

**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ  
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA**

**BEATRIZ APARECIDA LEPOLDINO DA SILVA**

**PIPELINE PARA EXTRAÇÃO AUTOMÁTICA DE GENES  
ASSOCIADOS À PGPR: UM ESTUDO DE CASO DE BACTÉRIAS DO  
GÊNERO *BACILLUS***

**DISSERTAÇÃO**

**CORNÉLIO PROCÓPIO**

**2023**

**BEATRIZ APARECIDA LEOPOLDINO DA SILVA**

**PIPELINE PARA EXTRAÇÃO AUTOMÁTICA DE GENES  
ASSOCIADOS À PGPR: UM ESTUDO DE CASO DE BACTÉRIAS DO  
GÊNERO BACILLUS**

**PIPELINE FOR AUTOMATIC EXTRACTION OF GENES  
ASSOCIATED WITH PGPR: A CASE STUDY OF *BACILLUS* BACTERIA.**

Dissertação apresentada como requisito parcial à obtenção do título de Mestra em Bioinformática, do Programa de Pós-Graduação em Bioinformática, da Universidade Tecnológica Federal do Paraná (UTFPR).

Orientador: Prof. Dr. Danilo Sipoli Sanches

**CORNÉLIO PROCÓPIO**

**2023**



**Ministério da Educação**  
**Universidade Tecnológica Federal do Paraná**  
**Campus Cornélio Procópio**



BEATRIZ APARECIDA LEOPOLDINO DA SILVA

**PIPELINE PARA EXTRAÇÃO AUTOMÁTICA DE GENES ASSOCIADOS À PGPR: UM ESTUDO DE CASO DE BACTÉRIAS DO GÊNERO BACILLUS**

Trabalho de pesquisa de mestrado apresentado como requisito para obtenção do título de Mestre Em Bioinformática da Universidade Tecnológica Federal do Paraná (UTFPR). Área de concentração: Bioinformática.

Data de aprovação: 18 de Agosto de 2023

Dr. Danilo Sipoli Sanches, Doutorado - Universidade Tecnológica Federal do Paraná

Dr. Admilton Goncalves De Oliveira Junior, Doutorado - Universidade Estadual de Londrina (Uel)

Dra. Glaucia Maria Bressan, Doutorado - Universidade Tecnológica Federal do Paraná

Documento gerado pelo Sistema Acadêmico da UTFPR a partir dos dados da Ata de Defesa em 18/08/2023.

## AGRADECIMENTOS

Inicialmente, agradeço à Deus por sempre iluminar meu caminho, me dando força, coragem e paciência para enfrentar cada obstáculo, pois o caminho nem sempre é fácil.

Agradeço aos meus pais, Luzia e Aparecido, por acreditarem na minha capacidade, pelo exemplo que sempre me deram de honestidade, dignidade, superação, força e amor. Obrigada por terem me proporcionado suporte financeiro e emocional para que fosse possível ingressar na Universidade, concluir minha graduação, e agora meu mestrado. Obrigada por serem minha base e meu refúgio nos momentos mais difíceis. Vocês são o motivo pelo qual eu nunca desisti. Agradeço também ao meu irmão, meu caçulinha Kauan, pelo carinho que sempre tem comigo e pela ajuda quando preciso. Amo vocês!

Às amigas que fiz no colégio e na graduação, e levarei para a vida. Obrigada por estarem comigo em cada momento, mesmo agora morando em cidades diferentes. Cada um de vocês tem um espacinho reservado no meu coração. Obrigada por me incentivarem, me apoiarem e se orgulharem de mim! Gratidão também, às amigas que fiz durante o mestrado, nossos compartilhamentos foram incríveis, principalmente nos momentos de desespero!

Ao meu amigo e namorado, Alisson, que conheci durante a graduação, nessa mesma universidade. Obrigada pela paciência e companheirismo, pois sempre estive ao meu lado, sobretudo nos piores momentos e, também, pelas contribuições durante o desenvolvimento da minha pesquisa. Obrigada por sempre acreditar em mim e me apoiar nas minhas decisões, por sempre me incentivar a correr atrás dos meus objetivos e não desistir jamais. Você foi o alicerce que eu precisava no fim da graduação e durante todo o meu mestrado. Amo muito você!

Aos professores da graduação que, de certa forma, marcaram minha vida. Em especial, aos professores: Dr. Anderson Paião, pelos preciosos ensinamentos, pois no início do curso foi o professor que mais me inspirou (e continua até hoje), você é incrível!; Profa. Dra. Michele Cristina Valentino por sempre me incentivar a não desanimar e seguir carreira acadêmica, além de me proporcionar incríveis experiências em trabalhos e projetos acadêmicos e, também, a Profa. Dra. Glaucia que, inclusive, foi minha coorientadora no Trabalho de Conclusão de Curso da graduação e, agora, minha banca de mestrado. Obrigada por todas as contribuições!

Ao professor Admilton, da Universidade Estadual de Londrina, por ter aceitado fazer parte da minha banca, e por todas as contribuições à minha pesquisa. Receber elogios do senhor, com toda certeza, deixou o dia da minha defesa ainda mais especial!

Ao meu orientador, professor Dr. Danilo Sipoli Sanches, pelas orientações e paciência durante a execução desta pesquisa, por ter depositado em mim a sua confiança e não me deixar desistir, por me apoiar e me incentivar a seguir minha carreira acadêmica. Obrigada também pelos elogios, isso nos faz querer buscar (ainda mais) nosso melhor a cada dia!

À Universidade Tecnológica Federal do Paraná e a Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo tempo em que fui bolsista.

À todos que, de alguma forma, contribuíram para a execução deste trabalho.

*“O conhecimento é uma aventura em aberto. O que significa que aquilo que saberemos amanhã é algo que desconhecemos hoje; e esse algo pode mudar as verdades de ontem.”*

Karl Popper

## RESUMO

SILVA, Beatriz Aparecida Leopoldino da. **Pipeline para Extração Automática de Genes Associados à PGPR: Um estudo de Caso de Bactérias do Gênero *Bacillus***. 2023. 136 f. Dissertação (Mestrado em Bioinformática) – Universidade Tecnológica Federal do Paraná. Cornélio Procópio, 2023.

A utilização de rizobactérias promotoras de crescimento em plantas (PGPR, do inglês “Plant Growth-Promoting Rizobacteria”) tornou-se uma solução sustentável para o controle de fitopatógenos e aumento da produtividade agrícola nas últimas décadas. Cada vez mais, é perceptível o aumento na quantidade de pesquisas nesta área e, conseqüentemente, também no número de novos resultados obtidos. Nesta pesquisa, foi proposto um pipeline para criar um banco de dados de genes que estão relacionados à PGPR, utilizando scripts na linguagem de programação Python com as funções do pacote Biopython e também com o software BLAST+, para auxiliar em futuras classificações, por algoritmos de aprendizado de máquina, de novas cepas de rizobactérias promotoras de crescimento em plantas. Por meio de alinhamentos de seqüências de aminoácidos, utilizando a ferramenta BLAST, analisa-se a presença dos genes englobados no banco de dados proposto, em genomas de rizobactérias do gênero *Bacillus*, classificadas pela literatura como PGPR. Com isso, é discutido o resultado encontrado e novas possíveis abordagens para pesquisas futuras.

**Palavras-chave:** Rizobactérias Promotoras de Crescimento em Plantas. *Pipeline*. Biopython. BLAST. Banco de Dados.

## ABSTRACT

SILVA, Beatriz Aparecida Leopoldino da. **Pipeline for Automatic Extraction of Genes Associated with PGPR: A Case Study of *Bacillus* Bacteria**. 2023. 136 p. Dissertation (Master's in Bioinformatic) – Universidade Tecnológica Federal do Paraná. Cornélio Procópio, 2023.

The utilization of Plant Growth-Promoting Rhizobacteria (PGPR) has evolved into a sustainable solution for phytopathogen control and increased agricultural productivity in recent decades. There is an evident surge in research in this field, resulting in a growing number of new findings. In this study, a pipeline has been proposed to establish a gene database related to PGPR. This is achieved through the use of Python scripts utilizing the functionalities of the Biopython package and the BLAST+ software. These tools aid in future classifications, by machine learning algorithms, of new strains of plant growth-promoting rhizobacteria. Through alignments of amino acid sequences using the BLAST tool, the presence of genes encompassed in the proposed database is analyzed within genomes of *Bacillus* genus rhizobacteria, as classified as PGPR in the literature. Consequently, the findings are deliberated upon, and potential new approaches for future research are discussed.

**Keywords:** Plant Growth-Promoting Rhizobacteria. Pipeline. Biopython. BLAST. Database.

## LISTA DE ILUSTRAÇÕES

Figura 1 – Representação da INSDC . . . . .	18
Figura 2 – Exemplo de Alinhamento Global e Local . . . . .	19
Figura 3 – Informações do Cabeçalho da Saída do BLAST . . . . .	22
Figura 4 – Descrições das Sequências do Banco de Dados que foram alinhados com a Sequência de Consulta . . . . .	22
Figura 5 – Informações do Alinhamento do BLAST . . . . .	23
Figura 6 – <i>Bacillus subtilis</i> . . . . .	39
Figura 7 – Pipeline Proposto . . . . .	58
Figura 8 – Primeira Análise do Cálculo ANI da cepa <i>Bacillus sp.</i> RZ2MS9 . . . . .	68
Figura 9 – Segunda Análise do Cálculo ANI da cepa <i>Bacillus sp.</i> RZ2MS9 . . . . .	69
Figura 10 – Árvore Filogenética representando cepas do gênero <i>Bacillus</i> . . . . .	70
Figura 11 – Representação Circular do Genoma da Bactéria <i>Bacillus velezensis</i> LABIM40	71
Figura 12 – Legenda da Figura 11 - Representação Circular da cepa <i>Bacillus velezensis</i> LABIM40 . . . . .	72
Figura 13 – Agrupamentos BGCs e Genes PGP na cepa <i>Bacillus velezensis</i> LABIM22 .	74
Figura 14 – Agrupamentos BGCs e Genes PGP na cepa <i>Bacillus velezensis</i> LABIM40 .	76
Figura 15 – Agrupamentos BGCs e Genes PGP na cepa <i>Bacillus velezensis</i> LABIM44 .	77
Figura 16 – Agrupamentos BGCs e Genes PGP na cepa <i>Bacillus velezensis</i> HNA3 . . .	78
Figura 17 – Agrupamentos BGCs e Genes PGP na cepa <i>Bacillus velezensis</i> Ag75 . . .	79
Figura 18 – Agrupamentos BGCs e Genes PGP na cepa <i>Bacillus velezensis</i> LDO2 . . .	79
Figura 19 – Agrupamentos BGCs e Genes PGP na cepa <i>Bacillus velezensis</i> FZB42 . . .	80
Figura 20 – Diagrama de Venn dos Genes Identificados pelo Pipeline . . . . .	86
Figura 21 – Diagrama de Venn dos Genes PGP das Quatro Cepas . . . . .	93
Figura 22 – Quantidade de Estudos Relacionados à PGPR na plataforma Google Scholar	97

## LISTA DE QUADROS

Quadro 1 – Programas do BLAST . . . . .	20
Quadro 2 – Genes associados à promoção de crescimento em plantas na cepa <i>B. velezensis</i> HNA3 . . . . .	41
Quadro 3 – Genes associados à promoção de crescimento em plantas na cepa <i>Bacillus velezensis</i> Ag75 . . . . .	43
Quadro 4 – Genes associados à promoção de crescimento em plantas na cepa <i>Bacillus velezensis</i> LDO2 . . . . .	46
Quadro 5 – Genes associados à PGP em <i>B. subtilis</i> BS87 e <i>B. megaterium</i> BM89 . . . . .	51

## LISTA DE TABELAS

Tabela 1 – Agrupamentos de Genes Biossintéticos em <i>Bacillus velezensis</i> LABIM22 com antiSMASH 7.0.0 . . . . .	73
Tabela 2 – Agrupamentos de Genes Biossintéticos em <i>Bacillus velezensis</i> LABIM40 com antiSMASH 7.0.0 . . . . .	75
Tabela 3 – Agrupamentos de Genes Biossintéticos em <i>Bacillus velezensis</i> LABIM44 com antiSMASH 7.0.0 . . . . .	76
Tabela 4 – Agrupamentos de Genes Biossintéticos com AntiSMASH 7.0.0 nas 19 Cepas	80
Tabela 5 – Quantidade de genes identificados das cepas HNA3, LABIM22, LABIM40 e LABIM44 . . . . .	85
Tabela 6 – Quantidade de genes PGP identificados em HNA3, LABIM22, LABIM40 e LABIM44 . . . . .	85
Tabela 7 – Resultado Alinhamento BLASTp <i>Bacillus velezensis</i> HNA3 . . . . .	89
Tabela 8 – Resultado Alinhamento BLASTp <i>Bacillus velezensis</i> Ag75 . . . . .	90
Tabela 9 – Resultado Alinhamento BLASTp <i>Bacillus velezensis</i> LDO2 . . . . .	91
Tabela 10 – Resultado da Comparação entre LABIMs e Cepas PGPRs <i>B. velezensis</i> . . .	92
Tabela 11 – Quantidade de Genes Banco de dados . . . . .	95
Tabela 12 – Resultados dos Alinhamento de Sequências do Banco de Dados com Genomas PGPRs . . . . .	96
Tabela 13 – Percentual de Genes PGPs nos Genomas PGPRs . . . . .	97

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>12</b>
1.1	OBJETIVOS	14
1.1.1	Objetivo Geral	14
1.1.2	Objetivos Específicos	14
1.1.3	Motivação	15
<b>2</b>	<b>BANCO DE DADOS E FERRAMENTAS</b>	<b>16</b>
2.1	BANCO DE DADOS BIOLÓGICOS	16
2.1.1	NCBI	16
2.1.1.1	GenBank	18
2.2	FERRAMENTAS COMPUTACIONAIS	18
2.2.1	BLAST	18
2.2.2	Biopython	24
2.2.3	AntiSMASH	26
2.2.4	OrthoANI	26
2.2.5	Gegenees	27
2.2.6	SplitsTree	27
2.2.7	BRIG	28
<b>3</b>	<b>REVISÃO DE LITERATURA</b>	<b>29</b>
3.1	FERTILIZANTES QUÍMICOS	29
3.2	PGPR	31
3.2.1	Gênero <i>Bacillus</i>	37
3.3	ESTUDOS RELACIONADOS	40
<b>4</b>	<b>MATERIAIS E MÉTODOS</b>	<b>53</b>
4.1	DADOS UTILIZADOS	53
4.2	PROPOSTA DE BANCO DE DADOS	54
4.3	GENÔMICA COMPARATIVA	55
4.3.1	Árvore Filogenética	55
4.3.2	Predição de Metabólitos Secundários com antiSMASH	56
4.4	PIPELINE PROPOSTO	57
4.4.1	Etapa 1 – Download e Complementação de Genes	59
4.4.2	Etapa 2 – Alinhamento BLASTp e Análise de Resultados	61
4.4.3	Etapa 3 – Comparação dos Genomas	62
4.4.4	Etapa 4 – Elaboração do Banco de Dados	63
<b>5</b>	<b>RESULTADOS</b>	<b>67</b>
5.1	RESULTADOS E ANÁLISE DA GENÔMICA COMPARATIVA	67
5.1.1	Análise e Árvore Filogenética	68
5.1.2	Análise antiSMASH e Representação Circular do Genoma	71
5.2	ANÁLISE DOS RESULTADOS DO PIPELINE PROPOSTO	82
5.2.1	Resultados Preliminares	83
5.2.2	Resultados Complementares	87

<b>6</b>	<b>CONCLUSÕES E PERSPECTIVAS . . . . .</b>	<b>98</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>103</b>
	<b>APÊNDICE A – GENES PGP <i>BACILLUS VELEZENSIS</i> HNA3 . . . .</b>	<b>114</b>
	<b>APÊNDICE B – GENES PGP <i>BACILLUS VELEZENSIS</i> AG75 . . . .</b>	<b>116</b>
	<b>APÊNDICE C – GENES PGP <i>BACILLUS VELEZENSIS</i> LDO2 . . . .</b>	<b>118</b>
	<b>APÊNDICE D – GENES PGP <i>BACILLUS VELEZENSIS</i> FZB42 . . . .</b>	<b>120</b>
	<b>APÊNDICE E – GENES PGP <i>BACILLUS SP.</i> RZ2MS9 . . . . .</b>	<b>124</b>
	<b>APÊNDICE F – RESULTADO ALINHAMENTO BLASTP <i>B. VELE-</i></b> <b><i>ZENSIS FZB42</i> . . . . .</b>	<b>134</b>

## 1 INTRODUÇÃO

Até meados da década de 1990, havia uma carência nas discussões relativas à necessidade do aumento da produtividade agrícola em escala global. Neste período, modelos computacionais utilizados nas simulações de equilíbrio geral, não apontavam relação entre o aumento populacional e a possível escassez de alimento, diferentemente do atual cenário (ITTERSUM *et al.*, 2013).

Segundo o último relatório da Organização das Nações Unidas (ONU), publicado em 17 de junho de 2019, a população mundial tende a aproximar-se de 10 bilhões de pessoas até o ano de 2050<sup>1</sup>. Para atender a demanda alimentar de toda essa população, a produção agrícola deverá aumentar proporcionalmente nos próximos anos. Além disso, como o Brasil é um importante produtor agrícola e um grande fornecedor mundial, o aumento constante na produção de alimentos, sem dúvidas, trará mais benefícios econômicos para o país. Entretanto, não é viável que o aumento de produção seja obtido por meio do aumento da área cultivada devido à existência de limitações. Dessa forma, a melhor maneira de expandir a produção de alimentos, é ampliando sua produtividade (VAN REES *et al.*, 2014).

De acordo com a Organização das Nações Unidas para a Alimentação e Agricultura (FAO – do inglês “*Food and Agriculture Organization of the United Nations*”), os agricultores deverão ter acesso a sementes resistentes, visando garantir o aumento na produção de alimentos, que deverá ser de, pelo menos, 50% nas próximas décadas<sup>2</sup>, principalmente devido a mudança climática que, segundo a maioria dos estudiosos do tema, será mais desafiadora para a produção agrícola, sobretudo em regiões tropicais. Assim, apesar da projeção do clima a longo prazo ser uma tarefa complexa e incerta, há uma tendência consensual por parte da comunidade científica de que, possivelmente, haverá uma elevação expressiva de dióxido de carbono ( $CO_2$ ) na atmosfera, o aquecimento do ar e uma mudança na distribuição de chuvas nas principais regiões agrícolas do mundo (ROSENZWEIG *et al.*, 2014).

A técnica mais conhecida e amplamente utilizada na agricultura convencional até os dias atuais, quando se diz respeito ao aumento da quantidade de nutrientes do solo e a produtividade agrícola, é o uso de fertilizantes. Estes são compostos químicos que constituem um dos principais insumos agrícolas, tendo como fontes de matéria-prima, produtos provenientes da petroquímica

<sup>1</sup> <https://www.un.org/development/desa/publications/world-population-prospects-2019-highlights.html>

<sup>2</sup> <https://news.un.org/pt/story/2021/11/1769342>

e da mineração (DIAS; FERNANDES, 2006; SOUZA *et al.*, 2015).

Todavia, os fertilizantes químicos têm se revelado um problema a nível mundial, devido aos impactos ambientais causados, tais como a degradação da qualidade do solo, a poluição das fontes de água e da atmosfera e o aumento da resistência de insetos considerados pragas para as plantações, em decorrência da sua produção e uso excessivo (ISHERWOOD, 2000; SOUZA *et al.*, 2018).

Além das questões ambientais, o fator econômico também é algo que chama atenção, uma vez que os custos de fabricação de fertilizantes químicos podem ser elevados. No caso do Brasil, por exemplo, os custos de fabricação são altos devido às grandes dificuldades logísticas representadas pela infraestrutura portuária insuficiente e pelos preços dos fretes, que podem sobrecarregar o custo da matéria-prima importada, refletindo diretamente no valor final desses insumos.

Diante deste cenário, a comunidade científica tem sido pressionada a desenvolver alternativas para o controle e/ou prevenção de patógenos, além da melhoria da produtividade (DIAS; FERNANDES, 2006; QUEIROZ, 2003). Nesse sentido, o uso de microrganismos na prática agrícola têm sido bastante estudado e vêm ganhando cada vez mais credibilidade, sobretudo, microrganismos colonizadores da rizosfera, isso porque, neste habitat, há uma maior diversidade de populações de bactérias, fungos, protozoários e nematoides. Estes microrganismos podem interagir com as plantas gerando efeitos benéficos, deletérios ou neutros (LYNCH, 1990).

Neste estudo, foram investigadas as rizobactérias, que colonizam a rizosfera, região onde interações entre solo, microrganismos e plantas ocorrem. Ainda, a rizosfera é uma região de intensa atividade microbiana, onde predominam bactérias que podem apresentar efeito benéfico, neutro ou deletério sobre as plantas hospedeiras. As bactérias que pertencem ao primeiro grupo têm sido designadas "*Plant Growth-Promoting Rhizobacteria* - PGPR". Assim, foi realizada uma pesquisa na literatura à respeito de PGPRs e proposto um pipeline que seja capaz de buscar informações nos bancos de dados de sequências genômicas disponíveis e, com isso, criar um banco de dados de genes relacionados à promoção de crescimento em plantas, a fim de auxiliar na análise *in silico* de genomas de bactérias que possam ser classificadas como PGPR. Assim, este trabalho foi estruturado da seguinte forma:

Após o Capítulo 1, é realizado no Capítulo 2, uma breve introdução aos bancos de dados biológicos, especificamente ao banco de dados primário GenBank, como também é apontado as principais ferramentas computacionais, que serão utilizadas no decorrer desta pesquisa: BLAST,

Biopython, antiSMASH, OrthoANI, Gegenees, SprintsTree e BRIG.

No Capítulo 3, é feita uma breve discussão à respeito da produção/uso de fertilizantes e os respectivos impactos causados, bem como sobre uma das tecnologias promissoras da atualidade, que é a utilização de rizobactérias promotoras de crescimento em plantas na agricultura. Além disso, é apresentado um estudo de alguns trabalhos dos últimos 20 anos relacionados à área de pesquisa de PGPRs, a fim de embasar o desenvolvimento desta investigação.

No Capítulo 4 é apresentado um esboço do pipeline proposto, além dos dados a serem utilizados para a implementação do mesmo.

Por fim, no Capítulo 5 são apresentados os resultados obtidos e, no Capítulo 6, a conclusão e possíveis perspectivas futuras.

## 1.1 OBJETIVOS

### 1.1.1 Objetivo Geral

Desenvolver um pipeline baseado na linguagem de programação Python e biblioteca Biopython, que identifica genes relacionadas à promoção de crescimento em plantas, em genomas de rizobactérias do gênero *Bacillus*, a partir de estudos da literatura, para elaborar um banco de dados de genes que auxilie na classificação de novos genomas PGPR, por meio de algoritmos de Aprendizado de Máquina.

### 1.1.2 Objetivos Específicos

- Buscar, na literatura, genes de bactérias do gênero *Bacillus* relacionados à promoção do crescimento em plantas;
- Estudar as ferramentas computacionais que tratam sequências genômicas;
- Desenvolver scripts na linguagem Python para busca de informações associadas à PGPRs em genomas depositados no GenBank;
- Criar um banco de dados com os genes relacionados a promoção de crescimento vegetal encontrados na literatura, a fim de corroborar com pesquisas posteriores.

### **1.1.3 Motivação**

Motiva-se este trabalho devido a importância desta área, que vem aumentando em virtude da preocupação atual com o crescimento populacional, previsto pela ONU para as próximas décadas, e uma possível escassez de alimentos, acrescido da busca por uma agricultura sustentável de maior impacto positivo, de forma a aumentar a produtividade de alimentos e diminuir os danos ambientais causados em decorrência da produção e/ou uso excessivo de fertilizantes químicos. Somado a isso, há o interesse em unir as informações relevantes das atuais pesquisas da área, de forma que seja possível desenvolver um pipeline eficiente baseado em diferentes ferramentas computacionais como Python, Biopython e BLAST, a fim de criar um banco de dados de genes envolvidos na promoção de crescimento em plantas, para auxiliar pesquisadores na identificação de possíveis PGPRs.

## 2 BANCO DE DADOS E FERRAMENTAS

Neste capítulo, realiza-se uma breve introdução aos bancos de dados biológicos, bem como às principais ferramentas computacionais utilizadas no decorrer desta pesquisa.

### 2.1 BANCO DE DADOS BIOLÓGICOS

Banco de dados (BD) são estruturas de dados que armazenam informações de diversos tipos, relacionando-as de forma a criar algum sentido e dar eficiência durante uma pesquisa ou estudo científico. Além disso, existem ainda os chamados bancos de dados biológicos (BDB) que, assim como os BDs, também armazenam informações, mas são específicos para armazenar dados biológicos, em sua maioria, derivados de experimentação (MARQUES, 2021).

Ainda segundo Marques (2021), os bancos de dados biológicos podem ser divididos em três categorias, sendo elas: BDBs primários, BDBs secundários (mais conhecidos como curados) e BDBs especializados. O primeiro se refere a bancos que não receberam curadoria e não foram analisados minuciosamente, podendo apresentar algumas informações redundantes. Já o segundo, se refere a bancos de dados que armazenam resultados de análises feitas a partir de dados primários e não apresentam redundância de informações, apresentando maior confiabilidade. Por fim, os BDBs especializados focam em estudos específicos e, portanto, possuem somente informações de um organismo ou dados de um mesmo tipo.

Os bancos de dados primários (também conhecidos como repositório de dados) são ambientes altamente organizados e "amigáveis", garantindo que usuários consultem uma enorme quantidade de dados biológicos produzidos por pesquisadores.

Na subseção a seguir, realiza-se uma breve introdução ao Centro Nacional de Informações sobre Biotecnologia, que é um repositório de dados biológicos, especificamente de biologia molecular.

#### 2.1.1 NCBI

O Centro Nacional de Informações sobre Biotecnologia (NCBI<sup>1</sup> - do inglês “*National Center for Biotechnology Information*”) formou-se no ano de 1988, nos Estados Unidos da

---

<sup>1</sup> <https://www.ncbi.nlm.nih.gov/>

América (EUA), como resultado do projeto de lei proposto pelo então senador Claude Pepper. O NCBI é uma divisão da Biblioteca Nacional da Medicina (NLM – do inglês “*National Library of Medicine*”), que pertence aos Institutos Nacionais da Saúde (NIH – do inglês “*National Institutes of Health*”), sendo um recurso nacional no que se diz respeito à informação disponível na área de biologia molecular. Ainda, o NCBI é responsável por criar bases de dados públicas, conduzir investigação em biologia computacional, desenvolver softwares para análise de dados genômicos e disseminar informações biomédicas. Cabe destacar que esse repositório é de uso gratuito e acolhe diversos bancos de dados biológicos, como genomas completos de diversas espécies, sequenciamento de cepas dos mais diversos organismos, entre outros (E-ESCOLA, 2005; MARQUES, 2021).

Entre o final da década de 1980 e a década de 1990, foram desenvolvidos os primeiros bancos de dados biológicos primários. Eles foram criados, inicialmente, para armazenar sequências que haviam sido determinadas de forma experimental. Naquela época, o sequenciamento de proteínas ocorria sequenciando um aminoácido por vez, e o sequenciamento do DNA também estava em fase inicial. Dessa forma, esses bancos de dados possuíam uma quantidade bastante limitada de sequências. Entretanto, com o surgimento do Sequenciamento de Nova Geração (NGS – do inglês “*Next-Generation Sequencing*”), uma técnica de sequenciamento de DNA e RNA automatizado, paralelo e de alto rendimento, um elevado número de dados de sequências biológicas vêm sendo geradas, isto é, esses repositórios de dados começaram a crescer exponencialmente (VICENTE, 2020).

Segundo Marques (2021), existem três bancos de dados primários: o “*DNA DataBank of Japan*” (DDBJ), hospedado pelo Centro Nacional de Genética (NCG - do inglês “*National Center for Genetics*”); o “*European Nucleotide Archive*” (ENA), hospedado pelo Laboratório Europeu de Biologia Molecular (EMBL, sigla para “*European Molecular Biology Laboratories*” e o GenBank, hospedado pelo NCBI. Estes três repositórios fazem parte da Colaboração Internacional de Banco de Dados de Sequências Nucleotídicas (INSDC – do inglês “*International Nucleotide Sequence Database Collaboration*”), como mostrado na Figura 1, e trocam dados diariamente. Dessa forma, qualquer que seja o repositório primário que uma determinada sequência seja depositada, a entrada também aparecerá nos outros bancos da INSDC, além disso, uma vez que forem depositadas, elas poderão ser acessadas livremente por usuários do mundo todo. Ademais, além de centralizarem informações biológicas sobre qualquer tipo de organismo, eles servem como fontes para outros bancos de dados (curados ou especializados).

**Figura 1 – Representação da INSDC**



Fonte: <https://www.ddbj.nig.ac.jp/about/insdc-e.html>

### 2.1.1.1 GenBank

O GenBank, banco de dados utilizado neste trabalho, originou-se no ano de 1982, a partir do então denominado Banco de Dados de Los Alamos ("*Los Alamos Sequence Database*"), criado por Walter Goad e alguns colaboradores do Laboratório Nacional de Los Alamos (LANL), no ano de 1979. A criação do Genbank foi financiada pelo NIH, pela Fundação Nacional da Ciência (NSF – do inglês "*National Science Foundation*"), pelo Departamento de Energia e pelo Departamento de Defesa dos Estados Unidos. De 1989 à 1992, o projeto do GenBank transicionou para o NCBI. O formato de um arquivo Genbank, além de permitir o armazenamento da sequência de DNA/proteína, também permite o armazenamento de diversas informações relevantes (MARQUES, 2021).

## 2.2 FERRAMENTAS COMPUTACIONAIS

Nesta seção, são abordadas as principais ferramentas utilizadas no desenvolvimento desta dissertação.

### 2.2.1 BLAST

Para o estudo realizado nesta seção, utilizou-se como base os trabalhos de Ferreira (2005), Madden (2013), Mount (2004) e Pearson (2013).

A principal ferramenta para obtenção dos alinhamentos é a Ferramenta de Pesquisa de Alinhamento Básico Local (BLAST, do inglês “*Basic Local Alignment Search Tool*”). Um alinhamento de sequências é realizado comparando duas ou mais sequências de ácidos nucleicos (DNA e RNA) ou aminoácidos (proteínas), com o objetivo de encontrar séries de caracteres individuais, ou padrões de caracteres, que ocorrem nas sequências em uma mesma ordem. Nesta pesquisa, em específico, inicialmente com o objetivo de encontrar genes bacterianos relacionados à promoção de crescimento em plantas, foi realizado um estudo direcionado ao alinhamento de proteínas.

Na Figura 2, duas sequências hipotéticas de proteínas são alinhadas de duas formas: no lado esquerdo da imagem, é realizado um alinhamento global das sequências, no qual os aminoácidos são identificados em sequência e os que são correspondentes estão identificados em azul; no lado direito da imagem, é realizado um alinhamento local, na qual é observado três trechos das mesmas sequências.

**Figura 2 – Exemplo de Alinhamento Global e Local**



**Fonte: Adaptado de Prosdocimi *et al.* (2002)**

Ao alinhar duas sequências, é realizada uma comparação e algumas informações podem ser coletadas dessa tarefa. O BLAST, em resumo, compara uma sequência de consulta (query) com outra sequência de referência, seja recém sequenciada ou com propriedades interessantes para investigação, ou ainda um banco de dados, no qual constam diversas sequências do interesse do pesquisador. O BLAST foi desenvolvido pelo NCBI (Centro Nacional para Informações Biotecnológicas, do inglês “National Center for Biotechnology Information”), com sua primeira

versão criada em 1990, sendo atualizada constantemente.

É importante ressaltar que, além de criar o BLAST, o próprio NCBI trabalha como banco de dados para comparações de sequências biológicas do BLAST, seja na versão Web, disponível em <https://blast.ncbi.nlm.nih.gov/Blast.cgi>, ou por linha de comando, utilizando os bancos de dados fornecidos pelo NCBI ou um banco de dados personalizado, construído a partir de comandos para criação de banco de dados do próprio BLAST.

O BLAST é subdividido em cinco programas, se diferenciando, principalmente, pelo tipo de sequências que se utiliza como consulta e no banco de dados, conforme consta no Quadro 1:

**Quadro 1 – Programas do BLAST**

<b>Programa</b>	<b>Consulta (<i>query</i>)</b>	<b>Banco de Dados</b>
BLASTN	Nucleotídeos	Nucleotídeos
BLASTP	Aminoácidos	Aminoácidos
BLASTX	Aminoácidos	Nucleotídeos traduzidos em aminoácidos
TBLASTN	Nucleotídeos traduzidos em aminoácidos	Aminoácidos
TBLASTX	Nucleotídeos traduzidos em aminoácidos	Nucleotídeos traduzidos em aminoácidos

**Fonte: Autoria Própria**

O BLAST tem como objetivo procurar por pares de segmentos de alta pontuação (HSP - do inglês “*high-scoring segment pair*”) existentes entre as sequências de consulta e as sequências do banco de dados. Por sua vez, define-se os HSPs como sendo pares de subsequências contíguas alinhadas da consulta e do banco de dados que possuem um alto valor de similaridade.

Vale ressaltar que o BLAST não realiza um alinhamento global das sequências, em fato, ele busca determinar os alinhamentos locais que sejam estatisticamente relevantes. O ponto chave da rapidez dessa ferramenta é o fato da mesma não explorar todo o espaço de busca para encontrar tais alinhamentos locais, utilizando-se de três heurísticas para refinar a sua busca por melhores candidatos a HSPs, resumidas a seguir.

Inicialmente, o BLAST, com o parâmetro W (Tamanho de Palavra Inicial) em padrão 3 para nucleotídeos e 11 para aminoácidos, realiza o alinhamento com fragmentos da sequência de consulta, fragmentos estes chamados W-mers, que nada mais são do que subsequências de tamanho W da sequência principal. Primeiramente, ele busca na sequência de referência ou no banco de dados, correspondências com a subsequência formada pela primeira letra da sequência até a W-ésima letra da sequência, em seguida realiza a busca com a subsequência da segunda

letra da sequência até a  $(W+1)$ -ésima letra da sequência, e assim por diante.

Com isso, o BLAST quantifica o “acerto” das sequências na busca considerando não somente palavras totalmente idênticas, porém utilizando o seguinte conceito de vizinhança. A vizinhança de uma subsequência é uma lista que contém a própria subsequência além de todas as outras subsequências, cuja pontuação é, no mínimo, um valor  $T$  quando compara com a mesma via uma matriz de pontuação. As principais matrizes de pontuação utilizadas são a BLOSUM62 e PAM200, sendo possível determinar qual matriz e versão será utilizada em linha de comando.

De uma forma resumida, as subsequências  $y$  que possuem, em relação a uma subsequência  $x$ , uma alta pontuação, dentro dessas matrizes, são as subsequências que estão mais próximas biologicamente relacionadas.

A segunda heurística que o BLAST utiliza é relacionada a extensão do alinhamento. A partir do alinhamento dos  $W$ -mers, é necessário tentar expandir esse alinhamento para encontrar os melhores candidatos a HSPs. Para isso, o BLAST determina um valor positivo para extensões assertivas e um valor negativo para extensões errôneas, até que o critério de parada da tarefa de extensão é atingido. O critério de parada é definido da seguinte forma: estabelece-se um valor  $X$  para representar o quanto é permitido à pontuação do alinhamento decair em relação a pontuação máxima obtida até então. Dessa forma, o alinhamento é interrompido caso o valor  $X$  atinja o valor estabelecido previamente. Nas versões padrão dos algoritmos, o valor de  $X$  para BLASTP é 15 e para BLASTN é 30.

Por fim, a terceira heurística tem como propósito determinar quais alinhamentos são os mais significativos para o resultado final. Para determinar essa avaliação, o BLAST estabelece um valor de corte  $S$  para eliminar sequências aleatórias ou de baixa pontuação. Com isso, o BLAST realiza o relatório dos alinhamentos e imprime, em um arquivo, os HSPs encontrados e suas respectivas pontuações.

Dessa forma, o BLAST retorna uma saída com informações sobre a similaridades dos alinhamentos encontrados. Nessa saída, existem três partes principais: a primeira (Figura 3) é um cabeçalho contendo as informações sobre a sequência de consulta e a sequência de referência ou banco de dados utilizado; a segunda (Figura 4) contém uma descrição de cada sequência do banco de dados alinhado com a sequência de consulta; e por fim, a terceira (Figura 5) mostra os alinhamentos da sequência de consulta com cada sequência que foi descrita na segunda parte. Nessa terceira parte do arquivo de saída, as sequências são pareadas e uma linha com informações sobre as correspondências fica entre elas.

**Figura 3 – Informações do Cabeçalho da Saída do BLAST**

```

1  BLASTP 2.10.1+
2
3
4  Reference: Stephen F. Altschul, Thomas L. Madden, Alejandro A.
5  Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J.
6  Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of
7  protein database search programs", Nucleic Acids Res. 25:3389-3402.
8
9
10 Reference for composition-based statistics: Alejandro A. Schaffer,
11 L. Aravind, Thomas L. Madden, Sergei Shavirin, John L. Spouge, Yuri
12 I. Wolf, Eugene V. Koonin, and Stephen F. Altschul (2001),
13 "Improving the accuracy of PSI-BLAST protein database searches with
14 composition-based statistics and other refinements", Nucleic Acids
15 Res. 29:2994-3005.
16
17
18
19 Database: database_ncbi/proteins_database.fasta
20 |         |         |         |         |         |         |         |
21 |         |         |         |         |         |         |         |
22 |         |         |         |         |         |         |         |
23 |         |         |         |         |         |         |         |
24 Query= ATL41633.1
25
26 Length=316

```

Fonte: Autoria Própria

**Figura 4 – Descrições das Sequências do Banco de Dados que foram alinhados com a Sequência de Consulta**

		Score	E
	Sequences producing significant alignments:	(Bits)	Value
27			
28			
29			
30	BBA74612.1 Bacillus velezensis S141 DNA, complete genome	650	0.0
31	BBA74612.1 Bacillus velezensis S141 DNA, complete genome	650	0.0
32	ABS72456.1 Bacillus amyloliquefaciens subsp. plantarum str. FZB42...	648	0.0
33	ABS72456.1 Bacillus amyloliquefaciens subsp. plantarum str. FZB42...	648	0.0
34	AWD89846.1 Bacillus velezensis strain QST713 chromosome, complete...	645	0.0
35	AWD89846.1 Bacillus velezensis strain QST713 chromosome, complete...	645	0.0
36	AWD89846.1 Bacillus velezensis strain QST713 chromosome, complete...	645	0.0
37	AWD89846.1 Bacillus velezensis strain QST713 chromosome, complete...	645	0.0
38	AFJ60103.1 Bacillus amyloliquefaciens Y2, complete genome	642	0.0
39	AFJ60103.1 Bacillus amyloliquefaciens Y2, complete genome	642	0.0
40	CBI41134.1 Bacillus amyloliquefaciens DSM7 complete genome	620	0.0
41	CBI41134.1 Bacillus amyloliquefaciens DSM7 complete genome	620	0.0
42	CAB11784.1 Bacillus subtilis subsp. subtilis str. 168 complete ge...	471	9e-169
43	AQZ88921.1 Bacillus subtilis strain NCIB 3610 chromosome, complet...	471	9e-169
44	CAB11784.1 Bacillus subtilis subsp. subtilis str. 168 complete ge...	471	9e-169
45	AQZ88921.1 Bacillus subtilis strain NCIB 3610 chromosome, complet...	471	9e-169
46	AFQ09470.1 Bacillus cereus FRI-35, complete genome	249	4e-81
47	AFQ09470.1 Bacillus cereus FRI-35, complete genome	249	4e-81
48	BBA76804.1 Bacillus velezensis S141 DNA, complete genome	35.8	0.062
49	BBA76804.1 Bacillus velezensis S141 DNA, complete genome	35.8	0.062

Fonte: Autoria Própria

**Figura 5 – Informações do Alinhamento do BLAST**

```

50
51
52 >BBA74612.1 Bacillus velezensis S141 DNA, complete genome
53 Length=329
54
55 Score = 650 bits (1676), Expect = 0.0, Method: Compositional matrix adjust.
56 Identities = 313/316 (99%), Positives = 314/316 (99%), Gaps = 0/316 (0%)
57
58 Query 1 MRDHEWKELSLFYSSVESSQKFLNKVYTESGMEEPKRYAFKNSERFIFFIKHAESFYKQAS 60
59 MRDHEWKELSLFYSSVESSQKFLNKVYTESGMEEPKRYAFKNSERFIFFIKHAESFYKQAS
60 Sbjct 14 MRDHEWKELSLFYSSVESSQKFLNKVYTESGMEEPKRYAFKNSERFIFFIKHAESFYKQAS 73
61
62 Query 61 LSPLEIKPILLFYGMSQLLKACLLTTDPSYPHASVLAHGVTARKRKKQNYFRDDEVKV 120
63 LSPLEIKPILLFYGMSQLLKACLLTTDPSYPHASVLAHGVTARKRKKQNYFRDDEVKV
64 Sbjct 74 LSPLEIKPILLFYGMSQLLKACLLTTDPSYPHASVLAHGVTARKRKKQNYFRDDEVKV 133
65
66 Query 121 QRNGLCIHVMKHLFGIAGLEEERYTMKKLLTAIPEFHTVFYFQEKKNWLRVELKDDLIL 180
67 QRNGLCIHVMKHLFGIAGLEEERYTMKKLLTAIPE HTVFYFQEKKNWLRVELKDDLI
68 Sbjct 134 QRNGLCIHVMKHLFGIAGLEEERYTMKKLLTAIPELHTVFYFQEKKNWLRVELKDDLIS 193
69
70 Query 181 VPEHAVIHYNMDSRFAEYINHHLKWSFLRKDNHNLVFEADPIDKEPWSSPNLLFHLEKN 240
71 VPEHAVIHYNMDSRFAEYINHHLKWSFLRKDNHNLVFEADPID+EPWSSPNLLFHLEKN
72 Sbjct 194 VPEHAVIHYNMDSRFAEYINHHLKWSFLRKDNHNLVFEADPIDREPWSSPNLLFHLEKN 253
73
74 Query 241 QYYIPSQRDHFHLPEMIIHYLVLYNIGMIARYETEWYELLTQHVSDDYVMIRRFLEVT 300
75 QYYIPSQRDHFHLPEMIIHYLVLYNIGMIARYETEWYELLTQHVSDDYVMIRRFLEVT
76 Sbjct 254 QYYIPSQRDHFHLPEMIIHYLVLYNIGMIARYETEWYELLTQHVSDDYVMIRRFLEVT 313
77
78 Query 301 ERKFPYYVSNFLKQYL 316
79 ERKFPYYVSNFLKQYL
80 Sbjct 314 ERKFPYYVSNFLKQYL 329
81
82

```

**Fonte: Autoria Própria**

A maior parte do arquivo de saída é onde está a terceira parte descrita anteriormente, que são os alinhamentos entre a sequência de consulta (*query*) e as sequências do Banco de Dados (Sbjct). Além disso, é nessa parte que estão descritos os resultados do alinhamento, como o *Score*, o *Expect (E-Value)* e o número e percentual de identidade entre aminoácidos/nucleotídeos (*Identities*).

O *Score* de um alinhamento é a pontuação que aquela comparação obteve pelo cálculo do BLAST, previamente explicado, tendo relação com a matriz de pontuação. Por sua vez, o *E-Value* retorna uma indicação de significância estatística de um alinhamento e reflete o tamanho do banco de dados e o sistema de *Score* utilizado. Com isso, quanto menor o *E-Value*, mais significativo é o pareamento, e isso retorna uma informação importante: se um alinhamento de sequências retornou 0.05, significa que a similaridade retornada pelo BLAST tem a chance de 5 em 100 (5%) de ter acontecido ao acaso.

Nesta pesquisa, o BLAST auxiliará na construção de um banco de dados com sequências genômicas de PGPRs, de modo que, a partir de alguns genomas encontrados na literatura, de

rizobactérias relacionadas à promoção de crescimento em plantas, seja possível acrescentar novas sequências encontradas por similaridade via BLAST.

### 2.2.2 Biopython

Biopython<sup>2</sup> é uma biblioteca composta por diversas ferramentas para facilitar o desenvolvimento de aplicações para a Bioinformática, utilizando a linguagem Python. Essa biblioteca compõe os chamados *Bio\* Projects*, que são projetos para o desenvolvimento de ferramentas destinadas à bioinformática e biologia molecular computacional, em várias linguagens de programação. Entre esses projetos estão o BioJava, BioPerl, BioSQL, BioDAS, entre outros. O projeto Biopython, em específico, é mantido pela associação internacional de desenvolvedores “*Open Bioinformatics Foundation*” (OBF), e é disponibilizado gratuitamente (COCK *et al.*, 2009; MARIANO *et al.*, 2016).

Segundo Mariano *et al.* (2016), o Biopython apresenta um conjunto de pacotes, módulos e classes que permitem realizar análises e anotações de sequências biológicas, além de diversos módulos separados utilizados para realizar análises de alinhamentos de sequências, estruturas de proteínas, genética de populações, filogenia, visualização de dados biológicos, dentre diversas outras funcionalidades. Ademais, também permite acessar bancos de dados biológicos online, como os bancos mantidos pelo NCBI.

O pacote “Bio” armazena a coleção completa de módulos fornecidas pelo Biopython para trabalhar com dados biológicos. A seguir, serão listados, e brevemente explicados, alguns dos pacotes e módulos que serão utilizados, a princípio, no decorrer deste trabalho.

O primeiro módulo a ser descrito, é o *Bio.SeqIO* que, em resumo, oferece funções para leitura e gravação de sequências, suportando diversos formatos, como por exemplo: abi, ace, clustal, embl, fasta, fastq, fastq-solexa, fastq-illumina, genbank ou gb, ig, imgt, nexus, phd, phylip, pir, seqxml, sff, swiss, tab, qual e uniprot-xml (MARIANO *et al.*, 2016).

Neste estudo em específico, será utilizado o módulo *Bio.SeqIO* para realizar as leituras de arquivos do genoma completo (formato genbank ou gb), disponibilizado no banco de dados do NCBI e, também, ao extrair informações desse arquivo gerado, será utilizado o módulo para escrever estas informações em um arquivo fasta.

Para realizar a busca do genoma no NCBI, é utilizado o segundo módulo que será descrito, chamado *Bio.Entrez*. Este módulo permite acessar os bancos de dados disponibilizados

<sup>2</sup> <https://biopython.org/>

pelo NCBI via WWW (sigla para “*World Wide Web*” ou, em português, “Rede de Alcance Mundial”).

Existem algumas funções disponíveis, seja para recuperar informações ou mesmo para ler dados específicos. Especificamente, resalta-se o uso da função *Efetch*, que realiza uma busca em um determinado banco de dados por um identificador. Os bancos de dados mais utilizados para buscas são: “*nucleotide*”, “*protein*”, “*pubmed*”, “*bioproject*”, “*biosample*”, “*genome*”, “*nucore*”, etc. Com isso, os identificadores são estritamente relacionados aos banco de dados requerido, como por exemplo, para o banco de dados “*protein*”, utiliza-se o identificador “*GI number*”, que é um número no qual está referenciado uma determinada sequência biológica (de proteínas, genomas completos ou DNA) nos bancos de dados do NCBI (SAYERS, 2009).

Desta forma, o módulo *Bio.Entrez* retorna uma lista com os dados gravados para o identificador e o banco de dados utilizado na busca, então utiliza-se o módulo *Bio.SeqIO*, primeiramente com a função *read()* para realizar a leitura dessas informações e, por fim, a função *write()* para escrever em um arquivo com os dados.

Com isso, tem-se o seguinte caminho: utiliza-se o módulo *Bio.Entrez* para realizar a busca de um genoma completo, com o banco de dados “*nucore*” e o indetificador “*GI number*” do genoma requerido, retorna um arquivo que utiliza-se o módulo *Bio.SeqIO* para ler e gravar as informações em formato “gb”.

O último módulo a ser descrito, denominado *Bio.Blast*, é utilizado para realizar buscas no BLAST de forma local (com informações contidas no próprio computador) ou pela internet. O interesse em utilizar esse módulo, está em, por meio do Python, realizar a busca BLAST de forma local sem recorrer ao terminal de comandos.

O módulo *Bio.Blast.Applications* oferece diversas funções para realizar essa busca desejada, como: *NcbiblastpCommandline* (para buscas em proteínas), *NcbiblastnCommandline* (para buscas em nucleotídeos), etc.

Para realizar um alinhamento BLAST de forma local, necessita-se ter um banco de dados para realizar a busca. Com isso, tem-se dois caminhos: o primeiro é utilizar um banco de dados disponibilizado pelo NCBI, disponíveis para download<sup>3</sup>, ou, ainda, criar um banco de dados próprio para o BLAST, por meio da função *NcbimakeblastdbCommandline*, a partir de um arquivo fasta ou multi-fasta (BLAST, 2008; CAMACHO *et al.*, 2009; MARIANO *et al.*, 2016)

Com esses módulos do pacote Biopython, é possível realizar alinhamentos de sequências

<sup>3</sup> <https://ftp.ncbi.nlm.nih.gov/blast/db/>

biológicas de forma local, a partir de dados recuperados direto do banco de dados do NCBI com um banco de dados customizado, por meio de scripts em linguagem Python, de forma automatizada e organizada.

### 2.2.3 AntiSMASH

O antiSMASH é um software abrangente, rápido e de fácil utilização para identificação, anotação e análise de *clusters* de genes de biossíntese de metabólitos secundários em genomas bacterianos e fúngicos. Ele pode detectar rapidamente todas as classes conhecidas de agrupamentos de genes biossintéticos (BGCs, do inglês “*biosynthetic gene clusters*”) de metabólitos secundários, fornecer anotações funcionais detalhadas de peptídeo sintetase não ribossomal (NRPS, do inglês “*non-ribosomal peptide synthetase*”) e policetídeo sintase (PKS, do inglês “*polyketide synthase*”) e prever a estrutura química de produtos NRPS/PKS com maior precisão do que os métodos existentes, podendo ser executado a partir de um servidor web, disponível em <http://antismash.secondarymetabolites.org/>, ou como uma versão autônoma em um computador desktop padrão. Além disso, esta ferramenta também foi equiparada com um módulo de análise comparativa de agrupamento de genes, a partir da construção de um banco de dados de todos os BGCs de metabólitos secundários conhecidos até então, no qual as semelhanças evolutivas entre um *cluster* de genes consultado e outros *clusters* de genes podem ser detectadas e visualizadas, possibilitando inferir rapidamente funções de genes e operóns com base na homologia (MEDEMA *et al.*, 2011).

Desde seu lançamento inicial em 2011, a ferramenta antiSMASH passou por atualizações, sendo o antiSMASH 7.0 a versão mais recente, e tornou-se a ferramenta computacional comumente utilizada para mineração de genomas microbianos para BGCs de metabólitos secundários, sendo considerada a “*gold standard*” (“padrão-ouro”) (BLIN *et al.*, 2021).

O antiSMASH é alimentado por várias ferramentas de código aberto, como NCBI BLAST+, HMmer 3, Muscle 3, FastTree, PySVG e JQuery SVG<sup>4</sup>.

### 2.2.4 OrthoANI

OrthoANI é uma ferramenta desenvolvida por Lee *et al.* (2016), implementada na linguagem de programação JAVA e disponibilizada em dois tipos diferentes de softwares: o

<sup>4</sup> <https://antismash.secondarymetabolites.org/#!/about>

primeiro deles é um software de interface gráfica do usuário (GUI, do inglês “*Graphical User Interface*”) que pode ser usado interativamente em ambientes de computador pessoal, denominado “Orthologous ANI Tool” (OAT). Já o segundo, denominado “OAT\_cmd”, é um software de linha de comando, que pode ser integrado ao próprio pipeline de bioinformática do usuário. Ambas as ferramentas de software estão disponíveis gratuitamente em <http://www.ezbiocloud.net/sw/oat>.

A ferramenta OrthoANI foi desenvolvida por Lee *et al.* (2016) com o objetivo de calcular a identidade média de nucleotídeos (ANI, do inglês “*average nucleotide identity*”), que foi proposta como um índice para estimar a similaridade genética entre pares de sequências de genomas de bactérias e determinar as suas classificações taxonômicas (RICHTER; ROSSELLÓ-MÓRA, 2009). A ferramenta utiliza o alinhamento BLASTn para encontrar blocos ortólogos em um par de sequências e, em seguida, calcula a identidade média considerando apenas alinhamentos de ortólogos recíprocos.

Assim, os autores concluem que, como o software está disponível nas versões GUI e linha de comando, o mesmo deve ser acessível a todos os níveis de microbiologistas e estudantes. No entanto, é importante destacar que o OrthoANI não deve ser usado para comparar genomas pertencentes a gêneros diferentes.

### 2.2.5 Gegenees

Gegenees é um software utilizado para análise comparativa de dados completos de sequências de genomas e outros dados de Sequência de Próxima Geração (NGS). Este programa pode, por exemplo, comparar centenas de genomas microbianos, fornecer visões gerais filogenômicas e definir assinaturas genômicas exclusivas para grupos-alvo especificados. Os genomas são fragmentados e comparados, todos contra todos, por um mecanismo de controle BLAST *multithread*. Além disso, alinhamentos prontos podem ser complementados com novos genomas sem recalculá-los os pontos de dados existentes (ÅGREN *et al.*, 2012).

### 2.2.6 SplitsTree

O SplitsTree, descrito em Huson e Bryant (2006) é um software bastante utilizado no estudo de filogenia. Dado um alinhamento de sequências, uma matriz de distância ou um conjunto de árvores, este programa calculará uma árvore ou uma rede filogenética usando métodos como decomposição dividida, rede vizinha, rede de consenso, métodos de super redes ou métodos para

computação de hibridização ou redes de recombinação simples. Sua versão atual é a SplitsTree4.

### 2.2.7 BRIG

O BLAST Ring Image Generator (BRIG) é um software baseado em uma interface gráfica de usuário, que permite gerar imagens representativas para múltiplas comparações de genomas de bactérias, sem um limite arbitrário no número de elementos comparados. Ela pode ser utilizada em Windows, Mac e Linux (ALIKHAN *et al.*, 2011).

Como saída, a ferramenta gera uma imagem com a similaridade entre uma sequência de referência central e outras sequências, fornecidas, por meio de um conjunto de sequências concêntricas, em que as correspondências do alinhamento BLAST indicam a identidade percentual definida. Além disso, a ferramenta também possibilita personalizar a exibição de diferentes tipos de dados e anotações, tornando-a uma abordagem cruzada para visualizar uma variedade de dados de comparação.

O software para download pode ser encontrado em seu repositório oficial <https://sourceforge.net/projects/brig/>. Destaca-se que, embora a instalação seja simples, é necessário instalar o NCBI BLAST+ ou BLAST *legacy* e o Java 1.6 ou superior como dependências para o seu funcionamento.

### 3 REVISÃO DE LITERATURA

Neste capítulo, realiza-se uma breve discussão a respeito do uso de fertilizantes químicos na agricultura convencional, bem como os impactos ambientais causados em decorrência da sua produção e/ou uso excessivo. Somado a isso, é feito um estudo sobre bactérias colonizadoras da rizosfera, conhecidas como Rizobactérias Promotoras de Crescimento em Plantas, consideradas como uma das tecnologias mais promissoras dos dias atuais para a melhoria da qualidade e produtividade agrícola. Além disso, foi desenvolvido um levantamento de produções científicas relevantes da área, presentes na literatura, bem como a existência de genes já identificados, envolvidos na promoção de crescimento de plantas.

#### 3.1 FERTILIZANTES QUÍMICOS

Segundo Dias e Fernandes (2006), os fertilizantes são definidos na legislação brasileira, Decreto 86.955, de 18 de fevereiro de 1982, como substâncias orgânicas ou minerais, naturais ou sintéticas, que fornecem um ou mais nutrientes para as plantas, cuja função é repor ao solo os nutrientes retirados a cada colheita, visando manter ou, até mesmo, expandir a produtividade. Os elementos químicos presentes na composição destes insumos, conforme a quantidade ou proporção, são divididos em dois subgrupos:

- **Macronutrientes:** carbono, hidrogênio, oxigênio, nitrogênio, fósforo, potássio, cálcio, magnésio e enxofre;
- **Micronutrientes:** boro, cloro, cobre, ferro, manganês, molibdênio, zinco, sódio, silício e cobalto.

Se não houver quantidades suficientes de quaisquer nutrientes listados acima, ainda que seja uma necessidade mínima, ocorrerá um prejuízo no crescimento e desenvolvimento da planta, acarretando na diminuição da produtividade daquela plantação. As deficiências mais comuns são dos macronutrientes nitrogênio (N), fósforo (P) e potássio (K). O nitrogênio é um importante componente das proteínas e da clorofila, sendo primordial para o aumento da produtividade agrícola. Já o fósforo é responsável pelos processos vitais das plantas, pelo armazenamento e uso de energia, pela promoção do crescimento das raízes e melhora da qualidade dos grãos, além de acelerar o processo de amadurecimento dos frutos. Por sua vez, o potássio tem como

função, o equilíbrio de cargas no interior das células vegetais, inclusive o controle da hidratação e das doenças da planta. Utilizando os fertilizantes químicos compostos por esses três elementos (chamados de fertilizantes NPK), o esperado é que ocorra um salto na produção agrícola, como observou-se nos séculos XIX e XX (DIAS; FERNANDES, 2006; EVENSON; GOLLIN, 2003).

No entanto, a produção de fertilizantes químicos, e seu uso recorrente, acarretam em grandes impactos ambientais, visto que uma grande parcela dos nutrientes aplicados nas lavouras através do uso destes insumos não são absorvidos pelas plantas, devido as características do solo, ocorrendo um desperdício de 60 a 90% desses nutrientes, conseqüentemente, poluindo o ambiente ao serem lixiviados para corpos d'água ou, ainda, podendo infiltrar no solo, chegando aos lençóis freáticos e mananciais. Quando o nitrogênio e o fósforo escoam para a água, eles são utilizados por cianobactérias e algas que poderão crescer descontroladamente, configurando uma eutrofização desses corpos d'água. A eutrofização causa diversos danos para o contexto econômico, ambiental e de saúde pública, entre eles, a falta de oxigênio na água, ocasionando a morte de peixes, alterações na rede trófica, entupimento de redes de captação de água, aumento de algas tóxicas, alterações na qualidade de água para consumo humano, atrapalham o tráfego náutico, entre outros (CARPENTER *et al.*, 1998; LAU; LANE, 2002; SOUZA *et al.*, 2015).

Um segundo problema é quando os derivados de nitrogênio volatizam para a atmosfera na forma de gases de efeito estufa, por meio da nitrificação e desnitrificação promovida por microorganismos, resultando em danos em escala global. Esses gases volatizados podem, ainda, retornar ao solo, incorporando nutrientes inadvertidamente em áreas naturais, levando a taxas de crescimentos anormais, desequilíbrios na ciclagem de nutrientes e alterações na biodiversidade. O óxido nitroso ( $N_2O$ ) é um dos gases de efeito estufa mais perigosos, uma vez que ele é estável e potente, permanecendo por uma média de 100 anos na atmosfera (BRAUN, 2007).

Por fim, um terceiro problema que pode ser listado é quanto à sua fabricação. A principal fonte de nitrogênio dos fertilizantes é a amônia ( $NH_3$ ), neste caso, sintética. A fonte de nitrogênio utilizada nos fertilizantes químicos é a atmosfera, a qual é composta por aproximadamente de 78% de  $N_2$  (nitrogênio atmosférico, dinitrogênio ou diazoto), um gás inerte que não é absorvido, tampouco utilizado por animais, fungos ou plantas, não participando da combustão e nem da respiração. Para que ocorra a quebra dos átomos desse gás, é necessária uma enorme quantidade de energia. Já no caso da produção do hidrogênio da amônia sintética, utiliza-se grandes quantidades de gás natural, sendo a forma mais viável de produção de hidrogênio livre. No entanto, o gás natural é um combustível fóssil não-renovável e limitado (SIMS *et al.*, 2011).

Já o fósforo e o potássio são produzidos a partir da mineração. O primeiro, precisa ser minerado de rochas e seus estoques também são limitados. Além disso, o processamento dessas rochas, a fim de obter nutrientes, pode acarretar na formação de rejeitos sólidos, normalmente mantidos em barragens, de forma que podem ocasionar uma contaminação do solo, do ar e de corpos d'água, em especial, no que diz respeito a lençóis freáticos. Já o potássio é minerado a partir de depósitos de antigos mares e lagos intracontinentais soterrados e, como uma fonte secundária, também é possível extrair o potássio de lagos salgados, como do mar Morto, por exemplo. Diferente do nitrogênio e do fósforo, o potássio, em geral, não causa danos detectáveis no meio ambiente (PROPEQ, 2020; RÖMHELD; KIRKBY, 2010).

Dessa forma, a agricultura atual, além de mostrar-se preocupada com a produção de alimentos, também almeja a sustentabilidade desse sistema de produção, ou seja, busca otimizar a utilização desses insumos agrícolas, de forma a minimizar os impactos ambientais, concomitantemente, melhorando a qualidade de vida e a saúde do consumidor. Assim, esse modelo de produção agrícola sustentável é altamente favorável à introdução de agentes de controle biológico de pragas e patógenos, bem como de microrganismos que agem no crescimento das plantas, seja pelo aumento de disponibilidade de nutrientes no solo, ou pela produção de fitohormônios (hormônio vegetal), e sideróforos<sup>1</sup> (QUEIROZ, 2003).

Nesse sentido, uma das tecnologias promissoras da atualidade, principalmente quanto à redução do uso de fertilizantes químicos e a sustentabilidade da produção agrícola, é o uso de Rizobactérias Promotoras de Crescimento de Plantas (PGPR), que são microrganismos capazes de aumentar o tamanho e a produtividade da planta, ao interagirem com essas cultivares, além de melhorar a saúde das mesmas (SOKOLOVA *et al.*, 2011).

### 3.2 PGPR

A camada fina de solo entorno das raízes, cujas propriedades físicas, químicas e biológicas são influenciadas pelo crescimento e atividade vegetal, é conhecida como rizosfera. Devido à alta quantidade de exsudatos<sup>2</sup> liberados pelas plantas, coexistem um grande número de microrganismo nesta região, sendo as bactérias os mais abundantes. Conseqüentemente, este é o local onde ocorre a maioria das interações entre plantas e microrganismos (LUSTER *et al.*, 2009; SAHARAN; NEHRA, 2011). Essas interações podem ser benéficas, nocivas ou neutras para a

<sup>1</sup> Compostos orgânicos que atuam na captação de ferro pelos microrganismos.

<sup>2</sup> Saída de líquidos orgânicos através das paredes e membranas celulares por lesão ou por inflamação.

planta, dependendo, às vezes, das consequências decorrentes das condições do solo (LYNCH, 1990).

Segundo Queiroz (2003) e Hayat *et al.* (2010), existe uma grande variedade de bactérias que vivem na rizosfera, as quais são chamadas de rizobactérias. Esses microrganismos ocorrem naturalmente em todos os tipos de plantas e, em sua maioria, são bactérias benéficas que interagem com as plantas pela produção de fitohormônios ou pela alteração da própria comunidade microbiana presente naquele ambiente. Algumas dessas bactérias são capazes de promover uma maior absorção de nutrientes e transferi-los para as plantas por diversos mecanismos, sendo denominadas de Rizobactérias Promotoras de Crescimento em Plantas (PGPR).

Embora as pesquisas com rizobactérias tenham sido iniciadas e utilizadas na agricultura ainda no século XIX, em países como a Rússia, Ucrânia e Índia (HARTHMAN, 2009), o termo só se popularizou a partir do ano de 1978, quando foram descritas, pela primeira vez, para cultivos de raízes, por Joseph W. Kloepper e Milton N. Schroth, quando o uso de resistência a antibióticos possibilitou o monitoramento de bactérias introduzidas no solo. Elas representam um subconjunto diversificado de bactérias colonizadoras da rizosfera, que induzem o aumento do crescimento da planta após a inoculação em sementes (ZABLOTOWICZ *et al.*, 1991).

Os primeiros relatórios que documentaram o aumento do crescimento de plantas, utilizando PGPRs, envolveram bactérias presentes na rizosfera de rabanetes (KLOEPPER, 1978), batata (BURR *et al.*, 1978; GEELS; SCHIPPERS, 1983; HOWIE; ECHANDI, 1983; KLOEPPER *et al.*, 1980) e beterraba sacarina (SUSLOW, 1982).

Do final da década de 1980 até o início da década de 1990, os estudos sobre as PGPRs continuaram em três direções adicionais. Em uma primeira área de pesquisa, foi realizado um trabalho significativo em “culturas sem raízes” como hospedeiros e foi possível verificar que a maioria das plantas cultivadas eram propícias à promoção de crescimento induzido por PGPRs. Já a segunda área, envolvia a caracterização de efeitos mais específicos dos PGPRs, isto é, efeitos além da promoção de crescimento. Tais estudos resultaram da designação de dois novos subgrupos de PGPRs, além de mostrarem que estas podem ser utilizadas como agentes de biocontrole. Por sua vez, a terceira área, denominada “modo de ação de trabalho”, revelou que algumas cepas de PGPRs podem promover o crescimento de plantas diretamente, isto é, quando não há microrganismos patogênicos ou deletérios (ZABLOTOWICZ *et al.*, 1991). A partir desses estudos, foi possível alcançar uma compreensão mais completa e detalhada a respeito das PGPRs quanto à promoção do crescimento de plantas e, conseqüentemente, permitiu-se o uso prático

desses microrganismos na agricultura na forma de bioinoculantes (JHA; SARAF, 2015).

A promoção do crescimento em plantas pode ocorrer direta ou indiretamente. Mecanismos diretos compreendem tanto a facilitação de nutrientes, a qual é dada pela fixação biológica de nitrogênio atmosférico, solubilização de fosfato e aquisição de ferro, usando sideróforos, quanto a modulação de níveis fitohormonais, através da produção de Ácido Indolacético – AIA, que acentuam o crescimento vegetal em vários estágios de desenvolvimento ou, ainda, ou pela redução do estresse hormonal etileno, empregando a enzima ACC desaminase (GLICK, 1995; PATERSON *et al.*, 2017; VESSEY, 2003).

Segundo Newton (2000), apesar de abundante, são poucos os organismos que são capazes de utilizar o nitrogênio atmosférico. Entre os organismos capazes de utilizar esse nitrogênio, estão alguns tipos de bactérias, chamadas de diazotróficas, as quais possuem a capacidade de transformar o dinitrogênio em amônia e incorporá-lo em suas moléculas orgânicas (aminoácidos e proteínas), processo conhecido como FBN (Fixação Biológica de Nitrogênio). A fixação biológica do nitrogênio, geralmente é categorizada como uma relação simbiótica entre a rizobactéria e a planta leguminosa, no entanto, também pode ocorrer em plantas não-leguminosas de forma assimbiótica, com rizobactérias de vida livre no solo (FRANCHE *et al.*, 2009).

O processo de fixação de nitrogênio é um tanto complicado e é realizado por meio do complexo enzimático nitrogenase, consistido em duas proteínas, dinitrogenase redutase possuindo ferro como cofator (proteína ferro-enxofre) e dinitrogenase, possuindo ferro e molibdênio como cofator (proteína ferro-molibdênio). A dinitrogenase redutase fornece elétrons e a dinitrogenase utiliza esses elétrons para converter nitrogênio em amônia (MOAT *et al.*, 1995). Para a síntese da nitrogenase, são necessários vários genes *nif* (do inglês, “*nitrogen fixation*”) na genética dos microrganismos, tais como *nifH*, *nifD*, *nifY*, *nifB*, *nifQ*, *nifE*, *nifN*, *nifX*, *nifS*, *nifV*, *nifW* e *nifZ* (FRANCHE *et al.*, 2009). As rizobactérias dos gêneros *Herbaspirillum*, *Azospirillum*, *Pseudomonas*, *Bradyrhizobium* e *Rhizobium* são exemplos de microrganismos que possuem funções de fixação de oxigênio (BALDANI *et al.*, 1986).

O fósforo é um nutriente essencial para o crescimento das plantas e pode ser encontrado na forma orgânica (derivado de microrganismos e plantas) ou inorgânica (fertilizantes), no entanto, é um dos nutrientes menos solúveis do ambiente, com menos de 5% do fosfato do solo estando disponível para as cultivares (PAUL; CLARK, 1989). Dessa forma, a adição de fertilizantes fosfatados é uma prática habitual na agricultura convencional. Entretanto, grande parte do fertilizante fosfatado aplicado ao solo é rapidamente imobilizada pelo ferro e pelo

alumínio, em solos ácidos, e por cálcio, em solos calcários, logo em seguida da aplicação, tornando-se, assim, indisponível para os vegetais (DOBBELAERE *et al.*, 2003; HOLFORD, 1997). Nesse sentido, destaca-se a importância das PGPRs, uma vez que algumas dessas bactérias são capazes de produzir sideróforos e solubilizar fosfato mineral insolúvel, através da produção de ácidos orgânicos que promovem uma acidez no solo, liberando ortofosfato, que é a forma do fósforo apropriada para o uso pelos vegetais (JONES, 1998). Existe uma enorme quantidade de bactérias solubilizadoras de fosfato no solo e na rizosfera, sendo este o local de maior concentração. As rizobactérias dos gêneros *Bacillus*, *Pseudomonas*, *Rhizobium*, *Enterobacter*, *Erwinia*, *Burkholderia*, *Achromobacter*, *Agrobacterium*, *Micrococcus*, *Aereobacter*, *Flavobacterium* e *Ochrobactrum* são exemplos de microrganismos solubilizadores de fosfato, sendo os três primeiros, os solubilizadores mais poderosos (BABALOLA, 2010; RODRÍGUEZ; FRAGA, 1999).

Quanto à aquisição de ferro, embora este seja um dos minerais mais abundantes na Terra, ele encontra-se relativamente indisponível no solo para a assimilação direta pelos microrganismos, com uma solubilidade extremamente baixa, uma vez que, no ambiente aeróbio, o ferro predominantemente existe como  $Fe^{3+}$  (férico) e é suscetível a formar hidróxidos e óxidos insolúveis. Para contornar este problema, em geral, os microrganismos presentes no solo adquirem ferro por meio de pequenas moléculas de baixo peso molecular denominadas de sideróforos que, por sua vez, se ligam ao ferro e são transportados de volta a célula microbiana e, assim, o ferro torna-se disponível para o desenvolvimento da bactéria (DOBBELAERE *et al.*, 2003; CABAJ; KOSAKOWSKA, 2009). Ao contrário dos fitopatógenos microbianos, as plantas não são prejudicadas pela perda de elementos fundamentais no organismo pelas PGPRs, neste caso, o ferro. Algumas delas podem capturar o complexo ferro-sideróforo, transportando-o para o interior de suas células que, posteriormente será liberado, ficando disponível para a planta utilizá-lo. Dessa forma, os sideróforos atuam como agentes solubilizantes de ferro inorgânico ou compostos orgânicos em condições limitantes de ferro disponível para as bactérias (CROWLEY *et al.*, 1988; MAHANTY *et al.*, 2017).

Os fitohormônios, por sua vez, desempenham um papel importantíssimo no crescimento e desenvolvimento de plantas e, por isso, são chamados de reguladores de crescimento. Eles podem ser classificados como auxinas, citocininas, giberilinas, ácido abscísico e etileno. As auxinas são responsáveis pela diferenciação celular, crescimento radicular, crescimento de frutos e controle da abscisão. As citocinas possuem como funções, a regulação do crescimento,

diferenciação celular e senescência<sup>3</sup> vegetal. As giberilinas são responsáveis pela divisão celular, interrupção da dormência e aumento do desenvolvimento dos frutos. Já o ácido abscísico (ABA), desempenha funções como a regulação da transpiração, quebra de dormência e desenvolvimento inicial das sementes. Por fim, o etileno ajuda no amadurecimento dos frutos, na promoção da abscisão de folhas, frutos e flores e influencia na expressão do sexo feminino (ARSHAD; FRANKENBERGER JR, 1997; MAHANTY *et al.*, 2017; RAVEN *et al.*, 1996).

Segundo Bianco *et al.* (2006) e Glick (2012), observou-se que os microrganismos que habitam a rizosfera também são capazes de sintetizar ou modular os níveis de hormônio vegetal em plantas hospedeiras ou associadas. Cabe destacar que a principal auxina das cultivares é o AIA, capaz de realizar processos celulares de desenvolvimento das plantas quando as condições para tal encontram-se desfavoráveis (KENDE; ZEEVAART, 1997; LAMBRECHT *et al.*, 2000). Em bactérias benéficas, o principal precursor para a biossíntese de AIA é o aminoácido triptofano, através do ácido indol-pirúvico (AIP). Quando secretado por essas bactérias, o AIA pode promover diretamente o crescimento da raiz pela estimulação do alongamento da célula vegetal ou divisão celular, pela influência da ACC desaminase.(MAHANTY *et al.*, 2017; PATTEN; GLICK, 2002). As rizobactérias dos gêneros *Pseudomonas*, *Bacillus*, *Arthrobacter*, *Enterobacter* e *Agrobacterium*, são exemplos de microrganismos que podem produzir compostos indólicos, como AIA, por exemplo (YUAN *et al.*, 2011).

O etileno é uma molécula orgânica produzida de forma espontânea pelas cultivares, necessário para seu desenvolvimento normal, atuando como regulador de crescimento, quando em baixas concentrações. A sua produção pode ser aumentada em condições de estresse, podendo ocasionar a inibição do crescimento radicular, ou seja, se torna prejudicial quando em altas concentrações. Nesse sentido, a enzima ACC desaminase é capaz de clivar o composto 1-aminociclopropano-1-carboxilato (ACC), que é o precursor imediato de etileno, reduzindo, dessa forma, os níveis elevados desse composto e, conseqüentemente os efeitos negativos às plantas (GLICK, 1995; GLICK, 2005; JACKSON, 1991).

Já os mecanismos indiretos envolvem a redução dos efeitos deletérios de um ou mais patógenos sobre o desenvolvimento e crescimento da planta, especificamente por meio de nutrientes, exclusão de nichos, indução de resistência sistêmica (IRS) e produção de metabólitos antimicrobianos. Nesse sentido, alguns compostos podem ser produzidos, vindo a atuarem como moléculas de biocontrole, portanto, influenciando indiretamente o crescimento das plantas

<sup>3</sup> Alterações decorrentes de processos fisiológicos do envelhecimento.

(GLICK, 1995; MAHANTY *et al.*, 2017; PATERSON *et al.*, 2017). Cabe destacar que, na maioria dos casos, bactérias do gênero *Bacillus* que estão envolvidos na IRS, também induzem a promoção do crescimento da planta (KLOEPPER *et al.*, 2004).

O objetivo da síntese dos diversos compostos e antibióticos é, quando secretados, evitar a proliferação de outros organismos, como os patógenos, por exemplo. Além da produção destes antibióticos, rizobactérias que atuam como biocontroladoras, podem produzir enzimas líticas, as quais podem provocar danos às estruturas celulares de vários organismos, especialmente os fungos. Além disso, os próprios sideróforos, já mencionados anteriormente como um mecanismo direto, também podem atuar como mecanismos de biocontrole, prevenindo que bactérias patogênicas adquiram ferro suficiente para seu desenvolvimento e proliferação. Nesse sentido, a produção de sideróforos se mostra um mecanismo muito eficiente, visto que estes possuem alta afinidade por ferro quando comparados à organismos patogênicos, inviabilizando a proliferação destes microrganismos nocivos, dada a insuficiência de ferro (GLICK, 2012; HERMOSA *et al.*, 2012; MAHANTY *et al.*, 2017).

As PGPR também podem atuar como indutoras de resistência sistêmica de plantas contra vários patógenos e estresses ambientais. Quando uma bactéria coloniza a raiz, moléculas que compõem a célula bacteriana ou por ela sintetizadas, agem como sinalizadores de novas rotas metabólicas. Com isso, essas moléculas ativam os genes promotores de compostos de defesa, fazendo com que ocorra a expressão da resistência sistêmica induzida (VAN LOON *et al.*, 1998). Dessa forma, as plantas são capazes de se protegerem contra fitopatógenos e agentes estressantes, produzindo diversos compostos antimicrobianos e enzimas de redução oxidativa, que se relacionam diretamente à defesa. Portanto, a biossíntese de enzimas relacionadas a indução de resistência sistêmica desempenha um papel crucial no grau de determinação da resistência do hospedeiro (THAKKER *et al.*, 2013).

Diante do exposto, pode-se concluir que as formas de ação das PGPR são diversas, e que nem todas elas possuem os mesmos mecanismos. Além disso, a eficiência desses microrganismos pode ser afetada de diversas maneiras, de forma que, com uma mesma bactéria, é possível obter diferentes resultados, quando sujeitas a condições distintas, o que torna um grande desafio compreender a interação entre PGPR e plantas. A eficiência de uma PGPR pode ser alterada pela quantidade de microrganismos inoculados, pelo veículo usado na inoculação, pelo estágio de desenvolvimento da cultivar no momento da inoculação, pelas propriedades físico-químicas do solo, pelas interações com a microbiota nativa, pelas interações com outras PGPRs, pelo estado

de saúde das plantas, pelo clima da região, entre muitos outros fatores (COSTA, 2012; VEJAN *et al.*, 2016). Em suma, as PGPRs podem afetar o crescimento vegetal por um ou mais mecanismos mencionados e, ainda, podem utilizar diferentes habilidades para promover o crescimento das plantas em diversos estágios durante o seu ciclo de vida (GLICK, 1995; GLICK *et al.*, 1999; PATERSON *et al.*, 2017).

A existência de tantos fatores capazes de alterar a eficiência de uma PGPR para um determinado vegetal pode tornar sua aplicação mais restrita a certas condições, além de tornar um grande desafio compreender sua interação com as plantas. Dessa forma, muito ainda deve ser estudado para que haja um entendimento concreto sobre a interação de PGPR com plantas e quais mecanismos estão envolvidos no efeito benéfico bacteriano, de modo a desenvolver bioinoculantes formulados com PGPRs que atuem consistentemente, mesmo em diferentes condições e culturas. Nesse sentido, a pesquisa nessa área sempre busca novos organismos que possuam maior eficiência ou inclusão de determinados nichos, como a resistência à metais pesados, por exemplo (COSTA, 2012; MACLEAN *et al.*, 2009).

A maioria das PGPRs relatadas são dos gêneros *Pseudomonas*, *Bacillus*, *Agrobacterium*, *Arthrobacter*, *Azospirillum*, *Acinetobacter*, *Azotobacter*, *Enterobacter*, *Bradyrhizobium*, *Burkholderia*, *Frankia*, *Rhizobium*, *Serratia* e *Thiobacillus*, sendo os dois primeiros os mais comumente estudados (GLICK, 1995; MELO, 1998; VESSEY, 2003). Neste estudo, em específico, as bactérias utilizadas serão as do gênero *Bacillus*.

### 3.2.1 Gênero *Bacillus*

As cepas de rizobactérias promotoras de crescimento de plantas do gênero *Bacillus* têm sido amplamente estudadas para melhorar o crescimento de plantas (CHOUDHARY; JOHRI, 2009; KLOEPPER *et al.*, 2004). Esses microrganismos pertencem à família *Bacilliaceae* e, atualmente, inclui diversas espécies conhecidas e classificadas, que diferem entre si em termos de características fenotípicas e fisiológicas (ROZADA, 2022).

Segundo Joung e Côté (2002), as bactérias do gênero *Bacillus* são Gram-positivas, aeróbicas ou anaeróbicas facultativas, em formato de bastonetes. Elas são capazes de formar estruturas dormentes estáveis, denominadas endósporos (ou esporos bacterianos), os quais são resistentes ao calor, frio, radiação, dessecação e desinfetantes e, portanto, responsáveis por garantir a sobrevivência da bactéria em períodos de estresse ambiental. Devido à essa habilidade de esporulação e o fácil cultivo de bactérias desse gênero, esse grupo de microrganismos são

atrativos para o uso prático como bioinoculantes, além de fornecerem uma solução para o problema de sobrevivência em formulações e em campo (TOYOTA, 2015).

As espécies de *Bacillus* são bastante utilizadas em diversos processos médicos, farmacêuticos, agrícolas e industriais que aproveitam a ampla gama de características fisiológicas e a capacidade de produzir várias enzimas, antibióticos e outros metabólitos que bactérias dessa espécie possuem. A bacitracina e a polimixina são dois exemplos de antibióticos bastante conhecidos e amplamente utilizados, obtidos de bactérias desse gênero (TURNBULL *et al.*, 1996).

Elas também podem ser aplicadas de forma a controlar biologicamente enfermidades nos vegetais, além de melhorar a produtividade das lavouras (NGUGI *et al.*, 2005; YAO *et al.*, 2006). A promoção de crescimento vegetal por rizobactérias do gênero *Bacillus* se dá, por exemplo, pela produção de fitohormônios, pela mobilização de fósforo, produção de sideróforos e antibióticos, inibição da síntese de etileno, indução de resistência das plantas contra patógenos e pela eliminação de microrganismos deletérios e de metabólitos tóxicos presentes na zona radicular. O controle desses fitopatógenos ocorre através da produção de antibióticos que atuam na supressão desses microrganismos na rizosfera (RATZ, 2014).

Algumas espécies de *Bacillus* estão sendo estudadas como possíveis bioinoculantes e vêm gerando bons resultados na agricultura no quesito promoção de crescimento, como é o caso dos *B. subtilis*, *B. amyloliquefaciens* e *B. velezensis*, entre outros (FERNÁNDEZ *et al.*, 2004).

Os *B. subtilis* são considerados agentes de controle biológico, por promover o desenvolvimento de plantas e prevenir doenças no solo. São caracterizados por serem Gram-positivos, produtores de endósporos resistentes a fatores físico-químicos prejudiciais, como temperaturas elevadas, radiação, entre outros, produzem enzimas hidrofílicas extracelulares que degradam polissacarídeos e antibióticos lipopeptídicos, como surfactina, bacitracina e polimixina (RAGAZZO-SÁNCHEZ *et al.*, 2011).

Já os *B. amyloliquefaciens* caracterizam-se por possuírem um genoma constituído por um grupo de genes relacionados à síntese de lipopeptídeos e policétidos com atividade antifúngica, antibacteriana e nematocida. Esses microrganismos podem associarem-se às raízes de plantas, estimulando o crescimento da mesma, além de favorecer a supressão de patógenos, criando condições ideais para a planta, possibilitando a produção de uma variedade de metabólitos secundários, que evitam o crescimento de outras bactérias competitivas dentro da rizosfera da planta (CHEN *et al.*, 2009).

Por fim, os *B. velezensis*, que são uma nova espécie do gênero *Bacillus*, descritos

inicialmente como sinônimos heterotípicos de *B. amyloliquefaciens*, são comumente encontrados no solo e caracterizam-se por apresentarem metabolismo aeróbio e serem capazes de produzir antibióticos com alta atividade antifúngica comprovada in vitro, além da capacidade de solubilizar fosfatos, produzir ácido indolacético e sideróforos que promovem o crescimento da cultivar (FERNÁNDEZ *et al.*, 2004; ZAID *et al.*, 2022).

Portanto, infere-se que o gênero *Bacillus* é um dos grupos mais comuns e facilmente cultiváveis, possuindo grande interesse em sistemas agrários devido ao seu enorme potencial em aumentar a produtividade agrícola e controlar doenças fitopatogênicas (OROZCO-MOSQUEDA *et al.*, 2020).

Na Figura 6, é possível observar uma representação de bactérias da espécie *Bacillus subtilis*.

**Figura 6 – *Bacillus subtilis***



**Fonte:** <https://www.pigprogress.net/specials/agp-benefits-ensured-by-bacillus-subtilis/>

Para esta pesquisa, foram utilizadas algumas cepas recém sequenciadas de *B. velezensis*, visto que esta espécie tem ganhado bastante atenção de pesquisadores da área, sendo consideradas como PGPRs vitais (YE *et al.*, 2018). Além disso, também foram utilizadas outras espécies de bactérias do gênero *Bacillus*, estudadas na literatura.

### 3.3 ESTUDOS RELACIONADOS

Com o avanço nos estudos de PGPR, mais análises genômicas têm sido realizadas com o intuito de comparar os efeitos dos genes envolvidos na promoção de crescimento de plantas e outros metabólitos secundários. Em suma, os pesquisadores isolam cepas de determinadas espécies de rizobactérias e realizam o sequenciamento, montagem e anotação genômica delas. Com isso, utilizam-se programas para identificar os genes das cepas e determinar quais codificam ou não proteínas. Com a utilização de alguns softwares, também é possível encontrar metabólitos secundários de um determinado genoma. A partir daí, os pesquisadores analisam as comparações.

Nesta etapa, o objetivo é investigar os genes de bactérias relacionados à produção de metabólitos secundários envolvidos, principalmente, na atividade de promoção de crescimento de cultivares. Assim, é realizado um estudo na literatura e uma discussão sobre os genes catalogados, de espécies do gênero *Bacillus*, relacionados à promoção de crescimento de plantas.

Em Zaid *et al.* (2022), os autores sequenciaram todo o genoma da cepa HNA3 (código de acesso GenBank [CP040881.1](#)), isolada da rizosfera de plantas e classificada anteriormente como *B. amyloliquefaciens*, com base em observações morfológicas e análise de sequência de rRNA 16S. Neste estudo, eles realizaram uma análise comparativa do genoma e reclassificaram esta cepa como *Bacillus velezensis* por análise do genoma central entre HNA3 e outras 74 cepas de *Bacillus* previamente definidas na árvore evolutiva (7 cepas de *B. paralicheniformis*, 21 cepas de *B. subtilis*, 18 cepas de *B. amyloliquefaciens* e 28 cepas de *B. velezensis*).

Os autores também realizaram uma análise comparativa da cepa HNA3 com outras três cepas de *Bacillus* de referência, incluindo *B. amyloliquefaciens* DSM7 (uma cepa não associada à plantas), *B. velezensis* FZB42 (rizobactéria associada à plantas) e *B. subtilis* 168. A análise entre os CDSs das quatro cepas em questão mostrou que 1.877 genes são compartilhados por todas as cepas selecionadas, 1.965 genes foram compartilhados entre HNA3 e *B. subtilis* 168, 3.505 genes foram compartilhados entre HNA3 e *B. velezensis* FZB42 e, por fim, 3358 genes compartilhados entre HNA3 e *B. amyloliquefaciens* DSM7. Além disso, constatou-se que HNA3 tinha 431 e 284 genes únicos relativos a *B. amyloliquefaciens* DSM7 e *B. velezensis* FZB42, respectivamente, e 184 genes de cópia única.

A partir da mineração do genoma, usando o software antiSMASH versão 5.1.0, foram identificados 12 agrupamentos de metabólitos secundários no genoma de *B. velezensis* HNA3. Com isso, Zaid *et al.* (2022) compararam os agrupamentos de genes entre a cepa HNA3 e as

outras três, onde foi possível constatar que 6 dos 12 agrupamentos são conservados pelas quatro, 2 são conservados por três cepas (HNA3, FZB42 e DSM7) e 2 foram identificados apenas na HNA3 e FZB42. Os autores destacam que, embora vários agrupamentos de genes tenham sido compartilhados pelas quatro cepas, a estrutura gênica delas se diferenciam, visto que um mesmo cluster de genes pode possuir menos genes em uma cepa do que em outra.

Os autores relatam, ainda, que o genoma da HNA3 possui vários genes que codificam proteínas que, possivelmente, estão associadas à atividade de promoção de crescimento, como genes putativos envolvidos na produção de AIA e solubilização de fosfato, por exemplo. No Quadro 2, é possível observar os 35 genes detectados pelos autores no genoma da cepa HNA3, possivelmente envolvidos na atividade de promoção do crescimento de plantas.

**Quadro 2 – Genes associados à promoção de crescimento em plantas na cepa *B. velezensis* HNA3**

Gene	Produto	Gene	Produto
<i>trpA</i>	<i>Tryptophan synthase alpha chain</i>	<i>flgM</i>	<i>Flagellar biosynthesis anti-sigma factor</i>
<i>trpB</i>	<i>Tryptophan synthase beta chain</i>	<i>dhaS</i>	<i>Putative aldehyde dehydrogenase</i>
<i>trpC</i>	<i>Indole-3-glycerol phosphate synthase</i>	<i>ysnE</i>	<i>N-acetyltransferase</i>
<i>trpD</i>	<i>Anthranilate phosphoribosyl transferase</i>	<i>phytase</i>	<i>Mineralize organic phosphorus enzyme</i>
<i>trpE</i>	<i>Anthranilate synthase component I</i>	<i>pksS</i>	<i>Polyketide biosynthesis cytochrome P450</i>
<i>trpF</i>	<i>Phosphoribosyl anthranilate isomerase</i>	<i>msmX</i>	<i>ABC-type spermidine transport systems</i>
<i>ywkB</i>	<i>Auxin efflux carrier</i>	<i>speG</i>	<i>Spermidine acetyltransferas</i>
<i>yhcX</i>	<i>Putative amidohydrolase</i>	<i>alsD</i>	<i>Acetolactate decarboxylase</i>
<i>speE</i>	<i>Spermidine synthase, polyamine metabolism</i>	<i>alsR</i>	<i>Transcriptional regulator</i>
<i>ilvB</i>	<i>Acetolactate synthase</i>	<i>ylbF</i>	<i>Controls biofilm development</i>
<i>bdhA</i>	<i>2,3-butanediol dehydrogenase</i>	<i>ymcA</i>	<i>Biofilm development</i>
<i>acuC</i>	<i>Acetoin dehydrogenase</i>	<i>iolU</i>	<i>Scyllo-inositol 2-dehydrogenase (NADP(+))</i>
<i>cysE</i>	<i>Serine O-acetyltransferase</i>	<i>slrR</i>	<i>Master regulator for biofilm formation</i>
<i>hfq</i>	<i>Hfq RNA chaperone</i>	<i>fliA</i>	<i>RNA polymerase sigma factor</i>
<i>sinR</i>	<i>Master regulator for biofilm formation</i>	<i>crr</i>	<i>Sugar-specific IIA component</i>
<i>luxS</i>	<i>S-ribosyl homocysteine lyase</i>	<i>wecB</i>	<i>UDP-N-acetylglucosamine 2-epimerase</i>
<i>rpoN</i>	<i>RNA polymerase sigma-54</i>	<i>tarA/tagA</i>	<i>Acetylglucosaminyl diphosphoundecaprenol N-acetyl-beta-D-mannosaminyl transferase</i>
<i>csrA</i>	<i>Carbon storage regulator</i>		

Fonte: Adaptado de Zaid *et al.* (2022)

Assim, os autores afirmam que a cepa HNA3 é uma PGPR prospectiva, com características genéticas diferentes de análogos em diversos ambientes biológicos, possuindo várias

características muito importantes para uma PGPR e concluem que ela pode ser formulada e preparada comercialmente para aplicação em campo, a fim de suprimir fitopatógenos e melhorar o desenvolvimento da cultura, seja sozinha ou como parte de associações microbianas.

Em Mosela *et al.* (2022), o objetivo foi selecionar e validar cepas de bactérias com tripla ação em milho e soja, sendo estas: promoção de crescimento em plantas, solubilização de fosfato e agente de biocontrole, a fim de desenvolver um inoculante microbiano multifuncional para a agricultura. Para tanto, selecionou-se linhagens bacterianas, isoladas da rizosfera do milho, com alta produção de AIA, solubilização de fosfato e atividade antifúngica contra fungos patogênicos do solo, sendo eles: *Rhizoctonia solani*, *Macrophomina phaseolina* e *Fusarium solani*.

A partir do estudo realizado e do experimento com milho em casa de vegetação, verificou-se que a cepa Ag75 (código GenBank [CP099465.1](#)), identificada como *Bacillus velezensis*, possuía um alto potencial para o aumento de biomassa tanto na parte da raiz como na parte aérea do milho e, também, no teor de fosfato da parte aérea dessa cultivar. Para os experimentos de campo, a inoculação da estirpe Ag75 aumentou a produtividade do milho e da soja em 17,8 e 26,5%, respectivamente, em relação ao controle (25 Kg  $P_2O_5$ ).

A rizobactéria *Bacillus velezensis* Ag75 também apresentou atividade antifúngica para os três fungos mencionados anteriormente, inibindo o crescimento micelial do *Rhizoctonia solani*, *Macrophomina phaseolina* e *Fusarium solani* em 44, 49 e 61%, respectivamente.

O genoma completo da cepa Ag75 foi sequenciado utilizando a plataforma Illumina NovaSeq 6000, no Instituto de Pesquisa do Câncer (IPEC), no município de Guarapuava, Paraná, Brasil. Uma série de montagens de novo foram realizadas por diferentes softwares, testando diversos parâmetros de montagem. As estratégias de montagem do genoma *CLC Genomics Workbench 11* e *IDBA Hybrid* demonstraram os melhores resultados para a montagem. Além disso, o genoma circular da estirpe Ag75 foi representado e comparado com outros genomas de referência usando o software BRIG.

Com o auxílio do servidor web antiSMASH 5.1.0, foram identificados 12 agrupamentos de genes biossintéticos, sendo seis deles ligados à síntese de *macrolactin*, *bacillaene*, *fengycin*, *difficifin*, *bacillibactin* e *bacilisyn*, apresentando 100% de similaridade, um ligado à síntese de *surfactin*, com 82% de similaridade e outro com 7% de similaridade, responsável pela síntese de *butyrosine A* e *B*. Além destes, quatro não apresentaram similaridade com o banco de dados. Ademais, com o software RAST (do inglês “*Rapid Annotation using Subsystem Technology*”), foi possível identificar um total de 47 genes relacionados à promoção do crescimento em plantas,

apresentados no Quadro 3.

**Quadro 3 – Genes associados à promoção de crescimento em plantas na cepa *Bacillus velezensis* Ag75**

<b>Gene</b>	<b>Produto</b>	<b>Gene</b>	<b>Produto</b>
<i>trpA</i>	<i>Tryptophan synthase subunit alpha</i>	<i>groEL</i>	<i>Heat shock protein 60 kDa family chaperone GroEL</i>
<i>trpB</i>	<i>Tryptophan synthase subunit beta</i>	<i>groES</i>	<i>Heat shock protein 10 kDa family chaperone GroES</i>
<i>trpC</i>	<i>Indole-3-glycerol phosphate synthase TrpC</i>	<i>cspC</i>	<i>Cold shock protein CspC</i>
<i>trpD</i>	<i>Anthranilate phosphoribosyltransferase</i>	<i>cspB</i>	<i>Cold shock-like protein CspB</i>
<i>trpE</i>	<i>Anthranilate synthase component I</i>	<i>cspD</i>	<i>Cold-shock protein CspD</i>
<i>ywdH</i>	<i>Aldehyde dehydrogenase</i>	<i>opuAA</i>	<i>Glycine/proline betaine ABC transporter ATP-binding protein OpuAA</i>
<i>ywkB</i>	<i>Auxin efflux carrier family protein</i>	<i>opuAB</i>	<i>Glycine/proline betaine ABC transporter permease subunit OpuAB</i>
<i>speE</i>	<i>Spermidine synthase</i>	<i>opuAC</i>	<i>Glycine/betaine ABC transporter</i>
<i>budA</i>	<i>Acetolactate decarboxylase</i>	<i>bcp</i>	<i>Thiol peroxidase</i>
<i>alsS</i>	<i>Acetolactate synthase</i>	<i>trpF</i>	<i>Phosphoribosylanthranilate isomerase</i>
<i>bdhA</i>	<i>2,3-Butanediol dehydrogenase</i>	<i>ysnE</i>	<i>GNAT family N-acetyltransferase</i>
<i>acoA</i>	<i>Acetoin dehydrogenase</i>	<i>amhX</i>	<i>Amidohydrolase</i>
<i>sigD</i>	<i>RNA polymerase sigma factor for flagellar operon and Biofilm formation</i>	<i>bioI</i>	<i>Biotin biosynthesis cytochrome P450</i>
<i>sinR</i>	<i>Master regulator of biofilm formation</i>	<i>bltD</i>	<i>Spermidine acetyltransferas</i>
<i>luxS</i>	<i>S-Ribosyl homocysteine lyase for Quorum sensing Biofilm formation</i>	<i>alsR</i>	<i>Transcriptional regulator</i>
<i>rpoN</i>	<i>RNA polymerase sigma-54 factor for Biofilm formation</i>	<i>ylbF</i>	<i>Controls biofilm development</i>
<i>csrA</i>	<i>Carbon storage regulator for Biofilm formation</i>	<i>ymcA</i>	<i>Biofilm development</i>
<i>flgM</i>	<i>Negative regulator of flagellin synthesis flgM for Biofilm formation</i>	<i>ioIW</i>	<i>Scyllo-inositol 2-dehydrogenase (NADP(1) involved in biofilm formation protein</i>
<i>wecB</i>	<i>UDP-N-acetylglucosamine 2-epimerase for Biofilm formation</i>	<i>ycdF</i>	<i>Glucose 1-dehydrogenase</i>
<i>gdh</i>	<i>Glucose 1-dehydrogenase</i>	<i>tuaD</i>	<i>UDP-glucose 6-dehydrogenase</i>
<i>zwf</i>	<i>Glucose-6-phosphate dehydrogenase</i>	<i>phzF</i>	<i>PhzF family phenazine biosynthesis isomerase</i>
<i>treP</i>	<i>PTS system trehalose-specific EIIBC component</i>	<i>opuD</i>	<i>Glycine betaine transporter OpuD</i>
<i>treR</i>	<i>Trehalose operon repressor</i>	<i>bsaA</i>	<i>Glutathione peroxidase</i>
<i>hslR</i>	<i>Ribosomal RNA binding protein involved in 50S recycling heat shock protein</i>		

**Fonte: Adaptado de Mosela et al. (2022)**

Por fim, os autores concluem que a cepa *Bacillus velezensis* Ag75 apresenta um grande potencial para o desenvolvimento de um inoculante multifuncional, combinando a capacidade de promover o crescimento das cultivares, solubilizar fosfato e atuar como agente de biocontrole, otimizando o uso de fertilizantes, de forma a contribuir para uma agricultura sustentável e com baixo custo.

Em Teixeira *et al.* (2021), o objetivo principal era avaliar a cepa *Bacillus velezensis* CMRP 4490 (LABIM22), quanto à sua capacidade de inibir patógenos de plantas do solo e aumentar o crescimento das cultivares. Este estudo contou com a avaliação do controle antifúngico *in vitro*, sequenciamento do genoma bacteriano, mineração de genes responsáveis pela síntese de metabólitos secundários, capacidade de colonização radicular da rizosfera e estudos em estufa para a avaliação da capacidade promotora de crescimento. A cepa LABIM22 foi isolada do solo da Universidade Estadual de Londrina (UEL), localizada no norte do Paraná, Brasil e é classificada como *B. velezensis*, um agente de biocontrole promissor.

A fim de revelar o potencial da cepa LABIM22 como agente de biocontrole contra fungos fitopatogênicos, os autores investigaram as propriedades genômicas e fisiológicas dessa estirpe. Especificamente, eles realizaram o sequenciamento do genoma, investigaram genes de metabolismo secundário, garimparam os genes ligados à regulação/formação de biofilme e testaram a capacidade da cepa de promover o crescimento vegetal.

Com a análise genômica, utilizando o servidor web antiSMASH 5.1.0, foi possível revelar que a estirpe LABIM22 possui 13 agrupamentos específicos de genes relacionados à biossíntese de metabólitos secundários, que desempenham funções importantes quanto à supressão de patógenos e à promoção do crescimento em plantas, e de genes relacionados à colonização da rizosfera. Dentre os agrupamentos encontrados, oito apresentaram similaridade com o banco de dados utilizado e são responsáveis pela síntese de *surfactin*, *butirosin* A e B, *macrolactin* H, *bacillaene*, *fengycin*, *difficidin*, *bacillibactin* e *bacilysin*. Os demais agrupamentos não apresentaram similaridade.

Os autores sugerem ainda que, devido à outros estudos descreverem o potencial para promoção de crescimento e supressão de patógenos para cepas de *B. velezensis*, como o caso das estirpes S141 e a FZB42, que apresentaram alta capacidade de promoção de crescimento, principalmente em soja, possivelmente, a cepa LABIM22 também possui tal capacidade, dada a alta similaridade genômica. Diversos outros estudos foram realizados com espécies de *Bacillus* intimamente relacionadas e sugeriram que esta cepa é uma boa candidata para agente de

biocontrole.

Assim, os autores concluem que a cepa *Bacillus velezensis* LABIM22 possui um potencial biotecnológico devido à alta similaridade genética com outras cepas do mesmo gênero já utilizadas como agentes de biocontrole e já estabelecidas no mercado. Além disso, eles afirmam que o estudo da maquinaria genética ligada ao metabolismo secundário também reforça o potencial biotecnológico desta estirpe ao encontrar agrupamentos de genes biossintéticos que estão relacionados à atividade antifúngica. Ademais, devido a capacidade promotora de crescimento desta cepa, descrita no estudo de Teixeira *et al.* (2021), eles classificam a *B. velezensis* LABIM22 como uma PGPR.

Em Chen *et al.* (2019), foi realizada uma análise da bactéria endofítica *Bacillus velezensis* LDO2 (código de acesso GenBank [CP029034.1](https://www.ncbi.nlm.nih.gov/nuclot/CP029034.1)), isolada da rizosfera do amendoim. Neste estudo avaliou-se a atividade antimicrobiana e antifúngica contra patógenos do amendoim e a promoção do crescimento do mesmo.

A atividade antimicrobiana e antifúngica da cepa LDO2 foi testada por ensaio de placa-cilindro e co-cultura em placas, respectivamente. O primeiro ensaio mostrou que a estirpe LDO2 inibiu o crescimento da bactéria patogênica *Ralstonia solanacearum*, conhecida como murcha bacteriana. Já no segundo ensaio, observou-se que a cepa inibiu significativamente o crescimento dos fungos *Alternaria tenuissima* (mal-das-folhas), *Aspergillus flavus* (tombamento), *Aspergillus niger* (podridão da coroa), *Fusarium oxysporum* (tombamento), *Fusarium moniliforme* (podridão da raiz), *Rhizoctonia solani* (tombamento e podridão da vagem), *Rhizopus sp.* (tombamento), com 78,97%, 80,77%, 79,74%, 81,03%, 81,28%, 79,23%, 75,64% da razão de inibição, respectivamente.

Ainda segundo os autores, a bactéria endofítica LDO2 inibiu significativamente o crescimento micelial de *Aspergillus flavus* e causou o surgimento de uma deformidade micelial. A partir dos resultados obtidos, foi possível verificar que a cepa LDO2 apresentou atividade antimicrobiana e antifúngica significativa para patógenos bacterianos e fúngicos do amendoim.

O endófito de amendoim também apresentou diversas características relacionadas à promoção de crescimento vegetal, como a síntese de sideróforos, a produção de compostos voláteis orgânicos promotores de crescimento e fitohormônios. Inclusive, os testes em casa de vegetação mostraram um aumento significativo na altura das plântulas do amendoim tratadas, assim como no peso seco das plântulas, no comprimento da raiz e no peso seco da raiz em relação às plantas de controle não inoculadas, indicando que a cepa LDO2 poderia promover o

crescimento de mudas de amendoim.

A relação completa de genes relacionados à promoção de crescimento vegetal, detectados na cepa LDO2 e relatada pelos autores, encontra-se no Quadro 4. Nele, observa-se um total de 43 genes possivelmente envolvidos na promoção de crescimento em plantas.

**Quadro 4 – Genes associados à promoção de crescimento em plantas na cepa *Bacillus velezensis* LDO2**

Gene	Produto	Gene	Produto
<i>alsS</i>	<i>Acetolactate synthase</i>	<i>alsD</i>	<i>Acetolactate decarboxylase</i>
<i>ilvB</i>	<i>Acetolactate synthase large subunit</i>	<i>ilvH</i>	<i>Acetolactate synthase small subunit</i>
<i>bdhA</i>	<i>2, 3-butanediol dehydrogenase</i>	<i>treA</i>	<i>Trehalose-6-phosphate hydrolase</i>
<i>treR</i>	<i>Tre operon transcriptional repressor</i>	<i>treP</i>	<i>Trehalose permease component</i>
<i>dhaS</i>	<i>Indol 3-acet-aldehyde dehydrogenase</i>	<i>phy</i>	<i>3-Phytase</i>
<i>yhcX</i>	<i>Nitrilase</i>	<i>ysnE</i>	<i>N-acetyltransferase</i>
<i>speE</i>	<i>Spermidine synthase</i>	<i>potA</i>	<i>Spermidine import ATP-binding protein</i>
<i>speA</i>	<i>Arginine decarboxylase</i>	<i>mdtJ</i>	<i>Spermidine export protein</i>
<i>speB</i>	<i>Agmatinase</i>	<i>moaB</i>	<i>Molybdenum cofactor synthesis protein</i>
<i>moaA</i>	<i>Molybdenum cofactor synthesis protein</i>	<i>moaE</i>	<i>Molybdenum cofactor synthesis protein</i>
<i>moaC</i>	<i>Molybdenum cofactor synthesis protein</i>	<i>nasE</i>	<i>Nitrite reductase small subunit</i>
<i>moaD</i>	<i>Molybdenum cofactor synthesis protein</i>	<i>nrgB</i>	<i>Nitrogen regulatory protein</i>
<i>nasD</i>	<i>Nitrite reductase large subunit</i>	<i>nrgA</i>	<i>Ammonium transporter</i>
<i>narG</i>	<i>Nitrate reductase</i>	<i>nark</i>	<i>Nitrate transporter</i>
<i>narH</i>	<i>Nitrate reductase</i>	<i>narP</i>	<i>Nitrate/nitrite response regulator</i>
<i>narI</i>	<i>Nitrate reductase</i>	<i>ktrB</i>	<i>Potassium uptake protein B</i>
<i>narJ</i>	<i>Nitrate reductase</i>	<i>corA</i>	<i>Mg<sup>2+</sup>/Co<sup>2+</sup> transport protein</i>
<i>ktrA</i>	<i>Potassium uptake protein A</i>	<i>dhb cluster</i> ( <i>dhbA</i> , <i>dhbB</i> , <i>dhbC</i> , <i>dhbE</i> , <i>dhbF</i> )	<i>Bacillibactin</i>
<i>ktrC</i>	<i>Potassium uptake protein C</i>		
<i>ktrD</i>	<i>Potassium uptake protein D</i>		
<i>mgtE</i>	<i>Mg<sup>2+</sup> transporter</i>		

Fonte: Adaptado de Chen *et al.* (2019)

Por meio da mineração do genoma da LDO2, com o auxílio do servidor web antiS-MASH v4.1.0, foram identificados 32 agrupamentos de genes biossintéticos. Dentre eles, quatro são antibióticos policetídeos (*butirosin*, *bacillaene*, *difficidin*, *macrolactin*), dois lipopeptídeos cíclicos (*fengycin*, *surfactin*), um sideróforo (*bacillibactin*) e um dipeptídeo (*bacilysin*). Segundo os autores, os quatro primeiros possuem excelente atividade antibacteriana, o lipopeptídeo *fengycin* foi reconhecido pela forte atividade antifúngica contra fungos filamentosos. Já o lipo-

peptídeo cíclico *surfactin* e o dipeptídeo *bacilysin* demonstraram ter atividades antibacteriana e antifúngica.

Por fim, os autores concluem que a estirpe *Bacillus velezensis* LDO2 possui alta capacidade para sintetizar metabólitos antimicrobianos que agem contra microrganismos patogênicos do amendoim, além de exibir várias características associadas à promoção do crescimento em plantas, sugerindo que a cepa possa ser um potencial agente de biocontrole para a produção de amendoim e uma fonte de compostos antimicrobianos para explorações futuras.

Em Chen *et al.* (2007), é feita uma análise comparativa do genoma da estirpe da bactéria promotora de crescimento vegetal *B. amyloliquefaciens* FZB42 com outras bactérias de referência da literatura. O ponto de partida desse estudo foram os resultados que mostraram que a estirpe *B. amyloliquefaciens* FZB42 simula a promoção de crescimento em plantas e também suprime os patógenos destas plantas (IDRISS *et al.*, 2002).

O estudo realizado comparou os seguintes organismos modelos: *B. amyloliquefaciens* FZB42, *B. subtilis* 168, *B. licheniformis*, *B. clausii*, *B. halodurans* e *B. cereus*. Os organismos *B. subtilis* 168 e *B. licheniformis* foram escolhidos por serem muito relacionados com *B. amyloliquefaciens* FZB42 e, por sua vez, a escolha dos demais organismos foi feita baseada nas relações filogenéticas da *B. amyloliquefaciens* FZB42.

Nesta pesquisa, os autores realizaram a identificação de 10 grupos de genes que estão envolvidos em interações planta-bactéria na cepa *B. amyloliquefaciens* FZB42. Estes grupos contêm dezenas de genes, cada um relacionado à uma interação específica, desde colonização de raízes até produção de fitohormônios. Devido à quantidade de informações organizadas pelos autores, os genes apresentados em Chen *et al.* (2007), com seus respectivos códigos de acesso, encontram-se no Apêndice E.

Em suma, os autores concluíram que a cepa *B. amyloliquefaciens* FZB42 é relacionada, mas não idêntica, a cepa *B. amyloliquefaciens* DSMZ7. Além disso, por meio do estudo da árvore filogenética, cujas informações estão contidas na Figura Suplementar 2 do artigo, concluíram que o *B. amyloliquefaciens* FZB42 representa uma ramificação distinta e claramente distante da *B. subtilis* 168.

Por fim, o estudo realizado conclui que o genoma da *B. amyloliquefaciens* FZB42 deve ter um alto valor para revelar interações complexas entre rizobactérias Gram-positivas e plantas, e também estudar a formação de biofilme e outros processos biológicos que foram perdidos ou atenuados durante o uso extensivo da *B. subtilis* 168 como genoma de referência das últimas

décadas.

Cabe ressaltar que, recentemente, a cepa *B. amyloliquefaciens* FZB42 foi reclassificada como *B. velezensis* FZB42 (TEIXEIRA *et al.*, 2021; ZAID *et al.*, 2022), agora com o código de acesso GenBank [CP000560.2](#).

Destaca-se, ainda, que os genes apresentados por Zaid *et al.* (2022), Mosela *et al.* (2022), Chen *et al.* (2019) e Chen *et al.* (2007) com seus códigos de acesso correspondentes, encontram-se no Apêndice A, Apêndice B, Apêndice C e Apêndice D, respectivamente.

Em Lu *et al.* (2022), um estudo recente, a cepa bacteriana CNBG-PGPR-1 ou BS96 ([CP090125.1](#)) foi isolada de um fertilizante orgânico líquido, naturalmente fermentado em Nanjing, Jiangsu, China.

O DNA genômico da referida cepa foi extraído, sequenciado, montado e anotado. A anotação foi realizada via pesquisa BLAST em seis bancos de dados de funções: banco de dados de proteínas não redundantes (NR), banco de dados de sequências de proteínas Swiss-Prot, banco de dados de famílias de proteínas (Pfam), banco de dados Clusters of Orthologous Groups (COG), bancos de dados Gene Ontology (GO) e bancos de dados Kyoto Encyclopedia of Genes and Genomes (KEGG).

Para a predição de agrupamentos de genes biossintéticos utilizou-se a ferramenta AntiSMASH 4.0, encontrando um total de oito BGCs, e a representação circular do genoma com anotação funcional foi plotado usando o software Circos.

Os genes anotados foram classificados em 21 categorias funcionais COG, e a maioria destes foi associada a funções como: transporte e metabolismo de aminoácidos; transporte e metabolismo de carboidratos; transcrição, replicação, recombinação e reparo; transporte e metabolismo de íons inorgânicos.

Com as análises realizadas, vários genes possivelmente envolvidos na promoção de crescimento em plantas foram identificados, incluindo 21 genes relacionados à quimiotaxia, 15 genes envolvidos no metabolismo do triptofano, principal pré-requisito para a síntese do ácido indol 3-acético; 20 genes envolvidos no metabolismo do enxofre, incluindo toda a via assimilatória de redução do sulfato, que estava relacionada à promoção do crescimento da planta e melhoria da tolerância da planta ao sal, além de 227 genes relacionados ao metabolismo do carbono, 202 genes relacionados à síntese de aminoácidos e 117 relacionados ao metabolismo energético. Destaca-se que as informações destes e outros genes relacionados à promoção de crescimento vegetal encontram-se na [Tabela Suplementar S1](#) do estudo realizado por Lu *et al.*

(2022). A tabela conta com um total de 448 genes, dos quais 14 não possuem nome ou alguma outra informação que seja possível identificá-los no arquivo GenBank, restando um total de 434 genes a serem explorados nesta pesquisa.

Por fim, os autores concluem que a análise genômica da cepa *Bacillus subtilis* BS96 mostrou que esta pode promover o crescimento vegetal por meio de mecanismos complexos via quimiotaxia bacteriana, produção de triptofano, metabolismo do enxofre e absorção de nutrientes.

Em Batista (2017), uma tese de doutorado, foi realizado um estudo genômico da cepa *Bacillus sp.* RZ2MS9, isolada originalmente da rizosfera de plantas saudáveis do guaranazeiro amazônico e utilizada em uma pesquisa anterior da mesma autora (BATISTA, 2012). Essa linhagem bacteriana pertence a coleção de cultura do Laboratório de Genética de Microrganismos “Prof. João Lúcio Azevedo”, da Escola Superior de Agricultura “Luiz de Queiroz” - Universidade de São Paulo (ESALQ/USP, Piracicaba - SP), depositada também na Coleção Brasileira de Microrganismos de Ambiente e Indústria (CBMAI), com número de acesso CBMAI1824.

O sequenciamento da estirpe RZ2MS9 foi realizado pelo Centro de Genômica Funcional da ESALQ/USP, Piracicaba – SP, utilizando a plataforma Illumina MiSeq (Illumina Inc., San Diego, CA). Para realizar o controle de qualidade dos dados brutos e a montagem final, utilizou-se o software SPAdes v.3.8.1. Para determinar as sequências codificadoras de proteínas, fez-se uso do NCBI “*Prokaryotic Genomes Automatic Annotation Pipeline*” v.3.3 (PGAAP) e, além disso, as características gerais do genoma, tais como rRNA, tRNA, CDS, entre outras, foram retiradas da anotação disponibilizada pelo GenBank.

Com a anotação finalizada, investigou-se os genes anotados a fim de identificar os possíveis candidatos a estarem envolvidos em funções de promoção de crescimento vegetal, utilizando genes já descritos prévia e comprovadamente na literatura, relacionados aos mecanismos de produção de AIA, quimiotaxia, motilidade, formação de biofilme, produção de compostos orgânicos voláteis, fixação biológica de nitrogênio, solubilização de fosfato, produção de sideróforos, entre outras atividades relevantes para as plantas.

A autora relata ainda, que a partir do genoma parcial da *Bacillus sp.* RZ2MS9, identificou-se vários genes relacionados às algumas das habilidades já comprovadas *in vitro* em Batista (2012), como a produção de AIA, fixação biológica de nitrogênio, solubilização de fosfato e produção de sideróforos, além de habilidades comprovadas *in vitro* no presente estudo, como a quimiotaxia. Adicionalmente, também foram identificadas outras habilidades não avaliadas *in vitro*, mas que podem estar relacionadas à promoção de crescimento de plantas,

segundo estudos da literatura.

A relação dos genes associados à promoção de crescimento em plantas, identificados na rizobactéria *Bacillus sp.* RZ2MS9, juntamente com seus respectivos códigos de acesso GenBank ou Swiss-Prot, encontra-se no Apêndice E, devido à grande quantidade de informações.

Acerca das discussões realizadas nesta pesquisa, foi analisado o efeito da inoculação do *Bacillus sp.* RZ2MS9 no desenvolvimento e produtividade do milho e da soja, além do potencial benéfico na promoção de crescimento destas cultivares, visto que esse efeito já havia sido analisado em condições de casa de vegetação em Batista (2012), onde verificou-se que as plantas inoculadas com essa bactéria tiveram incrementos expressivos na altura, peso seco da parte aérea e peso seco do sistema radicular, tanto em milho como em soja.

Por fim, a autora conclui que a linhagem bacteriana RZ2MS9 apresentou efeitos positivos sobre o desenvolvimento das plantas inoculadas, demonstrando um grande potencial em ser um agente de biocontrole. Destaca-se que a cepa da *Bacillus sp.* RZ2MS9 está depositada no GenBank, sob o código de acesso [CP049978.1](#).

Em Chandra *et al.* (2021) é realizada uma análise do genoma das cepas de duas bactérias promotoras de crescimento de plantas (*B. subtilis* BS87 e *B. megaterium* BM89). Destaca-se que o banco de dados da sequência do genoma das rizobactérias *B. subtilis* BS87 e *B. megaterium* BM89 foram depositados no módulo SRA (do inglês “*Sequence Read Archive*” – Arquivo de Leitura de Sequência) e podem ser acessadas através dos números de acesso SRR11870891 (PRJNA634628) para BS87 e SRR8887017 (PRJNA531071) para BM89.

O interesse em utilizar as rizobactérias do gênero *Bacillus* neste trabalho, se deu por um trabalho anterior dos mesmos autores (CHANDRA *et al.*, 2018), no qual constataram um potencial em promoção de plantas nessas espécies de bactérias entre outras que estavam presentes no estudo. Com isso, duas cepas se destacaram para essa nova análise e foram observadas as suas atividades de promoção de crescimento de plantas, no cultivo de cana-de-açúcar.

Os autores realizaram o sequenciamento de ambas as cepas utilizando a plataforma *Illumina HiSeq* 2500 e montaram o genoma de ambos os alvos. Utilizando o programa RAST, observaram os genes de codificação de proteínas e utilizaram o BLAST para comparar com o banco de dados do NCBI, de modo que as comparações foram positivas e retornaram um alto percentual de identidade.

Com o auxílio do “Gene Ontology” (GO), os genes encontrados na montagem foram revelados em ambas as cepas e suas informações de função molecular, componente celular e

processos biológicos foram atribuídas. Por meio dessas informações, conseguiram criar uma tabela informando a existência ou ausência de determinados genes em ambos os alvos, que estão relacionados à promoção de crescimento em plantas. Uma adaptação da tabela de genes das cepas *B. subtilis* BS87 e *B. megaterium* BM89, relacionados à promoção de crescimento em plantas, encontrada em Chandra *et al.* (2021), pode ser vista no Quadro 5, onde os sinais (“+”) e (“-“) representam, respectivamente, a existência ou ausência do gene em cada uma das cepas.

**Quadro 5 – Genes associados à PGP em *B. subtilis* BS87 e *B. megaterium* BM89**

Gene	Organismo		Gene	Organismo	
	<i>B. subtilis</i> (BS87)	<i>B. megaterium</i> (BM89)		<i>B. subtilis</i> (BS87)	<i>B. megaterium</i> (BM89)
<i>ycdF</i>	+	-	<i>ytnP</i>	+	-
<i>gdh</i>	+	-	<i>hsp15</i>	+	+
<i>zwf</i>	+	+	<i>GroEL</i>	+	+
<i>tuaD</i>	+	+	<i>GroES</i>	+	+
<i>ywqF</i>	+	+	<i>CspS family</i>	+	+
<i>yvhD</i>	+	-	<i>OpuA</i>	+	+
<i>chi113</i>	+	+	<i>OpuB</i>	+	+
<i>bmq 3912</i>	-	+	<i>OpuC</i>	+	+
<i>yqgL</i>	+	+	<i>OpuD</i>	+	+
<i>yzmL</i>	+	+	<i>Bcp</i>	+	-
<i>dhbE</i>	+	+	<i>BsaA</i>	+	+
<i>dhbB</i>	+	+	<i>YgaF</i>	+	+
<i>dhbF</i>	+	+	<i>YdbD</i>	+	-
<i>phzC</i>	+	-	<i>CotJC</i>	+	-
<i>phzF</i>	+	-	<i>KatE</i>	+	+
<i>phzO</i>	+	-	<i>NprB</i>	-	+
<i>treP</i>	+	-	<i>SodF</i>	+	+
<i>treA</i>	+	-	<i>SodS</i>	-	+
<i>treR</i>	+	-	<i>GabP</i>	-	+

**Fonte: Adaptado de Chandra *et al.* (2021)**

Por fim, após analisarem a presença de certos genes e vias metabólicas, os autores concluem que a cepa *B. subtilis* BS87 possui uma melhor capacidade de desempenho no quesito promoção de crescimento em plantas do que a cepa *B. megaterium* BM89.

Destaca-se que as cepas *B. subtilis* BS87 e *B. megaterium* BM89 não serão utilizadas neste estudo, uma vez que os dados das sequências genômicas de ambas foram depositadas como

Arquivo de Leitura de Sequência.

Para o desenvolvimento das próximas etapas desta dissertação, além dos estudos apresentados, outros genomas presentes na literatura, classificados como PGPRs, foram buscados e selecionados, o que será melhor explicado no capítulo posterior (Capítulo 4, seção 4.2), sendo eles: *Bacillus velezensis* CLA178 (CHEN *et al.*, 2020), *Bacillus velezensis* UCMB5113 (ELDAIM *et al.*, 2019), *Bacillus velezensis* SQR9 (HUANG *et al.*, 2022; ZHANG *et al.*, 2015), *Bacillus velezensis* SC60 (DONG *et al.*, 2022), *Bacillus velezensis* UCMB5140 (REVA *et al.*, 2020), *Bacillus velezensis* S141 (SIBPONKRUNG *et al.*, 2017), *Bacillus velezensis* WRN014 (WANG *et al.*, 2019), *Bacillus velezensis* YYC (YAN *et al.*, 2022), *Bacillus velezensis* TH-1 (HE *et al.*, 2023), *Bacillus velezensis* QST713 (PANDIN *et al.*, 2018), *Bacillus velezensis* NJN-6 (YUAN *et al.*, 2013), *Bacillus subtilis* UD1022 (ROSIER *et al.*, 2021) e *Bacillus pumilus* WP8 (KANG *et al.*, 2013).

## 4 MATERIAIS E MÉTODOS

Neste capítulo, na seção 4.1, realiza-se uma breve explicação à respeito dos dados utilizados. Na seção subsequente (seção 4.2), apresenta-se uma proposta de criação/construção de um banco de dados de genes relacionados à promoção de crescimento em plantas, identificados em cepas de rizobactérias promotoras de crescimento em plantas, a partir de informações encontradas na literatura. Na seção 4.3 explica-se os procedimentos adotados para realizar o cálculo ANI, construir a árvore filogenética, encontrar agrupamentos de genes biossintéticos e fazer a representação circular de genomas. Por fim, na seção 4.4, é apresentado o pipeline proposto.

### 4.1 DADOS UTILIZADOS

Para um estudo inicial, foram utilizadas as cepas das bactérias *Bacillus velezensis* LABIM22, *Bacillus velezensis* LABIM40 e *Bacillus velezensis* LABIM44, originalmente isoladas de um solo localizado na Universidade Estadual de Londrina (UEL), Paraná, Brasil, sequenciadas, montadas e anotadas por integrantes do Laboratório de Biotecnologia Microbiana (LABIM) da UEL, cujos códigos de acesso do GenBank são [CP045993.1](#), [CP023748.1](#) e [CP079719.1](#), respectivamente. Além disso, utilizou-se a cepa da rizobactéria *Bacillus velezensis* HNA3 ([CP040881.1](#)), descrita na seção 3.3, para buscar os genes PGPs.

Posteriormente, novas cepas de PGPRs foram adicionadas a esta pesquisa, são estas: *Bacillus velezensis* Ag75, *Bacillus velezensis* LDO2, *Bacillus velezensis* FZB42, *Bacillus subtilis* BS96 e *Bacillus sp.* RZ2MS9, com os respectivos códigos de acesso GenBank: [CP099465.1](#), [CP029034.1](#), [CP000560.2](#), [CP090125.1](#) e [CP049978.1](#), todas brevemente descritas na seção 3.3.

Também foram utilizados outros 13 genomas de bactérias do gênero *Bacillus*, classificados na literatura como PGPRs (seção 3.3), sendo estas: *Bacillus velezensis* CLA178 ([CP061087.1](#)), *Bacillus velezensis* UCMB5113 ([NC\\_022081.1](#)), *Bacillus velezensis* SQR9 ([CP006890.1](#)), *Bacillus velezensis* SC60 ([CP072311.1](#)), *Bacillus velezensis* UCMB5140 ([CP051463.1](#)), *Bacillus velezensis* S141 ([AP018402.1](#)), *Bacillus velezensis* WRN014 ([CP041361.1](#)), *Bacillus velezensis* YYC ([CP075055.1](#)), *Bacillus velezensis* TH-1 ([CP103770.1](#)), *Bacillus velezensis* QST713 ([CP025079.1](#)), *Bacillus velezensis* NJN-6, anteriormente classificada como *Bacillus amyloliquefaciens* ([CP007165.1](#)), *Bacillus pumilus* WP8

(CP010075.1) e *Bacillus subtilis* UD1022 (CP011534.1).

Ainda, utilizou-se outros oito genomas de diferentes espécies de *Bacillus*, para auxiliar na classificação da cepa *Bacillus sp.* RZ2MS9, sendo estes: *Bacillus cereus* FORC\_047 (GCF\_002220285.1), *Bacillus velezensis* JS25R (GCF\_000769555.1), *Bacillus amyloliquefaciens* GKT04 (GCF\_019396925.1), *Bacillus tequilensis* EA-CB0015 (GCF\_012225885.1), *Bacillus inaquosorum* A65.1 (GCF\_023238325.1), *Bacillus paralicheniformis* Bac84 (GCF\_002993925.1), *Bacillus haynesii* P19 (GCF\_014042035.1) e *Bacillus xiamenensis* VV3 (GCF\_001857925.1).

## 4.2 PROPOSTA DE BANCO DE DADOS

O produto final desta pesquisa é a proposta de criação de um banco de dados composto por genes envolvidos na promoção de crescimento em plantas, encontrados nas etapas anteriores.

Como o objetivo da criação desta base de dados é corroborar com pesquisas posteriores, como treinar um algoritmo de Aprendizado de Máquina capaz de classificar a bactéria de um referido genoma como uma possível PGPR, por exemplo, levou-se em consideração a necessidade da diversidade de amostras de genes que irão compor este banco de dados.

Nesse sentido, além das cepas *Bacillus velezensis* HNA3, *Bacillus velezensis* Ag75, *Bacillus velezensis* LDO2 e *Bacillus velezensis* FZB42, cepas de outras espécies de *Bacillus* também serão acrescentadas ao banco de dados: as cepas *Bacillus subtilis* BS96 e *Bacillus sp.* RZ2MS9, que são duas cepas encontradas, de espécies diferentes, cujos autores descrevem os genes em seus estudos.

Além disso, um estudo adicional será realizado, a fim de escolher alguns genomas da literatura, classificados como PGPRs, cujos autores não descreveram os genes promotores de crescimento vegetal, como ocorrido nos genomas estudados no Capítulo 3, seção 3.3.

Com os genomas PGPR selecionados, será realizado um alinhamento BLASTp, utilizando como arquivo de consulta (*query*), os genes PGP dos genomas das cepas *Bacillus velezensis* HNA3, Ag75, LDO2 e FZB42, além das cepas *Bacillus subtilis* BS96 e *Bacillus sp.* RZ2MS9. A finalidade desta etapa será verificar a quantidade de genes PGPs encontrados nos genomas PGPRs selecionados. Assim, caso uma quantidade significativa destes genes sejam encontrados em tais genomas, será possível inferir sobre a possibilidade de outros genomas, que contenham estes mesmos genes, também serem PGPRs.

Destaca-se que o objetivo é ter uma grande quantidade de genes PGPs descritos em

estudos da literatura, no entanto, são raras as pesquisas que trazem em seu escopo, a descrição e/ou identificação de cada um dos genes detectados no envolvimento de promoção de crescimento vegetal, dificultando o aumento da quantidade de amostras utilizadas para compor o banco de dados proposto. Dessa forma, é importante ressaltar que a alimentação deste banco de dados com novas amostras será contínua.

### 4.3 GENÔMICA COMPARATIVA

Optou-se por separar esta seção em duas subseções: na primeira encontra-se os procedimentos e ferramentas utilizadas para a análise de filogenia e construção da árvore filogenética, além do cálculo ANI para a cepa RZ2MS9. Já na segunda, explica-se as ferramentas utilizadas para encontrar os agrupamentos de genes biossintéticos e representar circularmente os genomas das cepas estudadas, de forma que exiba os agrupamentos BGCs e os genes envolvidos na promoção de crescimento vegetal de cada genoma.

#### 4.3.1 Árvore Filogenética

Primeiramente, utilizando o *software* OrthoANI, em sua versão GUI (OAT), realizou-se o cálculo ANI do genoma da bactéria *Bacillus sp.* RZ2MS9, juntamente com outros nove genomas, todos *Bacillus velezensis*: HNA3, Ag75, LDO2, FZB42, LABIM22, LABIM40, LABIM44, SQR9 e S141. Depois, este procedimento foi realizado novamente, utilizando um novo grupo, composto por oito genomas, de diferentes espécies de *Bacillus*: *B. cereus* FORC\_047, *B. velezensis* JS25R, *B. amyloliquefaciens* GKT04, *B. tequilensis* EA-CB0015, *B. inaquosorum* A65.1, *B. paralicheniformis* Bac84, *B. haynesii* P19 e *B. xiamenensis* VV3. Tais procedimentos foram realizados devido ao fato da espécie da estirpe RZ2MS9 não ter sido identificada no estudo realizado por Batista (2017), tampouco no NCBI, e esta informação é relevante para a análise do banco de dados, visto que a classificação da mesma pode ter influência nos resultados das comparações realizadas.

Posterior a isso, usando o *software* Gegenees, foram feitas comparações de todo o genoma da bactéria *Bacillus sp.* RZ2MS9 com as bactérias utilizadas no cálculo ANI. Os dados do Gegenees foram exportados para o SplitsTree para confecção da árvore usando o método da ligação média entre grupos (UPGMA, do inglês “*Unweighted Pair-Group Method using Arithmetic Averages*”).

Destaca-se que o grupo de oito genomas, de diferentes espécies, foi utilizado apenas para identificar a qual espécie a cepa *Bacillus sp.* RZ2MS9 pertence, não sendo utilizadas em nenhuma outra etapa da pesquisa, uma vez que não foi realizado um estudo da literatura para verificar se tais cepas possuem potencial promotor de crescimento vegetal.

#### 4.3.2 Predição de Metabólitos Secundários com antiSMASH

O servidor *web* antiSMASH 7.0.0 foi utilizado para prever os agrupamentos de genes biossintéticos (BGCs), bem como a posição e a função destes nos genomas das cepas *Bacillus velezensis* LABIM22, *Bacillus velezensis* LABIM40, *Bacillus velezensis* LABIM44, *Bacillus velezensis* HNA3, *Bacillus velezensis* Ag75, *Bacillus velezensis* LDO2 e *Bacillus velezensis* FZB42. As análises foram realizadas utilizando o número de acesso GenBank, do NCBI, como *input*.

Para cada uma das sete cepas, o respectivo genoma foi representado circularmente, identificando os agrupamentos de genes biossintéticos e os genes relacionados à promoção de crescimento em plantas. Para as quatro cepas PGPRs utilizadas, os genes PGPs identificados no genoma circular foram os descritos em seus respectivos estudos. Já para as cepas LABIMs, buscou-se por todos os genes descritos nos estudos das quatro cepas classificadas como PGPRs, e representou aqueles que foram encontrados.

Destaca-se que, para criar o arquivo com as informações das posições dos genes envolvidos na promoção de crescimento em plantas, para que pudessem ser utilizados na ferramenta BRIG como marcações personalizadas, criou-se um outro script (“localizacao\_genes.py”), disponível no repositório do [GitHub](#).

Ainda, realizou-se a análise antiSMASH dos outros 10 genomas da espécie *Bacillus velezensis*, a fim de comparar os agrupamentos BGCs de todas as 17 cepas dessa espécie, além da *Bacillus sp.* RZ2MS9 e a *Bacillus velezensis* BS96. As cepas *Bacillus velezensis* NJN-6, *Bacillus subtilis* UD1022 e *Bacillus pumilus* WP8 não foram analisadas nesta etapa porque foram adicionadas a esta pesquisa posteriormente, na etapa de análise do banco de dados.

Além disso, o genoma da cepa *Bacillus velezensis* LABIM40 foi representado circularmente e comparado com outros 17 genomas utilizados na pesquisa: *Bacillus velezensis* LABIM22, *Bacillus velezensis* LABIM44, *Bacillus velezensis* HNA3, *Bacillus velezensis* Ag75, *Bacillus velezensis* LDO2, *Bacillus velezensis* FZB42, *Bacillus velezensis* CLA178, *Bacillus velezensis* UCMB5113, *Bacillus velezensis* SQR9, *Bacillus velezensis* SC60, *Bacillus*

*velezensis* UCMB5140, *Bacillus velezensis* S141, *Bacillus velezensis* WRN014, *Bacillus velezensis* YYC, *Bacillus velezensis* TH-1, *Bacillus velezensis* QST713 e *Bacillus sp.* RZ2MS9, usando a ferramenta BRIG, descrita na subseção 2.2.7. Destaca-se que a cepa *Bacillus subtilis* BS96 não foi incluída nesta etapa, pois não havia sido encontrada e estudada até o momento desta análise.

#### 4.4 PIPELINE PROPOSTO

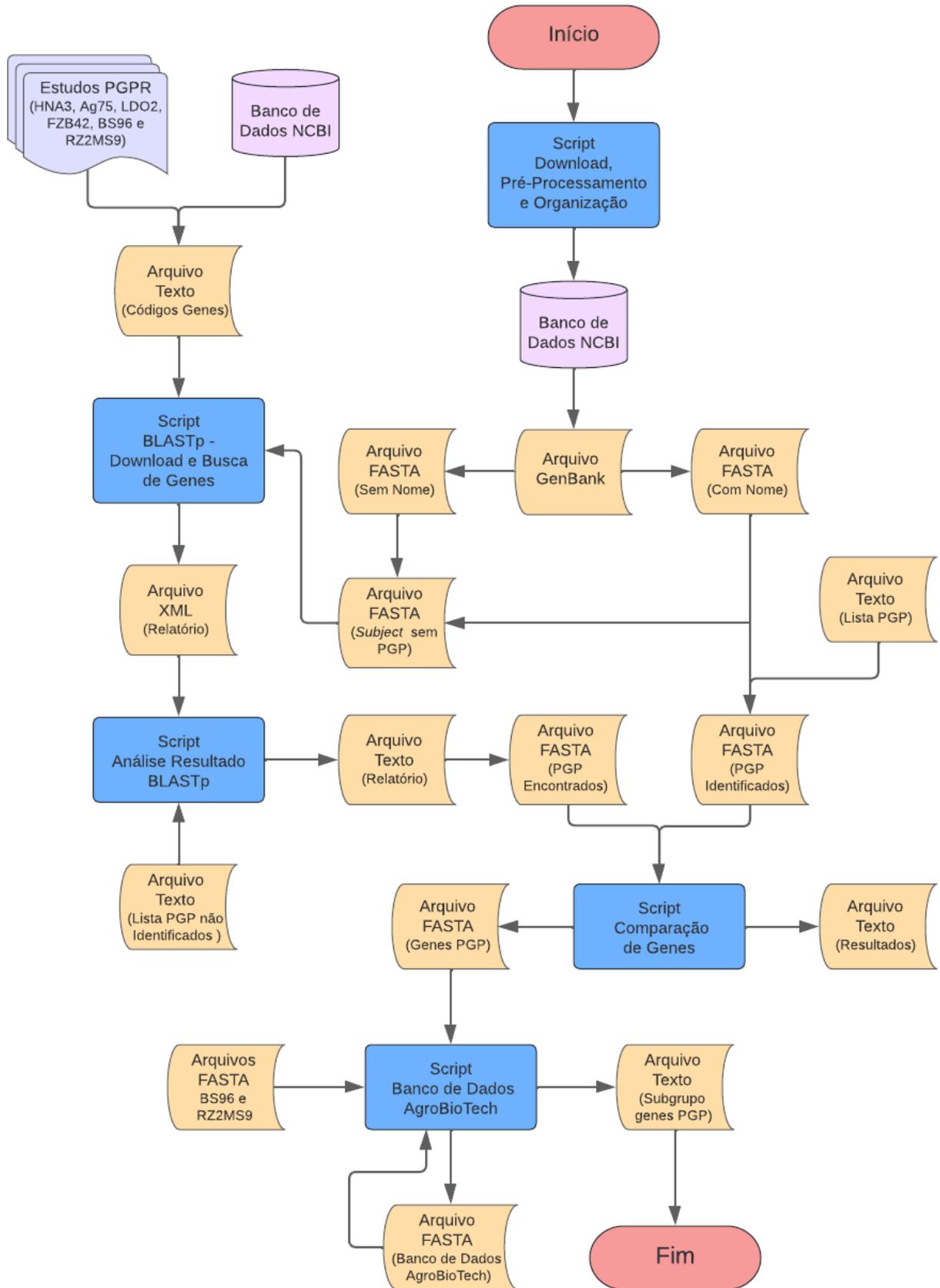
Um dos objetivos deste pipeline é utilizar genomas anotados e depositados no Genbank, hospedado no NCBI, de bactérias que estão relacionadas à promoção de crescimento de plantas, para auxiliar na análise *in silico* do genoma das cepas *Bacillus velezensis* LABIMs.

Com o auxílio de estudos científicos sobre outros genomas de bactérias classificadas como PGPRs, encontrados na literatura, o pipeline proposto buscará comparar os genes em comum entre esses genomas e as cepas LABIMs. Essa comparação é feita com o auxílio da biblioteca Biopython, que oferece diversas funções para lidar com sequências biológicas, em scripts na linguagem Python.

Ademais, o objetivo final, conforme mencionado na seção 4.2, é criar um banco de dados de genes envolvidos na promoção de crescimento em plantas. Assim, partindo das análises dos estudos realizados anteriormente, o pipeline proposto unirá os genes das cepas PGPRs, envolvidos na promoção de crescimento vegetal, e buscará esses genes, via alinhamento BLASTp, em outros genomas classificados pela literatura como PGPRs, cujos genes não foram descritos em seus respectivos estudos.

Na Figura 7, observa-se o fluxograma do pipeline proposto neste estudo.

Figura 7 – Pipeline Proposto



Fonte: Autoria Própria

Para facilitar a compreensão dos procedimentos metodológicos desenvolvidos, optou-se por organizá-los em quatro etapas. Dessa forma, cada passo do pipeline e os dados utilizados foram descritos nas subseções a seguir.

#### 4.4.1 Etapa 1 – Download e Complementação de Genes

Nesta primeira etapa, o objetivo é realizar o download dos genomas a serem estudados, e fazer a complementação dos genes, via *Entrez*, das seis cepas PGPRs *Bacillus velezensis* HNA3, Ag75, LDO2 e FZB42, *Bacillus subtilis* BS96, *Bacillus sp.* RZ2MS9, além das cepas *Bacillus velezensis* LABIM22, LABIM40 e LABIM44. É importante ressaltar que a etapa de complementação não foi necessária para a cepa RZ2MS9, uma vez que todos os seus genes foram descritos em Batista (2017) com seus respectivos códigos de acesso GenBank.

Conhecendo os códigos de acesso, define-se o primeiro script Python do pipeline: “Script Download, Pré-Processamento e Organização”, cujo objetivo é realizar o download das sequências genômicas em formato GenBank do banco de dados do NCBI, por meio do módulo do Biopython chamado *Entrez*, extrair as principais informações dos genes contidos nesse arquivo GenBank criado e, ainda, realizar uma recuperação de informações faltantes, também via *Entrez*.

O arquivo salvo em formato GenBank possui diversas informações sobre toda a anotação da cepa em questão, contudo, somente algumas serão relevantes para a próxima etapa do pipeline, entre elas: anotações de CDS, nome do gene, código RefSeq, o produto daquele gene, código *protein\_id* e, claro, a tradução do gene.

O código *protein\_id* é uma sequência de letras e números que identifica o determinado gene daquele genoma dentro do banco de dados do NCBI de forma única, ou seja, aquele código é exclusivo somente daquela anotação do gene. Por sua vez, o código RefSeq é a chave de acesso para o código *protein\_id* do gene de referência do banco de dados do NCBI. Esta referência, segundo Pruitt *et al.* (2002), oferece as seguintes características para justificar seu uso: não-redundância, ligação explícita com sequências de nucleotídeos e proteínas, atualizada para refletir o atual conhecimento dos dados biológicos e da sequência, validação dos dados e consistência do formato, série de acesso distinto, curadoria sempre realizada pela equipe do NCBI e colaboradores, com indicações de revisões registradas.

O nome do gene descrito no arquivo é uma sigla de, normalmente, 4 letras, sendo exemplos: *trpA*, *alsD*, *dhaS*, *msmX*, etc. Esta sigla oferece uma rápida chave de busca para que seja possível encontrá-la dentro de um determinado genoma. Por sua vez, o produto do gene é a

proteína que aquele gene traduz, por exemplo: o gene *trpA* produz “*tryptophan synthase subunit alpha*”. Por fim, a tradução é a sequência de aminoácidos daquele gene.

Ainda com este primeiro script, lê-se o arquivo GenBank e busca-se extrair as seguintes informações de cada gene: nome, descrição, *protein\_id* e RefSeq do gene. Em cada leitura, verifica-se a existência do nome do gene e se esta informação estiver ausente, utiliza-se o código RefSeq para pesquisá-lo em um genoma de referência via *Entrez*.

Todas as informações são salvas em um arquivo FASTA, identificado na Figura 7 como “Arquivo FASTA (Com Nome)”, cujo cabeçalho de cada gene contém as informações buscadas e a linha posterior, contém a sequência de aminoácidos do gene. Deste arquivo, juntamente com o arquivo Texto identificado na mesma figura como “Arquivo Texto (Lista PGP)”, será gerado um outro arquivo FASTA, descrito como “Arquivo FASTA (PGP Identificados)”. Os genes em que não foi possível identificar o nome, seja pelo arquivo GenBank ou via *Entrez* pelo código RefSeq, são salvos separadamente no “Arquivo FASTA (Sem Nome)” (Figura 7).

Além destes, dois outros arquivos são gerados: “Arquivo FASTA (Genes PGP)” e “Arquivo FASTA (Subject sem PGP)”. O primeiro arquivo FASTA, no qual constam informações e sequências genômicas dos genes presentes no arquivo de texto, identificados nos genomas das suas respectivas cepas, é gerado a partir de dois outros arquivos: “Arquivo FASTA (Com Nome)”, gerado por este mesmo script, e “Arquivo Texto (Lista PGP)”, um arquivo de texto criado manualmente com o nome dos genes relacionados à promoção de crescimento em plantas, relatados nos estudos selecionados. Já o segundo arquivo FASTA, identificado na Figura 7 como “Arquivo FASTA (Subject sem PGP)”, contém as informações de todos os genes presentes nos genomas das cepas selecionadas, com exceção dos genes relacionados à PGP já identificados anteriormente. Esse arquivo também é gerado a partir de dois outros arquivos criados por este script: “Arquivo FASTA (Com Nome)” e “Arquivo FASTA (Sem Nome)”, e é necessário para a próxima etapa desta pesquisa, que visa identificar os demais genes descritos nos estudos de cepas PGPRs selecionados, por meio do alinhamento de sequências.

Cabe destacar que os arquivos FASTA são criados separadamente para cada uma das cepas e, para as cepas LABIMs, o arquivo Texto gerado com os nomes dos genes relacionados à PGP, contém os nomes de todos os genes relacionados à PGP das quatro cepas PGPRs *Bacillus velezensis*, uma vez que pretende-se realizar uma comparação, utilizando os nomes dos genes, entre os genomas das cepas PGPRs *Bacillus velezensis* e as referidas cepas. Portanto, nesta etapa do script, será verificado a presença dos genes pertencentes a estes quatro genomas e,

consequentemente, o arquivo FASTA gerado conterá informações de genes das quatro cepas.

Dessa forma, o interesse desta próxima etapa é buscar as informações dos genes relacionados à PGP não identificados pelo script anterior. Para isso, realizou-se uma busca manual nos bancos de dados do NCBI de todos os genes pertencentes às cepas estudadas que não tiveram seus nomes identificados via Entrez e criou-se, manualmente, um arquivo Texto com os códigos destas sequências genômicas, para cada uma das quatro cepas (HNA3, Ag75, LDO2 e FZB42), identificado na Figura 7 como “Arquivo Texto (Códigos Genes)”.

#### 4.4.2 Etapa 2 – Alinhamento BLASTp e Análise de Resultados

Nesta segunda etapa, seguiu-se com a complementação dos genes, porém, o objetivo desta vez foi identificar, especificamente, os genes relacionados à promoção de crescimento em plantas, não identificados via *Entrez* nas cepas PGPRs selecionadas para estudo. Para as cepas LABIMs, esta etapa foi empregada visando identificar a presença dos demais genes relacionados à PGP das quatro estirpes *Bacillus velezensis*, possibilitando uma comparação por nome de gene, que será realizada na etapa seguinte.

Destaca-se que para a cepa *Bacillus subtilis* BS96, esta etapa de identificação dos genes, via alinhamento de sequências, foi realizado de forma diferente, devido ao fato de haver muitos genes descritos pelos autores com nomes iguais, mas isso será explicado com detalhes na subseção subseção 4.4.4.

Assim, o segundo script Python deste pipeline (“Script BLASTp - Download e Busca de Genes”) utiliza o arquivo Texto identificado na Figura 7 como “Arquivo Texto (Códigos Genes)”, para realizar o download das sequências de cada cepa, que serão salvas em um arquivo FASTA. Destaca-se que, para cada um dos sete genomas (HNA3, Ag75, LDO2, FZB42 e as três LABIMs), será gerado um arquivo FASTA separado, que será utilizado como arquivo de consulta (*query*) no alinhamento BLASTp com as sequências de aminoácidos dos genes presentes no “Arquivo FASTA (Subject sem PGP)”. Como saída, este script fornece um arquivo no formato XML, identificado na Figura 7 como “Arquivo XML (Relatório)”.

Já o terceiro script Python, identificado na Figura 7 como “Script Análise Resultado BLASTp”, analisa os resultados obtidos dos alinhamentos, a partir do arquivo gerado pelo script anterior, e gera dois novos arquivos: o primeiro é a conversão do relatório BLAST em formato Texto, visto que o formato XML é ótimo para análises em scripts, mas não amigável para leitura do usuário, enquanto o formato Texto pode ser lido e analisado facilmente. Já o segundo arquivo,

identificado como “Arquivo FASTA (PGP Encontrados)”, é uma compilação das sequências “*matches*” do relatório do alinhamento, em um arquivo multi-FASTA, no qual é escrito o nome do gene no início do cabeçalho, cujo motivo será explicado em uma etapa posterior, utilizando como guia um arquivo no formato Texto com os nomes dos genes que foram buscados no Alinhamento BLAST, identificado como “Arquivo Texto (Lista PGP não Identificados)”.

Com os alinhamentos BLAST realizados, espera-se que todos os genes relacionados à PGP, relatados nas pesquisas de Zaid *et al.* (2022), Mosela *et al.* (2022), Chen *et al.* (2019), Chen *et al.* (2007) e Lu *et al.* (2022), sejam identificados em seus respectivos genomas. Além disso, como todas as estirpes envolvidas nesta etapa são da espécie *Bacillus velezensis*, almeja-se que a maioria dos genes relacionados à PGP, das quatro cepas PGPRs (HNA3, Ag75, LDO2, FZB42), sejam encontrados nas cepas LABIMs.

#### 4.4.3 Etapa 3 – Comparação dos Genomas

Nesta etapa, o objetivo é realizar as comparações entre as sete cepas de *Bacillus velezensis* por meio do quarto script Python, denominado “Script Comparação de Genes”. Primeiramente, este script concatena os arquivos gerados pelo BLAST (“Arquivo FASTA (PGP Encontrados)”) e o arquivo gerado pelo primeiro script (“Arquivo FASTA (PGP Identificados)”), salvando-os em um arquivo multi-FASTA, identificado como “Arquivo FASTA (Genes PGP)”. Com isso, ele realiza a busca dos nomes dos genes dos arquivos FASTA criados nas etapas anteriores, e salva essas informações em uma variável conjunto (*set*) do Python. Esta variável *set* possui duas características particulares: todos seus elementos são únicos e não possuem ordenação. Além disso, como esta variável é uma referência para os conjuntos matemáticos, é possível realizar operações clássicas de conjuntos, como intersecção, união e subtração.

Dessa forma, realizou-se a intersecção dos conjuntos de genes, de cada um dos quatro genomas *Bacillus velezensis*, classificados como PGPRs, com o conjunto de genes de cada cepa LABIM, a fim de analisar a presença destes genes em todas as cepas LABIMs estudadas: LABIM22, LABIM40 e LAMBIM44. Ademais, também foi feita a comparação entre as cepas PGPRs (HNA3, Ag75, LDO2 e FZB42), para verificar se as mesmas compartilham genes relacionados à PGP entre si, uma vez que são da mesma espécie e, por fim, cria o arquivo identificado na Figura 7 como “Arquivo Texto (Resultados)”.

Destaca-se que, inicialmente, para as três etapas descritas até aqui (subseção 4.4.1, subseção 4.4.2 e subseção 4.4.3), escolheu-se apenas um genoma da literatura, classificado como

PGPR: *Bacillus velezensis* HNA3, isso porque era necessário estudar a estrutura do arquivo GenBank, para que fosse possível pensar na estrutura dos scripts de forma a atender os objetivos da pesquisa. Dessa forma, a partir deste único genoma, juntamente com as três cepas de estudo (LABIMs), foi possível desenvolver os scripts em suas primeiras versões. Com estes criados e os testes iniciais realizados, adequações necessárias foram feitas e, posteriormente, as outras cepas PGPRs foram adicionadas. Assim, novos estudos foram realizados e, sempre que necessário, alterações nos scripts foram feitas, até chegar em sua versão final.

Além disso, é importante enfatizar que as cepas *Bacillus sp.* RZ2MS9 e *Bacillus subtilis* BS96 também não foram utilizadas nessa etapa de comparação devido ao não pertencimento à espécie *Bacillus velezensis*. No entanto, um estudo posterior, que compare todas as cepas selecionadas para estudo, foi realizado.

#### **4.4.4 Etapa 4 – Elaboração do Banco de Dados**

Para atender o objetivo desta pesquisa, que (conforme mencionado na seção 4.2) é gerar como produto final um banco de dados de genes relacionados à promoção de crescimento em plantas, é necessário que esse banco contenha uma quantidade significativa de amostras de genes relacionados à PGP, de diferentes espécies de bactérias do gênero *Bacillus*.

Assim, o primeiro grupo de genes que irão compor o banco de dados, são das quatro cepas PGPRs *Bacillus velezensis* (HNA3, Ag75, LDO2 e FZB42). Destas, obteve-se um total de 245 genes envolvidos na promoção de crescimento. No entanto, destaca-se que alguns genes estão descritos no estudo de duas ou mais cepas, além dos genes em que aparecem com nomes diferentes, em diferentes cepas, mesmo sendo a mesma sequência. Para contornar isso, um alinhamento BLASTp foi realizado entre as quatro cepas, a fim de eliminar os genes altamente similares, cujo percentual mínimo de identidade utilizado foi de 90%, resultando em um total de 198 genes diferentes.

O segundo grupo de genes que constará no banco de dados são da cepa *Bacillus subtilis* BS96. Para esta cepa, foram descritos um total de 377 genes relacionados à PGP. No entanto, vários destes possuem nomes repetidos e/ou incompletos (tabela Suplementar S1 de Lu *et al.* (2022)), impossibilitando sua identificação no arquivo GenBank e, até mesmo, na pesquisa manual pelo NCBI. Para os genes não identificados com o primeiro script, uma busca nos bancos de dados do NCBI também foi necessária.

No entanto, como vários genes aparecem no arquivo disponibilizado por Lu *et al.* (2022)

com os mesmos nomes e o mesmos produtos, dificultando a busca, além de torná-la demorada e exaustiva, para auxiliar nesta tarefa de complementação, criou-se um script adicional, que realiza a busca pelo nome do gene, nos bancos de dados de proteínas do NCBI, utilizando como filtros o organismo (*Bacillus subtilis*) e o nome do gene por meio da função *Entrez Search* e gera dez sequências de genes. Para cada gene encontrado, o script realiza um alinhamento BLASTp com a sequência *subject* sendo o “Arquivo FASTA (Subject sem PGP)”, gerado pelo primeiro script do pipeline. O relatório BLAST é analisado e o script retorna a sequência da cepa BS96 com maior identidade no alinhamento com as sequências buscadas pelo *Entrez*. Este script gera um relatório dos alinhamentos, no formato XML, o qual é convertido para um arquivo no formato Texto, para melhor visualização dos resultados obtidos, além de um arquivo no formato FASTA, contendo todas estas sequências. O arquivo com as sequências similares encontradas será concatenado ao “Arquivo FASTA (PGP Identificados)”, gerado pelo primeiro script.

Destaca-se que, primeiramente, o script verifica os genes com nomes iguais para excluí-los, uma vez que estes resultam em um mesmo produto, segundo o arquivo disponibilizado em Lu *et al.* (2022), além de obter-se as mesmas sequências, como resultados, ao pesquisar nos bancos de dados do NCBI.

Por fim, o terceiro e último grupo, neste momento, são os genes da cepa *Bacillus sp.* RZ2MS9. Diferente dos estudos realizados por Chen *et al.* (2019), Chen *et al.* (2007) e Lu *et al.* (2022), onde os autores descreveram apenas o nome do gene e o seu produto, e por Mosela *et al.* (2022) e Zaid *et al.* (2022), que apresentaram um código de identificação específico, impossibilitando a busca nos bancos de dados do NCBI, Batista (2017) associa cada gene PGP da cepa *Bacillus sp.* RZ2MS9, ao seu respectivo código de acesso GenBank.

Dito isto, além do download do genoma da cepa RZ2MS9, descrito no início da primeira etapa (subseção 4.4.1), realizou-se os downloads de todos os genes descritos no Apêndice E. Para isso, utilizou-se um script adicional. Ressalta-se que este script não consta na Figura 7, mas pode ser encontrado no repositório online do [GitHub](#), junto com os demais scripts utilizados e todos os resultados obtidos.

Com o download dos genes finalizado, realiza-se um alinhamento BLASTp entre estes e o arquivo FASTA completo do genoma (FASTA *protein*), para confirmar que todos os 274 genes realmente estão presentes no genoma. Para isso, utilizou-se um percentual de identidade igual a 100%, uma vez que os os códigos disponibilizados por Batista (2017) são da própria cepa.

Após isso, como vários destes genes possuíam nomes iguais, um alinhamento BLASTp entre todos os genes dessa cepa foi realizado, a fim de verificar se existia alta similaridade entre eles, já que vários destes foram descritos pela autora contendo nomes iguais. Para isso, considerou-se como limiar de corte, o percentual de identidade igual à 90%.

Após o banco de dados ser criado, o próximo passo é utilizá-lo para verificar a quantidade de genes que estarão presentes em cepas classificadas na literatura como PGPRs, mas sem descrição de genes. Para isso, buscou-se por genomas PGPRs de diferentes espécies de *Bacillus*, cujos autores comprovaram o potencial promotor de crescimento das cepas e, ainda, disponibilizaram o código de acesso GenBank do genoma completo para download no NCBI.

Com a pesquisa realizada, um total de 14 genomas foram encontrados até o momento, sendo 12 deles da espécie *Bacillus velezensis*: *Bacillus velezensis* CMRP 4490 (LABIM22), *Bacillus velezensis* CLA178, *Bacillus velezensis* UCMB5113, *Bacillus velezensis* SQR9, *Bacillus velezensis* SC60, *Bacillus velezensis* UCMB5140, *Bacillus velezensis* S141, *Bacillus velezensis* WRN014, *Bacillus velezensis* YYC, *Bacillus velezensis* TH-1, *Bacillus velezensis* QST713 e *Bacillus velezensis* NJN-6, anteriormente classificado como *Bacillus amyloliquefaciens*; além de outras duas cepas de espécies diferentes: *Bacillus pumilus* WP8 e *Bacillus subtilis* UD1022.

Após a seleção dos genomas PGPRs encontrados na literatura, cria-se o último script do pipeline proposto. Este, tem duas funções importantes: a primeira é unir todos os genes PGPs catalogados até então em um único arquivo multi-FASTA, concatenando os três grupos de genes descritos anteriormente, e a segunda será realizar um alinhamento BLASTp entre o banco de dados criado (*query*) e cada um dos genomas PGPR selecionados.

Como resultado deste alinhamento, é gerado um arquivo no formato XML e um no formato Texto para cada genoma selecionado. Ademais, o script gera um arquivo no formato CSV para registrar quais genes do banco de dados de genes PGPs foram encontrados nos genomas comparados, com percentual de identidade igual ou superior a 90%. A partir deste CSV, outro arquivo Texto é gerado, contendo os genes encontrados em todos os genomas. Ainda, outras análises foram realizadas, utilizando os limiares 80%, 70% e 60%.

O objetivo deste último script é identificar o maior subgrupo de genes PGPs possível, dentro do banco de dados criado, que estejam presentes em todas as cepas selecionadas, de modo que essa presença possa indicar a relevância destes para o banco de dados criado. Com um subgrupo numeroso e resultados satisfatórios, este banco será de grande utilidade para estudos futuros, possibilitando inferir sobre a classificação de novos genomas PGPRs.

Destaca-se, novamente, que o banco de dados será continuamente alimentado a medida que novos estudos de genomas de PGPR, cujos genes envolvidos no processo de promoção de crescimento vegetal sejam identificados pelos autores, sejam publicados.

Como um estudo adicional, realizou-se uma busca por estudos sobre rizobactérias promotoras de crescimento em plantas, publicados até 2023, com o objetivo de analisar o período com a maior concentração de artigos publicados, e mapear os genomas PGPR estudados na literatura durante esse período, para que possam ser analisados em estudos futuros. Essa pesquisa foi realizada na plataforma Google Scholar<sup>1</sup>, no modo pesquisa avançada, utilizando como *features* de pesquisa: *bacillus pgpr* “*Plant Growth-Promoting Rhizobacteria*”.

---

<sup>1</sup> <https://scholar.google.com/>

## 5 RESULTADOS

Nesta pesquisa, foi proposto a elaboração de um pipeline a fim de buscar genes relacionados à promoção de crescimento em plantas, em rizobactérias encontradas na literatura, para a criação de um banco de dados de genes PGPs. Inicialmente utilizou-se a estirpe *Bacillus velezensis* HNA3, relatada em Zaid *et al.* (2022) e, posteriormente, estendeu-se às cepas *Bacillus velezensis* Ag75, LDO2 e FZB42, estudadas em Mosela *et al.* (2022), Chen *et al.* (2019) e Chen *et al.* (2007), respectivamente, para auxiliar no estudo *in silico* das cepas de *Bacillus velezensis* disponibilizadas pelo LABIM-UEL. Além destas, duas outras cepas foram estudadas: *Bacillus subtilis* BS96 e *Bacillus sp.* RZ2MS9, descritas em Lu *et al.* (2022) e Batista (2017), respectivamente, com o objetivo de corroborar com a elaboração de uma base de dados formada por genes de rizobactérias, envolvidos na promoção de crescimento vegetal, formado por genes de cepas de diferentes espécies de bactérias do gênero *Bacillus*. Destaca-se, ainda, que 13 novos genomas foram incluídos para auxiliar nas primeiras análises utilizando o banco de dados construído.

Além do pipeline proposto, realizou-se uma comparação filogenética da cepa RZ2MS9 com outras cepas, a representação circular dos genomas das cepas de estudo (LABIMs) e das quatro cepas PGPRs, indicando os genes envolvidos na promoção de crescimento vegetal, e a predição de metabólitos secundários presentes nas mesmas, também representados no genoma circular. Também realizou-se a predição de agrupamentos BGCs para outras 12 cepas, entre elas a *Bacillus subtilis* BS96 e a *Bacillus sp.* RZ2MS9, além de outras dez *Bacillus velezensis*. Esta análise teve como finalidade comparar os agrupamentos presentes em todas as cepas 19 analisadas.

Para melhor organização dos resultados, este capítulo está dividido em duas seções: na primeira consta a análise dos resultados obtidos na comparação filogenética, agrupamentos de genes biossintéticos das 19 cepas e a representação circular dos genomas gerados pela ferramenta BRIG; na segunda, é analisado os resultados obtidos em cada etapa de desenvolvimento do pipeline.

### 5.1 RESULTADOS E ANÁLISE DA GENÔMICA COMPARATIVA

Na primeira subseção (5.1.1), são mostrados os resultados obtidos pelos softwares OrthoANI (OAT), Gegenees e SplitsTree. Na segunda, os resultados dos softwares antiSMASH e

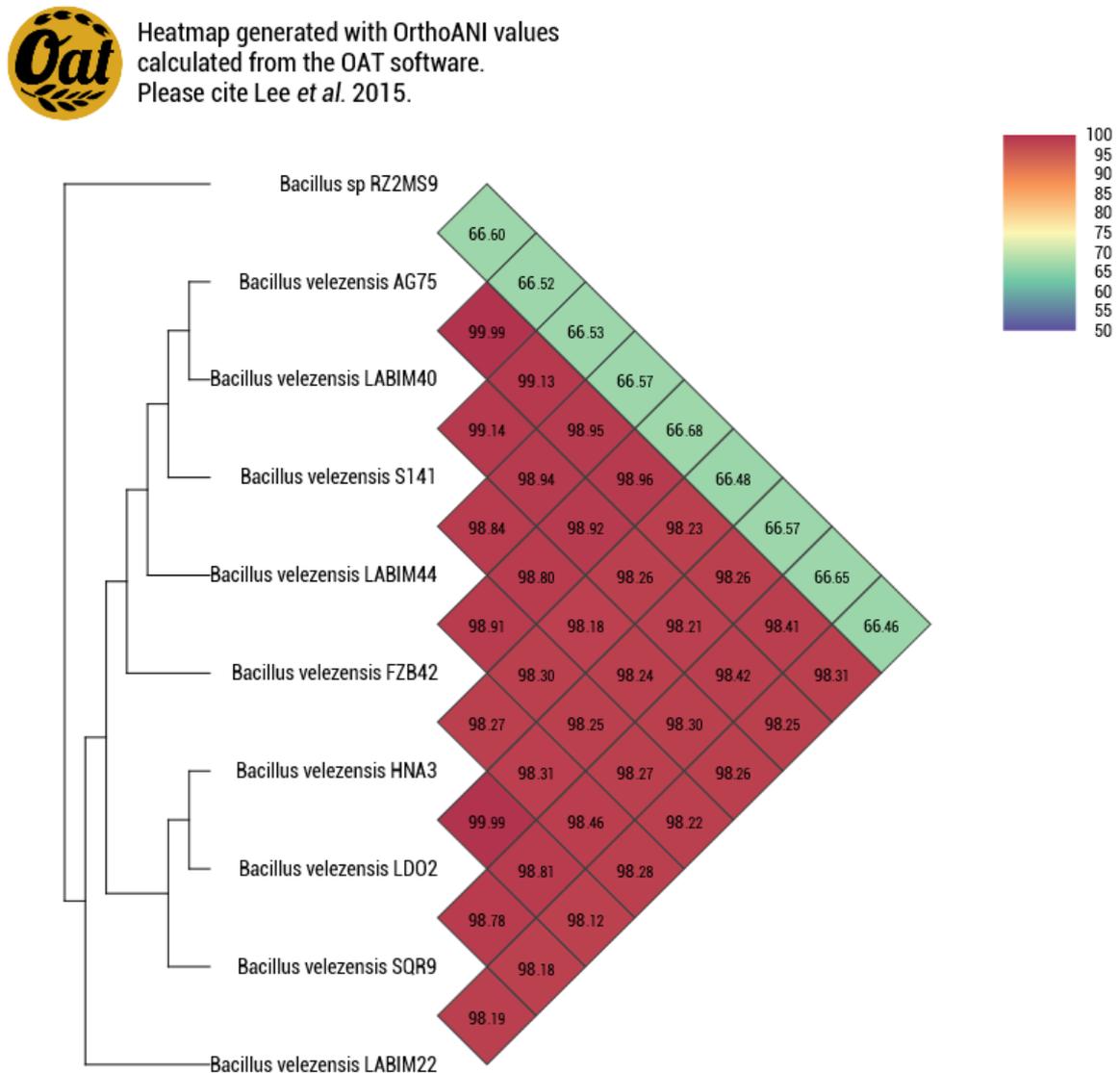
BRIG.

### 5.1.1 Análise e Árvore Filogenética

Primeiramente, utilizou-se a ferramenta OrthoANI (OAT) com a finalidade de identificar a espécie da cepa *Bacillus sp.* RZ2MS9.

Para realizar essa identificação, foram realizadas duas análises: na primeira, conforme Figura 8, foram utilizadas, além da RZ2MS9, bactérias somente da espécie *Bacillus velezensis*.

Figura 8 – Primeira Análise do Cálculo ANI da cepa *Bacillus sp.* RZ2MS9



Fonte: Autoria Própria

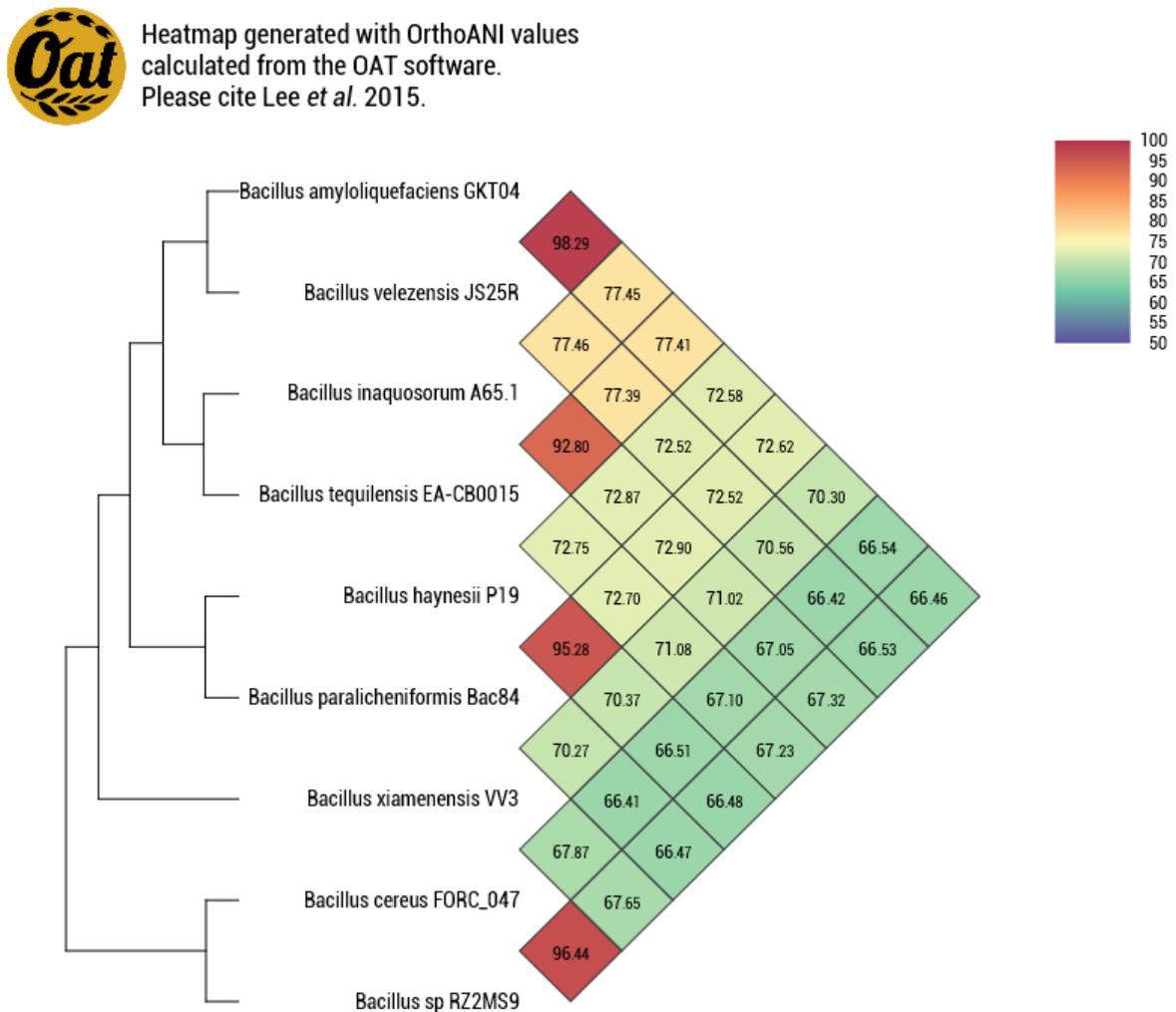
Analisando o Mapa de Calor gerado, conclui-se que a cepa RZ2MS9 realmente não

pertence a espécie *Bacillus velezensis*. Além disso, observa-se também que os resultados entre as cepas da espécie *Bacillus velezensis* foram todos acima do limite para demarcação de espécies, sendo 98,12% o menor valor obtido.

Como não foi possível descobrir a espécie da estirpe RZ2MS9 com a primeira análise realizada, escolheu-se, de forma aleatória, oito cepas de diferentes espécies de *Bacillus* para serem utilizadas na ferramenta OrthoANI, cujo resultado encontra-se na Figura 9.

Destaca-se, novamente, que este grupo dos oito genomas foram utilizados somente nesta etapa de classificação da cepa *Bacillus sp.* RZ2MS9, e não possui nenhuma relação com o banco de dados que será criado.

**Figura 9 – Segunda Análise do Cálculo ANI da cepa *Bacillus sp.* RZ2MS9**



**Fonte: Autoria Própria**

Agora, analisando a Figura 9, observa-se que o melhor valor ANI da cepa RZ2MS9 foi

com a cepa *Bacillus cereus* FORC\_047, cujo valor foi de 96,44%. Assim, dado que vários estudos recomendam um valor de ANI de aproximadamente 95-96% como limite para demarcação de espécies (CHUN; RAINEY, 2014; GORIS *et al.*, 2007; RICHTER; ROSSELLÓ-MÓRA, 2009), conclui-se que a cepa RZ2MS9 pertence a espécie *Bacillus cereus*.

O software Gegenees foi utilizada para comparar a RZ2MS9 com outros 25 genomas: as quatro cepas PGPRs, as três LABIMs, os dez genomas *Bacillus velezensis* PGPRs adicionais e os oito genomas de diferentes espécies de *Bacillus*.

Os resultados obtidos foram exportados para O software SplitsTree, para confecção da árvore filogenética, utilizando o método UPGMA (Figura 10).

**Figura 10 – Árvore Filogenética representando cepas do gênero *Bacillus***



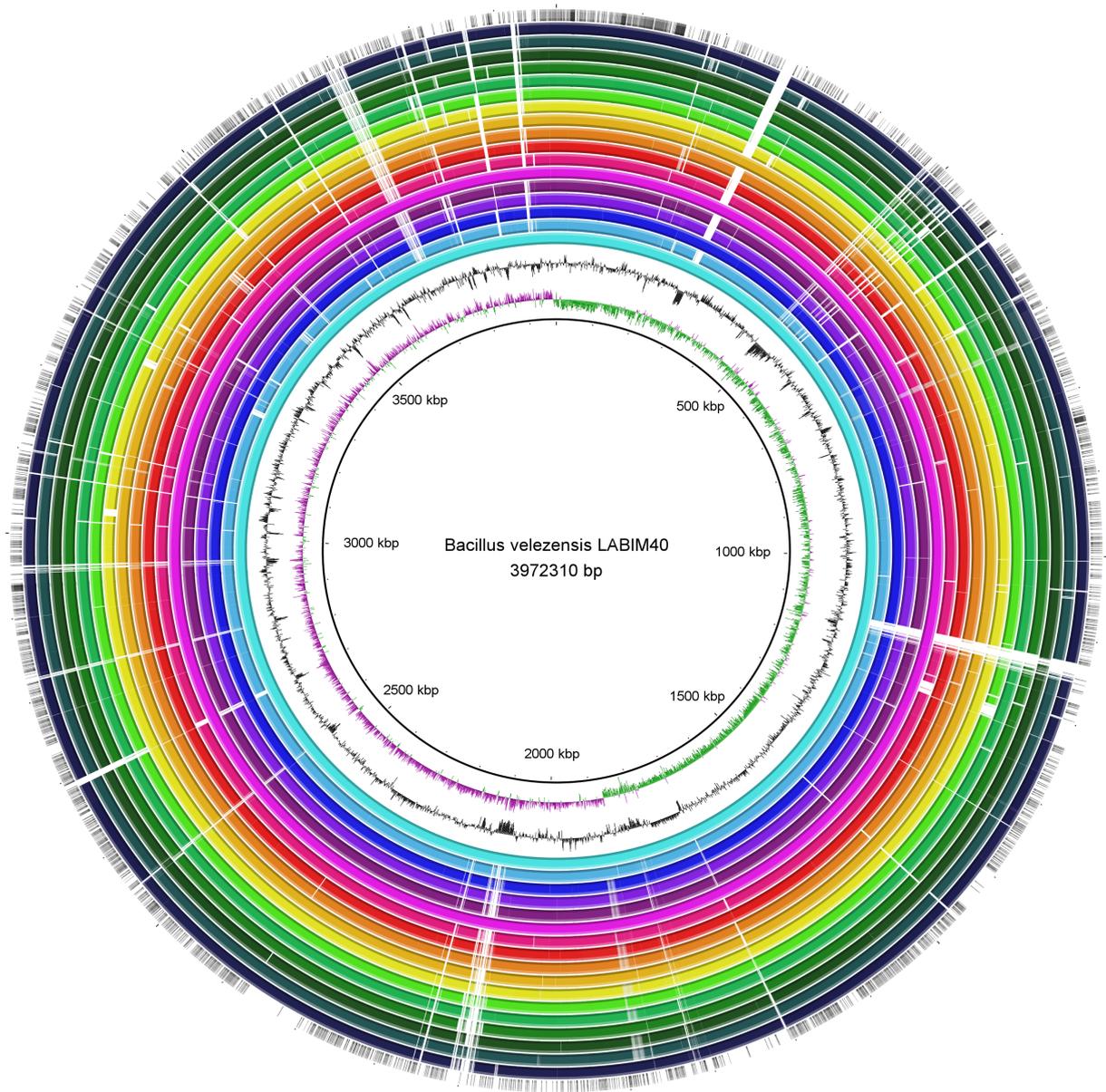
**Fonte: Autoria Própria**

Analisando a Figura 10, a única cepa semelhante a *Bacillus sp.* RZ2MS9 pertence a espécie *Bacillus cereus*, destacada na cor vermelha. Em verde, estão todas as cepas de *Bacillus velezensis* e *Bacillus amyloliquefaciens*. Em azul, *Bacillus tequilenses* e *Bacillus inaquosorum*. Em amarelo, *Bacillus haynesii* e *Bacillus paralicheniformis*. Por fim, sem destaque, a cepa da espécie *Bacillus xiamenensis*.

### 5.1.2 Análise antiSMASH e Representação Circular do Genoma

Primeiramente, realizou-se a representação circular do genoma da cepa *Bacillus velezensis* LABIM40, comparando-o com os outros 17 genomas, utilizando o software BRIG. Essa representação pode ser observada na Figura 11.

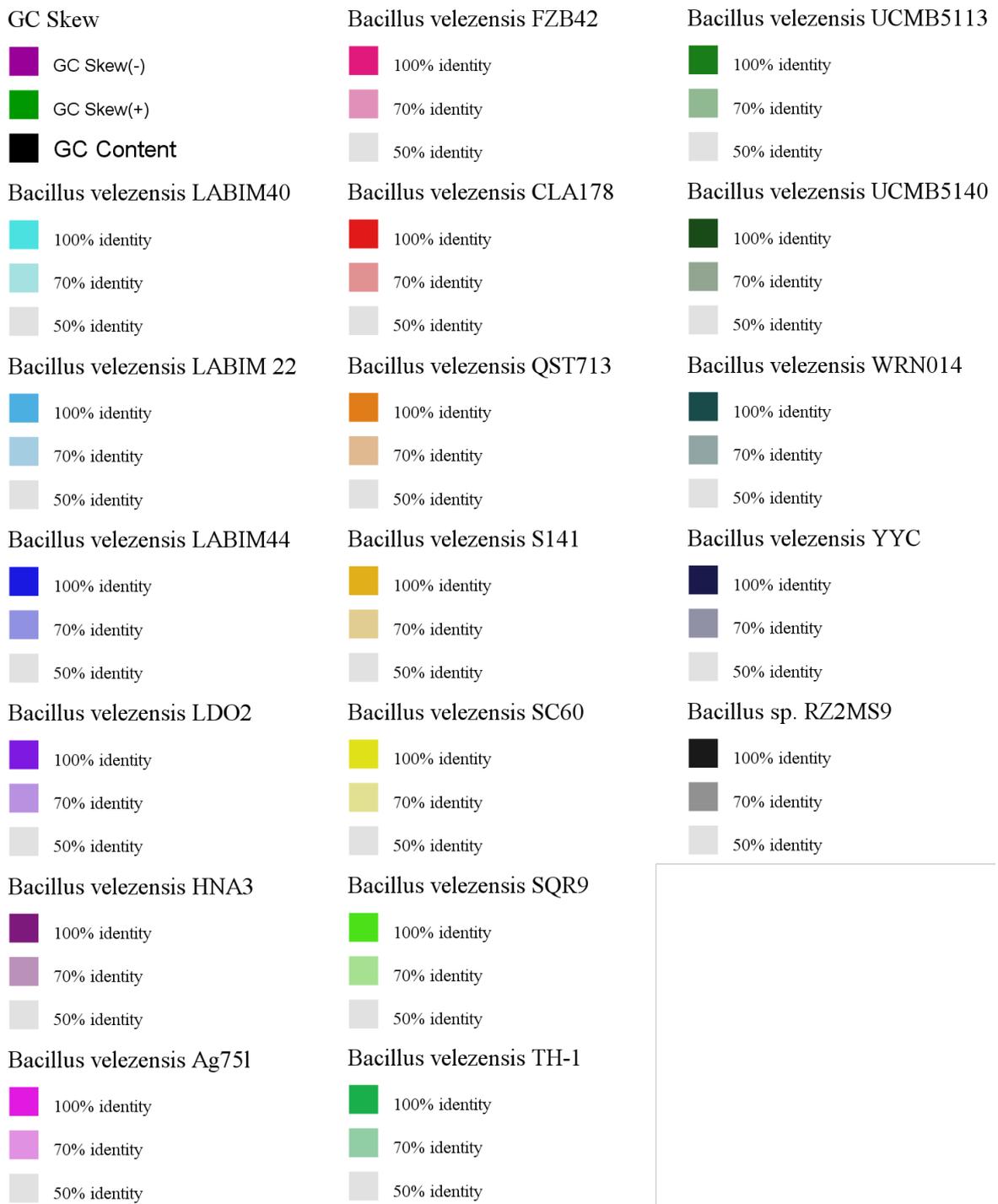
**Figura 11 – Representação Circular do Genoma da Bactéria *Bacillus velezensis* LABIM40**



**Fonte: Autoria Própria**

Destaca-se que, como foram utilizados 18 genomas, a legenda das cepas utilizadas ficou um pouco extensa, sendo necessário organizá-la separadamente, conforme Figura 12.

**Figura 12 – Legenda da Figura 11 - Representação Circular da cepa *Bacillus velezensis* LABIM40**



**Fonte: Autoria Própria**

Observa-se que a cepa RZ2MS9 é a única que difere consideravelmente das outras cepas, em diversas localizações do genoma, o que se deve ao fato de todas as outras cepas serem *Bacillus velezensis* e ela não.

Utilizou-se a ferramenta antiSMASH 7.0.0, em sua versão web, para encontrar os agrupamentos de genes biossintéticos (BGCs) nos genomas das rizobactérias *Bacillus velezensis* LABIM22, LABIM40 e LABIM44.

O servidor web antiSMASH 7.0.0 encontrou 13 BGCs no genoma da rizobactéria *Bacillus velezensis* LABIM22 Tabela 1. Dentre os 13 *clusters*, oito apresentaram similaridade com BGCs existentes no banco de dados. Eles estavam ligados a síntese de *surfactin*, *butirosin A/B*, *macrolactin H*, *bacillaene*, *fengycin*, *difficidin*, *bacillibactin* e *bacilysin*. Cinco dos 13 *clusters* não apresentaram nenhum *cluster* de similaridade no banco de dados.

**Tabela 1 – Agrupamentos de Genes Biossintéticos em *Bacillus velezensis* LABIM22 com antiSMASH 7.0.0**

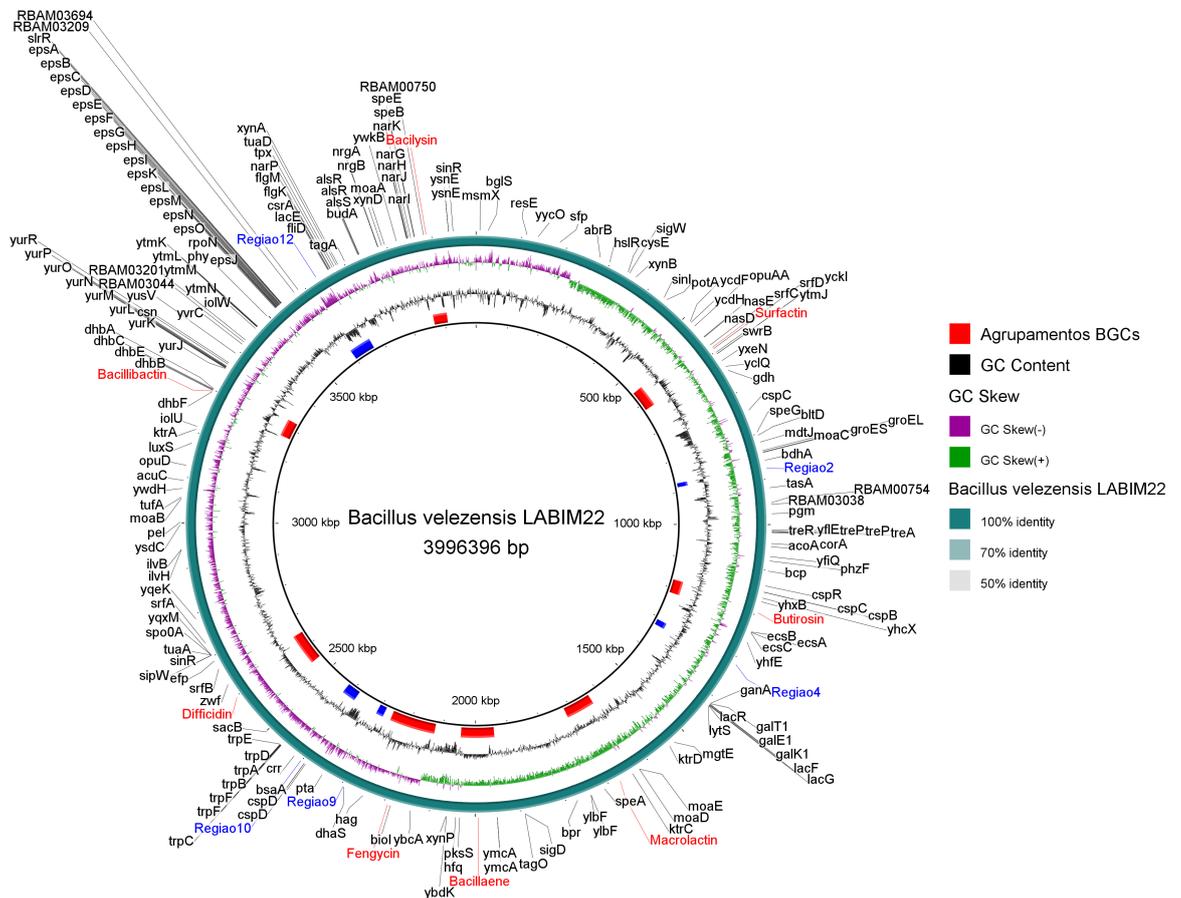
Região	Tipo	De (pb)	Para (pb)	Cluster Semelhante	Conhecido	Similaridade
Região 01	<i>NRPS</i>	556 211	620 549	<i>surfactin</i>	<i>NRP: Lipopeptide</i>	82%
Região 02	<i>phosphonate</i>	870 920	883 067			
Região 03	<i>PKS-like</i>	1 173 624	1 214 868	<i>butirosin A/B</i>	<i>Saccharide</i>	7%
Região 04	<i>terpene</i>	1 307 477	1 324 753			
Região 05	<i>transAT-PKS</i>	1 631 321	1 717 680	<i>macrolactin H</i>	<i>Polyketide</i>	100%
Região 06	<i>transAT-PKS, T3PKS, NRPS</i>	1 943 793	2 043 657	<i>bacillaene</i>	<i>Polyketide+NRP</i>	100%
Região 07	<i>NRPS, transAT-PKS, betalactone</i>	2 123 124	2 259 858	<i>fengycin</i>	<i>NRP</i>	100%
Região 08	<i>terpene</i>	2 283 916	2 305 799			
Região 09	<i>T3PKS</i>	2 381 420	2 422 520			
Região 10	<i>transAT-PKS</i>	2 549 542	2 643 330	<i>difficidin</i>	<i>Polyketide</i>	100%
Região 11	<i>NRP-metallophore, NRPS, RiPP-like</i>	3 270 676	3 222 469	<i>bacillibactin</i>	<i>NRP</i>	100%

Região 12	NRPS	3 599 775	3 665 117			
Região 13	other	3 867 537	3 908 955	<i>bacilysin</i>	Other	100%

Fonte: Autoria Própria

Com estes resultados, utilizou-se novamente o software BRIG para exibir os agrupamentos BGCs encontrados pelo antiSMASH e os genes PGP's identificados pelo pipeline (Figura 13).

Figura 13 – Agrupamentos BGCs e Genes PGP's na cepa *Bacillus velezensis* LABIM22



Fonte: Autoria Própria

Já para a cepa *Bacillus velezensis* LABIM40 e *Bacillus velezensis* LABIM44, o antiSMASH encontrou 12 BGCs (Tabela 2 e Tabela 3, respectivamente), dos quais oito apresentaram similaridade com o banco de dados de BGCs existentes. Destaca-se que os oito agrupamentos que apresentaram similaridade, assim como na LABIM22, são ligados a síntese de *surfactin*, *butirosin A/B*, *macrolactin H*, *bacillaene*, *fengycin*, *difficidin*, *bacillibactin* e *bacilysin*. Quatro

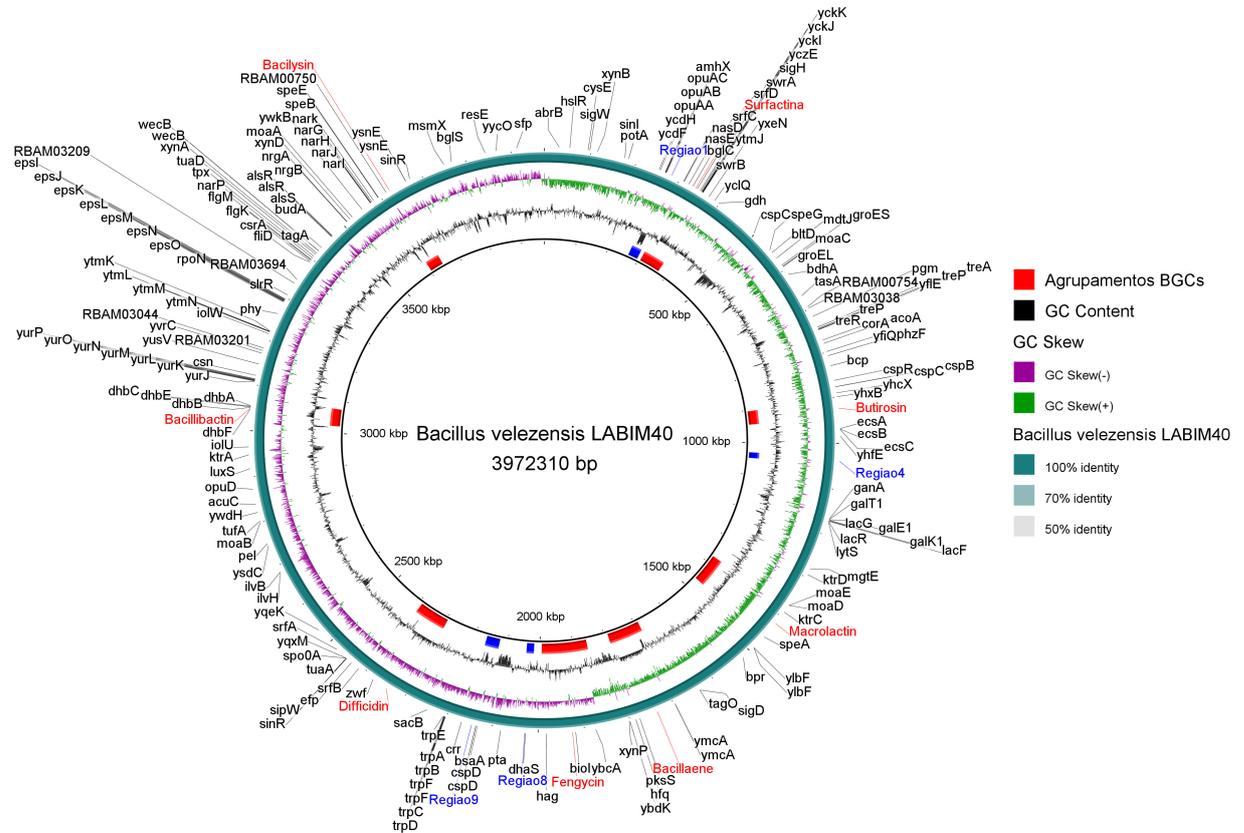
dos 12 *clusters* não apresentaram nenhum *cluster* de similaridade no banco de dados. Com os resultados obtidos, realizou a representação circular das cepas LABIM40 (Figura 14) e LABIM44 (Figura 15), juntamente com os agrupamentos BGCs, encontrados pelo antiSMASH, e os genes PGPs identificados pelo pipeline.

**Tabela 2 – Agrupamentos de Genes Biossintéticos em *Bacillus velezensis* LABIM40 com antiSMASH 7.0.0**

Região	Tipo	De (pb)	Para (pb)	Cluster Semelhante Conhecido		Similaridade
Região 01	<i>thiopeptide, LAP</i>	266 213	295 970			
Região 02	<i>NRPS</i>	306 792	371 874	<i>surfactin</i>	<i>NRP: Lipopeptide</i>	86%
Região 03	<i>PKS-like</i>	903 192	944 436	<i>butirosin A/B</i>	<i>Saccharide</i>	7%
Região 04	<i>terpene</i>	1 030 062	1 047 233			
Região 05	<i>transAT-PKS</i>	1 376 006	1 462 392	<i>macrolactin H</i>	<i>Polyketide</i>	100%
Região 06	<i>transAT-PKS, T3PKS, NRPS</i>	1 688 525	1 789 076	<i>bacillaene</i>	<i>Polyketide+NRP</i>	100%
Região 07	<i>NRPS, transAT-PKS, betalactone</i>	1 857 379	1 994 553	<i>fengycin</i>	<i>NRP</i>	100%
Região 08	<i>terpene</i>	2 017 590	2 039 473			
Região 09	<i>T3PKS</i>	2 123 150	2 164 250			
Região 10	<i>transAT-PKS</i>	2 296 150	2 164 250	<i>difficidin</i>	<i>Polyketide</i>	100%
Região 11	<i>NRP-metallophore, NRPS, RiPP-like</i>	3 022 525	3 074 313	<i>bacillibactin</i>	<i>NRP</i>	100%
Região 12	<i>other</i>	3 602 828	3 644 246	<i>bacilysin</i>	<i>Other</i>	100%

Fonte: Autoria Própria

**Figura 14 – Agrupamentos BGCs e Genes PGP na cepa *Bacillus velezensis* LABIM40**



Fonte: Autoria Própria

**Tabela 3 – Agrupamentos de Genes Biossintéticos em *Bacillus velezensis* LABIM44 com antiSMASH 7.0.0**

Região	Tipo	De (pb)	Para (pb)	Cluster Semelhante	Conhecido	Similaridade
Região 01	<i>thiopeptide, LAP</i>	287 889	317 422			
Região 02	<i>NRPS</i>	328 310	393 175	<i>surfactin</i>	<i>NRP: Lipopeptide</i>	91%
Região 03	<i>PKS-like</i>	970 642	1 011 886	<i>butirosin A/B</i>	<i>Saccharide</i>	7%
Região 04	<i>terpene</i>	1 097 511	1 114 682			
Região 05	<i>transAT-PKS</i>	1 419 116	1 506 956	<i>macrolactin H</i>	<i>Polyketide</i>	100%
Região 06	<i>transAT-PKS, T3PKS, NRPS</i>	1 726 478	1 827 124	<i>bacillaene</i>	<i>Polyketide+NRP</i>	100%



Ademais, utilizou-se o antiSMASH para identificar os BGCs presentes nas quatro cepas PGPRs *Bacillus velezensis*, a fim de representá-los no genoma circular, juntamente com os genes PGP descritos nos artigos das respectivas cepas. Dessa forma, quatro novas ilustrações foram geradas, uma para cada um dos quatro genomas de *Bacillus velezensis* PGPRs, utilizados nesta pesquisa. Destaca-se que, para as cepas HNA3, Ag75, LDO2 e FZB42, os resultados obtidos no antiSMASH encontram-se no repositório do [GitHub](#), além do script utilizado para criar os arquivos em formato Texto com as localizações de cada gene nos genomas das cepas e os arquivos gerados por ele.

Assim, a representação circular do genoma da bactéria *Bacillus velezensis* HNA3, *Bacillus velezensis* Ag75, *Bacillus velezensis* LDO2 e *Bacillus velezensis* FZB42 podem ser observados, respectivamente, nas Figuras 16, 17, 18 e 19.

**Figura 16 – Agrupamentos BGCs e Genes PGP na cepa *Bacillus velezensis* HNA3**

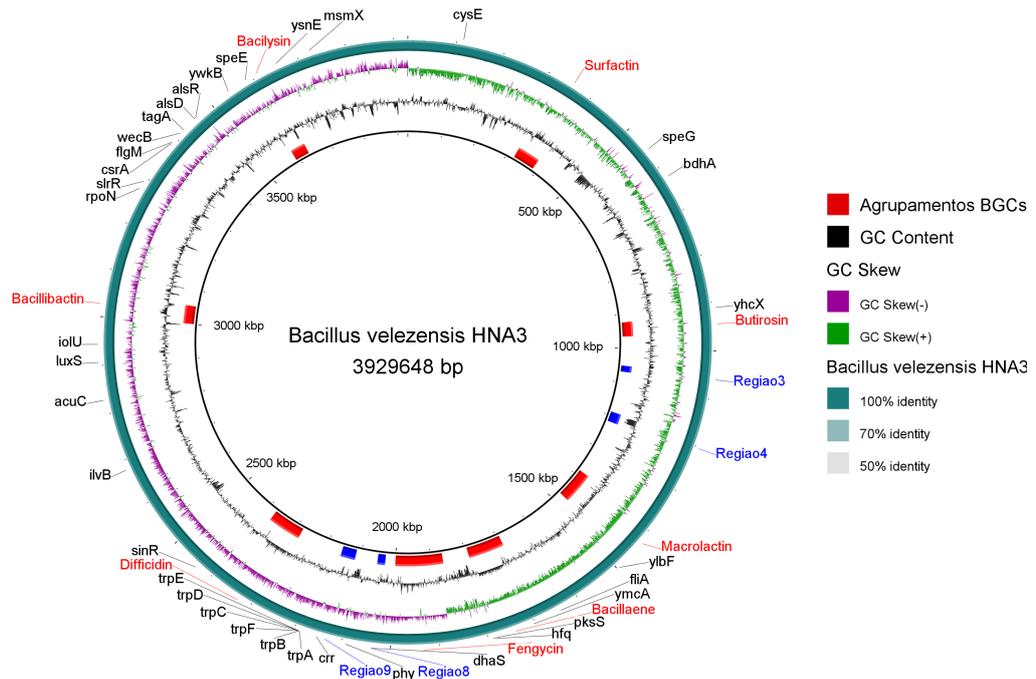
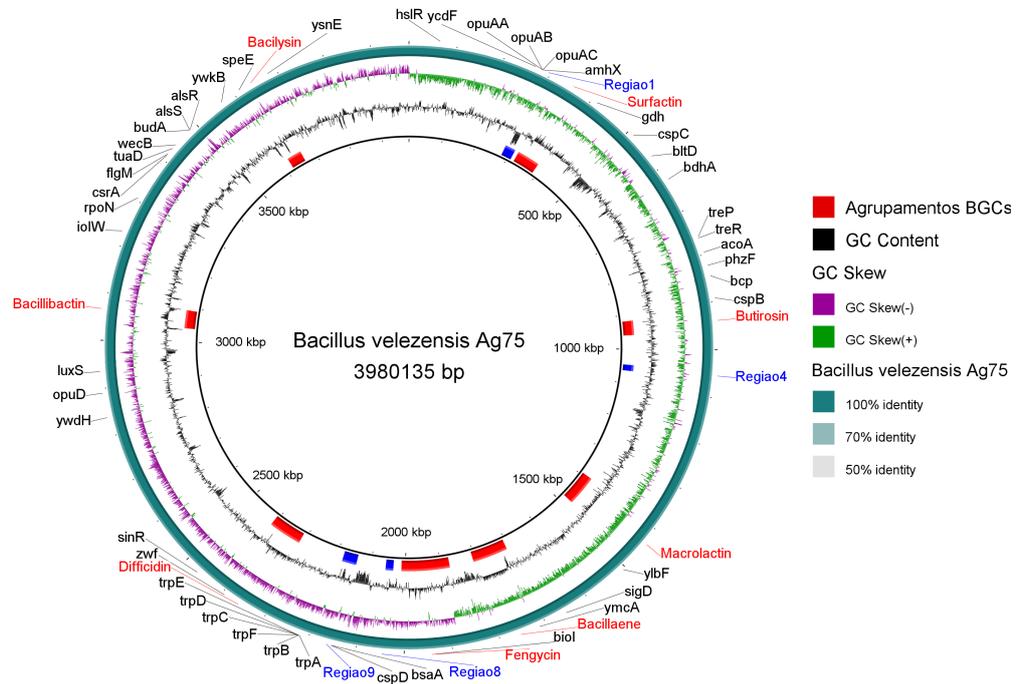
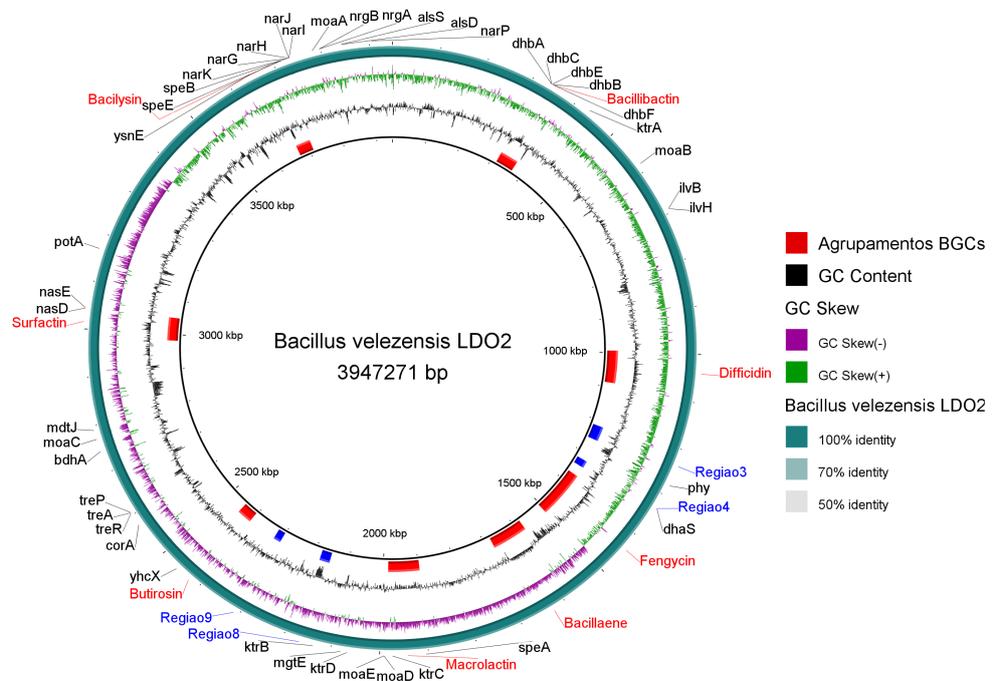


Figura 17 – Agrupamentos BGCs e Genes PGP na cepa *Bacillus velezensis* Ag75



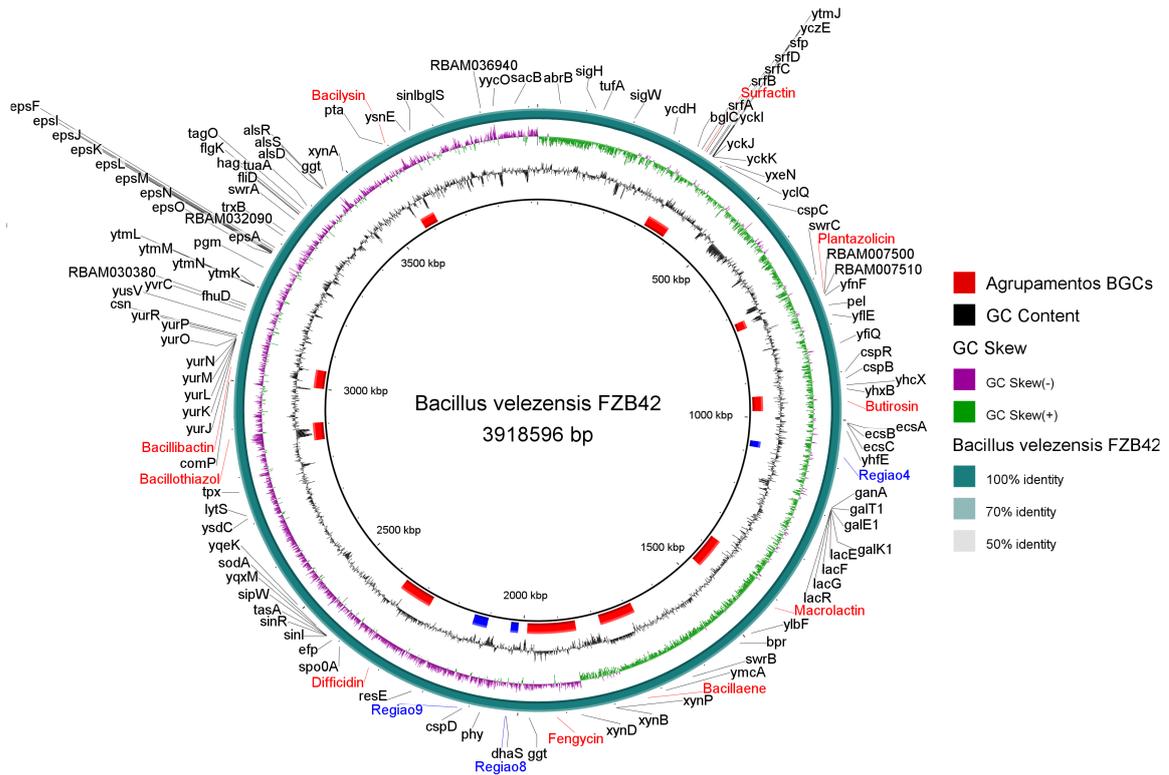
Fonte: Autoria Própria

Figura 18 – Agrupamentos BGCs e Genes PGP na cepa *Bacillus velezensis* LDO2



Fonte: Autoria Própria

**Figura 19 – Agrupamentos BGCs e Genes PGP na cepa *Bacillus velezensis* FZB42**



Fonte: Autoria Própria

Por fim, é realizada a predição de agrupamentos BGCs para outras 12 cepas, sendo dez *Bacillus velezensis*, uma *Bacillus subtilis* e uma *Bacillus sp.* A fim de facilitar a observação dos resultados e realizar uma comparação entre as cepas, as quantidade de BGCs e os BGCs semelhantes, de todas as 19 cepas analisadas, encontram-se organizados, de forma sucinta, na Tabela 4.

**Tabela 4 – Agrupamentos de Genes Biossintéticos com AntiSMASH 7.0.0 nas 19 Cepas**

Cepa	#Cluster	Clusters Similares Conhecidos
<i>B. velezensis</i> HNA3	12	1: surfactin, 2: butirosin A/B, 5: macrolactin H, 6: bacillaene, 7: fengycin, 10: difficidin, 11: bacillibactin, 12: bacilysin
<i>B. velezensis</i> Ag75	12	2: surfactin, 3: butirosin A/B, 5: macrolactin H, 6: bacillaene, 7: fengycin, 10: difficidin, 11: bacillibactin, 12: bacilysin
<i>B. velezensis</i> LDO2	12	1: bacillibactin, 2: difficidin, 5: fengycin, 6: bacillaene, 7: macrolactin H, 10: butirosin A/B, 11: surfactin, 12: bacilysin

<i>B. velezensis</i> FZB42	13	1: surfactin, 2: plantazolicin, 3: butirosin A/B, 5: macrolactin H, 6: bacillaene, 7: fengycin, 10: difficidin, 11: bacillothiazol A/B/C/D/E/F/G/H/I/J/K/L/M/N, 12: bacillibactin, 13: bacilysin
<i>B. subtilis</i> BS96	7	1: surfactin, 2: fengycin, 4: 1-carbapen-2-em-3-carboxylic acid, 5: bacillibactin, 6: subtilosin A, 7: bacilysin
<i>B. cereus</i> RZ2MS9	11	3: petrobactin, 4: bacillibactin, 6: fengycin, 10: molybdenum cofactor
<i>B. velezensis</i> LABIM22	13	1: surfactin, 3: butirosin A/B, 5: macrolactin H, 6: bacillaene, 7: fengycin, 10: difficidin, 11: bacillibactin, 13: bacilysin
<i>B. velezensis</i> LABIM40	12	2: surfactin, 3: butirosin A/B, 5: macrolactin H, 6: bacillaene, 7: fengycin, 10: difficidin, 11: bacillibactin, 12: bacilysin
<i>B. velezensis</i> LABIM44	12	2: surfactin, 3: butirosin A/B, 5: macrolactin H, 6: bacillaene, 7: fengycin, 10: difficidin, 11: bacillibactin, 12: bacilysin
<i>B. velezensis</i> CLA178	13	1: locillomycin A/B/C, 2: surfactin, 3: amyloliquecin GF610, 4: butirosin A/B, 6: macrolactin H, 7: bacillaene, 8: fengycin, 11: difficidin, 12: bacillibactin, 13: bacilysin
<i>B. velezensis</i> UCMB5113	14	1: andalusicin A/B, 2: surfactin, 4: butirosin A/B, 6: macrolactin H, 7: bacillaene, 8: fengycin, 11: difficidin, 13: bacillibactin, 14: bacilysin
<i>B. velezensis</i> SQR9	12	1: surfactin, 2: shuangdaolide A/B/C/D, 3: butirosin A/B, 5: macrolactin H, 6: bacillaene, 7: fengycin, 10: difficidin, 11: bacillibactin, 12: bacilysin
<i>B. velezensis</i> SC60	13	1: surfactin, 2: plantazolicin, 3: butirosin A/B, 5: macrolactin H, 6: bacillaene, 7: fengycin, 10: difficidin, 11: bacillibactin, 13: bacilysin
<i>B. velezensis</i> UCMB5140	12	1: surfactin, 2: butirosin A/B, 4: macrolactin H, 5: bacillaene, 6: fengycin, 9: difficidin, 10: bacillibactin, 12: bacilysin
<i>B. velezensis</i> S141	11	1: surfactin, 2: butirosin A/B, 5: macrolactin H, 6: bacillaene, 7: fengycin, 9: difficidin, 10: bacillibactin, 11: bacilysin
<i>B. velezensis</i> WRN014	13	1: locillomycin A/B/C, 2: surfactin, 3: butirosin A/B, 6: macrolactin H, 7: bacillaene, 8: fengycin, 11: difficidin, 12: bacillibactin, 13: bacilysin
<i>B. velezensis</i> YYC	14	1: andalusicin A/B, 2: surfactin, 3: plantazolicin, 4: butirosin A/B, 6: macrolactin H, 7: bacillaene, 8: fengycin, 11: difficidin, 12: bacillothiazol A/B/C/D/E/F/G/H/I/J/K/L/M/N, 13: bacillibactin, 14: bacilysin

<i>B. velezensis</i> TH-1	12	1: <i>surfactin</i> , 2: <i>butirosin</i> A/B, 5: <i>macrolactin</i> H, 6: <i>bacillaene</i> , 7: <i>fengycin</i> , 10: <i>difficidin</i> , 11: <i>bacillibactin</i> , 12: <i>bacilysin</i>
<i>B. velezensis</i> QST713	14	1: <i>locillomycin</i> A/B/C, 2: <i>surfactin</i> , 3: <i>butirosin</i> A/B, 5: <i>macrolactin</i> H, 6: <i>bacillaene</i> , 7: <i>fengycin</i> , 10: <i>difficidin</i> , 11: <i>bacillibactin</i> , 12: <i>subtilin</i> , 14: <i>bacilysin</i>

---

**Fonte: Autoria Própria**

Analisando a Tabela 4, observa-se que todas as cepas *Bacillus velezensis* possuem o *cluster* ligado a síntese de *surfactin*, *butirosin*, *macrolactin*, *bacillaene*, *fengycin*, *difficidin*, *bacilibactin* e *bacilysin*.

Observa-se, ainda, que a cepa *Bacillus subtilis* BS96 compartilha quatro *clusters* com todas as cepas *Bacillus velezensis*, sendo eles ligados a síntese de *surfactin*, *fengycin*, *bacilibactin* e *bacilysin*. Ainda, há dois *clusters* que são comuns a todas as cepas, incluindo a *B. cereus* RZ2MS9, sendo estes associados a síntese de *bacilibactin* e *fengycin*.

Por fim, há alguns *clusters* encontrados em apenas uma das 19 cepas, sendo estes ligados a síntese de *1-carbapen-2-em-3-carboxylic acid* e *subtilosin* (*B. subtilis* BS96), *petrobactin* e *molybdenum cofator* (*B. cereus* RZ2MS9), *amyloliquecin* GF610 (*B. velezensis* CLA178), *shuangdaolide* (*B. velezensis* SQR9) e *subtilin* (*B. velezensis* QST713).

Destaca-se que, em todas as 19 cepas, foram encontrados agrupamentos BGCs que não apresentaram similaridade com nenhum *cluster* do banco de dados. A cepa *B. cereus*, por exemplo, de 11 *clusters* encontrados, somente quatro apresentaram similaridade com o banco de dados.

## 5.2 ANÁLISE DOS RESULTADOS DO PIPELINE PROPOSTO

Este pipeline é constituído por cinco scripts na linguagem de programação Python que, em resumo, realizam as seguintes tarefas: fazem o download dos genomas completos utilizados em formato Genbank, identificam os nomes dos genes anotados nestes genomas e, nos casos omissos, fazem uma busca via *Entrez* com o auxílio do código RefSeq; complementam a identificação dos genes relacionados à PGP não identificados na etapa anterior por meio do algoritmo BLASTP, utilizando, como arquivo de consulta para o alinhamento, as sequências dos genes buscados manualmente no banco de dados de proteínas do NCBI; realizam a análise dos resultados obtidos com o alinhamento BLASTp; comparam os genes relacionados à PGP de

cada genoma analisado e exibe os resultados; por fim, concatena todas as sequências de genes PGPs, das seis cepas PGPRs, encontrados nas etapas anteriores, e realiza um alinhamento de sequências contra um grupo de 14 genomas classificados na literatura como PGPRs, além das cepas *Bacillus velezensis* LABIM40 e LABIM44.

Para compilar o “Script Download, Pré-Processamento e Organização”, por requerer um tempo de processamento maior que os demais scripts, optou-se por utilizar o Centro de Computação Científica e Tecnológica (CCCT-CP), da Universidade Tecnológica Federal do Paraná, Campus Cornélio Procópio (UTFPR-CP), o qual possui máquinas para computação de alto desempenho.

Nesta seção, são apresentados os resultados deste estudo, bem como os detalhes de cada etapa desenvolvida. Vale ressaltar que os scripts que constituem o pipeline proposto nesta pesquisa encontram-se no repositório online do [GitHub](#).

Como mencionado no Capítulo 4, primeiramente foi realizado um estudo somente com as cepas HNA3 e LABIMs, a fim de desenvolver um pipeline que atendesse os objetivos dessa pesquisa e analisar a eficiência do mesmo. Destaca-se que esse desenvolvimento ocorreu considerando apenas a cepa HNA3, uma vez que as cepas LABIMs seriam incluídas somente na etapa de comparação dos genomas. Partindo disso, optou-se por dividir esta seção em duas subseções: Resultados Preliminares (subseção 5.2.1), que considera apenas a cepa HNA3 e as cepas LABIMs, e Resultados Complementares (subseção 5.2.2), onde constam os resultados obtidos após adicionar as novas cepas PGPRs ao estudo e as análises do banco de dados construído.

### **5.2.1 Resultados Preliminares**

Primeiramente, destaca-se que a versão do pipeline utilizado nos estudos iniciais foi atualizada e, por isso, os scripts aqui descritos não necessariamente serão os mesmos da Figura 7, disponíveis no [GitHub](#).

O primeiro script, por exemplo, realizava apenas o download dos genomas, diferente da versão atual, que também realiza a complementação dos nomes de genes destas cepas, identificando, por meio do código RefSeq, o nome de genes não identificados no arquivo GenBank.

Ressalta-se, ainda, que os arquivos Genbank das quatro cepas aqui utilizadas (HNA3, LABIM22, LABIM40 e LABIM44), disponibilizados no NCBI, não continham o nome de todos

os genes, provavelmente devido à problemas de anotação.

No caso da cepa PGPR HNA3, dos 3 810 genes anotados, somente 800 continham nome, sobrando ainda um total de 3 010 genes sem nomes identificados. Além disso, dos 35 genes relacionados à PGP em Zaid *et al.* (2022), apenas nove deles foram identificados.

Desta forma, tornou-se necessário criar um script que, além de escrever em um arquivo FASTA os genes que possuíam nome, buscava no NCBI, via *Entrez*, os nomes dos demais genes por meio do código RefSeq. Com este script, para o genoma da estirpe HNA3, o número de genes com nomes identificados aumentou para 1 895 e o número de genes relacionados à PGP, aumentou para 19 (Quadro 2).

Assim, este script foi responsável por um aumento expressivo na quantidade de genes identificados. No entanto, apesar deste esforço, ainda restou um número considerável sem identificação de nome: um total de 1 915 genes, incluindo 16 genes relacionados à PGP (Quadro 3).

Para identificar os outros 16 genes, optou-se por realizar um alinhamento de sequências via BLASTp. Para isso, buscou-se no NCBI os genes relatados no Quadro 2 que ainda não haviam sido identificados pelo script anterior em genomas de bactérias, priorizando cepas PGPRs, como o caso da *Bacillus subtilis* 168 (CHEN *et al.*, 2007; CHANDRA *et al.*, 2021; ZAID *et al.*, 2022), por exemplo.

Para realizar o alinhamento, um arquivo FASTA com os genes identificados pelos scripts e por essa busca no NCBI, foi criado, totalizando 34 genes relacionados à PGP. Evidencia-se que um dos genes relatados em Zaid *et al.* (2022) não havia sido encontrado (*phytase*), por isso, 34 genes. Então, este novo script, selecionava do arquivo criado, os genes cujos nomes ainda não haviam sido identificados, e criava um arquivo FASTA de consulta para o alinhamento. Daí, o algoritmo BLASTp realizava sua busca no arquivo FASTA de consulta que continha todos os 1 915 genes com nomes não identificados.

Com isso, somente cinco genes continuaram sem identificação, eram eles: *fliA*, *alsD*, *ylbF*, *ymcA* e *speG*. Especificamente para os genes *fliA* e *speG*, o BLASTp encontrou alinhamentos com *E-value* em torno de  $10^{-5}$ , porém com baixa porcentagem de identidade (26%). Ao contrário de alinhamentos realizados com outros genes, como o gene *trpF*, por exemplo, que o BLASTp encontrou um alinhamento com *E-value* de  $10^{-156}$  e alta porcentagem de identidade (95%).

Por fim, para verificar se as cepas disponibilizadas pelo LABIM continham os mesmos genes relacionados à PGP da HNA3, foi utilizado os mesmos scripts, visto que estes foram

criados para serem utilizados para qualquer genoma. Na Tabela 5, é possível visualizar o número de genes encontrados para cada cepa, em todas as etapas do desenvolvimento do pipeline.

**Tabela 5 – Quantidade de genes identificados das cepas HNA3, LABIM22, LABIM40 e LABIM44**

	Total de Genes no Genoma	Genes Identificados sem Entrez	Genes Identificados com Entrez	Genes não Identificados
HNA3	3810	800	1895	1915
LABIM22	3786	922	1926	1860
LABIM40	3882	453	1746	2136
LABIM44	3905	1053	1993	1912

**Fonte: Autoria Própria**

Por sua vez, para a identificação dos genes relacionados à PGP, na Tabela 6, é apresentada a quantidade destes genes em cada etapa do desenvolvimento do pipeline.

**Tabela 6 – Quantidade de genes PGP identificados em HNA3, LABIM22, LABIM40 e LABIM44**

	Genes PGP sem Entrez	Genes PGP com Entrez	Genes PGP com BLASTp
HNA3	9	19	29
LABIM22	11	19	29
LABIM40	7	19	29
LABIM44	14	20	29

**Fonte: Autoria Própria**

Vale ressaltar que, na Tabela 6, após a utilização do *Entrez* no segundo script, foram identificadas os mesmos 19 genes para todas as cepas, com adição do gene *wecB* para a cepa LABIM44, sendo eles: *dhaS*, *luxS*, *hfq*, *flgM*, *acuC*, *sinR*, *trpA*, *cysE*, *ywkB*, *ilvB*, *csrA*, *trpD*, *trpB*, *speE*, *trpE*, *trpC*, *bdhA*, *rpoN*, *yhcX*.

Ainda, todas as cepas tiveram os mesmos 29 genes identificados com a utilização do alinhamento BLASTp do terceiro script, ou seja, foram identificados 10 outros genes, sendo eles: *trpF*, *ysnE*, *msmX*, *pksS*, *tarA/tagA*, *alsR*, *iolU*, *slrR*, *crr* e *wecB* (este último não precisou ser identificado nesta etapa para a cepa LABIM44, como explicado previamente). Desta forma, os mesmos 5 genes (*fliA*, *alsD*, *ylbF*, *ymcA* e *speG*) não foram identificados para nenhuma das cepas, pelos scripts do pipeline proposto.

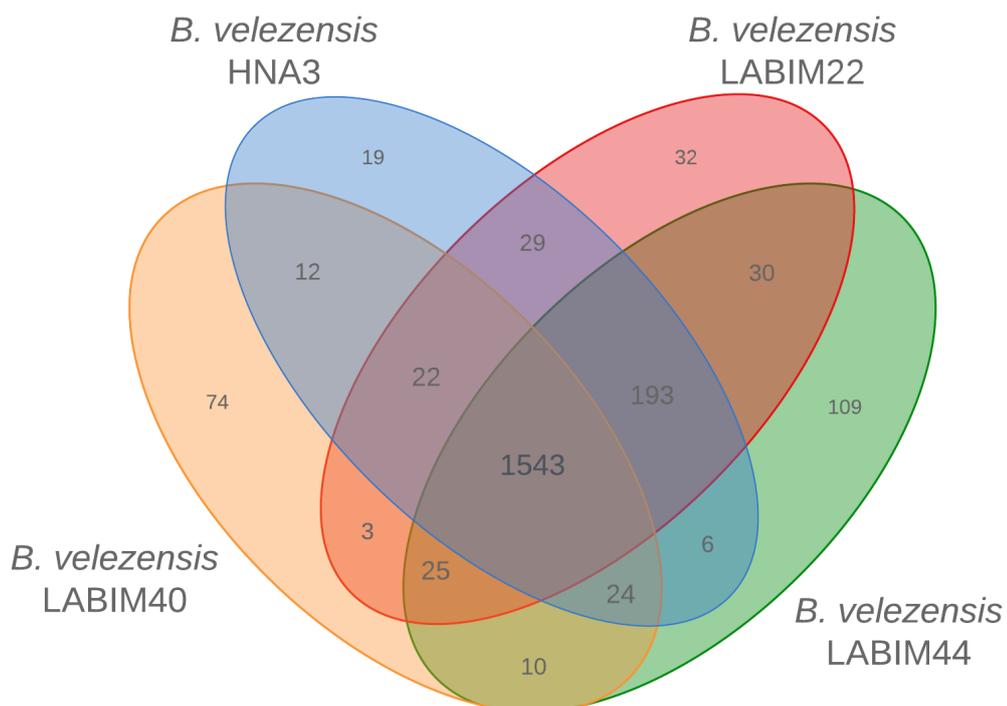
Portanto, pelo resultado obtido por esta versão do pipeline, os genomas das cepas disponibilizadas pelo LABIM compartilham os mesmos 29 genes relacionados à PGP com

o genoma da cepa HNA3. Destaca-se que as cepas analisadas são todas da espécie *Bacillus velezensis*, isto é, são evolutivamente próximas, o que pode, além do pequeno número de amostras, justificar o resultado obtido.

Devido à semelhança destes 29 genes relacionados à promoção de crescimento em plantas em todas as cepas LABIMs e HNA3, existe um indicativo de que as cepas disponibilizadas pelo LABIM possivelmente são PGPR. A partir disso, decidiu-se comparar todos os quatro genomas, a fim de identificar a quantidade de genes em comum entre eles, ressaltando enfaticamente que essa comparação foi feita somente com os genes identificados pelo pipeline, restando uma parte significativa de genes em cada genoma para ser identificada.

Destaca-se que, em cada genoma, existem multiplicidade de ocorrência de genes na anotação, por isso, a quantidade de genes a ser comparada é diferente do que foi apresentado na Tabela 5, uma vez que no cálculo das interseções dos conjuntos, as repetições não são mantidas. Para a cepa HNA3, foram utilizados na comparação, 1 848 dos 1 895 genes identificados após a busca via *Entrez*, para a LABIM22, foram 1 877 dos 1 926, para a LABIM40, 1 713 dos 1 746 e, por fim, para a LABIM44, 1 940 dos 1993 descritos na Tabela 5. Com isso, pode-se observar a quantidade de genes comuns entre os genomas, identificados por este pipeline, por meio de um diagrama de Venn, ilustrado na Figura 20.

**Figura 20 – Diagrama de Venn dos Genes Identificados pelo Pipeline**



**Fonte: Autoria Própria**

Conforme observado na Figura 20, os quatro genomas comparados compartilham entre si, em média, aproximadamente 84% dos genes identificados pelo pipeline. Ainda não se pode afirmar sobre a semelhança do genoma completo, pois existe um número significativo de dados faltantes. Contudo, essa informação reforça o indicativo de que as cepas LABIM aqui estudadas, podem ter funções iguais ou semelhantes as da cepa *Bacillus velezensis* HNA3.

Para acrescentar novas análises e discussões na continuação desta pesquisa, foram acrescentados a este estudo, novos genomas de PGPRs encontrados na literatura. Consequentemente, novas comparações entre todos os genomas foram realizadas, a fim de analisar se os genes presentes nos novos genomas, além dos já identificados até então, também estão presentes nas cepas LABIMs. Tais estudos podem ser verificados na subseção subsequente.

### 5.2.2 Resultados Complementares

Nesta etapa da pesquisa, cinco novas cepas foram consideradas, sendo elas: *Bacillus velezensis* Ag75, LDO2, FZB42, *Bacillus subtilis* BS96 e *Bacillus sp.* RZ2MS9. Além disso, todo o procedimento adotado para as estirpes Ag75, LDO2 e FZB42, foram reaplicados para as cepas HNA3, LABIM22, LABIM40 e LABIM44, uma vez que os scripts foram atualizados e, consequentemente, os resultados também.

Destaca-se que, unindo os genes PGPs das quatro cepas PGPRs selecionadas da literatura, obteve-se um total de 245 genes. No entanto, alguns genes estavam descritos nos estudos de duas ou mais destas cepas, totalizando 198 genes envolvidos na promoção de crescimento em plantas.

Inicialmente, foi realizado o mesmo processo para as cepas Ag75, LDO2 e FZB42, onde verificou-se o mesmo problema: alguns genes descritos pelos autores que estudaram as respectivas PGPRs, não foram identificados nos arquivos GenBank baixados do NCBI. Assim, necessitou-se da complementação de informações, a fim de identificar todos os genes relacionados à PGP.

Mesmo com a busca e complementação de genes via *Entrez*, alguns genes não puderam ser identificados. No caso da cepa *Bacillus velezensis* Ag75, 14 de 47 genes não foram encontrados, para a cepa *Bacillus velezensis* LDO2, 21 dos 43 permaneceram não identificados, para a cepa *Bacillus subtilis* BS96, foram encontrados 241 dos 377 genes e, por fim, para a cepa *Bacillus velezensis* FZB42, dos 120 genes descritos, foram encontrados apenas 36. No entanto, haviam oito destes genes cujos nomes foram descritos usando o "*locus\_tag*", possibilitando encontrá-los manualmente dentro do próprio genoma, restando 76 genes para serem identificados.

Dessa forma, um alinhamento de sequências também foi necessário para estas cepas.

Já para as LABIMs, dos 198 genes envolvidos na promoção de crescimento vegetal, nesta etapa foram identificados 84 genes para a *Bacillus velezensis* LABIM22, 79 genes para a *Bacillus velezensis* LABIM40 e 90 genes para a *Bacillus velezensis* LABIM44. Dessa forma, também realizou-se um alinhamento BLASTp, de forma a verificar se os demais genes estavam presentes nestas cepas.

Para realizar o alinhamento via BLASTp, foi preciso buscar pelos genes manualmente no NCBI. Inicialmente, a estratégia de busca foi a mesma utilizada na etapa dos estudos iniciais: buscar no banco de dados de genes do NCBI, uma vez que, ao digitar o nome do gene na barra de pesquisa do NCBI, a busca era realizada automaticamente dentro desse banco de dados específico. No entanto, alguns genes encontrados manualmente, ao realizar o alinhamento, tiveram resultados insatisfatórios, o que serviu de motivação para outra estratégia de busca. Uma das ideias que surgiram foi fazer essa busca no banco de dados Swiss-Prot, no entanto, os scripts deveriam ser reformulados para que isso fosse possível. Pesquisas foram realizadas para auxiliar no desenvolvimento da codificação, mas os resultados não foram positivos.

Dessa forma, uma estratégia mais simples foi adotada: pesquisar diretamente no banco de dados de proteínas do NCBI. Fazendo isso, novas sequências foram encontradas e, ao realizar o alinhamento BLASTp, os resultados foram excelentes. Com isso, uma nova busca foi realizada também para os genes da HNA3, uma vez que os resultados obtidos anteriormente não foram tão satisfatórios, além dos genes que não haviam sido encontrados ou obtiveram um alinhamento fraco.

Utilizando a nova estratégia de busca, todos os genes, dos três primeiros genomas PGPRs estudados, foram encontrados. Para o genoma da *Bacillus velezensis* FZB42, apenas três não foram identificados, pois os alinhamentos não foram positivos: *ybcA*, *ytmJ* e *ybdK*.

A seguir, na Tabela 7, Tabela 8 e Tabela 9, é possível observar os resultados obtidos no alinhamento BLASTp (percentual de identidade e *E-Value*), além dos códigos *protein-id* e RefSeq/GenBank de cada um dos genes. Para a cepa *Bacillus velezensis* FZB42, a tabela encontra-se em apêndice (Apêndice F), devido à quantidade de informações e, para a cepa *Bacillus subtilis* BS96, os resultados encontram-se no repositório online do [GitHub](#), uma vez que não houve tempo hábil para a organização dos resultados na forma de tabela, como realizado para as outras cepas estudadas anteriormente.

Assim, na Tabela 7, constam os 16 genes não identificados via *Entrez* na estirpe *Bacillus*

*velezensis* HNA3, mas encontrados via alinhamento de sequências de proteínas. Ressalta-se que todos os genes descritos no estudo de Zaid *et al.* (2022) foram encontrados por meio dos procedimentos adotados.

**Tabela 7 – Resultado Alinhamento BLASTp *Bacillus velezensis* HNA3**

Código RefSeq/ Genbank (Query)	RefSeq/Protein-id (Subject)	Nome do Gene	Percentual Identidade	E-Value
WP_012117858.1	QJC42419.1	<i>trpF</i>	95%	$2 \cdot 10^{-156}$
URM44301.1	QJC43788.1	<i>ysnE</i>	96%	$2 \cdot 10^{-110}$
URM46110.1	QJC42069.1	<i>pksS</i>	98%	0.0
USQ53343.1	QJC43590.1	<i>alsR</i>	99%	0.0
URM43645.1	QJC43101.1	<i>iolU</i>	99%	0.0
URM43968.1	QJC43436.1	<i>slrR</i>	99%	$2 \cdot 10^{-110}$
WP_012117833.1	QJC42378.1	<i>crr</i>	100%	$3 \cdot 10^{-122}$
WP_012118587.1	QJC43559.1	<i>tagA</i>	99%	0.0
WP_012118583.1	QJC43552.1	<i>wecB</i>	98%	0.0
BBA76563.1	QJC42309.1	<i>phy</i>	99%	0.0
AHZ17904.1	QJC43857.1	<i>msmX</i>	99%	0.0
QIR31908.1	QJC41018.1	<i>speG</i>	97%	$1 \cdot 10^{-130}$
URM44102.1	QJC43588.1	<i>alsD</i>	100%	0.0
BCU86156.1	QJC41859.1	<i>ylbF</i>	100%	$2 \cdot 10^{-95}$
URM46094.1	QJC42052.1	<i>ymcA</i>	100%	$3 \cdot 10^{-102}$
SIR80373.1	QJC41997.1	<i>ftiA</i>	100%	0.0

**Fonte: Autoria Própria**

Como pode ser observado, o percentual de identidade de todos os genes são excelentes, sendo 95% o menor valor obtido, correspondente ao alinhamento do gene *trpF*. Da mesma forma, os valores do *E-Value* também foram incrivelmente satisfatórios, sendo  $2 \cdot 10^{-95}$  o maior valor obtido, decorrente do alinhamento do gene *ylbF*. Destaca-se que, para o *E-value*, quanto mais próximo de zero, melhor.

Já na Tabela 8, é possível observar os 15 genes não identificados via *Entrez* na cepa da rizobactéria *Bacillus velezensis* Ag75. Assim como para a cepa HNA3, para esta estirpe também foram encontrados todos os genes envolvidos na promoção de crescimento em plantas, descritos no estudo de Mosela *et al.* (2022).

Ainda, o percentual de identidade de todos os genes da estirpe Ag75 foram excelentes, sendo 97% o menor valor obtido, correspondente ao alinhamento dos genes *bltD* e *phzF*. Da

mesma forma, os valores do *E-Value* também foram ótimos, sendo  $3 \cdot 10^{-102}$  o maior valor obtido, decorrente do alinhamento do gene *ymcA*.

**Tabela 8 – Resultado Alinhamento BLASTp *Bacillus velezensis* Ag75**

Código RefSeq/ Genbank (Query)	RefSeq/Protein-id (Subject)	Nome do Gene	Percentual Identidade	E-Value
URM44301.1	USQ53539.1	<i>ysnE</i>	98%	$2 \cdot 10^{-112}$
WP_011996289.1	USQ54049.1	<i>amhX</i>	99%	0.0
WP_012117684.1	USQ55443.1	<i>bioI</i>	99%	0.0
WP_012116949.1	USQ54296.1	<i>bltD</i>	97%	$2 \cdot 10^{-109}$
WP_014471118.1	USQ53343.1	<i>alsR</i>	98%	0.0
BCU86156.1	USQ55125.1	<i>ylbF</i>	99%	$3 \cdot 10^{-109}$
URM46094.1	USQ55315.1	<i>ymcA</i>	100%	$3 \cdot 10^{-102}$
URM43879.1	USQ53116.1	<i>iolW</i>	99%	0.0
WP_011996282.1	USQ54034.1	<i>ycdF</i>	99%	0.0
WP_012118579.1	USQ53301.1	<i>tuaD</i>	99%	0.0
WP_017419144.1	USQ54505.1	<i>phzF</i>	97%	0.0
WP_012118228.1	USQ52807.1	<i>apuD</i>	99%	0.0
WP_012117858.1	USQ52232.1	<i>trpF</i>	99%	$7 \cdot 10^{-164}$
NP_388484.1	USQ54335.1	<i>groEL</i>	98%	0.0
WP_012117811.1	USQ52160.1	<i>bsaA</i>	99%	$3 \cdot 10^{-122}$

**Fonte: A autoria Própria**

Na Tabela 9, observa-se os 21 genes não identificados via *Entrez* na cepa da rizobactéria *Bacillus velezensis* LDO2, estudada em Chen *et al.* (2019). Destaca-se que todos os genes PGPs descritos pelos autores, foram encontrados após estes procedimentos.

Nota-se que, assim como para as outras duas cepas PGPRs aqui estudadas, os resultados do alinhamento BLASTp para a cepa da PGPR *Bacillus velezensis* LDO2 também foram excelentes, sendo de 96% o menor percentual de identidade obtido (gene *ysnE*) e  $3 \cdot 10^{-74}$  o maior valor obtido no *E-Value* (gene *mdtJ*).

Tabela 9 – Resultado Alinhamento BLASTp *Bacillus velezensis* LDO2

Código RefSeq/ Genbank (Query)	RefSeq/Protein-id (Subject)	Nome do Gene	Percentual Identidade	E-Value
BCU87559.1	AWE15005.1	<i>dhbA</i>	100%	0.0
AHZ17188.1	AWE15008.1	<i>dhbB</i>	99%	0.0
UYP22140.1	AWE15006.1	<i>dhbC</i>	98%	0.0
AFZ91981.1	AWE15007.1	<i>dhbE</i>	99%	0.0
AFJ63359.1	AWE15009.1	<i>dhbF</i>	99%	0.0
URM44102.1	AWE18126.1	<i>alsD</i>	100%	0.0
BCU87210.1	AWE15327.1	<i>ilvH</i>	100%	$2 \cdot 10^{-122}$
AHZ14763.1	AWE17013.1	<i>treA</i>	98%	0.0
WP_014417186.1	AWE17014.1	<i>treP</i>	99%	0.0
WP_012117785.1	AWE15887.1	<i>phy</i>	98%	0.0
URM44301.1	AWE17928.1	<i>ysnE</i>	96%	$2 \cdot 10^{-110}$
CUX92117.1	AWE17506.1	<i>potA</i>	98%	0.0
BBA75191.1	AWE17165.1	<i>mdtJ</i>	100%	$3 \cdot 10^{-74}$
WP_007408054.1	AWE15232.1	<i>moaB</i>	99%	$2 \cdot 10^{-125}$
WP_007611116.1	AWE16410.1	<i>moaE</i>	100%	$4 \cdot 10^{-112}$
AHZ14339.1	AWE17396.1	<i>nasE</i>	99%	$1 \cdot 10^{-75}$
WP_003151223.1	AWE18078.1	<i>nrgA</i>	100%	0.0
AHZ17661.1	AWE18077.1	<i>nrgB</i>	99%	$8 \cdot 10^{-81}$
WP_007407632.1	AWE18003.1	<i>narK</i>	99%	0.0
CCG51474.1	AWE18178.1	<i>narP</i>	100%	$4 \cdot 10^{-172}$
QOY25331.1	AWE16487.1	<i>ktrB</i>	99%	0.0

Fonte: Autoria Própria

Já para a cepa FZB42, diferente das três estirpes anteriores, não foi possível encontrar todos os 120 genes descritos em Chen *et al.* (2007), uma vez que o gene *ybcA* não foi encontrado no NCBI para bactérias do gênero *Bacillus* e os genes *ytmJ* e *ybdK* que, embora tenham sido encontrados em várias sequências de bactérias da espécie *Bacillus velezensis*, os resultados do alinhamento BLASTp foram insatisfatórios, com percentual de identidade em torno de 57% e 26%, respectivamente.

No entanto, com exceção destes três genes, todos os demais foram encontrados, com ótimos resultados no alinhamento BLASTp, sendo de 96% o gene com menor percentual de identidade (gene *ysnE*) e de  $5.04 \cdot 10^{-43}$  o maior valor obtido para o *E-Value* (gene *cspD*). Conforme já mencionado, devido a quantidade de informações, esses e os demais resultados estão descritos no Apêndice F.

Além disso, esta complementação também foi feita para as cepas LABIMs, no mesmo script. Dessa forma, utilizou-se como base todos os genes pertencentes às quatro cepas *Bacillus velezensis* PGPRs.

De um total de 198 genes PGPs das quatro cepas PGPRs, restavam encontrar 114 para a cepa LABIM22, 119 para a LABIM40 e 108 para a LABIM44. Com o alinhamento BLASTp, foram identificados em cada uma destas três cepas, respectivamente, 108, 114 e 102 genes.

Destaca-se que, para a complementação da cepa *Bacillus subtilis* BS96, o procedimento foi um pouco diferente, visto que há vários genes descritos com o mesmo nome. Tal procedimento será explicado posteriormente.

A partir dos resultados obtidos das comparações realizadas, o script de comparação foi utilizado para comparar cada cepa *Bacillus velezensis* PGPR com cada uma das cepas LABIMs, utilizando os nomes dos genes, cujo resultado pode ser observado na Tabela 10.

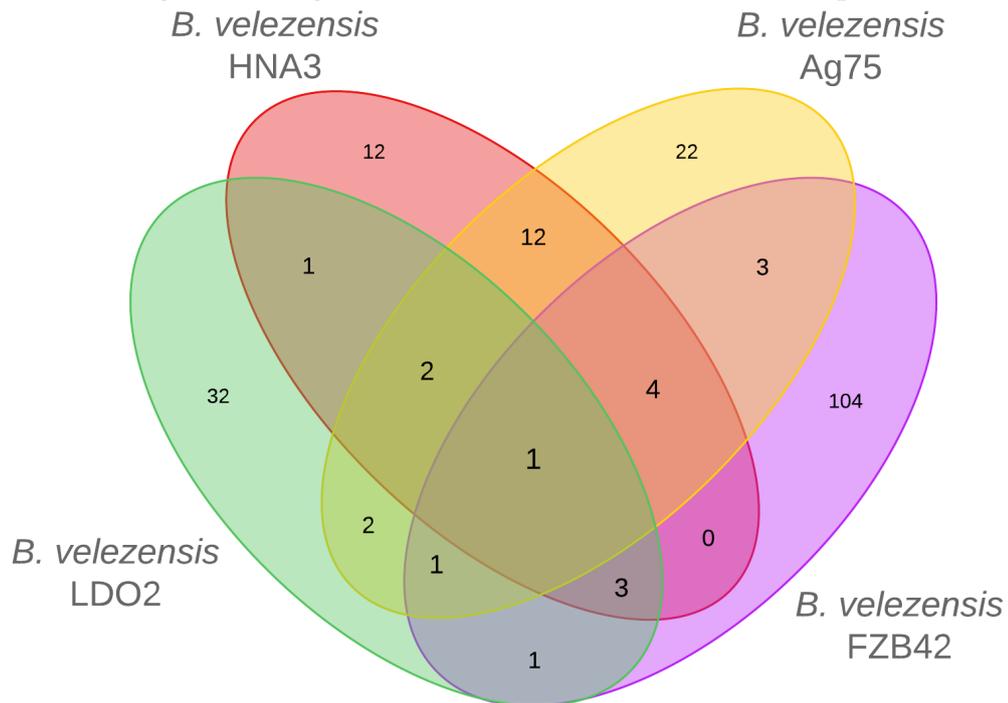
**Tabela 10 – Resultado da Comparação entre LABIMs e Cepas PGPRs *B. velezensis***

Cepas	<i>B. velezensis</i> HNA3 (35 genes PGP)	<i>B. velezensis</i> Ag75 (47 genes PGP)	<i>B. velezensis</i> LDO2 (43 genes PGP)	<i>B. velezensis</i> FZB42 (117 genes PGP)
<b>LABIM22</b>	31	44	40	106
<b>LABIM40</b>	32	45	40	105
<b>LABIM44</b>	32	45	40	104

Conforme pode ser observado na Tabela 10, a maioria dos genes descritos em cada um dos genomas PGPRs utilizados na comparação, foram encontrados em todas as cepas LABIMs. No mais, é importante enfatizar que esta comparação foi realizada utilizando o nome dos genes, e não a sequência de aminoácidos.

A comparação por nome de gene entre as quatro cepas PGPRs *Bacillus velezensis* também foi realizada, conforme mostra o diagrama da Figura 21.

**Figura 21 – Diagrama de Venn dos Genes PGP das Quatro Cepas**



Destaca-se que, dos 245 genes resultantes do somatório entre as quatro cepas, três não foram encontrados via *Entrez* ou busca manual, restando um total de 242 genes envolvidos na promoção de crescimento vegetal.

Analisando a Figura 21, percebe-se que vários genes estão presentes em duas ou mais cepas, inclusive em todas elas, como é o caso do gene *ysnE*. Dessa forma, dos 242 genes PGPs, ao realizar a comparação, resultou em 200 genes cujos nomes são diferentes.

No entanto, durante a etapa de complementação de genes, deparou-se com alguns casos em que um mesmo gene estava anotado em diferentes cepas, com diferentes nomes. Logo, um alinhamento de sequências também foi necessário, a fim de verificar se, além dos genes que aparecem em duas ou mais cepas, também existiam outros com alto percentual de similaridade entre si. Para isso, foi utilizado um limiar igual a 90%. Com isso, resultou em 198 genes diferentes, ou seja, além dos repetidos, haviam 2 pares de genes similares: o gene *alsD* das cepas HNA3, LDO2 e FZB42 com o gene *budA* da cepa Ag75; e o gene *fliA*, presente na estirpe HNA3, com o gene *sigD*, presente na Ag75.

Assim, dos 242 genes PGPs das quatro cepas: 35 da HNA3, 47 da Ag75, 43 da LDO2 e 117 da FZB42, 198 são distintos, compondo o primeiro grupo de genes PGPs do pipeline.

Já o segundo grupo, formado por genes PGPs pertencentes a *Bacillus subtilis* BS96, inicialmente imaginava-se que seriam encontrados um total de 377 genes. Com o primeiro script,

que realiza o download das sequências com suas respectivas informações, e faz uma busca via *Entrez* por meio do código RefSeq para encontrar o nome dos genes, do genoma todo, que não puderam ser identificados, foram encontrados um total de 241 genes.

Utilizando o script adicional criado, que exclui os genes cujos nomes são repetidos, realiza a busca, pelo nome do gene e o organismo, nos bancos de dados de proteínas do NCBI, via *Entrez Search*, e faz um alinhamento BLASTp entre as dez sequências encontradas para cada gene e o “Arquivo FASTA (Subject sem PGP)”, dos 136 genes faltantes, resultou em 104 não repetidos. Destes, 55 foram encontrados pelo script, com percentual mínimo de identidade igual a 89,35% (*opuBD*), e todos os outros 54 acima de 95%, e foram concatenados ao arquivo de genes PGPs gerado pelo primeiro arquivo do pipeline.

Ainda, tem-se que dos 49 genes não encontrados pelo script, 14 são arquivos em branco, isto é, não foi encontrada nenhuma sequência no banco de dados de proteínas do NCBI, são eles: *accB*, *bmpA*, *cydB*, *ddpD*, *ddpF*, *fliNY*, *fliY*, *glsA*, *mdlA*, *metQ*, *nupA*, *nupB*, *peb1B* e *peb1C*. Além disso, três não tiveram resultados tão ruins, mas ainda foi inferior ao limiar desejado, são eles: *gmuG*, *rapE* e *ytrC*, com 70,22%, 65,60% e 65,24%, respectivamente.

Ainda, tiveram 29 genes com percentual de identidade inferior a 60%, são eles: *bioY*, *cheY*, *cydA*, *desR*, *ecfT*, *flgE*, *flgG*, *ganP*, *ganQ*, *katE*, *lplA*, *lplB*, *lplC*, *lsrK*, *metB*, *motA*, *msmE*, *msmF*, *msmG*, *msmX*, *phrF*, *poxL*, *pps*, *rapA*, *rapH*, *tcyB*, *ycbA*, *ycbB* e *yidC*. Estes, e os demais resultados, encontram-se no [GitHub](#).

Como um total de 49 genes não foram identificados pelo script, outra estratégia foi realizada, a qual consistia em uma busca manual no “Arquivo FASTA (Subject sem PGP)” da cepa *Bacillus subtilis* BS96, verificando o *EC\_Number* ou o produto do gene. Assim, quatro outros genes foram encontrados: *accB*, *accC*, *cysH* e *phoA*. Este arquivo também foi concatenado ao arquivo com os demais genes PGPs encontrados nas cepa BS96.

Dessa forma, para o segundo grupo, foram obtidos 300 genes envolvidos na promoção de crescimento em plantas.

Por fim, para o formar o terceiro grupo de genes que compõem o banco de dados, primeiramente realizou-se o download de todas as sequências de genes descritas em Batista (2017) e, a fim de verificar se realmente todos os genes estavam presentes no genoma da cepa *Bacillus sp.* RZ2MS9, realizou-se um alinhamento de sequências, utilizando o percentual de similaridade igual a 100%, uma vez que os códigos fornecidos eram da própria anotação desta cepa. No entanto, seis genes não foram identificados: *spoIIIAD* (OGY06059.1), *yndD* (OGY01680.1),

*flgB* (OGY03710.1), *flgG* (OGY04640.1), *pilA* (OGY02999.1) e *ldh* (OGY02946.1).

Após excluir os seis genes não encontrados no genoma, a próxima etapa foi verificar se haviam genes repetidos entre os 268 restantes. Com a utilização do script adicional de alinhamento de sequências BLASTp, verificou-se que dois genes eram altamente similares, sendo eles: *gerPF* (OGY03157.1) e *gerPF* (OGY04214.1). Assim, ao final destes dois procedimentos, restaram um total de 267 genes, compondo o último grupo de genes envolvidos na promoção de crescimento em plantas, que compõem o banco de dados aqui proposto.

Assim como já mencionado anteriormente, o objetivo é continuar alimentando este banco de dados, para que, cada vez mais, o número de amostras de genes PGPs, de diferentes espécies de *Bacillus*, sejam acrescentadas, aumentando a relevância deste banco de dados para pesquisas futuras.

Destaca-se que um alinhamento de sequências não foi realizado entre os três grupos, uma vez que todos eles são de espécie de *Bacillus* diferentes. Dessa forma, as novas amostras que serão acrescentadas futuramente, a medida que novos estudos forem publicados, ou uma busca ainda mais minuciosa seja realizada, também deverão seguir este critério de separação por grupos.

Na Tabela 11, encontra-se, de forma resumida, a quantidade de genes de cada grupo do banco de dados.

**Tabela 11 – Quantidade de Genes Banco de dados**

<b>Grupos</b>	<b>Cepas</b>	<b>Total Genes</b>	<b>Total Genes</b>
<b>Genes</b>	<b>PGPRs</b>	<b>por Cepa</b>	<b>por Grupo</b>
Grupo 1	<i>B. velezensis</i> HNA3	35	198
	<i>B. velezensis</i> Ag75	47	
	<i>B. velezensis</i> LDO2	43	
	<i>B. velezensis</i> FZB42	117	
Grupo 2	<i>B. subtilis</i> BS96	377	300
Grupo 3	<i>B. cereus</i> RZ2MS9	274	267

**Fonte: Autoria Própria**

Com os três grupos de genes PGPs criados, o último script foi utilizado para concatenar todas essas sequências, e criar um arquivo FASTA, identificado na Figura 7 como “Arquivo FASTA (Banco de Dados AgroBioTech)”, que é o banco de dados proposto nesta pesquisa,

composto por 765 sequências de genes envolvidos na promoção de crescimento vegetal.

No entanto, antes de concatená-los, utilizou-se como banco de dados, somente os genes das cepas PGPRs *Bacillus velezensis*, realizando um alinhamento BLASTp com 12 dos 14 genomas selecionados, todos *Bacillus velezensis* e classificados pela literatura como PGPRs, com percentual de identidade igual a 90%, a fim de verificar quantos genes PGPs estas cepas da mesma espécie compartilham entre si. Com essa análise, dos 198 genes PGPs, foram encontrados, em todas as cepas, 188 genes. Utilizando o limiar igual a 80%, este número aumentou para 189 e se manteve igual para 70% e 60%.

Ainda com este último script, a partir do arquivo FASTA gerado (banco de dados final), realizou-se um alinhamento de sequências, primeiramente, somente com os 12 genomas *Bacillus velezensis*. A segunda análise ocorreu acrescentando a cepa *Bacillus subtilis* UD1022. Em uma terceira análise, adicionou-se a estirpe *Bacillus pumilus* WP8 e, por fim, acrescentou-se as cepas *Bacillus velezensis* LABIM40 e LABIM44. Essas análises foram realizadas considerando o percentual de identidade igual a 90%, 80%, 70% e 60%, cujos resultados encontram-se na Tabela 12.

**Tabela 12 – Resultados dos Alinhamento de Sequências do Banco de Dados com Genomas PGPRs**

Percentual Identidade	12 Cepas <i>B. velezensis</i>	<i>B. velezensis</i> (12)		<i>B. velezensis</i> (14)	
		<i>B. subtilis</i> (1)	<i>B. subtilis</i> (1)	<i>B. subtilis</i> (1)	<i>B. subtilis</i> (1)
90%	256	125	30	30	30
80%	376	299	133	133	133
70%	456	425	238	238	238
60%	493	479	336	336	336

**Fonte: Autoria Própria**

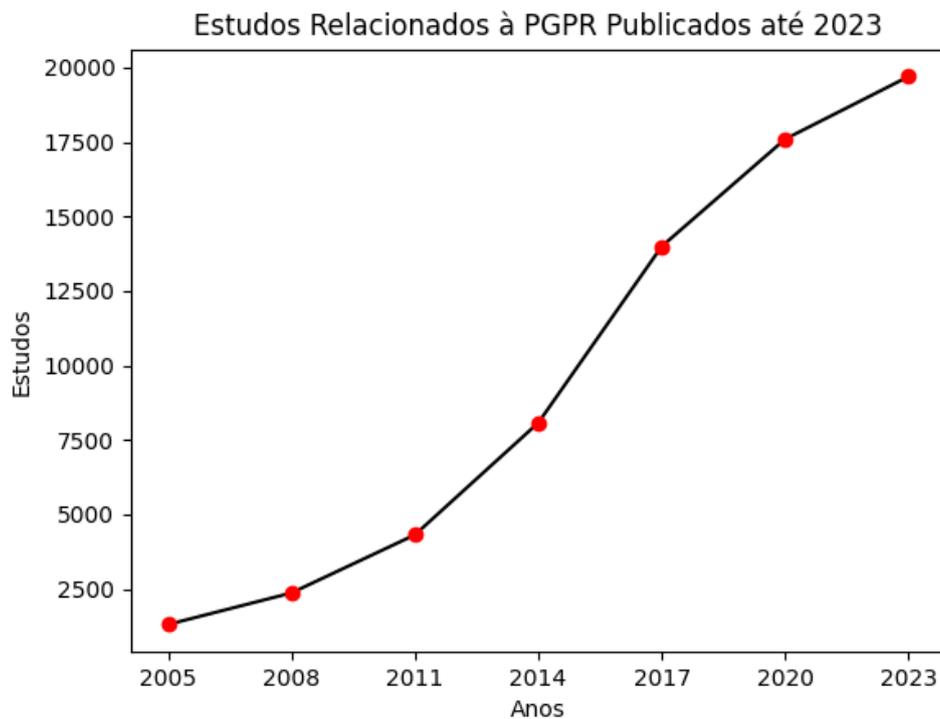
Na Tabela 13, encontram-se os mesmos resultados, mas como percentuais, para uma melhor visualização e compreensão dos resultados obtidos com as análises realizadas.

Ademais, a partir dos resultados obtidos ao realizar o estudo adicional de busca por estudos sobre rizobactérias promotoras de crescimento em plantas, publicados até 2023, com o objetivo de analisar o período de maior concentração de artigos publicados, optou-se por representar estes dados na forma de um gráfico, observado na Figura 22.

Tabela 13 – Percentual de Genes PGPs nos Genomas PGPRs

Percentual Identidade	12 Cepas	<i>B. velezensis</i> (12)		<i>B. velezensis</i> (12)	<i>B. velezensis</i> (14)
	<i>B. velezensis</i>	<i>B. subtilis</i> (1)	<i>B. subtilis</i> (1)	<i>B. subtilis</i> (1)	<i>B. subtilis</i> (1)
		<i>B. subtilis</i> (1)	<i>B. pumilus</i> (1)	<i>B. pumilus</i> (1)	<i>B. pumilus</i> (1)
90%	33,46%	16,34%		3,92%	3,92%
80%	49,15%	39,08%		17,39%	17,39%
70%	59,61%	55,56%		31,11%	31,11%
60%	64,44%	62,61%		43,92%	43,92%

Figura 22 – Quantidade de Estudos Relacionados à PGPR na plataforma Google Scholar



Fonte: Autoria Própria

Analisando a Figura 22, observa-se que, em relação à dez anos atrás, a quantidade de estudos publicados nessa área de pesquisa aumentou em três vezes. Nota-se, ainda, que entre 2014 e 2017 houve um aumento expressivo nessa quantidade.

Portanto, como um estudo futuro, objetiva-se realizar uma análise referente ao período dos últimos dez anos, a fim de mapear os genomas PGPR estudados na literatura durante esse período e, quando necessário, realizar a montagem e anotação destes genomas.

## 6 CONCLUSÕES E PERSPECTIVAS

O uso de microrganismos na prática agrícola têm sido bastante estudado e vêm ganhando cada vez mais credibilidade, principalmente no que se diz respeito ao aumento de produtividade e a diminuição dos impactos ambientais causados por fertilizantes químicos, denominada agricultura sustentável. Assim, como mostra a Figura 22, o interesse por essa área cresceu muito na última década, e o número de estudos publicados triplicou. Dessa forma, a proposta de um banco de dados de genes envolvidos na promoção de crescimento em plantas, pode corroborar em novas pesquisas de classificação *in silico* de novas possíveis PGPRs, por meio de algoritmos de Aprendizado de Máquina. Para isso, é imprescindível criar um banco de dados com um número significativo de amostras, de diversas espécies do gênero estudado (*Bacillus*).

No entanto, fazer uma mineração textual de genes foi uma tarefa exaustiva, visto que são raros os autores que descrevem quais são os genes envolvidos na promoção de crescimento vegetal, identificados na cepa estudada. Durante esta pesquisa, foram encontrados um total de sete estudos, totalizando oito genomas, dos quais somente seis puderam ser utilizados, pois dois deles, estudados em Chandra *et al.* (2021), tiveram seus genomas depositados como Arquivo de Leitura de Sequência (módulo SRA do NCBI).

Outra dificuldade encontrada foi em relação à identificação dos genes no genoma, pois algumas vezes, os nomes dos genes descritos pelos pesquisadores não condizia com a anotação do genoma, sendo necessária uma busca via alinhamento de sequências. Em alguns casos, os genes descritos não foram encontrados, principalmente para a cepa *Bacillus subtilis* BS96 que, inclusive, possui vários genes com nomes repetidos, dificultando ainda mais a identificação no genoma. Já para a cepa *Bacillus velezensis* FZB42, apenas três não foram encontrados e para as cepas *Bacillus velezensis* HNA3, Ag75 e LDO2, todos os genes descritos pelos autores, foram identificados.

Ainda em relação às dificuldades encontradas, a busca por artigos cujas bactérias são classificadas como PGPRs também não foi fácil, uma vez que vários artigos publicados descrevem a cepa, mas não identificam seu código de acesso e, muitas vezes, ao procurar somente pelo nome da cepa nos bancos de dados do NCBI, o resultado aparece como inexistente. Somado à isso, tem os casos em que os códigos são descritos, mas a forma de depósito do genoma no NCBI impossibilita a utilização do mesmo nesta pesquisa, pois são depositados como Arquivo de Leitura de Sequência (módulo SRA), por exemplo, como é o caso das cepas

estudadas em Chandra *et al.* (2021). Há também casos em que somente o genoma parcial foi depositado, ou casos em que é necessário entrar em contato com o autor para solicitar os dados de sequenciamento utilizados.

Diante disso, concomitante ao interesse de incorporar novas estirpes PGPRs à este estudo, destaca-se que uma etapa de montagem e anotação genômica se fará necessária, pois, desta forma, será ampliado o número de genomas PGPR aplicados por este estudo, enriquecendo os resultados obtidos. Ressalta-se, ainda, que esta mesma tarefa também poderá ser realizada para os genomas estudados em Chandra *et al.* (2021), possibilitando que os genes PGPs descritos pelos autores, sejam incorporados ao banco de dados.

Quanto aos resultados obtidos pelo pipeline, o primeiro script se mostrou muito eficaz ao realizar a busca, via *Entrez*, utilizando o código RefSeq, nos bancos de dados do NCBI. Como visto nos resultados, essa etapa do pipeline encontrou um número considerável dos genes PGPs buscados. Vale ressaltar que este script só não encontra todos os genes PGPs pelo fato de que as anotações de genomas depositadas no NCBI nem sempre identificam todas as informações necessárias dos genes.

Por sua vez, o segundo script foi essencial para encontrar os genes que ainda não haviam sido identificados no primeiro script. Os resultados dessa etapa foram analisados pelo terceiro script e, para todos os genomas incluídos nesta etapa do pipeline, a grande maioria dos seus genes descritos como PGPs nos artigos da literatura foram encontrados.

Dessa forma, ao utilizar os resultados obtidos nestes três scripts acima descritos, o quarto script realiza a comparação dos genes em comum em cada uma das estirpes utilizadas, possibilitando a visualização do Diagrama de Venn com a quantidade de genes encontrados.

Por fim, o quinto e último script, cria o banco de dados de genes relacionados a PGP e analisa a existência, via BLAST, destes genes nas cepas PGPRs selecionadas para esta etapa. Os resultados foram positivos e, mesmo com espécies que não estão próximas na árvore filogenética, foi encontrada uma grande quantidade de genes, sobretudo quando é diminuído o limiar de corte de identidade dos resultados desses alinhamentos.

Acrescido a este último script, foram utilizados os genes PGPs de outros dois genomas da espécie *Bacillus* (RZ2MS9 e BS96), cujos artigos descrevem, de forma específica, os seus genes relacionados à PGP. Por tal razão, a criação e utilização de scripts adicionais foram necessários, contudo, todos os scripts estão depositados no [GitHub](#) para serem analisados e/ou utilizados, se necessário.

Em relação ao banco de dados, utilizando um percentual de identidade igual a 90%, observou-se que para as cepas da espécie *Bacillus velezensis*, obteve-se um maior número de sequências similares em relação ao total de genes PGPs considerados (188/198), o que é intuitivo, uma vez que são da mesma espécie. Ainda, realizou-se outros três testes, utilizando os limiares iguais a 80%, 70% e 60%. Com isso, obteve-se um total de 189 genes em todas as análises, ou seja, mesmo diminuindo o percentual de similaridade, a quantidade encontrada não se alterou.

Dessa forma, é possível observar que, para os alinhamentos realizados somente com as cepas *Bacillus velezensis*, utilizando como banco de dados somente os genes das cepas PGPRs, também *Bacillus velezensis*, encontrou-se uma alta quantidade de genes similares em todos os 12 genomas, e usando os quatro valores distintos como limiar. Com isso, conclui-se que a verificação de presença dos genes do banco de dados proposto para estas cepas, por serem da mesma espécie, gerou resultados relevantes, já que apenas nove genes, dos 198, não são compartilhados por todos os genomas analisados.

Agora, conforme observado nas Tabelas 12 e 13, realizando o alinhamento de sequências com o banco de dados completo e os 12 genomas *Bacillus velezensis*, obteve-se um total de 256 genes. Ao acrescentar o *Bacillus subtilis*, de 765 genes, 125 foram encontrados em todas as 13 cepas (16,34%). Provavelmente este número não foi tão expressivo devido ao fato de comparar diferentes espécies de *Bacillus*, e considerar um percentual de similaridade alto (90%) para estes alinhamentos.

Adicionando o *Bacillus pumilus* nessa comparação, o total de genes, que antes era 125 de 765, passou a ser 30 de 765, ou seja, apenas 3,92%, reforçando o indicativo de que, possivelmente, o limiar de corte considerado para comparar espécies diferentes, foi muito alto.

Nesse sentido, realizou-se outras análises, considerando diferentes valores para este limiar. Primeiramente, utilizou-se o percentual igual a 80%. Com isso, houve um aumento expressivo na quantidade de genes, sendo de aproximadamente 47% o aumento de genes encontrados quando realizado o alinhamento do banco de dados somente com as 12 cepas *Bacillus velezensis* e de 39% ao acrescentar a *Bacillus subtilis*. No entanto, o maior aumento foi em relação ao alinhamento considerando todas as cepas selecionadas, cujo valor foi de quase cinco vezes maior.

Para a análise com o limiar igual a 70%, os valores das quantidades de genes continuaram aumentando. Em relação a primeira análise (90%), para a comparação do banco de dados com as 12 *Bacillus velezensis*, o aumento foi de aproximadamente 78%. Acrescentando a *Bacillus subtilis* e, posteriormente, a *Bacillus pumilus*, os valores aumentaram aproximadamente

em três e oito vezes, respectivamente.

Por fim, utilizando o limiar igual a 60%, na última análise realizada, em relação ao limiar de 90%, para o alinhamento com as 12 *Bacillus velezensis*, o aumento foi de 92,6%. Já para os alinhamentos acrescentando a cepa *Bacillus subtilis* e, posteriormente, *Bacillus pumilus*, a quantidade aumentou, respectivamente, em quatro e onze vezes, aproximadamente, sendo este último, o grupo cujo aumento foi mais expressivo ao variar o limiar de identidade. Ainda, tem-se que, acrescentando as cepas *Bacillus velezensis* LABIM40 e LABIM44, os resultados não foram alterados em relação ao grupo formado pelas 14 cepas PGPRs.

Assim, analisando os resultados obtidos, verifica-se que, a medida que novas cepas foram acrescentadas ao estudo, com estas sendo de diferentes espécies de *Bacillus*, a quantidade de genes PGPs compartilhados entre todas elas diminuiu. Além disso, a medida que o limiar de corte se torna mais flexível, variando entre 60 e 90%, mais genes vão sendo encontrados. Tais fatos reforçam a necessidade do banco de dados conter uma quantidade satisfatória de amostras, de espécies variadas de *Bacillus*. Caso essa variedade não aconteça, o banco poderá ser eficiente para inferir sobre a classificação PGPR de algumas espécies de *Bacillus*, como a *Bacillus velezensis*, que os resultados foram excelentes, mas não sobre outras espécies do mesmo gênero.

Destaca-se que todos os resultados aqui discutidos, com maiores detalhes, bem como todos os códigos e arquivos base utilizados, encontram-se no repositório do [GitHub](#). No mais, algumas perguntas ainda ficam em aberto, destacam-se elas:

- Pela observação dos resultados apresentados nesta pesquisa, como encontrar as características dos genomas de rizobactérias de diferentes espécies?
- Quais espécies de bactérias do gênero *Bacillus* considerar?
- O banco de dados será eficiente para a classificação de todas espécies de *Bacillus*? Ou bactérias não tão próximas evolutivamente, mesmo sendo do mesmo gênero, não poderão ser utilizadas?
- Qual o percentual mínimo de identidade a ser considerado para verificar a similaridade de genes em diferentes espécies de um mesmo gênero?
- Como discutido anteriormente no texto, alguns autores não depositam o genoma completo das cepas PGPR no NCBI, por vezes, somente os arquivos de alinhamento. Com isso, o

quão válido é realizar a montagem e anotação dos genomas, utilizando os arquivos de alinhamento disponibilizados pelos autores?

Portanto, tem-se como a principal perspectiva futura, analisar, de outras formas, como extrair as características que fazem com que uma rizobactéria seja classificada como promotora de crescimento em plantas, de modo a gerar mais informações para o banco de dados e, com isso, construir um modelo de Aprendizado de Máquina para classificar as bactérias PGPR, além de buscar responder todos os questionamentos que ficaram em aberto.

## REFERÊNCIAS

- ÅGREN, J. *et al.* Gegenees: fragmented alignment of multiple genomes for determining phylogenomic distances and genetic signatures unique for specified target groups. **PloS one**, Public Library of Science San Francisco, USA, v. 7, n. 6, p. e39107, 2012.
- ALIKHAN, N. *et al.* Blast ring image generator (brigs): simple prokaryote genome comparisons. **BMC genomics**, BioMed Central, v. 12, n. 1, p. 1–10, 2011.
- ARSHAD, M.; FRANKENBERGER JR, W. T. Plant growth-regulating substances in the rhizosphere: microbial production and functions. **Advances in agronomy**, Elsevier, v. 62, p. 45–151, 1997.
- BABALOLA, O. O. Beneficial bacteria of agricultural importance. **Biotechnology letters**, Springer, v. 32, n. 11, p. 1559–1570, 2010.
- BALDANI, J. I. *et al.* Characterization of herbaspirillum seropedicae gen. nov., sp. nov., a root-associated nitrogen-fixing bacterium. **International Journal of Systematic and Evolutionary Microbiology**, Microbiology Society, v. 36, n. 1, p. 86–93, 1986.
- BATISTA, B. D. **Promoção de crescimento em milho (*Zea mays L.*) por rizobactérias associadas á cultura do guaranazeiro (*Paullinia cupana var. sorbilis*)**. 2012. Tese (Doutorado) — Universidade de São Paulo, 2012.
- BATISTA, B. D. **Promoção de crescimento vegetal por *Bacillus sp.* RZ2MS9: dos genes ao campo**. 2017. Tese (Doutorado) — Universidade de São Paulo, 2017.
- BIANCO, C. *et al.* Indole-3-acetic acid improves escherichia coli's defences to stress. **Archives of Microbiology**, Springer, v. 185, n. 5, p. 373–382, 2006.
- BLAST. Building a blast database with your (local) sequences. *In*: NCBI (Ed.). Bethesda, Maryland, USA: NCBI, 2008. p. 51–52.
- BLIN, K. *et al.* antiSMASH: 6.0: improving cluster detection and comparison capabilities. **Nucleic acids research**, Oxford University Press, v. 49, n. W1, p. W29–W35, 2021.
- BRAUN, E. **Reactive nitrogen in the environment: too much or too little of a good thing**. Paris, França: UNEP/Earthprint, 2007.

BURR, T. J.; SCHROTH, M. N.; SUSLOW, T. Increased potato yields by treatment of seed pieces with specific strains of *Pseudomonas fluorescens* and *P. putida*. **Phytopathology**, v. 68, n. 9, p. 1377–1383, 1978.

CABAJ, A.; KOSAKOWSKA, A. Iron-dependent growth of and siderophore production by two heterotrophic bacteria isolated from brackish water of the southern Baltic Sea. **Microbiological Research**, Elsevier, v. 164, n. 5, p. 570–577, 2009.

CAMACHO, C. *et al.* Blast+: architecture and applications. **BMC Bioinformatics**, v. 10, n. 421, 2009.

CARPENTER, S. R. *et al.* Nonpoint pollution of surface waters with phosphorus and nitrogen. **Ecological Applications**, Wiley Online Library, v. 8, n. 3, p. 559–568, 1998.

CHANDRA, A.; CHANDRA, P.; TRIPATHI, P. Whole genome sequence insight of two plant growth-promoting bacteria (*B. subtilis* BS87 and *B. megaterium* BM89) isolated and characterized from sugarcane rhizosphere depicting better crop yield potentiality. **Microbiological Research**, Elsevier, v. 247, p. 126733, 2021.

CHANDRA, P.; TRIPATHI, P.; CHANDRA, A. Isolation and molecular characterization of plant growth-promoting *Bacillus* spp. and their impact on sugarcane (*Saccharum* spp. hybrids) growth and tolerance towards drought stress. **Acta Physiologiae Plantarum**, Springer, v. 40, n. 11, p. 1–15, 2018.

CHEN, L. *et al.* Antimicrobial, plant growth-promoting and genomic properties of the peanut endophyte *Bacillus velezensis* ldo2. **Microbiological Research**, Elsevier, v. 218, p. 41–48, 2019.

CHEN, L. *et al.* *Bacillus velezensis* cla178-induced systemic resistance of *Rosa multiflora* against crown gall disease. **Frontiers in Microbiology**, Frontiers Media SA, v. 11, p. 587667, 2020.

CHEN, X. H. *et al.* Comparative analysis of the complete genome sequence of the plant growth-promoting bacterium *Bacillus amyloliquefaciens* FZB42. **Nature Biotechnology**, Nature Publishing Group, v. 25, n. 9, p. 1007–1014, 2007.

CHEN, X. H. *et al.* Genome analysis of *Bacillus amyloliquefaciens* FZB42 reveals its potential for biocontrol of plant pathogens. **Journal of Biotechnology**, Elsevier, v. 140, n. 1–2, p. 27–37, 2009.

CHOUDHARY, D. K.; JOHRI, B. N. Interactions of *Bacillus* spp. and plants—with special reference to induced systemic resistance (ISR). **Microbiological Research**, Elsevier, v. 164, n. 5, p. 493–513, 2009.

CHUN, J.; RAINEY, F. A. Integrating genomics into the taxonomy and systematics of the bacteria and archaea. **International journal of systematic and evolutionary microbiology**, Microbiology Society, v. 64, n. Pt\_2, p. 316–324, 2014.

COCK, P. J. A. *et al.* Biopython: freely available python tools for computational molecular biology and bioinformatics. **Bioinformatics**, Oxford University Press, v. 25, n. 11, p. 1422–1423, 2009.

COSTA, P. B. da. **Isolamento e Caracterização de Bactérias Promotoras de Crescimento Vegetal de Lavouras Experimentais de Arroz sob Diferentes Níveis de Fertilização**. 2012. Dissertação (Mestrado) — Instituto de Biociências, Universidade Federal do Rio Grande do Sul, 2012.

CROWLEY, D. E.; REID, C. P. P.; SZANISZLO, P. J. Utilization of microbial siderophores in iron acquisition by oat. **Plant Physiology**, American Society of Plant Biologists, v. 87, n. 3, p. 680–685, 1988.

DIAS, V. P.; FERNANDES, E. **Fertilizantes: uma visão global sintética**. 2006.

DOBBELAERE, S.; VANDERLEYDEN, J.; OKON, Y. Plant growth-promoting effects of diazotrophs in the rhizosphere. **Critical reviews in plant sciences**, Taylor & Francis, v. 22, n. 2, p. 107–149, 2003.

DONG, X. *et al.* The genome of bacillus velezensis sc60 provides evidence for its plant probiotic effects. **Microorganisms**, MDPI, v. 10, n. 4, p. 767, 2022.

E-ESCOLA. **A NCBI (National Center for Biotechnology Information)**. 2005. Acesso em: 06 abr. 2022. Disponível em: <http://e-escola.tecnico.ulisboa.pt/biologia/a+ncai+national+center+for+biotechnology+information>.

EL-DAIM, I. A. Abd; BEJAI, S.; MEIJER, J. Bacillus velezensis 5113 induced metabolic and molecular reprogramming during abiotic stress tolerance in wheat. **Scientific Reports**, Nature Publishing Group UK London, v. 9, n. 1, p. 16282, 2019.

EVENSON, R. E.; GOLLIN, D. Assessing the impact of the green revolution, 1960 to 2000. **science**, American Association for the Advancement of Science, 2003.

FERNÁNDEZ, A. *et al.* Nuevo aislado de Bacillus y su utilización para el control de hongos fitopatógenos. **Agroecologia. net**, v. 6, p. 50–57, 2004.

FERREIRA, Y. C. **Estratégias de Paralelismo na Busca por Similaridades em Sequências de DNA e Proteínas Junto ao MPIBLAST**. 2005. Dissertação (Mestrado) — Instituto Alberto

Luiz Coimbra de Pós-Graduação e Pesquisa em Engenharia - Universidade Federal do Rio de Janeiro, 2005.

FRANCHE, C.; LINDSTRÖM, K.; ELMERICH, C. Nitrogen-fixing bacteria associated with leguminous and non-leguminous plants. **Plant and soil**, Springer, v. 321, n. 1, p. 35 – 59, 2009.

GEELS, F. P.; SCHIPPERS, B. Reduction of yield depressions in high frequency potato cropping soil after seed tuber treatments with antagonistic fluorescent pseudomonas spp. **Journal of Phytopathology**, Wiley Online Library, v. 108, n. 3-4, p. 207–214, 1983.

GLICK, B. R. The enhancement of plant growth by free-living bacteria. **Canadian journal of microbiology**, NRC Research Press Ottawa, Canada, v. 41, n. 2, p. 109–117, 1995.

GLICK, B. R. Modulation of plant ethylene levels by the bacterial enzyme acc deaminase. **FEMS microbiology letters**, Blackwell Publishing Ltd Oxford, UK, v. 251, n. 1, p. 1–7, 2005.

GLICK, B. R. Plant growth-promoting bacteria: mechanisms and applications. **Scientifica**, Hindawi, v. 2012, 2012.

GLICK, B. R. *et al.* **Biochemical and Genetic Mechanisms Used by Plant Growth Promoting Bacteria**. Londres: Imperial College Press, 1999.

GORIS, J.; KONSTANTINIDIS, K. T.; KLAPPENBACH, J. A.; COENYE, T.; VANDAMME, P.; TIEDJE, J. M. Dna–dna hybridization values and their relationship to whole-genome sequence similarities. **International journal of systematic and evolutionary microbiology**, Society for General Microbiology, v. 57, n. 1, p. 81–91, 2007.

HARTHMANN, O. E. L. **Microbioliza ção de Sementes com Rizobactérias na Produção de Cebola**. 2009. Tese (Doutorado) — Universidade Federal do Paraná, 2009.

HAYAT, R. *et al.* Soil beneficial bacteria and their role in plant growth promotion: a review. **Annals of microbiology**, Springer, v. 60, n. 4, p. 579–598, 2010.

HE, P. *et al.* Complete genome sequence of bacillus velezensis th-1, a candidate biocontrol bacterium from china. **Molecular Plant-Microbe Interactions**, Am Phytopath Society, v. 36, n. 5, p. 305–308, 2023.

HERMOSA, R. *et al.* Plant-beneficial effects of trichoderma and of its genes. **Microbiology**, Microbiology Society, v. 158, n. 1, p. 17–25, 2012.

HOLFORD, I. C. R. Soil phosphorus: its measurement, and its uptake by plants. **Soil Research**, CSIRO Publishing, v. 35, n. 2, p. 227–240, 1997.

HOWIE, W. J.; ECHANDI, E. Rhizobacteria: Influence of cultivar and soil type on plant growth and yield of potato. **Soil Biology and Biochemistry**, Elsevier, v. 15, n. 2, p. 127–132, 1983.

HUANG, R. *et al.* Identification of adhesins in plant beneficial rhizobacteria bacillus velezensis sqr9 and their effect on root colonization. **Molecular Plant-Microbe Interactions**, Am Phytopath Society, v. 35, n. 1, p. 64–72, 2022.

HUSON, D. H.; BRYANT, D. Application of phylogenetic networks in evolutionary studies. **Molecular biology and evolution**, Oxford University Press, v. 23, n. 2, p. 254–267, 2006.

IDRISS, E. E. *et al.* Extracellular phytase activity of *Bacillus amyloliquefaciens* FZB45 contributes to its plant-growth-promoting effect. **Microbiology**, v. 148, n. 7, p. 2097–2109, 2002.

ISHERWOOD, K. F. Mineral fertilizer use and the environment by International Fertilizer Industry Association. **Revised Edition, Paris**, 2000.

ITTERSUM, M. K. VAN *et al.* Yield gap analysis with local to global relevance – A review. **Field Crops Research**, Elsevier, v. 143, p. 4 – 17, 2013.

JACKSON, M. B. Ethylene in root growth and development. *In: The plant hormone ethylene*. EUA: CRC Press, 1991. p. 159 – 181.

JHA, C. K.; SARAF, M. Plant growth promoting rhizobacteria (PGPR): a review. **Journal of Agricultural Research and Development**, v. 5, n. 2, p. 108 – 119, 2015.

JONES, D. L. Organic acids in the rhizosphere – a critical review. **Plant and soil**, Springer, v. 205, n. 1, p. 25 – 44, 1998.

JOUNG, K. B.; CÔTÉ, J. C. Evaluation of ribosomal RNA gene restriction patterns for the classification of *Bacillus* species and related genera. **Journal of applied microbiology**, Wiley Online Library, v. 92, n. 1, p. 97 – 108, 2002.

KANG, Y. *et al.* A possible mechanism of action of plant growth-promoting rhizobacteria (pgpr) strain bacillus pumilus wp8 via regulation of soil bacterial community structure. **The Journal of general and applied microbiology**, Applied Microbiology, Molecular and Cellular Biosciences Research Foundation, v. 59, n. 4, p. 267–277, 2013.

KENDE, H.; ZEEVAART, J. The Five “Classical” Plant Hormones. **The plant cell**, American Society of Plant Biologists, v. 9, n. 7, p. 1197, 1997.

KLOEPPER, J. W. Plant growth-promoting rhizobacteria on radishes. *In: Proc. of the 4th Internet. Conf. on Plant Pathogenic Bacter, Station de Pathologie Vegetale et Phytobacteriologie*. Franca: INRA, 1978. v. 2, p. 879 – 882.

KLOEPPER, J. W. *et al.* Enhanced plant growth by siderophores produced by plant growth-promoting rhizobacteria. **Nature**, Nature Publishing Group, v. 286, n. 5776, p. 885 – 886, 1980.

KLOEPPER, J. W.; RYU, C.; ZHANG, S. Induced systemic resistance and promotion of plant growth by bacillus spp. **Phytopathology**, Am Phytopath Society, v. 94, n. 11, p. 1259 – 1266, 2004.

LAMBRECHT, M. *et al.* Indole-3-acetic acid: a reciprocal signalling molecule in bacteria–plant interactions. **Trends in microbiology**, Elsevier, v. 8, n. 7, p. 298 – 300, 2000.

LAU, S. S. S.; LANE, S. N. Biological and chemical factors influencing shallow lake eutrophication: a long-term study. **Science of the Total Environment**, Elsevier, v. 288, n. 3, p. 167–181, 2002.

LEE, I.; KIM, Y. O.; PARK, S.; CHUN, J. Orthoani: an improved algorithm and software for calculating average nucleotide identity. **International journal of systematic and evolutionary microbiology**, Microbiology Society, v. 66, n. 2, p. 1100–1103, 2016.

LU, S. *et al.* Complete genome sequence of bacillus subtilis cnbg-pgpr-1 for studying the promotion of plant growth. **Molecular Plant-Microbe Interactions**, Am Phytopath Society, v. 35, n. 12, p. 1115–1119, 2022.

LUSTER, J. *et al.* Sampling, defining, characterising and modeling the rhizosphere – the soil science tool box. **Plant and Soil**, Springer, v. 321, n. 1, p. 457 – 482, 2009.

LYNCH, J. M. Introduction: Some consequences of microbial rhizosphere complete for plant and soil. **The rhizosphere**, John Wiley & Sons Ltd., p. 1 – 10, 1990.

MACLEAN, D.; JONES, J. D. G.; STUDHOLME, D. J. Application of ‘next-generation’ sequencing technologies to microbial genetics. **Nature Reviews Microbiology**, Nature Publishing Group, v. 7, n. 4, p. 96 – 97, 2009.

MADDEN, T. The BLAST Sequence Analysis Tool. *In: NCBI (Ed.). Bethesda, Maryland, USA: NCBI, 2013. p. 361 – 370.*

MAHANTY, T. *et al.* Biofertilizers: a potential approach for sustainable agriculture development. **Environmental Science and Pollution Research**, Springer, v. 24, n. 4, p. 3315 – 3335, 2017.

MARIANO, D. *et al.* **Introdução à Programação para Bioinformática com Biopython**. Minas Gerais: Amazon, Brasil, 2016.

MARQUES, G. **Bancos de Dados Biológicos (parte I): o NCBI**. 2021. Acesso em: 06 abr. 2022. Disponível em: <https://medium.com/omixdata/bancos-de-dados-biol%C3%B3gicos-parte-i-o-ncbi-c16dfc1b0a84>.

MEDEMA, M. H. *et al.* antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. **Nucleic acids research**, Oxford University Press, v. 39, n. suppl\_2, p. W339–W346, 2011.

MELO, I. S. de. Rizobactérias Promotoras de Crescimento de Plantas: Descrição e Potencial de uso na Agricultura. *In*: MELO, Itamar Soares de; AZEVEDO, João Lúcio de (Ed.). **Ecologia Microbiana**. Jaguariúna: Embrapa Meio Ambiente, 1998. cap. 3, p. 87 – 116.

MOAT, A. G.; FOSTER, J. W.; SPECTOR, M. P. Nitrogen metabolism. *In*: **Microbial Physiology**. 4. ed. New York: John Wiley & Sons, 1995. cap. 14, p. 475 – 502.

MOSELA, M. *et al.* Bacillus velezensis strain ag75 as a new multifunctional agent for biocontrol, phosphate solubilization and growth promotion in maize and soybean crops. **Scientific reports**, Nature Publishing Group, v. 12, n. 1, p. 1–12, 2022.

MOUNT, D. W. **Bioinformatics: Sequence and Genome Analysis**. 2. ed. New York, USA: Cold Spring Harbor Laboratory Press, 2004.

NEWTON, W. E. Nitrogen fixation in perspective. *In*: **Nitrogen fixation: from molecules to crop productivity**. Dordrecht: Springer, 2000. p. 3 – 8.

NGUGI, H. K. *et al.* Effect of flower-applied serenade biofungicide (*Bacillus subtilis*) on pollination-related variables in rabbiteye blueberry. **Biological control**, Elsevier, v. 33, n. 1, p. 32 – 38, 2005.

OROZCO-MOSQUEDA, M. del C.; GLICK, B. R.; SANTOYO, G. Acc deaminase in plant growth-promoting bacteria (PGPB): an efficient mechanism to counter salt stress in crops. **Microbiological research**, Elsevier, v. 235, p. 126439, 2020.

PANDIN, C. *et al.* Complete genome sequence of bacillus velezensis qst713: A biocontrol agent that protects agaricus bisporus crops against the green mould disease. **Journal of Biotechnology**, Elsevier, v. 278, p. 10–19, 2018.

PATERSON, J. *et al.* The contribution of genome mining strategies to the understanding of active principles of pgpr strains. **FEMS microbiology ecology**, Oxford University Press, v. 93, n. 3, 2017.

PATTEN, C. L.; GLICK, B. R. Role of pseudomonas putida indoleacetic acid in development of the host plant root system. **Applied and environmental microbiology**, Am Soc Microbiol, v. 68, n. 8, p. 3795–3801, 2002.

PAUL, E. A.; CLARK, F. E. Reduction and transport of nitrate. *In: Soil Microbiology and Biochemistry*. San Diego, Califórnia: Academic Press, 1989. p. 147 – 163.

PEARSON, W. R. Selecting the right similarity-scoring matrix. **Current protocols in bioinformatics**, v. 43, 2013.

PROPEQ. **Fertilizantes: processo produtivo e considerações ambientais**. 2020. Acesso em: 06 abr. 2022. Disponível em: <https://propeq.com/fertilizantes/>.

PROSDOCIMI, F. *et al.* Bioinformática: manual do usuário. **Biotecnologia Ciência & Desenvolvimento**, v. 29, p. 12–25, 2002.

PRUITT, K. *et al.* The Reference Sequence (RefSeq) Database. *In: NCBI (Ed.)*. Bethesda, Maryland, USA: NCBI, 2002. p. 307 – 328.

QUEIROZ, B. P. V. de. **Isolamento e Seleção de Rizobactérias para a Promoção de Crescimento e Controle de Phytophthora Parasítica em Citros**. 2003. Tese (Doutorado) — Instituto de Biociências, Universidade Estadual Paulista, 2003.

RAGAZZO-SÁNCHEZ, J. A. *et al.* Selección de cepas de Bacillus spp. productoras de antibióticos aisladas de frutos tropicales. **Revista Chapingo. Serie horticultura**, Universidad Autónoma Chapingo, v. 17, n. SPE1, p. 5 – 11, 2011.

RATZ, R. J. **Uso de Bacterias do Gênero Bacillus como Promotoras de Crescimento para a Cultura do Milho e da Soja**. 2014. Dissertação (Mestrado) — Centro de Engenharias e Ciências Exatas, Universidade Estadual do Oeste do Paraná - UNIOESTE, 2014.

RAVEN, P. H.; EVERT, R. F.; EICHHORN, S. E. Regulação do crescimento e do desenvolvimento: Os hormônios vegetais. *In: Biologia Vegetal*. 5. ed. Rio de Janeiro: Editora Guanabara Koogan, 1996. cap. 25, p. 509 – 533.

REVA, O. N. *et al.* Complete genome sequence and epigenetic profile of bacillus velezensis ucmb5140 used for plant and crop protection in comparison with other plant-associated bacillus strains. **Applied microbiology and biotechnology**, Springer, v. 104, p. 7643–7656, 2020.

RICHTER, M.; ROSSELLÓ-MÓRA, R. Shifting the genomic gold standard for the prokaryotic species definition. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 106, n. 45, p. 19126–19131, 2009.

RODRÍGUEZ, H.; FRAGA, R. Phosphate solubilizing bacteria and their role in plant growth promotion. **Biotechnology Advances**, Elsevier, v. 17, n. 4 - 5, p. 319 – 339, 1999.

RÖMHELD, V.; KIRKBY, E. A. Research on potassium in agriculture: needs and prospects. **Plant and soil**, Springer, v. 335, n. 1, p. 155 – 180, 2010.

ROSENZWEIG, C. *et al.* Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. **Proceedings of the national academy of sciences**, National Acad Sciences, v. 111, n. 9, p. 3268 – 3273, 2014.

ROSIER, A.; B., Pascale B.; BAIS, H. P. Quorum quenching activity of the pgpr bacillus subtilis ud1022 alters nodulation efficiency of sinorhizobium meliloti on medicago truncatula. **Frontiers in Microbiology**, Frontiers Media SA, v. 11, p. 596299, 2021.

ROZADA, C. **Bacillus spp.** São Paulo, 2022. Acesso em: 02 mai. 2022. Disponível em: <http://gmicsesalq.com.br/genero-bacillus-spp/>.

SAHARAN, B. S.; NEHRA, V. Plant growth promoting rhizobacteria: a critical review. **Life Sci Med Res**, v. 21, n. 1, p. 30, 2011.

SAYERS, E. Sample Applications of the E-utilities. *In*: NCBI (Ed.). Bethesda, Maryland, USA: NCBI, 2009. p. 27–38.

SIBPONKRUNG, S. *et al.* Genome sequence of bacillus velezensis s141, a new strain of plant growth-promoting rhizobacterium isolated from soybean rhizosphere. **Genome Announcements**, Am Soc Microbiol, v. 5, n. 48, p. 10–1128, 2017.

SIMS, R. E. H. *et al.* “Energy-smart” food for people and climate: issue paper. Food and Agriculture Organization of the United Nations (FAO), 2011.

SOKOLOVA, M. G.; AKIMOVA, G. P.; VAISHLYA, O. B. Effect of phytohormones synthesized by rhizosphere bacteria on plants. **Applied Biochemistry and Microbiology**, Springer, v. 47, n. 3, p. 274 – 278, 2011.

SOUZA, A. K. R.; MORASSUTI, C. Y.; DE DEUS, W. B. Poluição do ambiente por metais pesados e utilização de vegetais como bioindicadores. **Acta Biomédica Brasiliensia**, Faculdade de Ciências Biológicas e da Saúde, v. 9, n. 3, p. 95 – 106, 2018.

SOUZA, R. de; AMBROSINI, A.; PASSAGLIA, L. M. P. Plant growth-promoting bacteria as inoculants in agricultural soils. **Genetics and molecular biology**, SciELO Brasil, v. 38, p. 401 – 419, 2015.

SUSLOW, T.V. Rhizobacteria of sugar beets: effect of seed application and root colonization on yield. **Phytopathology**, v. 72, p. 199 – 206, 1982.

TEIXEIRA, G. M. *et al.* Genomic insights into the antifungal activity and plant growth-promoting ability in *Bacillus velezensis* CMRP 4490. **Frontiers in microbiology**, Frontiers, v. 11, p. 3495, 2021.

THAKKER, J. N.; PATEL, S.; DHANDHUKIA, P. C. Induction of defense-related enzymes in banana plants: Effect of live and dead pathogenic strain of *fusarium oxysporum* f. sp. cubense. **International Scholarly Research Notices**, Hindawi, v. 2013, 2013.

TOYOTA, K. *Bacillus*-related spore formers: attractive agents for plant growth promotion. **Microbes and Environments**, Japanese Society of Microbial Ecology/Japanese Society of Soil Microbiology . . . , v. 30, n. 3, p. 205 – 207, 2015.

TURNBULL, P. C. B.; KRAMER, J. M.; MELLING, J. *Bacillus*. **Medical microbiology**, The University of Texas Medical Branch at Galveston Galveston, 1996.

VAN LOON, L. C.; BAKKER, P. A. H. M.; PIETERSE, C. M. J. Systemic resistance induced by rhizosphere bacteria. **Annual review of phytopathology**, v. 36, n. 1, p. 453 – 483, 1998.

VAN REES, H. *et al.* Leading farmers in south east australia have closed the exploitable wheat yield gap: prospects for further improvement. **Field Crops Research**, Elsevier, v. 164, p. 1 – 11, 2014.

VEJAN, P. *et al.* Role of plant growth promoting rhizobacteria in agricultural sustainability — a review. **Molecules**, Multidisciplinary Digital Publishing Institute, v. 21, n. 5, p. 573, 2016.

VESSEY, J. K. Plant growth promoting rhizobacteria as biofertilizers. **Plant and soil**, Springer, v. 255, n. 2, p. 571 – 586, 2003.

VICENTE, E. J. **Introdução à Bioinformática - do DNA à proteína**: Capítulo 2: Introdução aos Bancos de Dados Primários (“Primary Data Banks”) – Roteiro Teórico Prático. 2020. Disponível em: [https://edisciplinas.usp.br/pluginfile.php/6311518/mod\\_resource/content/6/2.%20Cap%202.%20Bete.Bioinformatica.2019.06.03.pdf](https://edisciplinas.usp.br/pluginfile.php/6311518/mod_resource/content/6/2.%20Cap%202.%20Bete.Bioinformatica.2019.06.03.pdf).

WANG, J. *et al.* Complete genome sequencing of *bacillus velezensis* wrn014, and comparison with genome sequences of other *bacillus velezensis* strains. Korean Society for Microbiology and Biotechnology, 2019.

YAN, Y. *et al.* Complete genome sequence of *bacillus velezensis* yyc, a bacterium isolated from the tomato rhizosphere. **Archives of Microbiology**, Springer, v. 204, p. 1–5, 2022.

YAO, A. V. *et al.* Effect of FZB 24 *Bacillus subtilis* as a biofertilizer on cotton yields in field tests. **Archives of Phytopathology and Plant Protection**, Taylor & Francis, v. 39, n. 4, p. 323 – 328, 2006.

YE, M. *et al.* Characteristics and application of a novel species of *Bacillus*: *Bacillus velezensis*. **ACS chemical biology**, v. 13, n. 3, p. 500 – 505, 2018.

YUAN, C. *et al.* Effect of different fertilization treatments on indole-3-acetic acid producing bacteria in soil. **Journal of Soils and Sediments**, Springer, v. 11, n. 2, p. 322 – 329, 2011.

YUAN, J. *et al.* Plant growth-promoting rhizobacteria strain *Bacillus amyloliquefaciens* njn-6-enriched bio-organic fertilizer suppressed fusarium wilt and promoted the growth of banana plants. **Journal of agricultural and food chemistry**, ACS Publications, v. 61, n. 16, p. 3774–3780, 2013.

ZABLOTOWICZ, R. M. *et al.* Plant growth promotion mediated by bacterial rhizosphere colonizers. *In: The rhizosphere and plant growth*. Dordrecht: Springer, 1991. p. 315 – 326.

ZAID, D. S. *et al.* Comparative Genome Analysis Reveals Phylogenetic Identity of *Bacillus velezensis* HNA3 and Genomic Insights into Its Plant Growth Promotion and Biocontrol Effects. **Microbiology spectrum**, Am Soc Microbiol, v. 10, n. 1, p. e02169 – 21, 2022.

ZHANG, N. *et al.* Whole transcriptomic analysis of the plant-beneficial rhizobacterium *Bacillus amyloliquefaciens* sqr9 during enhanced biofilm formation regulated by maize root exudates. **BMC genomics**, Springer, v. 16, p. 1–20, 2015.

**APÊNDICE A – GENES PGP *BACILLUS VELEZENSIS* HNA3**

<b>RefSeq</b>	<b>Nome do Gene</b>	<b>Produto</b>
NP_390144.1	<i>trpA</i>	<i>Tryptophan synthase alpha chain</i>
NP_390145.2	<i>trpB</i>	<i>Tryptophan synthase beta chain</i>
WP_007409461.1	<i>trpC</i>	<i>Indole-3-glycerol phosphate synthase</i>
NP_390148.2	<i>trpD</i>	<i>Anthranilate phosphoribosyl transferase</i>
NP_390149.1	<i>trpE</i>	<i>Anthranilate synthase component I</i>
NP_389813.1	<i>dhaS</i>	<i>Putative aldehyde dehydrogenase</i>
NP_391585.1	<i>ywkB</i>	<i>Auxin efflux carrier</i>
NP_388806.2	<i>yhcX</i>	<i>Putative amidohydrolase</i>
NP_391630.1	<i>speE</i>	<i>Spermidine synthase, polyamine metabolism</i>
NP_390709.1	<i>ilvB</i>	<i>Acetolactate synthase</i>
NP_388505.1	<i>bdhA</i>	<i>2,3-butanediol dehydrogenase</i>
NP_390849.1	<i>acuC</i>	<i>Acetoin dehydrogenase</i>
NP_387974.1	<i>cysE</i>	<i>Serine O-acetyltransferase</i>
NP_389616.1	<i>hfq</i>	<i>Hfq RNA chaperone</i>
NP_390341.1	<i>sinR</i>	<i>Master regulator for biofilm formation</i>
NP_390945.1	<i>luxS</i>	<i>S-ribosyl homocysteine lyase</i>
WP_016937868.1	<i>rpoN</i>	<i>RNA polymerase sigma-54</i>
WP_003151402.1	<i>csrA</i>	<i>Carbon storage regulator</i>
NP_391423.1	<i>flgM</i>	<i>Flagellar biosynthesis anti-sigma factor</i>
WP_012117858.1	<i>trpF</i>	<i>Phosphoribosyl anthranilate isomerase</i>
WP_041482135.1	<i>ysnE</i>	<i>N-acetyltransferase</i>
BBA76563.1	<i>phytase</i>	<i>Mineralize organic phosphorus enzyme</i>
URM46110.1	<i>pksS</i>	<i>Polyketide biosynthesis cytochrome P450</i>
AHZ17904.1	<i>msmX</i>	<i>ABC-type spermidine transport systems</i>
QIR31908.1	<i>speG</i>	<i>Spermidine acetyltransferas</i>
URM44102.1	<i>alsD</i>	<i>Acetolactate decarboxylase</i>
USQ53343.1	<i>alsR</i>	<i>Transcriptional regulator</i>
BCU86156.1	<i>ylbF</i>	<i>Controls biofilm development</i>
URM46094.1	<i>ymcA</i>	<i>Biofilm development</i>
URM43645.1	<i>iolU</i>	<i>Scyllo-inositol 2-dehydrogenase (NADP(+))</i>
URM43968.1	<i>slrR</i>	<i>Master regulator for biofilm formation</i>
SIR80373.1	<i>fliA</i>	<i>RNA polymerase sigma factor</i>
WP_012117833.1	<i>crr</i>	<i>Sugar-specific IIA component</i>

*(Continua na próxima página)*

*(Continuação)*

<b>RefSeq</b>	<b>Nome do Gene</b>	<b>Produto</b>
WP_012118583.1	<i>wecB</i>	<i>UDP-N-acetylglucosamine 2-epimerase</i>
WP_012118587.1	<i>tarA/tagA</i>	<i>Acetylglucosaminyl diphosphoundecaprenol N-acetyl-beta-D-mannosaminyl transferase</i>

**Fonte: Adaptado de Zaid et al. (2022)**

## APÊNDICE B – GENES PGP *BACILLUS VELEZENSIS* AG75

<b>Código Ref-Seq</b>	<b>Nome do Gene</b>	<b>Produto</b>
NP_390144.1	<i>trpA</i>	<i>Tryptophan synthase subunit alpha</i>
WP_014305305.1	<i>trpB</i>	<i>Tryptophan synthase subunit beta</i>
WP_007409461.1	<i>trpC</i>	<i>Indole-3-glycerol phosphate synthase TrpC</i>
NP_390148.2	<i>trpD</i>	<i>Anthranilate phosphoribosyltransferase</i>
NP_390149.1	<i>trpE</i>	<i>Anthranilate synthase component I</i>
NP_391675.2	<i>ywdH</i>	<i>Aldehyde dehydrogenase</i>
NP_391585.1	<i>ywkB</i>	<i>Auxin efflux carrier family protein</i>
NP_391630.1	<i>speE</i>	<i>Spermidine synthase</i>
NP_391481.1	<i>budA</i>	<i>Acetolactate decarboxylase</i>
NP_391482.2	<i>alsS</i>	<i>Acetolactate synthase</i>
NP_388505.1	<i>bdhA</i>	<i>2,3-Butanediol dehydrogenase</i>
NP_388687.1	<i>acoA</i>	<i>Acetoin dehydrogenase</i>
NP_389529.1	<i>sigD</i>	<i>RNA polymerase sigma factor for flagellar operon and Biofilm formation</i>
NP_390341.1	<i>sinR</i>	<i>Master regulator of biofilm formation</i>
NP_390945.1	<i>luxS</i>	<i>S-Ribosyl homocysteine lyase for Quorum sensing Biofilm formation</i>
WP_015418263.1	<i>rpoN</i>	<i>RNA polymerase sigma-54 factor for Biofilm formation</i>
WP_003151402.1	<i>csrA</i>	<i>Carbon storage regulator for Biofilm formation</i>
NP_391423.1	<i>flgM</i>	<i>Negative regulator of flagellin synthesis flgM for Biofilm formation</i>
WP_012118583.1	<i>wecB</i>	<i>UDP-N-acetylglucosamine 2-epimerase for Biofilm formation</i>
NP_388275.1	<i>gdh</i>	<i>Glucose 1-dehydrogenase</i>
NP_390266.2	<i>zwf</i>	<i>Glucose-6-phosphate dehydrogenase</i>
WP_014417186.1	<i>treP</i>	<i>PTS system trehalose-specific EIIBC component</i>
NP_388663.1	<i>treR</i>	<i>Trehalose operon repressor</i>
NP_387940.1	<i>hslR</i>	<i>Ribosomal RNA binding protein involved in 50S recycling heat shock protein</i>
NP_388484.1	<i>groEL</i>	<i>Heat shock protein 60 kDa family chaperone GroEL</i>
NP_388483.2	<i>groES</i>	<i>Heat shock protein 10 kDa family chaperone GroES</i>
WP_003224086.1	<i>cspC</i>	<i>Cold shock protein CspC</i>
WP_003728273.1	<i>cspB</i>	<i>Cold shock-like protein CspB</i>
WP_003230776.1	<i>cspD</i>	<i>Cold-shock protein CspD</i>
NP_388180.2	<i>opuAA</i>	<i>Glycine/proline betaine ABC transporter ATP-binding protein OpuAA</i>
NP_388181.1	<i>opuAB</i>	<i>Glycine/proline betaine ABC transporter permease subunit OpuAB</i>
NP_388182.1	<i>opuAC</i>	<i>Glycine/betaine ABC transporter</i>

(Continua na próxima página)

(Continuação)

<b>RefSeq</b>	<b>Nome do Gene</b>	<b>Produto</b>
WP_007610192.1	<i>bcp</i>	<i>Thiol peroxidase</i>
WP_012117858.1	<i>trpF</i>	<i>Phosphoribosylanthranilate isomerase</i>
URM44301.1	<i>ysnE</i>	<i>GNAT family N-acetyltransferase</i>
WP_011996289.1	<i>amhX</i>	<i>Amidohydrolase</i>
WP_012117684.1	<i>bioI</i>	<i>Biotin biosynthesis cytochrome P450</i>
WP_012116949.1	<i>bltD</i>	<i>Spermidine acetyltransferas</i>
WP_014471118.1	<i>alsR</i>	<i>Transcriptional regulator</i>
BCU86156.1	<i>ylbF</i>	<i>Controls biofilm development</i>
URM46094.1	<i>ymcA</i>	<i>Biofilm development</i>
URM43879.1	<i>iolW</i>	<i>Scyllo-inositol 2-dehydrogenase (NADP(1) involved in biofilm formation protein</i>
WP_011996282.1	<i>ycdF</i>	<i>Glucose 1-dehydrogenase</i>
WP_012118579.1	<i>tuaD</i>	<i>UDP-glucose 6-dehydrogenase</i>
WP_017419144.1	<i>phzF</i>	<i>PhzF family phenazine biosynthesis isomerase</i>
WP_012118228.1	<i>apuD</i>	<i>Glycine betaine transporter OpuD</i>
WP_012117811.1	<i>bsaA</i>	<i>Glutathione peroxidase</i>

**Fonte: Adaptado de Mosela et al. (2022)**

**APÊNDICE C – GENES PGP *BACILLUS VELEZENSIS* LDO2**

<b>Código Ref-Seq</b>	<b>Nome do Gene</b>	<b>Produto</b>
NP_391482.2	<i>alsS</i>	<i>Acetolactate synthase</i>
NP_390709.1	<i>ilvB</i>	<i>Acetolactate synthase large subunit</i>
NP_388505.1	<i>bdhA</i>	<i>2, 3-butanediol dehydrogenase</i>
NP_388663.1	<i>treR</i>	<i>Tre operon transcriptional repressor</i>
NP_389813.1	<i>dhaS</i>	<i>Indol 3-acet-aldehyde dehydrogenase</i>
NP_388806.2	<i>yhcX</i>	<i>Nitrilase</i>
NP_391630.1	<i>speE</i>	<i>Spermidine synthase</i>
NP_389346.1	<i>speA</i>	<i>Arginine decarboxylase</i>
NP_391629.1	<i>speB</i>	<i>Agmatinase</i>
NP_391551.1	<i>moaA</i>	<i>Molybdenum cofactor synthesis protein</i>
WP_007408857.1	<i>moaC</i>	<i>Molybdenum cofactor synthesis protein</i>
WP_016936285.1	<i>moaD</i>	<i>Molybdenum cofactor synthesis protein</i>
NP_388212.1	<i>nasD</i>	<i>Nitrite reductase large subunit</i>
NP_391609.2	<i>narG</i>	<i>Nitrate reductase</i>
NP_391608.1	<i>narH</i>	<i>Nitrate reductase</i>
NP_391606.1	<i>narI</i>	<i>Nitrate reductase</i>
NP_391607.1	<i>narJ</i>	<i>Nitrate reductase</i>
NP_390987.1	<i>ktrA</i>	<i>Potassium uptake protein A</i>
NP_389334.1	<i>ktrC</i>	<i>Potassium uptake protein C</i>
NP_389233.1	<i>ktrD</i>	<i>Potassium uptake protein D</i>
NP_389213.1	<i>mgtE</i>	<i>Mg<sup>2+</sup> transporter</i>
NP_388681.1	<i>corA</i>	<i>Mg<sup>2+</sup> /Co<sup>2+</sup> transport protein</i>
BCU87559.1	<i>dhbA</i>	<i>Bacillibactin</i>
AHZ17188.1	<i>dhbB</i>	<i>Bacillibactin</i>
UYP22140.1	<i>dhbC</i>	<i>Bacillibactin</i>
AFZ91981.1	<i>dhbE</i>	<i>Bacillibactin</i>
AFJ63359.1	<i>dhbF</i>	<i>Bacillibactin</i>
URM44102.1	<i>alsD</i>	<i>Acetolactate decarboxylase</i>
BCU87210.1	<i>ilvH</i>	<i>Acetolactate synthase small subunit</i>
AHZ14763.1	<i>treA</i>	<i>Trehalose-6-phosphate hydrolase</i>
WP_014417186.1	<i>treP</i>	<i>Trehalose permease component</i>
WP_012117785.1	<i>phy</i>	<i>3-Phytase</i>
WP_041482135.1	<i>ysnE</i>	<i>N-acetyltransferase</i>

*(Continua na próxima página)*

(Continuação)

<b>RefSeq</b>	<b>Nome do Gene</b>	<b>Produto</b>
CUX92117.1	<i>potA</i>	<i>Spermidine import ATP-binding protein</i>
BBA75191.1	<i>mdtJ</i>	<i>Spermidine export protein</i>
WP_007408054.1	<i>moaB</i>	<i>Molybdenum cofactor synthesis protein</i>
WP_007611116.1	<i>moaE</i>	<i>Molybdenum cofactor synthesis protein</i>
AHZ14339.1	<i>nasE</i>	<i>Nitrite reductase small subunit</i>
WP_003151223.1	<i>nrgB</i>	<i>Nitrogen regulatory protein</i>
AHZ17661.1	<i>nrgA</i>	<i>Ammonium transporter</i>
WP_007407632.1	<i>nark</i>	<i>Nitrate transporter</i>
CCG51474.1	<i>narP</i>	<i>Nitrate/nitrite response regulator</i>
QOY25331.1	<i>ktrB</i>	<i>Potassium uptake protein B</i>

**Fonte: Adaptado de Chen et al. (2019)**

**APÊNDICE D – GENES PGP *BACILLUS VELEZENSIS* FZB42**

<b>RefSeq/Protein-id</b>	<b>Gene</b>	<b>Produto</b>
WP_003328596.1	<i>abrB</i>	<i>AbrB/MazE/SpoVT family DNA-binding domain-containing protein</i>
NP_391481.1	<i>alsD</i>	<i>Acetolactate decarboxylase</i>
WP_014471118.1	<i>alsR</i>	<i>LysR family transcriptional regulator</i>
NP_391482.2	<i>alsS</i>	<i>acetolactate synthase AlsS</i>
ABS72758.1	<i>bglC</i>	<i>glycosyl hydrolase family protein</i>
ABS75948.1	<i>bglS</i>	<i>glycosyl hydrolase family protein</i>
WP_007409712.1	<i>bpr</i>	<i>DUF3823 domain-containing protein</i>
WP_015483632.1	<i>comP</i>	<i>histidine kinase</i>
ABS75305.1	<i>csn</i>	<i>chitosanase</i>
WP_003155376.1	<i>cspB</i>	<i>cold-shock protein</i>
WP_003240492.1	<i>cspC</i>	<i>cold shock protein CspC</i>
WP_003230776.1	<i>cspD</i>	<i>Cold-shock protein CspD</i>
WP_007610253.1	<i>cspR</i>	<i>tRNA (uridine(34)/cytosine(34)/5- carboxymethylaminomethyluridine (34)-2'-O)- methyltransferase TrmL</i>
ABS74268.1	<i>dhaS</i>	<i>aldehyde dehydrogenase family</i>
ABS73392.1	<i>ecsA</i>	<i>ABC transporter ATP-binding protein</i>
ABS73393.1	<i>ecsB</i>	<i>ABC transporter permease</i>
WP_012117211.1	<i>ecsC</i>	<i>EcsC family protein</i>
NP_390325.2	<i>efp</i>	<i>elongation factor P</i>
WP_014418936.1	<i>epsA</i>	<i>hypothetical protein</i>
ABS75497.1	<i>epsB</i>	<i>CpsD/CapB family tyrosine-protein kinase</i>
WP_007407392.1	<i>epsC</i>	<i>polysaccharide biosynthesis protein</i>
WP_012118513.1	<i>epsD</i>	<i>glycosyltransferase family 4 protein</i>
ABS75494.1	<i>epsE</i>	<i>glycosyltransferase</i>
WP_015240636.1	<i>epsF</i>	<i>glycosyltransferase family 1 protein</i>
ABS75492.1	<i>epsG</i>	<i>EpsG family protein</i>
WP_015418269.1	<i>epsH</i>	<i>glycosyltransferase family 2 protein</i>
ABS75490.1	<i>epsI</i>	<i>pyruvyl transferase</i>
WP_012118507.1	<i>epsJ</i>	<i>glycosyltransferase</i>
ABS75488.1	<i>epsK</i>	<i>MATE family efflux transporter</i>
ABS75487.1	<i>epsL</i>	<i>sugar transferase</i>
WP_017418872.1	<i>epsM</i>	<i>acetyltransferase</i>
WP_003151568.1	<i>epsN</i>	<i>aminotransferase class I/II-fold pyridoxal phosphate-dependent enzyme</i>

*(Continua na próxima página)*

(Continuação)

RefSeq	Nome do Gene	Produto
WP_003151569.1	<i>epsO</i>	<i>pyruvyl transferase</i>
WP_012118570.1	<i>flgK</i>	<i>flagellar hook-associated protein FlgK</i>
WP_016937780.1	<i>fliD</i>	<i>flagellar hook-associated protein 2</i>
WP_003154945.1	<i>galE1</i>	<i>UDP-glucose 4-epimerase GalE</i>
WP_012117334.1	<i>galK1</i>	<i>galactokinase</i>
WP_016936496.1	<i>galT1</i>	<i>UDP-glucose-hexose-1-phosphate uridylyltransferase</i>
WP_017417565.1	<i>ganA</i>	<i>arabinogalactan endo-1,4-beta-galactosidase</i>
NP_391491.1	<i>ggt</i>	<i>gamma-glutamyltransferase</i>
WP_014470327.1		
WP_003151403.1	<i>hag</i>	<i>flagellin</i>
WP_016936493.1	<i>lacE</i>	<i>PTS lactose transporter subunit IIBC</i>
WP_003154940.1	<i>lacF</i>	<i>PTS lactose/cellobiose transporter subunit IIA</i>
WP_015239655.1	<i>lacG</i>	<i>6-phospho-beta-galactosidase</i>
WP_017417571.1	<i>lacR</i>	<i>DeoR/GlpR transcriptional regulator</i>
WP_015417977.1	<i>lytS</i>	<i>sensor histidine kinase</i>
WP_012117054.1	<i>pel</i>	<i>pectate lyase</i>
NP_391271.1	<i>pgm</i>	<i>2,3-bisphosphoglycerate-independent phosphoglycerate mutase</i>
WP_088037351.1	<i>phy</i>	<i>3-Phytase</i>
NP_391646.1	<i>pta</i>	<i>phosphate acetyltransferase</i>
WP_012117040.1	RBAM_007500	<i>class I SAM-dependent methyltransferase</i>
WP_013351328.1	RBAM_007510	<i>LTA synthase family protein</i>
ABS73139.1	RBAM_007540	<i>hypothetical protein</i>
WP_007410044.1	RBAM_030380	<i>YvrJ family protein</i>
ABS75375.1	RBAM_030440	<i>iron-hydroxamate ABC transporter</i>
NP_391359.1	RBAM_032010	<i>thioredoxin-disulfide reductase</i>
WP_007614019.1	RBAM_032090	<i>1-(5-phosphoribosyl)-5-((5-phosphoribosylamino) methylideneamino) imidazole-4- carboxamide</i>
WP_016937403.1	RBAM_036940	<i>RDD family protein</i>
WP_012117884.1	<i>resE</i>	<i>cell wall metabolism sensor histidine kinase WalK</i>
WP_014419358.1	<i>sacB</i>	<i>glycoside hydrolase family 68</i>
WP_014469731.1	<i>sfp</i>	<i>4'-phosphopantetheinyl transferase superfamily protein</i>
NP_387979.1	<i>sigH</i>	<i>RNA polymerase sporulation sigma factor SigH</i>
NP_388054.1	<i>sigW</i>	<i>RNA polymerase sigma factor SigW</i>
WP_003150917.1	<i>sinI</i>	<i>DNA-binding anti-repressor SinI</i>
WP_007612543.1		

(Continua na próxima página)

(Continuação)

RefSeq	Nome do Gene	Produto
ABS74654.2	<i>sinR</i>	<i>helix-turn-helix domain-containing protein</i>
WP_003153100.1	<i>sipW</i>	<i>signal peptidase I</i>
ABS74694.1	<i>sodA</i>	<i>superoxide dismutase</i>
NP_390302.1	<i>spo0A</i>	<i>sporulation transcription factor Spo0A</i>
WP_057766256.1	<i>srfA</i>	<i>surfactin non-ribosomal peptide synthetase SrfAA</i>
WP_057766253.1	<i>srfB</i>	<i>surfactin non-ribosomal peptide synthetase SrfAB</i>
WP_032872840.1	<i>srfC</i>	<i>surfactin non-ribosomal peptide synthetase SrfAC</i>
WP_007609310.1	<i>srfD</i>	<i>surfactin biosynthesis thioesterase SrfAD</i>
WP_015251121.1	<i>swrA</i>	<i>Swarming motility protein SwrAA</i>
WP_014305010.1	<i>swrB</i>	<i>Swarming motility protein SwrB</i>
ABS73096.1	<i>swrC</i>	<i>efflux RND transporter permease subunit</i>
ABS75598.1	<i>tagO</i>	<i>undecaprenyl/decaprenyl-phosphate alpha-N-acetylglucosaminyl 1-phosphate transferase</i>
ABS74655.1	<i>tasA</i>	<i>spore coat protein</i>
ABS75000.1	<i>tpx</i>	<i>thiol peroxidase</i>
WP_014721880.1	<i>tuaA</i>	<i>exopolysaccharide biosynthesis polyprenyl glycosylphosphotransferase</i>
NP_387994.1	<i>tufA</i>	<i>elongation factor Tu</i>
ABS75708.1	<i>xynA</i>	<i>1,4-beta-xylanase</i>
ABS74096.1	<i>xynB</i>	<i>glycoside hydrolase family 43 protein</i>
ABS74178.1	<i>xynD</i>	<i>carbohydrate-binding protein</i>
ABS74095.1	<i>xynP</i>	<i>MFS transporter</i>
ABS72710.1	<i>ycdH</i>	<i>zinc ABC transporter substrate-binding protein</i>
ABS72775.1	<i>yckI</i>	<i>amino acid ABC transporter ATP-binding protein</i>
ABS72776.2	<i>yckJ</i>	<i>amino acid ABC transporter permease</i>
ABS72777.1	<i>yckK</i>	<i>amino acid ABC transporter substrate-binding protein</i>
WP_003156329.1	<i>yclQ</i>	<i>siderophore ABC transporter substrate-binding</i>
WP_014720709.1	<i>yczE</i>	<i>YitT family protein</i>
WP_003155518.1	<i>yfiQ</i>	<i>acyltransferase</i>
WP_017419102.1	<i>yflE</i>	<i>LTA synthase family protein</i>
ABS73315.1	<i>yhcX</i>	<i>GNAT family N-acetyltransferase</i>
WP_015417161.1	<i>yhfE</i>	<i>M42 family metallopeptidase</i>
WP_012117166.1	<i>yhxB</i>	<i>phospho-sugar mutase</i>
ABS73848.1	<i>ylbF</i>	<i>YlbF family regulator</i>
ABS74049.1	<i>ymcA</i>	<i>Biofilm development</i>
ABS74753.1	<i>yqeK</i>	<i>HD domain-containing protein</i>

(Continua na próxima página)

(Continuação)

<b>RefSeq</b>	<b>Nome do Gene</b>	<b>Produto</b>
WP_014305409.1	<i>yqxM</i>	<i>amyloid fiber anchoring/assembly protein TapA</i>
ABS74946.1	<i>ysdC</i>	<i>M42 family metallopeptidase</i>
WP_012118742.1	<i>ysnE</i>	<i>N-acetyltransferase</i>
ABS72775.1	<i>ytmJ</i>	<i>amino acid ABC transporter ATP-binding protein</i>
ABS75414.1	<i>ytmK</i>	<i>amino acid ABC transporter substrate-binding protein</i>
WP_007409997.1	<i>ytmL</i>	<i>amino acid ABC transporter permease</i>
ABS75412.1	<i>ytmM</i>	<i>amino acid ABC transporter permease</i>
ABS75411.1	<i>ytmN</i>	<i>amino acid ABC transporter ATP-binding protein</i>
WP_015240524.1	<i>yurJ</i>	<i>sn-glycerol-3-phosphate ABC transporter ATP-binding protein UgpC</i>
WP_003151898.1	<i>yurK</i>	<i>transcriptional regulator PhoB</i>
WP_007613586.1	<i>yurL</i>	<i>fructoselysine 6-kinase</i>
WP_015240526.1	<i>yurM</i>	<i>carbohydrate ABC transporter permease</i>
ABS75296.2	<i>yurN</i>	<i>sugar ABC transporter permease</i>
ABS75297.2	<i>yurO</i>	<i>extracellular solute-binding protein</i>
ABS75298.1	<i>yurP</i>	<i>SIS domain-containing protein</i>
WP_007410093.1	<i>yurR</i>	<i>FAD-binding oxidoreductase</i>
WP_003151820.1	<i>yusV</i>	<i>ABC transporter ATP-binding protein</i>
ABS75362.1	<i>yvrC</i>	<i>ABC transporter substrate-binding protein</i>
ABS72792.1	<i>yxeN</i>	<i>amino acid ABC transporter permease</i>
WP_007614938.1	<i>yycO</i>	<i>hypothetical protein</i>

**Fonte: Adaptado de Chen et al. (2007)**

## APÊNDICE E – GENES PGP *BACILLUS SP. RZ2MS9*

Códigos de Acesso		Nome Gene	Produto
GenBank	RefSeq ou Swiss-Prot		
OGY04772.1	WP_000387193.1	<i>sigH</i>	<i>RNA polymerase sporulation-specific sigma factor</i>
OGY05779.1	WP_000197753.1	<i>sigG</i>	<i>RNA polymerase sporulation-specific sigma factor</i>
OGY05780.1	WP_001062677.1	<i>sigE</i>	<i>RNA polymerase sporulation-specific sigma factor</i>
OGY05943.1	WP_002014922.1	<i>sigF</i>	<i>RNA polymerase sporulation-specific sigma factor</i>
OGY06203.1	P62180.1	<i>sigK</i>	<i>RNA polymerase sporulation-specific sigma factor</i>
OGY06746.1	WP_003161566.1	<i>Spo0A</i>	<i>Response regulator, stage 0 sporulation protein A</i>
OGY02638.1	WP_000398594.1	<i>Spo0F</i>	<i>Response regulator, stage 0 sporulation protein F</i>
OGY03170.1	WP_000486620.1	<i>Spo0M</i>	<i>Sporulation-control protein</i>
OGY06303.1	WP_016130289.1	<i>Spo0B</i>	<i>Proteína B de esporulação de estágio 0 ( fosfotransferase de iniciação de esporulação)</i>
OGY05944.1	WP_001243393.1	<i>spoIIAB</i>	<i>Stage II sporulation protein AB (anti-sigma F factor)</i>
OGY05945.1	WP_000059133.1	<i>spoIIAA</i>	<i>Stage II sporulation protein AA (anti-sigma F factor antagonist)</i>
OGY05961.1	WP_000269098.1	<i>spoIIM</i>	<i>Stage II sporulation protein M</i>
OGY05781.1	WP_002165147.1	<i>spoIIGA</i>	<i>Stage II sporulation protein GA (sporulation sigma-E factor processing peptidase)</i>
OGY04739.1	WP_001123538.1	<i>spoIIE</i>	<i>Stage II sporulation protein E</i>
OGY03361.1	WP_018765483.1	<i>spoIIP</i>	<i>Stage II sporulation protein P</i>
OGY03519.1	WP_001999658.1	<i>spoIIP</i>	<i>Stage II sporulation protein P</i>
OGY03050.1	WP_002097913.1	<i>spoIIR</i>	<i>Stage II sporulation protein R</i>
OGY02686.1	WP_016078202.1	<i>spoIID</i>	<i>Stage II sporulation protein D</i>
OGY02690.1	WP_001214926.1	<i>spoIIQ</i>	<i>Stage II sporulation protein Q</i>
OGY02692.1	WP_016112657.1	<i>spoIIID</i>	<i>Stage III sporulation protein D</i>
OGY06055.1	WP_016077257.1	<i>spoIIIAH</i>	<i>Stage III sporulation protein AH</i>
OGY06056.1	WP_002173413.1	<i>spoIIIAG</i>	<i>Stage III sporulation protein AG</i>
OGY06057.1	WP_001161027.1	<i>spoIIIAF</i>	<i>Stage III sporulation protein AF</i>
OGY06058.1	WP_000946108.1	<i>spoIIIAE</i>	<i>Stage III sporulation protein AE</i>
OGY06059.1	WP_000478632.1	<i>spoIIIID</i>	<i>Stage III sporulation protein AD</i>
OGY06060.1	WP_000020914.1	<i>spoIIIIAC</i>	<i>Stage III sporulation protein AC</i>
OGY06061.1	WP_015000953.1	<i>spoIIIIAB</i>	<i>Stage III sporulation protein AB</i>
OGY06062.1	WP_016718462.1	<i>spoIIIIAA</i>	<i>Stage III sporulation protein AA</i>
OGY03838.1	WP_000416519.1	<i>spoIVA</i>	<i>Stage IV sporulation protein A</i>
OGY06041.1	WP_002164433.1	<i>spoIVB</i>	<i>Stage IV sporulation protein B</i>
OGY06308.1	WP_016086140.1	<i>spoIVFB</i>	<i>Stage IV sporulation protein FB</i>

(Continua na próxima página)

(Continuação)

Códigos de Acesso		Nome Gene	Produto
GenBank	RefSeq ou Swiss-Prot		
OGY06309.1	WP_016086139.1	<i>spoIVFA</i>	Stage IV sporulation protein FA
OGY06274.1	WP_016135368.1	<i>spoVB</i>	Stage V sporulation protein B
OGY05790.1	WP_001266233.1	<i>spoVD</i>	Stage V sporulation protein D
OGY05934.1	WP_012095221.1	<i>spoVAF</i>	Stage V sporulation protein AF
OGY05935.1	WP_001254416.1	<i>spoVAE</i>	Stage V sporulation protein AE
OGY05936.1	WP_016111265.1	<i>spoVAE</i>	Stage V sporulation protein AE
OGY05937.1	WP_012261648.1	<i>spoVAD</i>	Stage V sporulation protein AD
OGY05938.1	WP_002088449.1	<i>spoVAC</i>	Stage V sporulation protein AC
OGY05939.1	WP_000572932.1	<i>spoVAB</i>	Stage V sporulation protein AB
OGY05940.1	WP_000438369.1	<i>spoVAA</i>	Stage V sporulation protein AA
OGY05568.1	WP_000438196.1	<i>spoVK</i>	Stage V sporulation protein K
OGY05653.1	WP_000404341.1	<i>spoVS</i>	Stage V sporulation protein S
OGY05733.1	WP_003328987.1	<i>spoVM</i>	Stage V sporulation protein M
OGY04487.1	WP_002150265.1	<i>spoVB</i>	Stage V sporulation protein B
OGY04725.1	WP_013054864.1	<i>spoVG</i>	Stage V sporulation protein G
OGY04731.1	WP_000648302.1	<i>spoVT</i>	Stage V sporulation protein T
OGY03282.1	WP_000427804.1	<i>spoVS</i>	Stage V sporulation protein S
OGY02830.1	WP_017796500.1	<i>spoVAE</i>	Stage V sporulation protein AE
OGY02831.1	WP_000938953.1	<i>spoVAD</i>	Stage V sporulation protein AD
OGY02832.1	WP_000095399.1	<i>spoVAC</i>	Stage V sporulation protein AC
OGY04589.1	WP_001202693.1	<i>spoVR</i>	Stage V sporulation protein R
OGY04554.1	WP_001046260.1	<i>cotJA</i>	Spore coat protein JA
OGY04555.1	WP_016107874.1	<i>cotJB</i>	Spore coat protein JB
OGY04556.1	WP_011983759.1	<i>cotJC</i>	Spore coat protein JC
OGY05647.1	WP_002173051.1	<i>cotE</i>	Spore coat protein E
OGY02101.1	WP_000051528.1	<i>cotB</i>	Spore coat protein B
OGY02483.1	WP_000361676.1	<i>cotX</i>	Spore coat protein X
OGY02485.1	WP_000332789.1	<i>cotX</i>	Spore coat protein X
OGY02945.1	WP_001002981.1	<i>cotF</i>	Spore coat protein CotF
OGY03372.1	WP_002171536.1	<i>cotF</i>	Spore coat protein CotF
OGY06712.1	WP_000512121.1	<i>cotF</i>	Spore coat protein CotF
OGY03381.1	WP_016717275.1	<i>cotH</i>	Spore coat protein CotH
OGY02215.1	Q81I94.1	<i>sspK</i>	Small acid-soluble spore protein K (minor)
OGY03675.1	WP_001025741.1	<i>sspH</i>	Small acid-soluble spore protein H (minor)

(Continua na próxima página)

(Continuação)

Códigos de Acesso		Nome Gene	Produto
GenBank	RefSeq ou Swiss-Prot		
OGY04721.1	WP_003194225.1	<i>sspF</i>	<i>Small acid-soluble spore protein F (minor alpha/beta-type SASP)</i>
OGY05464.1	Q81AG1.1	<i>sspN</i>	<i>Small acid-soluble spore protein N (minor)</i>
OGY05470.1	WP_016105827.1	<i>sspO</i>	<i>Small acid-soluble spore protein O (minor)</i>
OGY05472.1	WP_016086814.1	<i>sspP</i>	<i>Small acid-soluble spore protein P (minor)</i>
OGY06408.1	Q817I4.1	<i>sspI</i>	<i>Small acid-soluble spore protein I (minor)</i>
OGY01571.1	WP_016099335.1	<i>gerQ</i>	<i>Spore germination protein Q</i>
OGY01749.1	WP_000136669.1	<i>gerKA</i>	<i>Spore germination protein KA</i>
OGY02269.1	WP_017152885.1	<i>gerPF</i>	<i>Spore germination protein PF</i>
OGY03157.1	WP_017152885.1	<i>gerPF</i>	<i>Spore germination protein PF</i>
OGY04209.1	WP_001111191.1	<i>gerPA</i>	<i>Spore germination protein PA</i>
OGY04210.1	P0A3T6.1	<i>gerPB</i>	<i>Spore germination protein PB</i>
OGY04211.1	O68685.1	<i>gerPC</i>	<i>Spore germination protein PC</i>
OGY04212.1	WP_001052803.1	<i>gerPD</i>	<i>Spore germination protein PD</i>
OGY04213.1	WP_016088925.1	<i>gerPE</i>	<i>Spore germination protein PE</i>
OGY04214.1	WP_001141566.1	<i>gerPF</i>	<i>Spore germination protein PF</i>
OGY04826.1	WP_000827519.1	<i>gerD</i>	<i>Spore germination protein D</i>
OGY06342.1	WP_002160327.1	<i>gerM</i>	<i>Spore germination protein M</i>
OGY06345.1	WP_000659484.1	<i>gerE</i>	<i>LuxR family transcriptional regulator of spore coat protein</i>
OGY01679.1	WP_000801985.1	<i>yndE</i>	<i>Spore germination protein</i>
OGY01680.1	WP_000891262.1	<i>yndD</i>	<i>Spore germination protein</i>
OGY05035.1	WP_000790846.1	<i>yndF</i>	<i>Spore germination protein</i>
OGY05036.1	WP_000628604.1	<i>yndE</i>	<i>Spore germination protein</i>
OGY05037.1	WP_000119244.1	<i>yndD</i>	<i>Spore germination protein</i>
OGY05251.1	WP_000421999.1	<i>mcp</i>	<i>Methyl-accepting chemotaxis protein</i>
OGY05371.1	WP_000481867.1	<i>mcp</i>	<i>Methyl-accepting chemotaxis protein</i>
OGY04230.1	WP_000273768.1	<i>mcp</i>	<i>Methyl-accepting chemotaxis protein</i>
OGY03407.1	WP_000040625.1	<i>mcp</i>	<i>Methyl-accepting chemotaxis protein</i>
OGY02875.1	WP_002068236.1	<i>mcp</i>	<i>Methyl-accepting chemotaxis protein</i>
OGY02907.1	WP_001012193.1	<i>mcp</i>	<i>Methyl-accepting chemotaxis protein</i>
OGY02908.1	WP_016110909.1	<i>mcp</i>	<i>Methyl-accepting chemotaxis protein</i>
OGY02933.1	WP_012261940.1	<i>mcp</i>	<i>Methyl-accepting chemotaxis protein</i>
OGY02116.1	WP_016716342.1	<i>mcp</i>	<i>Methyl-accepting chemotaxis protein</i>
OGY02139.1	WP_016121779.1	<i>mcp</i>	<i>Methyl-accepting chemotaxis protein</i>

(Continua na próxima página)

(Continuação)

Códigos de Acesso		Nome Gene	Produto
GenBank	RefSeq ou Swiss-Prot		
OGY01706.1	WP_000495008.1	<i>mcp</i>	<i>Methyl-accepting chemotaxis protein</i>
OGY01807.1	WP_002194354.1	<i>mcp</i>	<i>Methyl-accepting chemotaxis protein</i>
OGY01824.1	WP_006925075.1	<i>mcp</i>	<i>Methyl-accepting chemotaxis protein</i>
OGY03723.1	WP_016082189.1	<i>cheA</i>	<i>Two-component system, chemotaxis family, sensor kinase CheA</i>
OGY03697.1	WP_001983058.1	<i>cheV</i>	<i>Two-component system, chemotaxis family, response regulator CheV</i>
OGY03719.1	WP_002194228.1	<i>cheR</i>	<i>Chemotaxis protein methyltransferase</i>
OGY04334.1	WP_000429022.1	<i>cheR</i>	<i>Chemotaxis protein methyltransferase CheR</i>
OGY03724.1	WP_000940575.1	<i>cheY</i>	<i>Two-component system, chemotaxis family, response regulator CheY</i>
OGY04333.1	WP_016716693.1	<i>cheA</i>	<i>Two-component system, chemotaxis family, sensor kinase CheA</i>
OGY06357.1	WP_000344862.1	<i>motA</i>	<i>Chemotaxis protein/Flagellar motor component MotA</i>
OGY06356.1	WP_001256737.1	<i>motB</i>	<i>Chemotaxis protein/Flagellar motor component MotB</i>
OGY03726.1	WP_000502152.1	<i>motA</i>	<i>Chemotaxis protein/Flagellar motor component MotA</i>
OGY03725.1	WP_000596050.1	<i>motB</i>	<i>Chemotaxis protein/Flagellar motor component MotB</i>
OGY03713.1	WP_000929288.1	<i>fliD</i>	<i>Flagellar hook-associated protein 2</i>
OGY03708.1	WP_000701659.1	<i>fliE</i>	<i>Flagellar hook-basal body complex protein FliE</i>
OGY03707.1	WP_000411360.1	<i>fliF</i>	<i>Flagellar M-ring protein FliF</i>
OGY03706.1	WP_000883356.1	<i>fliG</i>	<i>Flagellar motor switch protein FliG</i>
OGY03704.1	WP_001064657.1	<i>fliI</i>	<i>Flagellum-specific ATP synthase</i>
OGY03691.1	WP_000012369.1	<i>fliM</i>	<i>Flagellar motor switch protein FliM</i>
OGY03690.1	WP_000680020.1	<i>fliNY</i>	<i>Flagellar motor switch protein FliN/FliY</i>
OGY03688.1	WP_001220578.1	<i>fliP</i>	<i>Flagellar biosynthetic protein FliP</i>
OGY03687.1	WP_001098084.1	<i>fliQ</i>	<i>Flagellar biosynthetic protein FliQ</i>
OGY03686.1	WP_002186228.1	<i>fliR</i>	<i>Flagellar biosynthetic protein FliR</i>
OGY03712.1	WP_002194234.1	<i>fliS</i>	<i>Flagellar protein FliS</i>
OGY03692.1	WP_002011999.1	<i>fliN</i>	<i>Flagellar motor switch/type III secretory pathway protein FliN</i>
OGY03702.1	WP_000616461.1	<i>fliK</i>	<i>Flagellar hook-length control protein FliK</i>
OGY03705.1	WP_000051695.1	<i>fliH</i>	<i>Flagellar biosynthesis protein FliH</i>
OGY03722.1	WP_001114989.1	<i>fliNY</i>	<i>Flagellar motor switch protein FliN/FliY</i>
OGY06590.1	WP_000716605.1	<i>fliF</i>	<i>Flagellar biosynthesis protein FliF</i>
OGY06717.1	WP_002137233.1	<i>fliD</i>	<i>Flagellar capping protein FliD</i>
OGY03684.1	WP_000472544.1	<i>fliA</i>	<i>Flagellar biosynthesis protein FliA</i>

(Continua na próxima página)

(Continuação)

Códigos de Acesso		Nome Gene	Produto
GenBank	RefSeq ou Swiss-Prot		
OGY03685.1	WP_002202515.1	<i>flhB</i>	<i>Flagellar biosynthesis protein FlhB</i>
OGY03683.1	WP_000446160.1	<i>flhF</i>	<i>Flagellar biosynthesis protein FlhF</i>
OGY03158.1	WP_016092667.1	<i>flgL</i>	<i>Flagellin and related hook-associated protein FlgL</i>
OGY03710.1	WP_001113116.1	<i>flgB</i>	<i>Flagellar basal-body rod protein FlgB</i>
OGY03709.1	WP_016094399.1	<i>flgC</i>	<i>Flagellar basal-body rod protein FlgC</i>
OGY04640.1	WP_002202513.1	<i>flgG</i>	<i>Flagellar basal-body rod protein FlgG</i>
OGY03700.1	WP_000588325.1	<i>flgE</i>	<i>Flagellar hook protein FlgE</i>
OGY03715.1	WP_001239716.1	<i>flgK</i>	<i>Flagellar hook-associated protein 1 FlgK</i>
OGY03714.1	WP_001266398.1	<i>flgL</i>	<i>Flagellar hook-associated protein 3 FlgL</i>
OGY03701.1	WP_002202543.1	<i>flgD</i>	<i>Flagellar basal-body rod modification protein FlgD</i>
OGY06628.1	WP_001141369.1	<i>luxS</i>	<i>luxS autoinducer biosynthesis</i>
OGY05307.1	WP_000216572.1	<i>ahlD</i>	<i>N-acyl homoserine lactone hydrolase</i>
OGY03387.1	WP_002181641.1	<i>pilA</i>	<i>Pilus assembly protein, major pilin PilA</i>
OGY02999.1	WP_000289354.1	<i>pilA</i>	<i>Pilus assembly protein, major pilin PilA</i>
OGY06075.1	WP_000111890.1	<i>fimV</i>	<i>Pilus assembly protein FimV</i>
OGY06143.1	WP_000825421.1	<i>fimV</i>	<i>Pilus assembly protein FimV</i>
OGY03694.1	WP_002086453.1	<i>fliC</i>	<i>Flagellin</i>
OGY03695.1	WP_001219723.1	<i>fliC</i>	<i>Flagellin</i>
OGY04675.1	WP_000690973.1	<i>lbs</i>	<i>Lipopolysaccharide biosynthesis protein</i>
OGY05366.1	WP_001087530.1	<i>lbs</i>	<i>Lipopolysaccharide biosynthesis protein</i>
OGY03577.1	WP_000854511.1	<i>pgsA</i>	<i>Poly-gamma-glutamate synthesis protein</i>
OGY05163.1	WP_000887601.1	<i>pgsA</i>	<i>Poly-gamma-glutamate synthesis protein</i>
OGY03582.1	WP_002165980.1	<i>alsS</i>	<i>Acetolactate synthase small subunit</i>
OGY03583.1	WP_000392596.1	<i>alsS</i>	<i>Acetolactate synthase, large subunit, biosynthetic type</i>
OGY03946.1	WP_002193239.1	<i>alsS</i>	<i>Acetolactate synthase small subunit</i>
OGY03947.1	WP_000095870.1	<i>alsS</i>	<i>Acetolactate synthase, large subunit, biosynthetic type</i>
OGY04497.1	WP_017561828.1	<i>alsS</i>	<i>Acetolactate synthase large subunit</i>
OGY04496.1	WP_000215002.1	<i>alsD</i>	<i>Alpha-acetolactate decarboxylase</i>
OGY01716.1	WP_016112154.1	<i>bdh</i>	<i>(R,R)-butanediol dehydrogenase</i>
OGY01551.1	WP_000424691.1	<i>speE</i>	<i>Spermidine synthase</i>
OGY04534.1	WP_000897996.1	<i>speE</i>	<i>Spermidine synthase</i>
OGY04054.1	WP_002011483.1	<i>potC</i>	<i>ABC-type spermidine/putrescine transport system, permease component II</i>
OGY04056.1	WP_000720332.1	<i>potA</i>	<i>ABC-type spermidine/putrescine transport systems, ATPase components</i>

(Continua na próxima página)

(Continuação)

Códigos de Acesso		Nome Gene	Produto
GenBank	RefSeq ou Swiss-Prot		
OGY04055.1	WP_016082399.1	<i>potB</i>	ABC-type spermidine/putrescine transport system, permease component I
OGY05886.1	WP_002088373.1	<i>pot</i>	ABC-type spermidine/putrescine transport systems, ATPase components
OGY01816.1	WP_000571400.1	<i>pot</i>	ABC-type spermidine/putrescine transport systems, ATPase components
OGY04053.1	WP_016109576.1	<i>potD</i>	Spermidine/putrescine-binding periplasmic protein
OGY06562.1	WP_016512227.1	<i>moaC</i>	Molybdenum cofactor biosynthesis enzyme
OGY06598.1	Q816R0.1	<i>moaB</i>	Molybdopterin biosynthesis enzyme MoaB
OGY03299.1	WP_000573688.1	<i>moaD</i>	Molybdopterin converting factor, small subunit
OGY03303.1	WP_000545013.1	<i>moaA</i>	Molybdenum cofactor biosynthesis enzyme MoaA
OGY05418.1	WP_000628471.1	<i>moaD</i>	Molybdopterin converting factor, small subunit
OGY05423.1	WP_016111392.1	<i>moaA</i>	Molybdenum cofactor biosynthesis enzyme MoaA
OGY06558.1	WP_000621790.1	<i>moaD</i>	Molybdopterin converting factor, small subunit
OGY06564.1	WP_001156506.1	<i>moaA</i>	Molybdenum cofactor biosynthesis enzyme MoaA
OGY04866.1	WP_000007159.1	<i>modB</i>	ABC-type molybdate transport system, permease component
OGY03500.1	WP_016090856.1	<i>modA</i>	ABC-type molybdate transport system, periplasmic component
OGY04867.1	WP_016716276.1	<i>modA</i>	ABC-type molybdate transport system, periplasmic component
OGY03301.1	WP_016079239.1	<i>moeA</i>	Molybdopterin biosynthesis enzyme
OGY05420.1	WP_016123213.1	<i>moeA</i>	Molybdopterin biosynthesis enzyme
OGY05421.1	WP_016080703.1	<i>moeB</i>	Molybdopterin biosynthesis adenyltransferase
OGY06561.1	WP_016077197.1	<i>moeA</i>	Molybdopterin biosynthesis enzyme
OGY03864.1	WP_002171129.1	<i>fer</i>	Ferredoxin
OGY03967.1	WP_000041153.1	<i>fldA</i>	Flavodoxin
OGY05394.1	WP_001062757.1	<i>fldA</i>	Flavodoxin
OGY02276.1	WP_000594477.1	<i>fix</i>	Signal transduction histidine kinase involved in nitrogen fixation and metabolism regulation
OGY03883.1	WP_000797534.1	<i>fix</i>	Signal transduction histidine kinase involved in nitrogen fixation and metabolism regulation
OGY05122.1	WP_000839485.1	<i>fix</i>	Signal transduction histidine kinase involved in nitrogen fixation and metabolism regulation
OGY06511.1	WP_000023856.1	<i>fix</i>	Signal transduction histidine kinase involved in nitrogen fixation and metabolism regulation
OGY03306.1	WP_002177941.1	<i>narJ</i>	Nitrate reductase assembly protein NarJ, required for insertion of molybdenum cofactor

(Continua na próxima página)

(Continuação)

Códigos de Acesso		Nome Gene	Produto
GenBank	RefSeq ou Swiss-Prot		
OGY03307.1	WP_000692658.1	<i>narH</i>	<i>Nitrate reductase beta subunit</i>
OGY03308.1	WP_018764556.1	<i>narG</i>	<i>Nitrate reductase alpha subunit</i>
OGY03305.1	WP_002065103.1	<i>narI</i>	<i>Nitrate reductase gamma subunit</i>
OGY03298.1	WP_016134287.1	<i>narK</i>	<i>Nitrate/nitrite transporter NarK</i>
OGY02309.1	WP_000814376.1	<i>narK</i>	<i>Nitrate/nitrite transporter NarK</i>
OGY03290.1	WP_002085947.1	<i>nirB</i>	<i>Nitrite reductase (NADH) large subunit</i>
OGY03291.1	WP_000616667.1	<i>nirD</i>	<i>Nitrite reductase (NADH) small subunit</i>
OGY04034.1	WP_000893016.1	<i>nirC</i>	<i>Nitrite transporter NirC</i>
OGY01621.1	WP_000047628.1	<i>nos</i>	<i>Nitric oxide synthase, oxygenase domain</i>
OGY06705.1	Q81EP4.1	<i>ldh</i>	<i>L-lactate dehydrogenase</i>
OGY03481.1	Q81EP4.1	<i>ldh</i>	<i>L-lactate dehydrogenase</i>
OGY02946.1	WP_002109008.1	<i>ldh</i>	<i>L-lactate dehydrogenase</i>
OGY06437.1	WP_002122975.1	<i>ldh</i>	<i>Isocitrate dehydrogenase</i>
OGY06362.1	WP_000678351.1	<i>sdhC</i>	<i>Succinate dehydrogenase</i>
OGY06360.1	WP_001291827.1	<i>sdhB</i>	<i>Succinate dehydrogenase</i>
OGY06361.1	WP_000676742.1	<i>sdhA</i>	<i>Succinate dehydrogenase</i>
OGY06555.1	WP_006924883.1	<i>gdh</i>	<i>Glucose 1-dehydrogenase</i>
OGY03892.1	WP_000380275.1	<i>pit</i>	<i>Inorganic phosphate transporter, PiT family</i>
OGY06136.1	WP_001242550.1	<i>pstA</i>	<i>ABC-type phosphate transport system, permease component</i>
OGY06137.1	WP_016120812.1	<i>pstC</i>	<i>ABC-type phosphate transport system, permease component</i>
OGY01672.1	WP_000994130.1	<i>pstA</i>	<i>ABC-type phosphate transport system, permease component</i>
OGY01840.1	WP_000676681.1	<i>pstC</i>	<i>ABC-type phosphate transport system, permease component</i>
OGY06138.1	WP_000871821.1	<i>pstS</i>	<i>ABC-type phosphate transport system, periplasmic component</i>
OGY01673.1	WP_016083492.1	<i>pstS</i>	<i>ABC-type phosphate transport system, periplasmic component</i>
OGY06135.1	WP_002088644.1	<i>pstB</i>	<i>ABC-type phosphate transport system, ATPase component</i>
OGY05543.1	WP_000047911.1	<i>phnE</i>	<i>ABC-type phosphate/phosphonate transport system, permease component</i>
OGY05544.1	WP_002160868.1	<i>phnE</i>	<i>ABC-type phosphate/phosphonate transport system, permease component</i>
OGY05545.1	Q81A96.2	<i>phnC</i>	<i>ABC-type phosphate/phosphonate transport system, ATPase component</i>

(Continua na próxima página)

(Continuação)

Códigos de Acesso		Nome Gene	Produto
GenBank	RefSeq ou Swiss-Prot		
OGY05546.1	WP_000913375.1	<i>phnD</i>	ABC-type phosphate/phosphonate transport system, periplasmic component
OGY06210.1	WP_000714904.	<i>alk</i>	Alkaline phosphatase
OGY02594.1	WP_000935915.1	<i>alk</i>	Alkaline phosphatase
OGY06134.1	WP_002015145.1	<i>phr</i>	Phosphate uptake regulator
OGY06169.1	WP_000840525.1	<i>phoH</i>	Phosphate starvation-inducible protein PhoH and related proteins
OGY05960.1	WP_000392651.1	<i>fur</i>	Ferric uptake regulator
OGY03430.1	WP_001261821.1	<i>sid</i>	Siderophore synthetase componente
OGY03431.1	WP_001163328.1	<i>sid</i>	Siderophore synthetase componente
OGY05615.1	WP_016108952.1	<i>fevA</i>	ABC-type $Fe^{3+}$ -siderophores transport system, ATPase component
OGY05617.1	WP_016108954.1	<i>fevP</i>	ABC-type $Fe^{3+}$ -siderophore transport system, permease component
OGY06388.1	WP_017674005.1	<i>fevP</i>	ABC-type $Fe^{3+}$ -siderophore transport system, permease component
OGY06389.1	WP_002092271.1	<i>fevS</i>	ABC-type $Fe^{3+}$ -hydroxamate transport system, periplasmic component
OGY01563.1	WP_000388278.1	<i>fevP</i>	ABC-type $Fe^{3+}$ -siderophore transport system, permease component
OGY01766.1	WP_012260297.1	<i>fevA</i>	ABC-type $Fe^{3+}$ -siderophores transport system, ATPase component
OGY05319.1	WP_000902706.1	<i>fevP</i>	ABC-type $Fe^{3+}$ -siderophore transport system, permease component
OGY05318.1	WP_002199770.1	<i>fevP</i>	ABC-type $Fe^{3+}$ -siderophore transport system, permease component
OGY05616.1	WP_000920450.1	<i>fevP</i>	ABC-type $Fe^{3+}$ -siderophore transport system, permease component
OGY06229.1	WP_002082127.1	<i>fevP</i>	ABC-type $Fe^{3+}$ -siderophore transport system, permease component
OGY02093.1	WP_000983751.1	<i>fevP</i>	ABC-type $Fe^{3+}$ -siderophore transport system, permease component
OGY02094.1	WP_002006758.1	<i>fevP</i>	ABC-type $Fe^{3+}$ -siderophore transport system, permease component
OGY01767.1	WP_000760037.1	<i>fevP</i>	ABC-type $Fe^{3+}$ -siderophore transport system, permease component
OGY01768.1	WP_000925208.1	<i>fevP</i>	ABC-type $Fe^{3+}$ -siderophore transport system, permease component
OGY01562.1	WP_001260757.1	<i>fevP</i>	ABC-type $Fe^{3+}$ -siderophore transport system, permease component

(Continua na próxima página)

(Continuação)

Códigos de Acesso		Nome Gene	Produto
GenBank	RefSeq ou Swiss-Prot		
OGY06228.1	WP_000626363.1	<i>fevA</i>	ABC-type $Fe^{3+}$ -siderophores transport system, ATPase component
OGY06387.1	WP_000403746.1	<i>fevA</i>	ABC-type $Fe^{3+}$ -siderophores transport system, ATPase component
OGY01561.1	WP_000212060.1	<i>fevA</i>	ABC-type $Fe^{3+}$ -siderophores transport system, ATPase component
OGY03668.1	WP_016136606.1	<i>afuB</i>	ABC-type $Fe^{3+}$ transport system, permease component
OGY03669.1	WP_000714679.1	<i>afuA</i>	ABC-type $Fe^{3+}$ transport system, periplasmic component
OGY03670.1	WP_000351780.1	<i>afuC</i>	ABC-type $Fe^{3+}$ transport systems, ATPase component
OGY04018.1	WP_000425407.1	<i>afuB</i>	ABC-type $Fe^{3+}$ transport system, permease component
OGY04019.1	WP_002202702.1	<i>afuC</i>	ABC-type $Fe^{3+}$ transport systems, ATPase component
OGY04020.1	WP_000750887.1	<i>afuA</i>	ABC-type $Fe^{3+}$ transport system, periplasmic component
OGY02722.1	WP_001225604.1	<i>afuA</i>	ABC-type $Fe^{3+}$ transport system, periplasmic component
OGY06643.1	WP_000060512.1	<i>feoA</i>	$Fe^{2+}$ transport system protein FeoA
OGY01674.1	WP_000252982.1	<i>feoA</i>	$Fe^{2+}$ transport system protein FeoA
OGY01676.1	WP_017150021.1	<i>feoB</i>	$Fe^{2+}$ transport system protein FeoB
OGY06642.1	WP_016116800.1	<i>feoB</i>	$Fe^{2+}$ transport system protein FeoB
OGY02857.1	WP_000749429.1	<i>fevS</i>	ABC-type enterochelin transport system, periplasmic component
OGY02858.1	WP_016107228.1	<i>fevP</i>	ABC-type enterochelin transport system, permease component
OGY02859.1	WP_000631135.1	<i>fevP</i>	ABC-type enterochelin transport system, permease component
OGY02860.1	WP_014966664.1	<i>fevA</i>	ABC-type enterochelin transport system, ATPase component
OGY05380.1	WP_017560596.1		Outer membrane receptor for ferrienterochelin and colicins
OGY04107.1	WP_016082431.1	<i>trpF</i>	Phosphoribosylanthranilate isomerase
OGY04108.1	WP_001052803.1	<i>trpC</i>	Indole-3-glycerol phosphate synthase
OGY04109.1	A9VJV9.1	<i>trpD</i>	Anthranilate phosphoribosyltransferase
OGY04110.1	WP_000636314.1	<i>trpG</i>	Anthranilate synthase component II
OGY04111.1	WP_002181772.1	<i>trpE</i>	Anthranilate synthase component I

(Continua na próxima página)

(Continuação)

Códigos de Acesso		Nome Gene	Produto
GenBank	RefSeq ou Swiss-Prot		
OGY04105.1	WP_016133917.1	<i>trpA</i>	<i>Tryptophan synthase alpha chain</i>
OGY04106.1	WP_001105024.1	<i>trpB</i>	<i>Tryptophan synthase beta chain</i>
OGY01885.1	WP_014654562.1	<i>ipdC</i>	<i>Indolepyruvate decarboxylase</i>
OGY03397.1	WP_016104606.1	<i>mao</i>	<i>Monoamine oxidase</i>
OGY03480.1	WP_000528025.1	<i>mao</i>	<i>Monoamine oxidase</i>
OGY03040.1	WP_000808507.1	<i>patA</i>	<i>Aminotransferase</i>
OGY05879.1	WP_000434584.1	<i>patA</i>	<i>Aminotransferase</i>
OGY05905.1	WP_016118577.1	<i>mtnE</i>	<i>Aminotransferase</i>
OGY02002.1	WP_000285660.1	<i>Ami</i>	<i>Amidase</i>
OGY03497.1	WP_001044906.1		<i>Amidase</i>

Fonte: Adaptado de Batista (2017)

**APÊNDICE F – RESULTADO ALINHAMENTO BLASTP *B. VELEZENSIS* FZB42**

<b>Código RefSeq/ Genbank (Query)</b>	<b>RefSeq/Protein-id (Subject)</b>	<b>Nome do Gene</b>	<b>Percentual Identidade</b>	<b>E-Value</b>
CAA43955.1	ABS72486.1	<i>abrB</i>	99%	$4.18 \cdot 10^{-63}$
URM44102.1	ABS75646.1	<i>alsD</i>	100%	0.0
USQ53343.1	ABS75648.1	<i>alsR</i>	100%	0.0
AHZ15494.1	ABS73876.1	<i>bpr</i>	99%	0.0
QNV52080.1	ABS75208.1	<i>comP</i>	99%	0.0
UYP23744.1	ABS73301.1	<i>cspB</i>	100%	$1.07 \cdot 10^{-43}$
WP_003230776.1	ABS74369.1	<i>cspD</i>	98%	$5.04 \cdot 10^{-43}$
URM45355.1	ABS73285.2	<i>cspR</i>	99%	$8.85 \cdot 10^{-121}$
AHZ14997.1	ABS73394.1	<i>ecsC</i>	99%	$1.99 \cdot 10^{-174}$
QAR58334.1	ABS75498.1	<i>epsA</i>	100%	$3.88 \cdot 10^{-176}$
QAR58332.1	ABS75496.1	<i>epsC</i>	100%	0.0
ASB54741.1	ABS75495.1	<i>epsD</i>	99%	0.0
URM43962.1	ABS75493.1	<i>epsF</i>	99%	0.0
URM43960.1	ABS75491.1	<i>epsH</i>	99%	0.0
QHK62906.1	ABS75489.1	<i>epsJ</i>	99%	0.0
ASB67230.1	ABS75486.1	<i>epsM</i>	99%	$2.56 \cdot 10^{-152}$
ASB67229.1	ABS75485.1	<i>epsN</i>	100%	0.0
URM43954.1	ABS75484.1	<i>epsO</i>	100%	0.0
OOI00461.1	ABS75580.1	<i>fliD</i>	100%	0.0
CDG25445.1	ABS73577.1	<i>galE1</i>	99%	0.0
BBA75777.1	ABS73578.1	<i>galK1</i>	99%	0.0
CDG25444.1	ABS73576.1	<i>galT1</i>	99%	0.0
AFZ90262.1	ABS73575.1	<i>ganA</i>	99%	0.0
AFZ92343.1	ABS75581.1	<i>hag</i>	100%	0.0
QNV54176.1	ABS73579.1	<i>lacE</i>	100%	0.0
BCU85855.1	ABS73580.1	<i>lacF</i>	100%	$2.30 \cdot 10^{-73}$
AFZ90269.1	ABS73582.1	<i>lacR</i>	100%	0.0
BCU87272.1	ABS74955.1	<i>lytS</i>	100%	0.0
BCU85421.1	ABS73156.1	<i>pel</i>	100%	0.0
BBA77731.1	ABS75458.1	<i>pgm</i>	100%	0.0
WP_012117785.1	ABS74326.1	<i>phy</i>	100%	0.0
AHZ16420.1	ABS74486.1	<i>resE</i>	100%	0.0
AHZ18054.1	ABS76090.1	<i>sacB</i>	99%	0.0

(Continua na próxima página)

(Continuação)

<b>Código RefSeq/ Genbank (Query)</b>	<b>RefSeq/Protein-id (Subject)</b>	<b>Nome do Gene</b>	<b>Percentual Identidade</b>	<b>E-Value</b>
CCP20395.1	ABS72773.1	<i>sfp</i>	100%	$5.85 \cdot 10^{-171}$
AHZ16596.1	ABS74656.1	<i>sipW</i>	100%	$3.56 \cdot 10^{-140}$
CCF03839.1	ABS72764.1	<i>srfA</i>	98%	0.0
WP_057766253.1	ABS72765.1	<i>srfB</i>	99%	0.0
WP_032872840.1	ABS72767.1	<i>srfC</i>	98%	0.0
CCF03842.1	ABS72768.1	<i>srfD</i>	100%	0.0
BBA77851.1	ABS75567.1	<i>swrA</i>	99%	$9.44 \cdot 10^{-171}$
AGZ56375.1	ABS73995.1	<i>swrB</i>	100%	$1.46 \cdot 10^{-111}$
AFZ89691.1	ABS73096.1	<i>swrC</i>	100%	0.0
AHZ17570.1	ABS75606.2	<i>tuaA</i>	100%	$2.33 \cdot 10^{-163}$
AHZ14142.1	ABS72561.1	<i>tufA</i>	100%	0.0
CDG24574.1	ABS72710.1	<i>ycdH</i>	100%	0.0
CAE02640.1	ABS72775.1	<i>yckI</i>	100%	0.0
CDH94099.1	ABS72776.2	<i>yckJ</i>	100%	$8.91 \cdot 10^{-171}$
AHZ14363.1	ABS72777.1	<i>yckK</i>	100%	0.0
URM44892.1	ABS72807.1	<i>yckQ</i>	100%	0.0
URM44861.1	ABS72774.1	<i>yczE</i>	100%	$1.33 \cdot 10^{-155}$
QJC91660.1	ABS73231.1	<i>yfiQ</i>	99%	0.0
AKL75384.1	ABS73176.1	<i>yflE</i>	100%	0.0
URM45476.1	ABS73405.1	<i>yhfE</i>	100%	0.0
AHZ14918.1	ABS73321.1	<i>yhxB</i>	100%	0.0
BBA76077.1	ABS73848.1	<i>ylbF</i>	100%	$4.70 \cdot 10^{-110}$
URM46094.1	ABS74049.1	<i>ymcA</i>	100%	$2.59 \cdot 10^{-102}$
QDF56567.1	ABS74657.1	<i>yqxM</i>	99%	$7.55 \cdot 10^{-169}$
CDH94746.1	ABS74946.1	<i>ysdC</i>	100%	0.0
URM44301.1	ABS75867.1	<i>ysnE</i>	96%	$1 \cdot 10^{-110}$
CDH97056.1	ABS75414.1	<i>ytmK</i>	99%	0.0
AFJ63546.1	ABS75413.1	<i>ytmL</i>	100%	$3.99 \cdot 10^{-175}$
AHZ17365.1	ABS75412.1	<i>ytmM</i>	99%	$8.93 \cdot 10^{-170}$
AFJ63544.1	ABS75411.1	<i>ytmN</i>	99%	0.0
URM43773.1	ABS75292.1	<i>yurJ</i>	99%	0.0
GJJ27101.1	ABS75293.1	<i>yurK</i>	100%	0.0
AHZ17250.1	ABS75294.1	<i>yurL</i>	100%	0.0
URM43775.1	ABS75295.1	<i>yurM</i>	100%	0.0
URM43776.1	ABS75296.2	<i>yurN</i>	100%	0.0

(Continua na próxima página)

(Continuação)

<b>Código RefSeq/ Genbank (Query)</b>	<b>RefSeq/Protein-id (Subject)</b>	<b>Nome do Gene</b>	<b>Percentual Identidade</b>	<b>E-Value</b>
GJJ27105.1	ABS75297.2	<i>yurO</i>	99%	0.0
AHZ17254.1	ABS75298.1	<i>yurP</i>	98%	0.0
BCT29397.1	ABS75302.1	<i>yurR</i>	99%	0.0
URM43815.1	ABS75337.1	<i>yusV</i>	100%	0.0
URM44470.1	ABS76048.1	<i>yycO</i>	99%	0.0