificadas 5 regiões de

clusters biossintéticos, contendo 1 *cluster* para produção de PKS do tipo III, 1 para produção de betalactona, 2 para NRPs e 1 para produção de furano (Figura 13).

Dentre estes 5 *clusters*, 3 deles apresentaram uma baixa similaridade com BGCs com funções já determinadas depositados nos banco de dados. A região 3.1 apresenta similaridade de 7% com o *cluster* responsável pela produção de microansamicina uma família de metabólitos secundários bacterianos que mostram atividade antimicrobiana (HILL, STAUNTON, 2010); a região 11.1 apresenta similaridade de 31% com o *cluster* responsável pela produção de estenotricina que possui atividade antimicrobiana (LIU et al., 2014); e a região 16.1 apresenta similaridade de 14% com o *cluster* responsável pela produção de metilenomicina A, um antibiótico produzido pela *Streptomyces coelicolor* (BRIAN et al., 1996) (Figura 13). As outras 2 regiões não apresentaram similaridade alguma com o banco de dados.

Estes resultados podem indicar que esses BGCs possivelmente estão relacionados a vias de biossíntese de novas moléculas ou que pertencem a vias não caracterizadas de moléculas já conhecidas. Além disso, estes dados revelam que a *Arthrobacter sp.* possui um inexplorado arsenal metabólico.









Region	Type	From	To	Most similar known cluster		Similarity
Region 1.1	T3PKS 	251,099	292,235			
Region 3.1	betalactone 	373,799	399,103	microansamycin 	Polyketide	7%
Region 4.1	NRPS-like 	269,717	312,497			
Region 11.1	NAPAA 	62,275	96,261	stenothricin 	NRP:Cyclic depsipeptide	31%
Region 16.1	furan 	46,901	66,239	methylenomycin A 	Other	14%

Figura 13: Imagem retirada do resultado da análise de BGCS e metabólitos secundário do antiSMASH da *Arthrobacter sp.*.

4.5.2. KEGG Automatic Annotation Server (KAAS)

4.5.2.1. Análise da *Priestia megaterium*

Foram encontradas 25 rotas metabólicas. Das rotas previstas, 16 são provenientes do metabolismo primário ou central como glicólise e ciclo do citrato e 6 são provenientes do processamento de informações genéticas como a replicação do DNA e recombinação homóloga (Tabela 5).

Também foram preditas 2 rotas com falsos positivos como a via de resistência à insulina e a via da cardiomiopatia diabética (Tabela 5). Tais vias não estão presentes em bactérias, logo não foram consideradas relevantes assim como as do metabolismo primário.

Não foram identificadas rotas relacionadas com a produção de antimicrobianos.

Tabela 5: Resultado da análise de rotas metabólicas da *Priestia megaterium* pelo KAAS.

	Rota metabólica	Enzimas encontradas	Nomenclatura
Metabolismo primário/central	Glicólise/Gluconeogênese	1	piruvato
	Ciclo do citrato (TCA)	1	piruvato carboxilase
	Metabolismo da Frutose e Manose	1	Componente IIA da enzima fosfotransferase específica de manitol
	Metabolismo de açúcares de aminoácidos e nucleotídeos	1	transaminase de glutamina-frutose-6-fosfato (isomerização)
	Metabolismo do Piruvato	2	piruvato e piruvato carboxilase
	Metabolismo de glioxilato e dicarboxilato	1	formato desidrogenase
	Metabolismo do propanoato	1	metilmalonato-semialdeído desidrogenase (acilação)
	Vias de fixação de carbono em procariotos	2	piruvato e piruvato carboxilase
	Metabolismo do metano	1	formato desidrogenase
	Metabolismo do Nitrogênio	1	glutamato sintase (NADPH)
	Metabolismo de Glicerolípídeo	1	glicerol quinase
	Purina	3	fosforibosil formilglicinamidina sintase, guanosina monofosfato (GMP) sintetase e inosina-5'-monofosfato (IMP) desidrogenase

	Metabolismo da Alanina, aspartato e glutamato	3	glutamato sintase (NADPH), aspartato amônia-liase e transaminase de glutamina-frutose-6-fosfato (isomerização)
	Degradação de valina, leucina e isoleucina	1	metilmalonato-semialdeído desidrogenase (acilação)
	Metabolismo da fenilalanina	1	2-enoato redutase
	Sistema de fosfotransferase	1	Componente IIA da enzima fosfotransferase específica de manitol
Outras	Droga - outras enzimas	2	inosina-5'-monofosfato (IMP) desidrogenase e guanosina monofosfato (GMP) sintetase
Processamento de informações genéticas	RNA polimerase	1	Subunidade β' de RNA polimerase dirigida por DNA
	Replicação DNA	1	Subunidade α da DNA polimerase III
	Reparo da excisão de nucleotídeo	1	fator de acoplamento de reparo de transcrição (superfamília II helicase) (MFD)
	Reparo de mismatch	1	Subunidade α da DNA polimerase III
	Recombinação homóloga	1	Subunidade α da DNA polimerase III
	Transportadores ABC	1	Proteína de ligação ao substrato do sistema de transporte de D-metionina (MetQ)
Falso positivo	Cardiomiopatia diabética	1	transaminase de glutamina-frutose-6-fosfato (isomerização)
	Resistência à insulina	1	proteína quinase serina/treonina

4.5.2.2. Análise *Arthrobacter* sp.

Foram encontradas 49 rotas metabólicas. Das rotas preditas, 29 são provenientes do metabolismo primário ou central como glicólise e ciclo do citrato e 6 são provenientes do processamento de informações genéticas como a degradação de RNA e necroptose (Tabela 6).

Também foram preditas 12 rotas com falsos positivos como a via da doença de Parkinson e via da secreção pancreática (Tabela 6). Tais vias não estão presentes em bactérias, logo não foram consideradas relevantes assim como as do metabolismo primário.

Não foram identificadas rotas relacionadas com a produção de antimicrobianos.

Tabela 6: Resultado da análise de rotas metabólicas da *Arthrobacter sp.* pelo KAAS.

Grupo	Rota metabólica	Enzimas encontradas	Nomenclatura
Metabolismo primário/central	Glicólise/Gliconeogênes e	3	piruvato desidrogenase, aldeído desidrogenase e acetil-CoA sintetase
	Ciclo do citrato (TCA)	2	isocitrato desidrogenase e piruvato desidrogenase
	Metabolismo da galactose	1	beta-galactosidase
	Metabolismo do ascorbato e do aldarato	1	aldeído desidrogenase
	Metabolismo do amido/sacarose	3	glicogênio fosforilase, isoamilase e enzima ramificação 1,4-alfa-glucana
	Metabolismo do Piruvato	3	piruvato desidrogenase, acetil-CoA sintetase e aldeído desidrogenase
	Metabolismo do Glioxilato/dicarboxilato	2	acetil-CoA sintetase e glicina desidrogenase
	Metabolismo do Propanoato	1	acetil-coa sintetase
	Via da Fixação de carbono	2	acetil-CoA sintetase e isocitrato desidrogenase
	Metabolismo do Metano	2	subunidade A da metilamina glutamato N-metil transferase e acetil-CoA sintetase

	Metabolismo do Nitrogênio	1	glutamato sintetase
	Metabolismo do enxofre	1	Sulfidrilase de O-succinil homoserina
	Degradação de Ácido graxo	2	acil-CoA desidrogenase e aldeído desidrogenase
	Metabolismo de glicerolípídeos	1	aldeído desidrogenase
	Metabolismo de Esfingolípídeos	1	beta-galactosidase
	Metabolismo da purina	2	fosforibosil formilglicinamidina sintase e Adenilosuccinato liase
	Metabolismo da pirimidina	1	Fosfato de carbamoil sintase
	Metabolismo da Alanina, Aspartato e Glutamato	4	aspartato amônia-liase, Adenilosuccinato liase, glutamato sintetase e Fosfato de carbamoil sintase
	Metabolismo da Glicina/serina/treonina	1	glicina desidrogenase
	Metabolismo da cisteína e metionina	1	Sulfidrilase de O-succinil homoserina
	Degradação da Valina, leucina e isoleucina	2	acil-CoA desidrogenase e aldeído desidrogenase
	Degradação da lisina	2	aminobutiraldeído desidrogenase e aldeído desidrogenase
	Metabolismo da Arginina e prolina	2	aminobutiraldeído desidrogenase e aldeído desidrogenase
	Metabolismo da histidina	1	aldeído desidrogenase
	Metabolismo do triptofano	1	aldeído desidrogenase
	Metabolismo da Beta-alanina	2	aminobutiraldeído desidrogenase e aldeído desidrogenase
	Metabolismo da Glutamina	1	isocitrato desidrogenase
	Degradação de outros glicanos	1	beta-galactosidase

	Biossíntese de Pantotenato e CoA	1	aldeído desidrogenase
Possíveis rotas de interesse	Degradação do limoneno e do pineno	1	aldeído desidrogenase
	Degradação de cloroalcano e cloroalqueno	1	aldeído desidrogenase
Processamento de informações genéticas	RNA polimerase	1	Subunidade β' de RNA polimerase dirigida por DNA
	Degradação de RNA	2	polirribonucleotídeo nucleotíditransferase (PNPase) e chaperona molecular DnaK
	Reparo por excisão de nucleotídeos	1	fator de acoplamento de reparo de transcrição (superfamília II helicase) (MFD)
	Transportadores ABC	2	transportadores ABC, subfamília B, bombas de resistência a múltiplas drogas (EfrA/B) e proteína de ligação de ATP do sistema de transporte de glutatona (GsiA)
	Peroxisomo	1	isocitrato desidrogenase (IDH)
	Necroptose	1	glicogênio fosforilase
Falsos positivos	Formação de biofilme (<i>Escherichia coli</i>)	1	glicogênio fosforilase
	Via da sinalização de insulina	1	glicogênio fosforilase
	Via da sinalização de glucagon	1	glicogênio fosforilase
	Via da sinalização dos receptores ativados por proliferadores de peroxissoma (PPARs)	1	acil-CoA desidrogenase
	Secreção pancreática	1	carboxipeptidase A1
	Digestão e absorção proteica	1	carboxipeptidase A1
	Regulação de longevidade - múltiplas espécies	1	Subunidade CLPB de ligação de ATP da protease CLP dependente de ATP

	Metabolismo central do carbono em câncer	1	isocitrato desidrogenase
	Infecção por <i>Salmonella</i>	1	tiorredoxina
	Tuberculose	1	chaperona molecular DnaK
	Doença de Parkinson	1	tiorredoxina
	Resistência à insulina	1	proteína quinase serina/treonina

4.5.3. deepBGC

Os resultados do deepBGC (Tabela 7, 8, 9 e 10) foram analisados com o parâmetro de pontuação mínima para a detecção de *clusters* biossintéticos entre 0,5 e 1, e estes resultados foram separados e tabelados por intervalos de de ~0,1 de pontuação.

4.5.3.1. Análise da *Priestia megaterium*

Foram encontrados 46 BGCs que estão classificados quanto às suas funções e classes (Tabela 7 e 8). Entre o *score* 0,5-0,6 estão contidos 22 BGCs sendo 16 com função antibacteriana e 6 com função não determinada. E quanto a classe, 2 são policetídeos, 1 RiPP, 17 possuem classe não determinada e 2 são de outras classes.

Entre o *score* 0,61-0,7 estão contidos 10 BGCs sendo 6 com função antibacteriana, 1 citotóxico e 3 com função não determinada. E quanto a classe, 3 são policetídeos, 2 RiPPs, 4 possuem classe não determinada e 1 pertence a outra classe.

Entre o *score* 0,71-0,8 estão contidos 8 BGCs sendo 3 com função antibacteriana e 5 com função não determinada. E quanto a classe, 7 possuem classe não determinada e 1 pertence a outra classe.

Entre o *score* 0,81-0,9 estão contidos 4 BGCs sendo 1 com função antibacteriana e 3 com função não determinada. E quanto a classe, 1 é um sacarídeo e 3 possuem classe não determinada.

Entre o score 0,91-1 estão contidos 2 BGCs sendo 1 com função antibacteriana e 1 com função não determinada. E quanto a classe, 1 é um RiPP e o outro pertence a classe dos sacarídeos.

Tabela 7: Resultado da análise de função de BGCs da *Priestia megaterium* pelo deepBGC.

Score	Quantidade de BGCs	Função		
		Antibacteriana	Função não determinada	Citotóxica
(0,5 - 0,6)	22	16	6	-
(0,61 - 0,7)	10	6	3	1
(0,71 - 0,8)	8	3	5	-
(0,81 - 0,9)	4	1	3	-
(0,91 - 1)	2	1	1	-
Total	46	27	18	1

Tabela 8: Resultado da análise de classe de BGCs da *Priestia megaterium* pelo deepBGC.

Score	Quantidade de BGCs	Classe				
		Classe não determinada	Policetídeo	RiPP	Sacarídeo	Outra
(0,5 - 0,6)	22	17	2	1	-	2
(0,61 - 0,7)	10	4	3	2	-	1
(0,71 - 0,8)	8	7	-	-	-	1
(0,81 - 0,9)	4	3	-	-	1	-
(0,91 - 1)	2	-	-	1	1	-
Total	46	31	5	4	2	4

4.5.3.2. Análise da *Arthrobacter* sp.

Foram encontrados 60 BGCs que estão classificados quanto às suas funções e classes (Tabela 9 e 10). Entre o score 0,5-0,6 estão contidos 39 BGCs sendo 35 com função antibacteriana e 4 com função não determinada. E quanto a classe, 2 são policetídeos, 3 RiPPs, 2 terpenos, 1 NRP, 30 possuem classe não determinada e 1 é de outra classe.

Entre o score 0,61-0,7 estão contidos 9 BGCs sendo 4 com função antibacteriana, 1 inibitório e 4 com função não determinada. E quanto a classe, 1 é policetídeo, 1 sacarídeo e 7 possuem classe não determinada.

Entre o score 0,71-0,8 estão contidos 4 BGCs sendo 1 com função antibacteriana e 3 com função não determinada. E quanto a classe, 2 possuem classe não determinada, 1 RiPP e 1 pertence a outra classe.

Entre o score 0,81-0,9 estão contidos 5 BGCs sendo 1 com função citotóxica e 4 com função não determinada. E quanto a classe, 1 é um sacarídeo e 4 possuem classe não determinada.

Entre o score 0,91-1 estão contidos 3 BGCs sendo os 3 com função antibacteriana. E quanto a classe, 1 é um policetídeo e 2 possuem classe não determinada.

Tabela 9: Resultado da análise de função de BGCs da *Arthrobacter sp.* pelo deepBGC.

Score	Quantidade de BGCs	Função			
		Antibacteriana	Função não determinada	Citotóxica	Inibitória
(0,5 - 0,6)	39	35	4	-	-
(0,61 - 0,7)	9	4	4	-	1
(0,71 - 0,8)	4	1	3	-	-
(0,81 - 0,9)	5	-	4	1	-
(0,91 - 1)	3	3	-	-	-
Total	60	43	15	1	1

Tabela 10: Resultado da análise de classe de BGCs da *Arthrobacter sp.* pelo deepBGC.

Score	Quantidade de BGCs	Classe						
		Classe não determinada	Terpeno	NRP	Policetídeo	RiPP	Sacarídeo	Outra
(0,5 - 0,6)	39	30	2	1	2	3	-	1
(0,61 - 0,7)	9	7	-	-	-	1	1	-
(0,71 - 0,8)	4	2	-	-	-	1	-	1
(0,81 - 0,9)	5	4	-	-	-	-	1	-
(0,91 - 1)	3	2	-	-	1	-	-	-
Total	60	45	2	1	3	5	2	2

5. CONCLUSÃO

A partir de um isolamento feito de amostras de solo e com um sequenciamento e montagem de genomas de qualidade foi possível realizar as análises *in silico* para a predição de regiões de *clusters* biossintéticos (BGCs) que produzissem metabólitos secundários de interesse.

A análise do antiSMASH encontrou possíveis regiões com a presença de BGCs que quando comparadas com os resultados do deepBGC identificou-se que as regiões abordadas eram em sua maioria localizadas na mesma região genômica. Com isso, é possível verificar que apesar da utilização de ferramentas diferentes para a análise de BGCs que utilizam parâmetros e metodologias divergentes, elas conseguem prever os BGCs em regiões parecidas. Entretanto, a ferramenta deepBGC foi capaz de identificar uma quantidade maior de BGCs devido a utilização de *deep learning* e redes neurais. Por isso, o deepBGC é capaz de identificar não só as BGCs em comparação com o banco de dados, mas também identificar novos BGCs e metabólitos secundários.

Porém, uma das dificuldades encontradas é a comparação com o banco de dados, pois devido a grande quantidade de metabólitos existentes ainda há muito a ser caracterizado e depositado nos bancos de dados para que as análises sejam cada vez mais eficazes quanto a classificação da função e classe dos metabólitos. Entretanto, a análise *in silico* em conjunto com a bioprospecção de novos organismos, está em constante avanço para explorar o campo de metabólitos secundários.

Contudo, a análise *in silico* auxilia quanto ao gerenciamento do tempo e diminuição dos custos de pesquisas *in vitro* para detectar produtos naturais, pois com as predições das análises de bioinformática é possível fazer uma melhor escolha da amostra que será analisada em laboratório. Logo, após as análises *in silico* tem-se um melhor parâmetro para escolher quais espécies passarão por uma análise manual com experimentos *in vitro* para a validação dos resultados encontrados e aproveitamento do potencial biológico dos microrganismos.

6. OBJETIVOS FUTUROS

A partir dos resultados e das conclusões deste trabalho os objetivos futuros são: acessar a conservação dos genes envolvidos na biossíntese de metabólitos secundários em outros microrganismos dos gêneros *Priestia* e *Arthrobacter* que possuem genoma sequenciado; testar os isolados em laboratório sob diferentes tipos de parâmetros como fontes de carbono, ausência e presença de luz, diferentes concentrações de CO₂ e diferentes temperaturas, no intuito de avaliar a alteração da produção de metabólitos secundários; realizar testes laboratoriais para identificar a capacidade antimicrobiana da *Priestia megaterium* e *Arthrobacter sp.* e mediante a comprovação da atividade antimicrobiana, planeja-se analisar por cromatografia quais são os metabólitos responsáveis por tal atividade; realizar a mesma metodologia deste trabalho para analisar os outros genomas sequenciados em busca de metabólitos secundários e antimicrobianos.

7. REFERÊNCIAS

Albarano L, Esposito R, Ruocco N, Costantini M. Genome Mining as New Challenge in Natural Products Discovery. Mar Drugs. 2020 Apr 9;18(4):199. doi: 10.3390/md18040199. PMID: 32283638; PMCID: PMC7230286.

Allmann S, Baldwin IT. Insects betray themselves in nature to predators by rapid isomerization of green leaf volatiles. Science. 2010 Aug 27;329(5995):1075-8. doi: 10.1126/science.1191634. PMID: 20798319.

Andrews, S. FastQC A Quality control tool for high throughput sequence data <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.

Anthony M. Bolger, Marc Lohse, Bjoern Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics, Volume 30, Issue 15, 1 August 2014, Pages 2114–2120, <https://doi.org/10.1093/bioinformatics/btu170>.

Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, Camarero JA, Campopiano DJ, Challis GL, Clardy J, Cotter PD, Craik DJ, Dawson M, Dittmann E, Donadio S, Dorrestein PC, Entian KD, Fischbach MA, Garavelli JS, Göransson U, Gruber CW, Haft DH, Hemscheidt TK, Hertweck C, Hill C, Horswill AR, Jaspars M,

Kelly WL, Klinman JP, Kuipers OP, Link AJ, Liu W, Marahiel MA, Mitchell DA, Moll GN, Moore BS, Müller R, Nair SK, Nes IF, Norris GE, Olivera BM, Onaka H, Patchett ML, Piel J, Reaney MJ, Rebuffat S, Ross RP, Sahl HG, Schmidt EW, Selsted ME, Severinov K, Shen B, Sivonen K, Smith L, Stein T, Süssmuth RD, Tagg JR, Tang GL, Truman AW, Vederas JC, Walsh CT, Walton JD, Wenzel SC, Willey JM, van der Donk WA. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep.* 2013 Jan;30(1):108-60. doi: 10.1039/c2np20085f. PMID: 23165928; PMCID: PMC3954855.

AUCH, Alexander F.; VON JAN, Mathias; KLENK, Hans-Peter; GÖKER, Markus. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Standards In Genomic Sciences*, [S.L.], v. 2, n. 1, p. 117-134, 2010.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012 May;19(5):455-77. doi: 10.1089/cmb.2012.0021. Epub 2012 Apr 16. PMID: 22506599; PMCID: PMC3342519.

Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res.* 2013 Jan;41(Database issue):D36-42. doi: 10.1093/nar/gks1195. Epub 2012 Nov 27. PMID: 23193287; PMCID: PMC3531190.

Bian G, Ma T, Liu T. In Vivo Platforms for Terpenoid Overproduction and the Generation of Chemical Diversity. *Methods Enzymol.* 2018;608:97-129. doi: 10.1016/bs.mie.2018.04.025. Epub 2018 Jul 6. PMID: 30173775.

Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, Weber T. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* 2021 Jul 2;49(W1):W29-W35. doi: 10.1093/nar/gkab335. PMID: 33978755; PMCID: PMC8262755.

Bloudoff K, Schmeing TM. Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity. *Biochim Biophys Acta Proteins Proteom*. 2017 Nov;1865(11 Pt B):1587-1604. doi: 10.1016/j.bbapap.2017.05.010. Epub 2017 May 16. PMID: 28526268.

Brian P, Riggle PJ, Santos RA, Champness WC. Global negative regulation of *Streptomyces coelicolor* antibiotic synthesis mediated by an *absA*-encoded putative signal transduction system. *J Bacteriol*. 1996 Jun;178(11):3221-31. doi: 10.1128/jb.178.11.3221-3231.1996. PMID: 8655502; PMCID: PMC178074.

Buffalo,V. (2014) Scythe—a Bayesian adapter trimmer (version 0.994BETA). <https://github.com/vsbuffalo/scythe>.

CAMPOS, Caio Cézar Barbosa et al. Diversidade de clusters gênicos biossintéticos em *Streptomyces* sp. MAD39, isolado com potencial para biocontrole de patógenos de cadeias agrícolas regionais. 1º Simpósio de Biotecnologia da UFAM, [S. l.], 3 mar. 2022. Estratégias Biotecnológicas para o desenvolvimento sustentável da Amazônia, p. 7-7.

Challis, G. L. (2005). A Widely Distributed Bacterial Pathway for Siderophore Biosynthesis Independent of Nonribosomal Peptide Synthetases. *ChemBioChem*, 6(4), 601–611. doi:10.1002/cbic.200400283.

CHENG, Cheng; HUA, Zi-Chun. Lasso Peptides: Heterologous Production and Potential Medical Application. *Frontiers in Bioengineering and Biotechnology*, [s. l.], ano 2020, v. 8, 28 set. 2020. DOI 10.3389/fbioe.2020.571165. Disponível em: <https://www.frontiersin.org/article/10.3389/fbioe.2020.571165>. Acesso em: 9 maio 2022.

Cox, M.P., Peterson, D.A. & Biggs, P.J. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* 11, 485 (2010). <https://doi.org/10.1186/1471-2105-11-485>.

Dinghua Li, Chi-Man Liu, Ruibang Luo, Kunihiko Sadakane, Tak-Wah Lam, MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph, *Bioinformatics*, Volume 31, Issue 10, 15 May 2015, Pages 1674–1676, <https://doi.org/10.1093/bioinformatics/btv033>.

Don Cowan, Quinton Meyer, William Stafford, Samson Muyanga, Rory Cameron, Pia Wittwer, Metagenomic gene discovery: past, present and future, *Trends in Biotechnology*, Volume 23, Issue 6, 2005, Pages 321-329, ISSN 0167-7799, <https://doi.org/10.1016/j.tibtech.2005.04.001>.

Du L, Sánchez C, Shen B. Hybrid peptide-polyketide natural products: biosynthesis and prospects toward engineering novel molecules. *Metab Eng.* 2001 Jan;3(1):78-95. doi: 10.1006/mben.2000.0171. PMID: 11162234.

Eddy SR. Profile hidden Markov models. *Bioinformatics*. 1998;14(9):755-63. doi: 10.1093/bioinformatics/14.9.755. PMID: 9918945.

Fábio Barros, Cedenir de Quadros, Mário Maróstica Jr, Gláucia Pastore. (2007). Surfactina: Propriedades químicas, tecnológicas e funcionais para aplicações em alimentos. *Quimica Nova - QUIM NOVA*. 30. 10.1590/S0100-40422007000200031.

FILHO, Carlos Augusto Spartaco Astolfi; SILVA, Maria de Fátima Nunes; BIGI, Mendes Acácio. Bioprospecção e biotecnologia. *Parcerias Estratégicas*, Brasília, ano 2014, v. 19, ed. 38, p. 45-80.

GOBBETTI, M.; RIZZELLO, C.G. *Arthrobacter*: Academic Press. *Encyclopedia of Food Microbiology*, [S. l.], n. 2, p. 69-76, 2014.

GOMES, Elisângela Soares. Análise in silico de um novo "cluster" de PKS II Metagenômico. Orientador: Eliana Gertrudes de Macedo Lemos. 2011. 135 p.

Dissertação (Mestrado em Microbiologia Agropecuária) - Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, Jaboticabal, 2011.

Grininger M. The role of the iterative modules in polyketide synthase evolution. *Proc Natl Acad Sci U S A*. 2020 Apr 21;117(16):8680-8682. doi: 10.1073/pnas.2004190117. Epub 2020 Apr 14. PMID: 32291336; PMCID: PMC7183167.

Gupta RS, Patel S, Saini N, Chen S. Robust demarcation of 17 distinct *Bacillus* species clades, proposed as novel Bacillaceae genera, by phylogenomics and comparative genomic analyses: description of *Robertmurraya kyonggiensis* sp. nov. and proposal for an emended genus *Bacillus* limiting it only to the members of the *Subtilis* and *Cereus* clades of species. *Int J Syst Evol Microbiol*. 2020 Nov;70(11):5753-5798. doi: 10.1099/ijsem.0.004475. Epub 2020 Oct 27. Erratum in: *Int J Syst Evol Microbiol*. 2020 Dec;70(12):6531-6533. PMID: 33112222.

Hannigan GD, Prihoda D, Palicka A, Soukup J, Klempir O, Rampula L, Durcak J, Wurst M, Kotowski J, Chang D, Wang R, Piizzi G, Temesi G, Hazuda DJ, Woelk CH, Bitton DA. A deep learning genome-mining strategy for biosynthetic gene cluster prediction. *Nucleic Acids Res*. 2019 Oct 10;47(18):e110. doi: 10.1093/nar/gkz654. PMID: 31400112; PMCID: PMC6765103.

HILL, Alison M.; STAUNTON, James. 1.10 - Type I Modular PKS. Elsevier: *Comprehensive Natural Products II*, [s. l.], p. 385-452, 2010.

Hyatt, D., Chen, GL., LoCascio, P.F. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11, 119 (2010). <https://doi.org/10.1186/1471-2105-11-119>.

ISAH, Murtala Bindawa et al. Terpenoids as Emerging Therapeutic Agents: Cellular Targets and Mechanisms of Action against Protozoan Parasites. Elsevier: *Studies in Natural Products Chemistry*, [s. l.], v. 59, p. 227-250, 2018.

Joana Barbosa, Tânia Caetano, Sónia Mendo. Class I and Class II Lanthipeptides Produced by *Bacillus* spp. *Journal of Natural Products* 2015 78 (11), 2850-2866. DOI: 10.1021/np500424y.

Julian D. Hegemann, Marcel Zimmermann, Xiulan Xie, and Mohamed A. Marahiel. Lasso Peptides: An Intriguing Class of Bacterial Natural Products. *Accounts of Chemical Research* 2015 48 (7), 1909-1919 DOI: 10.1021/acs.accounts.5b00156.

KALLIO, P. Type II aromatic polyketide biosynthetic tailoring enzymes: diversity and adaptation in *Streptomyces* secondary metabolism. 2008. 49p. University of Turku, Turku, Finland, 2008.

Kongkham B, Prabakaran D, Puttaswamy H. Opportunities and challenges in managing antibiotic resistance in bacteria using plant secondary metabolites. *Fitoterapia*. 2020 Nov;147:104762. doi: 10.1016/j.fitote.2020.104762. Epub 2020 Oct 16. PMID: 33069839.

Kuzuyama, T. Biosynthetic studies on terpenoids produced by *Streptomyces*. *J Antibiot* 70, 811–818 (2017). <https://doi.org/10.1038/ja.2017.12>.

Li Cao, Truc Do, A James Link, Mechanisms of action of ribosomally synthesized and posttranslationally modified peptides (RiPPs), *Journal of Industrial Microbiology and Biotechnology*, Volume 48, Issue 3-4, April 2021, kuab005, <https://doi.org/10.1093/jimb/kuab005>.

Liu WT, Lamsa A, Wong WR, Boudreau PD, Kersten R, Peng Y, Moree WJ, Duggan BM, Moore BS, Gerwick WH, Linington RG, Pogliano K, Dorrestein PC. MS/MS-based networking and peptidogenomics guided genome mining revealed the stenothricin gene cluster in *Streptomyces roseosporus*. *J Antibiot (Tokyo)*. 2014 Jan;67(1):99-104. doi: 10.1038/ja.2013.99. Epub 2013 Oct 23. PMID: 24149839; PMCID: PMC3919142.

Mariana Uenojo, Mário Roberto, Mário Junior, Gláucia Pastore, Pastore. (2007). Carotenóides: Propriedades, aplicações e biotransformação para formação de

compostos de aroma. Quim. Nova. 30. 616-622.
10.1590/S0100-40422007000300022.

Meier-Kolthoff, J.P., Göker, M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat Commun 10, 2182 (2019). <https://doi.org/10.1038/s41467-019-10210-3>.

Minoru Kanehisa, Yoko Sato, Masayuki Kawashima, Miho Furumichi, Mao Tanabe, KEGG as a reference resource for gene and protein annotation, Nucleic Acids Research, Volume 44, Issue D1, 4 January 2016, Pages D457–D462, <https://doi.org/10.1093/nar/gkv1070>.

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 2007 Jul;35(Web Server issue):W182-5. doi: 10.1093/nar/gkm321. Epub 2007 May 25. PMID: 17526522; PMCID: PMC1933193.

NASCIMENTO, Edson Alexandre. Análise genômica e metabolômica de uma linhagem de *Streptomyces galbus* isolada do semiárido paraibano. Orientador: Samuel Paulo Cibulski. 2020. 151 p. Trabalho de conclusão de curso (Bacharelado em Biotecnologia) - Universidade Federal da Paraíba, João Pessoa, 2020.

NETT, Markus; IKEDAB, Haruo; MOORE, Bradley S.. Genomic basis for natural product biosynthetic diversity in the actinomycetes. Natural Product Reports, The Royal Society of Chemistry, ano 2009, v. 26, n. 11, 26 nov. 2009. Genomics, p. 1362-1384.

Nodwell JR. Novel links between antibiotic resistance and antibiotic production. J Bacteriol. 2007 May;189(10):3683-5. doi: 10.1128/JB.00356-07. Epub 2007 Mar 23. PMID: 17384190; PMCID: PMC1913332.

Piro, V.C., Faoro, H., Weiss, V.A. et al. FGAP: an automated gap closing tool. BMC Res Notes 7, 371 (2014). <https://doi.org/10.1186/1756-0500-7-371>.

Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 2007 Jan;35(Database issue):D61-5. doi: 10.1093/nar/gkl842. Epub 2006 Nov 27. PMID: 17130148; PMCID: PMC1716718.

REYES, Beverly A. S. et al. *Selected Phyto and Marine Bioactive Compounds: Alternatives for the Treatment of Type 2 Diabetes.* Elsevier: *Studies in Natural Products Chemistry*, [s. l.], v. 55, p. 111-143, 12 jun. 2018.

REZENDE, Cássia Cristina et al. *Microrganismos multifuncionais: utilização na agricultura.* Research, Society and Development, [s. l.], v. 10, n. 2, p. 1-15, 30 nov. 2021.

Risdian C, Mozef T, Wink J. Biosynthesis of Polyketides in *Streptomyces*. *Microorganisms.* 2019 May 6;7(5):124. doi: 10.3390/microorganisms7050124. PMID: 31064143; PMCID: PMC6560455.

Rodriguez-R LM, Gunturu S, Harvey WT, Rosselló-Mora R, Tiedje JM, Cole JR, Konstantinidis KT. The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. *Nucleic Acids Res.* 2018 Jul 2;46(W1):W282-W288. doi: 10.1093/nar/gky467. PMID: 29905870; PMCID: PMC6031002.

Satria A Kautsar, Kai Blin, Simon Shaw, Jorge C Navarro-Muñoz, Barbara R Terlouw, Justin J J van der Hooft, Jeffrey A van Santen, Vittorio Tracanna, Hernando G Suarez Duran, Victòria Pascal Andreu, Nelly Selem-Mojica, Mohammad Alanjary, Serina L Robinson, George Lund, Samuel C Epstein, Ashley C Sisto, Louise K Charkoudian, Jérôme Collemare, Roger G Linington, Tilmann Weber, Marnix H Medema, MIBiG 2.0: a repository for biosynthetic gene clusters of known function, *Nucleic Acids Research*, Volume 48, Issue D1, 08 January 2020, Pages D454–D458, <https://doi.org/10.1093/nar/gkz882>.

Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014 Jul 15;30(14):2068-9. doi: 10.1093/bioinformatics/btu153. Epub 2014 Mar 18. PMID: 24642063.

SILVA, Mariana Aguiar et al. Microrganismos multifuncionais: uma revisão. In: OLIVEIRA, Antonella Carvalho (ed.). *Sistemas de Produção nas Ciências Agrárias 2*. Ponta Grossa: Atena, 2021. cap. 19, p. 218-227. ISBN 978-65-5706-812-0.

Smith KA. Laws, leaders, and legends of the modern National Library of Medicine. *J Med Libr Assoc*. 2008 Apr;96(2):121-33. doi: 10.3163/1536-5050.96.2.121. PMID: 18379667; PMCID: PMC2268223.

Tang L, Li M, Wu FX, Pan Y, Wang J. MAC: Merging Assemblies by Using Adjacency Algebraic Model and Classification. *Front Genet*. 2020 Jan 31;10:1396. doi: 10.3389/fgene.2019.01396. PMID: 32082361; PMCID: PMC7005248.

Tanja Magoč, Steven L. Salzberg, FLASH: fast length adjustment of short reads to improve genome assemblies, *Bioinformatics*, Volume 27, Issue 21, 1 November 2011, Pages 2957–2963, <https://doi.org/10.1093/bioinformatics/btr507>.

Thirumurugan, D., Cholarajan, A., & Vijayakumar, S. S. R. a. (2018). An Introductory Chapter: Secondary Metabolites. In R. Vijayakumar, & S. S. Raja (Eds.), *Secondary Metabolites - Sources and Applications*. IntechOpen. <https://doi.org/10.5772/intechopen.79766>.

Travis A. Pemberton, Mengbin Chen, Golda G. Harris, Wayne K. W. Chou, Lian Duan, Mustafa Köksal, Alex S. Genshaft, David E. Cane, and David W. Christianson. Exploring the Influence of Domain Architecture on the Catalytic Function of Diterpene Synthases *Biochemistry* 2017 56 (14), 2010-2023 DOI: 10.1021/acs.biochem.7b00137.

Ventola CL. The antibiotic resistance crisis: part 1: causes and threats. *P T*. 2015 Apr;40(4):277-83. PMID: 25859123; PMCID: PMC4378521.

Walsh CT, Chen H, Keating TA, Hubbard BK, Losey HC, Luo L, Marshall CG, Miller DA, Patel HM. Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Curr Opin Chem Biol.* 2001 Oct;5(5):525-34. doi: 10.1016/s1367-5931(00)00235-0. PMID: 11578925.

Yu Peng, Henry C. M. Leung, S. M. Yiu, Francis Y. L. Chin, IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth, *Bioinformatics*, Volume 28, Issue 11, 1 June 2012, Pages 1420–1428, <https://doi.org/10.1093/bioinformatics/bts174>.

Zhu S, Hegemann JD, Fage CD, Zimmermann M, Xie X, Linne U, Marahiel MA. Insights into the Unique Phosphorylation of the Lasso Peptide Paeninodin. *J Biol Chem.* 2016 Jun 24;291(26):13662-78. doi: 10.1074/jbc.M116.722108. Epub 2016 May 5. PMID: 27151214; PMCID: PMC4919450.

"Autorizo a reprodução e divulgação total deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte."

"A inclusão deste trabalho na Biblioteca Digital foi aprovada pela Comissão Coordenadora do Curso em sua 63ª Sessão Extraordinária, realizada em 06/06/2022"