

**UNIVERSIDADE FEDERAL DE MINAS GERAIS**

**Instituto de Ciências Biológicas**

**Programa de Pós-Graduação em Bioinformática**

Gabriel Camargos Gomes

**ANÁLISES COMPARATIVAS E FUNCIONAIS DE DUAS LINHAGENS DE  
BACILLUS PARALICHENIFORMIS ISOLADAS DO SOLO DA BAHIA REVELAM  
HOMOGENEIDADE DA ESPÉCIE E METABÓLITOS BIOATIVOS DE POTENCIAL  
BIOTECNOLÓGICO**

Belo Horizonte

2025

Gabriel Camargos Gomes

**ANÁLISES COMPARATIVAS E FUNCIONAIS DE DUAS LINHAGENS DE  
BACILLUS PARALICHENIFORMIS ISOLADAS DO SOLO DA BAHIA REVELAM  
HOMOGENEIDADE DA ESPÉCIE E METABÓLITOS BIOATIVOS DE POTENCIAL  
BIOTECNOLÓGICO**

Dissertação apresentada ao Programa Interunidades de Pós-graduação em Bioinformática, do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, como parte dos requisitos para obtenção do título de mestre.

ORIENTADOR: PROF. DR. VASCO AZEVEDO

Belo Horizonte

2025

043

Gomes, Gabriel Camargos.

Análises comparativas e funcionais de duas linhagens de *Bacillus paralicheniformis* isoladas do solo da Bahia revelam homogeneidade da espécie e metabólitos bioativos de potencial biotecnológico [manuscrito] / Gabriel Camargos Gomes. – 2025.

93 f. : il. ; 29,5 cm.

Orientador: Prof. Dr. Vasco Azevedo.

Dissertação (mestrado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa Interunidades de Pós-Graduação em Bioinformática.

1. Bioinformática. 2. *Bacillus licheniformis*. 3. Genômica. 4. Biotecnologia. I. Azevedo, Vasco Ariston de Carvalho II. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. III. Título.

CDU: 573:004



UNIVERSIDADE FEDERAL DE MINAS GERAIS  
FOLHA DE APROVAÇÃO

*Gabriel Camargos Gomes*

**"Análises comparativas e funcionais de duas linhagens de *Bacillus paralicheniformis* isoladas do solo da Bahia revelam homogeneidade da espécie e metabólitos bioativos de potencial biotecnológico"**

Dissertação aprovada pela banca examinadora constituída pelos Professores:

Prof. Vasco Ariston de Carvalho Azevedo - Orientador  
UFMG

Prof. Aristóteles Góes Neto  
UFMG

Prof. Luis Gustavo Carvalho Pacheco  
UFBA

Prof. Luis Cláudio Lima de Jesus  
INRAE

Belo Horizonte, 29 de julho de 2025.



Documento assinado eletronicamente por **Luis Gustavo Carvalho Pacheco, Usuário Externo**, em 29/07/2025, às 16:50, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Luís Cláudio Lima de Jesus, Usuário Externo**, em 29/07/2025, às 16:55, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Vasco Ariston de Carvalho Azevedo, Professor do Magistério Superior**, em 29/07/2025, às 16:55, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **Aristoteles Goes Neto, Professor do Magistério Superior**, em 30/07/2025, às 09:41, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



A autenticidade deste documento pode ser conferida no site [https://sei.ufmg.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **4412564** e o código CRC **D8BE52AD**.

---

Referência: Processo nº 23072.245729/2025-21

SEI nº 4412564



UNIVERSIDADE FEDERAL DE MINAS GERAIS

**ATA**

INSTITUTO DE CIÊNCIAS BIOLÓGICAS  
PROGRAMA INTERUNIDADES DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA

**DEFESA DE DISSERTAÇÃO**  
*Gabriel Camargos Gomes*

Às quatorze horas do dia **29 de julho de 2025**, reuniu-se, por videoconferência através do aplicativo Zoom, a Comissão Examinadora de Dissertação, indicada pelo Colegiado do Programa, para julgar, em exame final, o trabalho intitulado: "**Análises comparativas e funcionais de duas linhagens de *Bacillus paralicheniformis* isoladas do solo da Bahia revelam homogeneidade da espécie e metabólitos bioativos de potencial biotecnológico**", requisito para obtenção do grau de Mestre em **Bioinformática**. Abrindo a sessão, o Presidente da Comissão, **Dr. Vasco Ariston de Carvalho Azevedo**, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra ao candidato, para apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa do candidato. Logo após, a Comissão se reuniu, sem a presença do candidato e do público, para julgamento e expedição de resultado final. Foram atribuídas as seguintes indicações:

| <b>Professor(a)/Pesquisador(a)</b>    | <b>Instituição</b> | <b>Indicação</b> |
|---------------------------------------|--------------------|------------------|
| Dr. Vasco Ariston de Carvalho Azevedo | UFMG               | Aprovado         |
| Dr. Aristóteles Góes Neto             | UFMG               | Aprovado         |
| Dr. Luis Gustavo Carvalho Pacheco     | UFBA               | Aprovado         |
| Dr. Luis Cláudio Lima de Jesus        | INRAE              | Aprovado         |

Pelas indicações, o candidato foi considerado: **Aprovado**

O resultado final foi comunicado publicamente ao candidato pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora.

**Belo Horizonte, 29 de julho de 2025.**



Documento assinado eletronicamente por **Luis Gustavo Carvalho Pacheco, Usuário Externo**, em 29/07/2025, às 16:49, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Luís Cláudio Lima de Jesus, Usuário Externo**, em 29/07/2025, às 16:54, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **Vasco Ariston de Carvalho Azevedo, Professor do Magistério Superior**, em 29/07/2025, às 16:55, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **Aristoteles Goes Neto, Professor do Magistério Superior**, em 30/07/2025, às 09:41, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



A autenticidade deste documento pode ser conferida no site [https://sei.ufmg.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **4412540** e o código CRC **A65E5CF3**.

---

## **AGRADECIMENTOS**

Gostaria de expressar minha sincera gratidão ao Professor Dr. Vasco Azevedo, meu orientador, pela orientação científica, confiança e apoio constante ao longo desta jornada. À mestra Eduarda Guimarães, minha coorientadora, agradeço profundamente pela dedicação, conselhos valiosos e contribuições fundamentais para o desenvolvimento deste trabalho. Agradeço à agência de fomento FAPEMIG pelo suporte financeiro, essencial para a realização deste projeto, bem como ao Laboratório de Genética Celular e Molecular (LGCM) e à pós-graduação da Bioinformática, que me acolheram, provendo infraestrutura e um ambiente colaborativo que permitiram o avanço das análises e reflexões científicas. Sou especialmente grato aos colegas de laboratório e parceiros de pesquisa do LGCM e da pós-graduação, que caminharam comigo ao longo desta trajetória: Gabriela, Rhayane, Andria, Luis, Leandro, Fernanda, Bernardo, Janaína, Jailan, Marcus, Talita, Arun e Lucas. Obrigado pelas discussões produtivas, colaborações em artigos e capítulos, apoio técnico e, sobretudo, pela parceria no dia a dia acadêmico. Agradeço também aos meus pais, Alexandre e Lígia, por todo o amor, incentivo e suporte incondicional. Sua confiança e exemplo foram pilares fundamentais durante toda a minha formação acadêmica, desde a graduação em Ciências Biológicas até o fim do mestrado. Por fim, deixo um agradecimento muito especial à minha namorada Ludmila, que esteve ao meu lado em todos os momentos, oferecendo carinho, paciência, motivação e apoio emocional nas fases mais desafiadoras. Sua presença fez toda a diferença. A todos que contribuíram, direta ou indiretamente, para esta conquista, o meu mais sincero obrigado.

## RESUMO

O gênero *Bacillus* é um grupo de bactérias amplamente estudadas, e abriga espécies que impactam diversas indústrias. Muitas espécies do grupo *Bacillus* são estudadas pelo potencial biotecnológico e industrial, envolvendo a produção de metabólitos secundários, o uso como agentes de biocontrole em plantas, reduzindo potencialmente a utilização de agrotóxicos químicos convencionais, além da associação à dieta de animais de consumo, como galinhas, porcos e bois. Entretanto, devido a divergências entre técnicas de identificação taxonômica como o MALDI-TOF, e outras mais recentes e precisas, como o Next-Generation Sequencing (NGS), têm sido observadas incongruências na distinção de espécies de *Bacillus*, atribuídas ao alto nível de semelhança entre os organismos do grupo, principalmente no caso de espécies recentes e pouco estudadas, como *B. paralicheniformis*. A recente descrição do grupo, em 2015, reflete a escassez de estudos de filogenômicos e comparativos que analisem a espécie isoladamente e contribuam para o entendimento do seu comportamento genético. Nesse contexto, este estudo buscou identificar e caracterizar, a partir do NGS, os isolados BAC30 e BAC220, sendo ambos identificados como *B. paralicheniformis* e incluídos em análises comparativas e filogenômicas com outras 28 linhagens da espécie, a fim de analisar o repertório genético do grupo. Em seguida, foi feita a anotação funcional dos isolados BAC30 e BAC220, com foco no potencial biotecnológico da espécie. Os resultados demonstraram que a identificação dos isolados evidencia a relevância da técnica como padrão ouro na identificação taxonômica. Apesar da grande semelhança entre as linhagens de *B. paralicheniformis*, as mesmas mostram um pangenoma aberto, com alta proporção de genes acessórios em contraste com genes centrais, evidenciando a ubiquidade da espécie, com propensão à aquisição de novos genes que favoreçam a sobrevivência em diversos nichos e contribuam para a simbiose com outros organismos. As análises funcionais dos isolados revelaram poucos genes de resistência, produção de metabólitos relevantes, como bacitracina, riboflavina e heme, além de genes relacionados com a fixação de nitrogênio em plantas. Mesmo que sejam necessárias análises *in vitro* para comprovar a expressão gênica, os isolados mostram potencial de aplicação no contexto biotecnológico, seja na produção isolada desses metabólitos e/ou atuando como bactéria de biocontrole em modelos *in vivo*, principalmente com plantas, mas também com animais de consumo e humanos.

**Palavras-chave:** Bioinformática; NGS; *Bacillus paralicheniformis*; Genômica comparativa; Pangenômica; Biotecnologia;

## ABSTRACT

The genus *Bacillus* encompasses a widely studied group of bacteria, harboring species that significantly impact various industries. Many species are analyzed for their biotechnological and industrial potential, including the production of secondary metabolites, their use as biocontrol agents in plants, potentially reducing conventional chemical pesticide use, and their association with the diets of livestock like chickens, pigs, and cattle. However, due to divergences between taxonomic identification techniques such as MALDI-TOF and more recent, precise methods like Next-Generation Sequencing (NGS), inconsistencies in *Bacillus* species distinction have been observed. This is attributed to the high degree of similarity among organisms within the group, especially for recently described and less-studied species such as *B. paralicheniformis*. The group's recent description in 2015 highlights the scarcity of phylogenomic and comparative studies that analyze this species in isolation and contribute to understanding its genetic behavior. In this context, the present study aimed to identify and characterize isolates BAC30 and BAC220 using NGS, which were both confirmed as *B. paralicheniformis* and included in comparative and phylogenomic analyses with 28 other strains of the species to examine the group's genetic repertoire. Subsequently, functional annotation of BAC30 and BAC220 was performed, focusing on the species' biotechnological potential. Results showed that the isolates' identification highlights the relevance of NGS as the gold standard for taxonomic identification. Despite the high similarity among *B. paralicheniformis* strains, they exhibit an open pangenome with a high proportion of accessory genes, contrasting with core genes. This indicates the species' ubiquity and its propensity for acquiring new genes that favor survival in diverse niches and contribute to symbiosis with other organisms. Functional analyses of the isolates revealed few resistance genes, and demonstrated the production of relevant metabolites like bacitracin, riboflavin, and heme, along with genes related to nitrogen fixation in plants. Although *in vitro* analyses are necessary to confirm gene expression, these isolates show potential for biotechnological application, either in the isolated production of these metabolites and/or by acting as biocontrol bacteria in *in vivo* models, primarily with plants, but also with livestock and humans.

**Keywords:** Bioinformatics; NGS; *Bacillus paralicheniformis*; Comparative genomics; Pangenomics; Biotechnology;

## LISTA DE FIGURAS

### INTRODUÇÃO

**Figura 1:** Linha do tempo da descrição de algumas das linhagens do grupo *Bacillus subtilis sensu lato*.....20

### CAPÍTULO I

**Figura 1:** Análises taxonômicas e filogenômicas de espécies de *Bacillus*.....46

**1a:** Análise taxonômica de 16 genomas de *Bacillus*, utilizando o pyANI

**1b:** Análise filogenômica de 30 genomas de *B. paralicheniformis*, utilizando o software Orthofinder

**Figura 2:** Mapa circular e curva de rarefação dos genomas de *B. paralicheniformis*.....48

**2a:** Mapa circular de 30 genomas, utilizando o software Proksee

**2b:** Curva de rarefação de 30 genomas, utilizando o software PPanGGolin

**Figura 3:** Esquemático de vias metabólicas relacionadas à resistência a Beta-lactâmicos a partir dos genomas dos isolados BAC30 e BAC220 utilizando o software KEGG.....49

**Figura 4:** Esquemático de vias metabólicas relacionadas com o metabolismo de nitrogênio e produção de heme, a partir dos genomas dos isolados BAC30 e BAC220 utilizando o software KEGG.....51

**Figura 5:** Identificação de regiões associadas com síntese de peptídeos não-ribossomais (NRPS) nos isolados BAC30 (6a) e BAC220 (6b) a partir do software Anti-SMASH.....52

## LISTA DE TABELAS

|   |    |
|---|----|
| <b>Tabela 1:</b> Informações sobre os 28 genomas completos de <i>Bacillus paralicheniformis</i> obtidos pelo NCBI Datasets..... | 41 |
|---|----|

## LISTA DE ABREVIATURAS

|           |  |
|-----------|--|
| AMCs      | Antimicrobial Compounds                                    |
| PGPB      | Plant Growth-Promoting Bacteria                            |
| EPS       | Exopolissacarídeos   |
| MALDI-TOF | Matrix-Assisted Laser Desorption/Ionization Time-of-Flight |
| NCBI      | National Center for Biotechnology Information              |
| WGS       | Whole Genome Sequencing                                    |
| EFSA      | European Food Safety Authority                             |
| MRSA      | Methicillin resistant Staphylococcus aureus                |
| PGPR      | Plant Growth-Promoting Rhizobacteria                       |
| DNA       | Deoxyribonucleic Acid                                      |
| RNA       | Ribonucleic Acid   |
| rRNA      | Ribosomal Ribonucleic Acid                                 |
| OLC       | Overlap-layout-consensus                                   |
| CDS       | Coding DNA Sequence  |
| RAST      | Rapid Annotations using Subsystems Technology              |
| RefSeq    | Reference Sequence   |
| PGAP      | Prokaryotic Genome Annotation Pipeline                     |
| BGC       | Biosynthetic Gene Cluster                                  |
| NRPS      | Non Ribosomal Peptide Synthesis                            |
| dDDH      | Digital DNA-DNA Hybridization                              |
| ANI       | Average Nucleotide Identification                          |
| NGS       | Next Generation Sequencing                                 |
| TYGS      | Type (Strain) Genome Server                                |
| GTDB-Tk   | Genome Taxonomy Database Toolkit                           |

|       |   |
|-------|---|
| BLAST | Basic Local Alignment Search Tool       |
| NJ    | Neighbour Joining                       |
| ML    | Maximum Likelihood                      |
| COG   | Clusters of Orthologous Groups          |
| KEGG  | Kyoto Encyclopedia of Genes and Genomes |

## **APRESENTAÇÃO DA DISSERTAÇÃO DE MESTRADO**

### **Colaboradores**

Este trabalho foi realizado sob a orientação do Prof. Dr. Vasco Azevedo e a coorientação da Mestre Eduarda Guimarães Sousa, no Laboratório de Genética Celular e Molecular (LGCM), vinculado ao Departamento de Genética, Ecologia e Evolução do Instituto de Ciências Biológicas (ICB) da Universidade Federal de Minas Gerais (UFMG), em Belo Horizonte.

O LGCM é um grupo de pesquisa com forte atuação em bioinformática aplicada à genômica bacteriana. Reconhecido como um dos pioneiros nessa área no Brasil, o laboratório tem desenvolvido projetos abrangentes em genômica, transcriptômica, além de iniciativas voltadas à criação de vacinas e métodos diagnósticos, consolidando-se como uma referência no estudo de microrganismos. Esta pesquisa contou com apoio financeiro da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).

## PREFÁCIO

O delineamento da dissertação está organizado nas seguintes seções:

**I - INTRODUÇÃO:** Busca explicar sobre o gênero *Bacillus* e sua importância, a partir da primeira descrição de *B. subtilis*, a separação das espécies, bem como, o impacto geral do gênero em indústrias como a agropecuária, farmacêutica e alimentícia. Posteriormente, aprofunda-se na espécie *B. paralicheniformis*, foco do estudo, mostrando sua relevância no contexto biotecnológico. Por fim, contextualiza a utilização da bioinformática, assim como pontos importantes sobre identificação taxonômica, análises comparativas e genômica bacteriana.

**II - JUSTIFICATIVA:** Algumas das linhagens da espécie *Bacillus paralicheniformis* já vinham sendo estudadas, porém reconhecidas como *B.licheniformis*. Apenas em 2015, houve distinção de ambas espécies, onde nota-se a escassez de estudos que abordem a espécie *B. paralicheniformis* como um todo e de forma isolada, com análises comparativas e filogenômicas que contribuam para o entendimento de seu repertório genético. Consideram-se também as divergências na identificação taxonômica por métodos como o MALDI-TOF, em relação aos mais recentes, como os derivados de *Whole Genome Sequencing* (WGS), levando à necessidade de análises taxonômicas mais aprofundadas da espécie. Ademais, a espécie *B. paralicheniformis* demonstra potencial biotecnológico, com produção de compostos antimicrobianos (AMCs) e auxílio comprovado no crescimento de plantas, sendo, portanto, necessárias mais estudos que comprovem o potencial biotecnológico da espécie.

**III - OBJETIVO DA DISSERTAÇÃO:** Realizar análises de bioinformática em linhagens de *B. paralicheniformis* para ajudar a compreender o panorama genético da espécie, bem como montar e anotar o genoma de dois isolados do solo da Bahia sequenciados por NGS, a fim de avaliar o potencial biotecnológico dessas linhagens *in silico*.

**IV - CAPÍTULO I:** O artigo de pesquisa intitulado como “Comparative and Functional Analyses of *Bacillus paralicheniformis* Strains BAC30 and BAC220 by WGS Uncover Species Homogeneity and Biotechnological Potential” analisa o repertório gênico de linhagens de *Bacillus paralicheniformis*, assim como faz a anotação funcional de dois isolados da espécie.

**V - MATERIAL SUPLEMENTAR:** Lista todos os materiais suplementares, como tabelas e figuras que foram feitas para a confecção de cada artigo de pesquisa.

**VI - DISCUSSÃO GERAL:** Sumariza a discussão ao redor dos resultados da dissertação.

**VII - CONCLUSÃO GERAL:** Sumariza os principais achados da dissertação.

**VIII - PERSPECTIVAS:** Traz as principais ideias a serem exploradas para a continuidade deste trabalho.

**IX - REFERÊNCIAS BIBLIOGRÁFICAS:** Lista as referências utilizadas no desenvolvimento deste trabalho.

**X - APÊNDICE:** Lista os trabalhos realizados durante esse período de mestrado (2023 - 2025).

## SUMÁRIO

|  |           |
|--|-----------|
| <b>1. INTRODUÇÃO.....</b>  | <b>19</b> |
| 1.1 Gênero <i>Bacillus</i> .....                                 | 19        |
| 1.2 Espécie <i>Bacillus paralicheniformis</i> .....              | 22        |
| 1.2.1. Taxonomia e classificação.....                            | 22        |
| 1.2.2. Produção de metabólitos secundários.....                  | 23        |
| 1.2.3. Simbiose com modelos <i>in vivo</i> .....                 | 24        |
| 1.3 Fundamentos e aplicações em bioinformática e genômica.....   | 25        |
| 1.4 Abordagens de identificação e caracterização bacteriana..... | 28        |
| <b>2. JUSTIFICATIVA DA PESQUISA DA DISSERTAÇÃO.....</b>          | <b>32</b> |
| <b>3. OBJETIVO DA DISSERTAÇÃO.....</b>                           | <b>33</b> |
| 3.1 Geral.....   | 33        |
| 3.2 Específicos.....   | 33        |
| <b>4. CAPÍTULO 1.....</b>  | <b>35</b> |
| <b>5. DISCUSSÃO GERAL.....</b>                                   | <b>70</b> |
| <b>6. CONCLUSÃO GERAL.....</b>                                   | <b>71</b> |
| <b>7. PERSPECTIVAS.....</b>                                      | <b>72</b> |
| <b>8. REFERÊNCIAS BIBLIOGRÁFICAS.....</b>                        | <b>73</b> |
| <b>9. APÊNDICE.....</b>  | <b>86</b> |
| 9.1. Análises em andamento.....                                  | 86        |
| 9.2. Produção científica no mestrado.....                        | 86        |
| 9.3. Totais de produção.....                                     | 89        |
| 9.4. Artigos científicos.....                                    | 90        |
| 9.5. Capítulo de livro.....                                      | 92        |

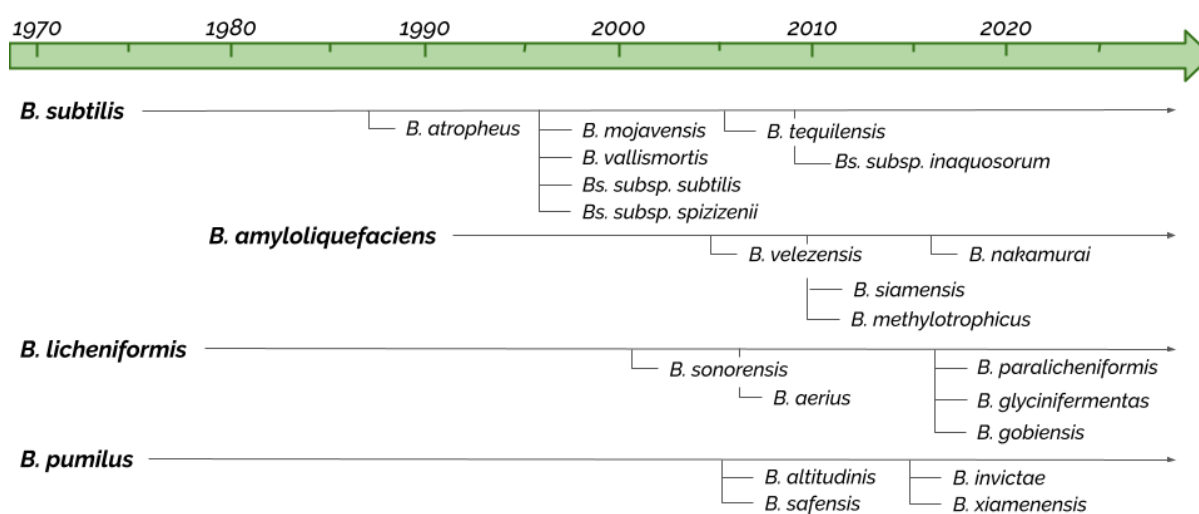
## 1. INTRODUÇÃO

### 1.1 Gênero *Bacillus*

O gênero *Bacillus*, pertencente ao reino *Bacteria*, filo *Firmicutes*, classe *Bacilli*, ordem *Bacillales* e família *Bacillaceae*, foi primeiramente citado em 1835, pelo o cientista Christian Gottfried sendo, inicialmente, denominada *Vibrio subtilis*. Posteriormente, o microbiologista Ferdinand Julius Cohn, em 1872, renomeou o organismo como *Bacillus subtilis* (Stülke et al., 2023), dando origem ao gênero. A partir desse momento, e ainda hoje, esse conjunto é reconhecido como um grupo de bactérias Gram-positivas, ubíquas, em formato de bacilo e formadoras de endósporos (Shu & Yang, 2017). Os endósporos são formados em condições desfavoráveis, como falta de nutrientes, quando a bactéria entra em estado latente, com parede celular espessa e extremamente resistente a estresses ambientais até a retomada de condições favoráveis (Tan & Ramamurthi, 2014). Por volta de 1960 a 1970, a espécie *B. subtilis* foi nomeada como organismo de referência para bactérias Gram-positivas, sendo desde a sua descoberta modelo para estudos de divisão celular, secreção de proteínas, motilidade bacteriana, formação de biofilmes, produção de metabólitos, entre outros (Kovács, 2019). Devido ao avanço nas técnicas de taxonomia, principalmente com estudos do gene 16S rRNA, e reconhecimento de novos marcadores filogenéticos, os critérios de classificação têm sido constantemente revisados, resultando na reclassificação de diversos membros do grupo *Bacillus* e na adição de várias espécies e subespécies. Após as classificações, o grupo foi dividido em vários subgrupos, sendo os de maior relevância *Subtilis sensu lato* e *Cereus sensu lato* (Bhandari et al., 2013).

O grupo *Subtilis sensu lato*, inicialmente conhecido como *Bacillus sensu stricto*, era inicialmente composto pelas espécies *B. subtilis*, *B. licheniformis*, *B. pumilus* e *B. amyloliquefaciens* (Blanco Crivelli et al., 2024), e posteriormente foram adicionadas novas espécies e subespécies (**Figura 1**). As bactérias pertencentes a este grupo são de grande interesse industrial, devido à sua diversidade de características fisiológicas, como a elevada capacidade de produção de compostos antimicrobianos (AMCs). Além disso, estas espécies apresentam uma distribuição ubíqua, sendo achadas no solo, ambientes aquáticos, alimentos fermentados, na microbiota intestinal de animais e na rizosfera de plantas (Caulier et al., 2019). Na indústria agrícola, muitas linhagens do grupo *Subtilis* são reconhecidas como *Plant Growth-Promoting Bacteria* (PGPB), onde podem ser encontradas em uma relação simbiótica

com plantas, muitas dessas de relevância industrial, como feijão (Kumar et al., 2012) e açaí (Garcias et al., 2025), por exemplo. Essas bactérias mostram efeitos benéficos comprovados, que incluem a inibição de patógenos relevantes para plantas, como *Fusarium graminearum* (Petkova et al., 2024) e *Fusarium oxysporum* (Russi et al., 2022), e a estimulação do crescimento da planta, através da secreção de metabólitos, aumento da resistência a estresses osmóticos (Chen et al., 2016) e indução da produção de fitohormônios, como as giberelinas (Kang et al., 2019). Além disso, considera-se também a administração de linhagens de *Bacillus* em modelos *in vivo*, já que, em modelos *in vitro*, nota-se a inibição do crescimento de patógenos como *Escherichia coli* e *Salmonella Typhimurium*, além de demonstrar efeitos anti-inflamatórios e antioxidantes (Pézsza et al., 2022). Considerando a produção de nutrientes por essas linhagens, incluindo exopolissacarídeos (EPS), vitaminas e coenzimas, os mesmos podem atuar como pós-bióticos, auxiliando no crescimento de outros organismos benéficos presentes na microbiota intestinal, como *Lactobacillus* (Cai et al., 2020) e *Enterococcus* (Li et al., 2023).



**Figura 1:** Linha do tempo da descoberta de algumas das espécies pertencentes ao grupo *Bacillus subtilis sensu lato*. A proximidade entre espécies nos clados não indica relação filogenética, mas sim períodos de descrição semelhantes.

Já o grupo *Bacillus cereus sensu lato* é composto por pelo menos doze espécies, onde as mais relevantes incluem principalmente *B. anthracis* e *B. cereus*, mas também *B. thuringiensis*, *B. mycoides*, *B. pseudomycoides*, *B. weihenstephanensis*, *B. cytotoxicus*, *B. wiedmanni*, e *B. toyonensis*, entre outras recém-descobertas (Yossa et al., 2022). Esse grupo é conhecido por abrigar bactérias patogênicas presentes principalmente em alimentos, que

incluem carnes em todos os estados (crua, intacta ou processada), além de produtos de origem aviária, incluindo carne crua de aves e produtos processados com carne e ovos (Carroll et al., 2022). O consumo desses produtos contaminados está associado à intoxicação alimentar causada por linhagens de *B. cereus*, podendo causar duas formas distintas de gastroenterite, a emética e a diarréica (Ehling-Schulz, Fricker & Scherer, 2004). Além disso, esporos de membros desse grupo podem ser encontrados em produtos cosméticos, causando infecções oculares graves, que podem resultar em perda parcial ou total da visão (Mursalin, Livingston & Callegan, 2020). A patogênese de *Bacillus anthracis* é historicamente conhecida e amplamente estudada, já que essa bactéria, causadora da doença zoonótica antraz, já foi reconhecida como potencial arma biológica (Moayeri et al., 2015). A doença antraz afeta a pele e/ou órgãos internos, e quando os esporos de *B. anthracis* são inalados podem ser fatais, principalmente devido às toxinas produzidas por esse organismo (Guichard, Nizet & Bier, 2012). Todavia, a espécie *B. thuringiensis*, é explorada pelas propriedades entomo-pesticidas, já que as proteínas produzidas durante as fases exponencial e vegetativa não são consideradas tóxicas para humanos e servem como inseticidas naturais (Gupta, Kumar & Kaur, 2021), atuando na inibição de lepidópteros, como *Euprosterina elaeasa* (Plata-Rueda et al., 2020), coleópteros, como *Leptinotarsa decemlineata* (Naimov et al., 2001) e dípteros, como *Aedes aegypti* (Alba-Tercedor & Vilchez., 2023). Entretanto, o descarte inadequado ou dispersão de *B. thuringiensis* em ecossistemas pode ter impacto negativo, já que as toxinas podem afetar organismos não-alvo na teia alimentar (Belousova et al., 2021).

A taxonomia atual do gênero *Bacillus* vem sendo continuamente revista e debatida, considerando que métodos tradicionalmente utilizados, como análises bioquímicas, *Matrix-Assisted Laser Desorption/Ionization Time-of-Flight* (MALDI-TOF) e o sequenciamento do gene 16S rRNA, são eficazes para classificações ao nível de gênero, mas apresentam limitações na distinção precisa entre espécies, devido à elevada similaridade genômica e bioquímica entre os organismos (Janda et al., 2007), somada, por muito tempo, à escassez de genomas disponíveis para comparação (Patel et al., 2020). Contudo, com o avanço e a redução de custos das tecnologias de sequenciamento e análise genômica, o número de genomas do gênero *Bacillus* depositados em bancos de dados, como o National Center for Biotechnology Information (NCBI) e o European Molecular Biology Laboratory – European Bioinformatics Institute (EMBL-EBI), tem crescido exponencialmente (Land et al., 2015). Esse cenário possibilita análises taxonômicas mais robustas e precisas, essenciais para o entendimento das relações ecológicas e evolutivas do grupo (Maughan & Auwera, 2011).

Nesse contexto, a identificação por MALDI-TOF continua sendo considerada uma ferramenta de melhor custo-benefício para o monitoramento ambiental e a identificação rotineira de isolados, especialmente de espécies patogênicas (Baek et al., 2021). Por outro lado, métodos acurados, como o sequenciamento genômico completo (Whole Genome Sequencing – WGS), vêm sendo apontados como os mais indicados e padrão ouro para a descoberta e caracterização de novas linhagens, por permitirem uma análise mais aprofundada e detalhada (Mazhari et al., 2025).

## 1.2 Espécie *Bacillus paralicheniformis*

### 1.2.1. Taxonomia e classificação

A espécie *Bacillus licheniformis*, descrita por volta de 1900, foi posteriormente relatada na produção de enzimas, como  $\alpha$ -amilase, penicillinase, pentosanase, entre outros (Rey et al., 2004), e produção de antimicrobiano natural Bacitracina (Cai et al., 2020). Além disso, essa bactéria foi aplicada em plantas, promovendo efeitos pró-crescimento, e auxiliando na mitigação de efeitos de seca (Akhtar et al., 2020), e em estudos *in vivo*, com leitões, demonstrou-se eficaz como bactéria de biocontrole, inibindo o patógeno *Salmonella spp.* (Barba-Vidal et al., 2017), e em galinhas, inibiu o patógeno *Clostridium perfringens*, além de promover a modulação da microbiota e redução de citocinas pró-inflamatórias (Xiao et al., 2024).

Até então, as linhagens pertenciam apenas a uma espécie, principalmente devido ao alto nível de similaridade entre os organismos, aliado à escassez de marcadores filogenéticos até então disponíveis. Em 2015, observou-se que a espécie compreendia, na verdade, duas linhagens filogeneticamente próximas, mas distintas (Dunlap et al., 2015), o que resultou na reclassificação de muitas bactérias anteriormente identificadas como *B. licheniformis*, agora como *B. paralicheniformis* (Du et al., 2019). Nesse contexto, novos marcadores específicos para *B. paralicheniformis*, como, por exemplo, genes codificadores de fengicina, que podem estar presentes exclusivamente nos genomas da espécie e ausentes em *B. licheniformis* (He et al., 2023). Isso se deve ao fato de que o sequenciamento do gene 16S rRNA tem se mostrado, por vezes, insuficiente para a diferenciação entre essas duas espécies, ressaltando a importância do sequenciamento completo do genoma (WGS) como uma abordagem mais confiável para essa distinção (EFSA, 2024). Essa distinção é especialmente relevante devido à aplicação potencial dessas linhagens como produtoras de enzimas alimentícias. Em 2021, a Autoridade Europeia para a Segurança dos Alimentos (EFSA) passou a exigir a identificação dessas espécies por meio de WGS (EFSA, 2021). Um exemplo dessa problemática pode ser

visto no isolado BAC220 — objeto do presente estudo — anteriormente identificados como *B. subtilis* (Reis et al., 2022), mas que, após análises taxonômicas baseadas em sequenciamento pela plataforma Illumina HiSeq, foram reclassificados como *B. paralicheniformis*. Ademais, embora já existam estudos comparativos, como os de pangenômica (Du et al., 2019; Asif et al., 2023), ainda se observa uma escassez de trabalhos que abordem *B. paralicheniformis* de forma abrangente em análises taxonômicas e filogenômicas, comparando as próprias linhagens em um conjunto formado exclusivamente por genomas da espécie.

### 1.2.2. Produção de metabólitos secundários

Muitos estudos reconhecem o potencial biotecnológico da espécie, e parte disso é baseado na sua produção de diversos metabólitos secundários. *In silico*, é notada a presença de genes responsáveis pela produção de metabólitos relevantes nas linhagens de *B. paralicheniformis*, como enzimas e compostos antimicrobianos (AMCs). Entre eles a bacitracina se mostra como composto de maior relevância, sendo produzido predominantemente pelas espécies *B. paralicheniformis*, *B. licheniformis* e *B. subtilis* (Zhu et al., 2023), assim como bacilibactina, fengicina, bacilicina, exopolissacarídeos (EPS) e outros, que variam entre linhagens de *B. paralicheniformis* (Harwood et al., 2018) (Rao et al., 2024). A atividade inibitória da espécie, derivada desses metabólitos, já mostra atividades comprovadas *in vitro*, sendo capazes de inibir patógenos bacterianos como o *Staphylococcus aureus* meticilina-resistente (MRSA) (Sermkaew et al., 2024), *Escherichia coli* e *Listeria monocytogenes* (Iqbal et al., 2023), além de inibir também fungos fitopatogênicos como *Botrytis cinerea* (Ruiz et al., 2024), *Magnaporthe grisea*, *Fusarium solani* e *Alternaria alternata* (Ashajyothi et al., 2024). Quanto a produção isolada de compostos, a linhagem *B. paralicheniformis* AUMC B-516 se mostrou capaz de produzir L-asparaginase através da fermentação (Saleh et al., 2025), e essa enzima é de grande interesse da indústria farmacêutica, já que é utilizada para tratamento de pacientes com leucemia linfoblástica aguda (LLA) (Egler et al., 2016).

Diante disso, espécies do gênero *Bacillus*, como o *B. paralicheniformis*, *B. subtilis*, entre outros, representam uma alternativa para o uso e produção dessa enzima, já que a L-asparaginase é atualmente produzida principalmente por bactérias como *Escherichia coli* e *Erwinia chrysanthemi* (Ashok et al., 2019). Ademais, espécies de plantas mutantes, como tomate, que expressam a proteína BpXynC, retirada de *B. paralicheniformis* NMSW12, mostraram que essa proteína é capaz de induzir a resposta imune, aumentando a resistência a

fitopatógenos como *Sclerotinia sclerotiorum*, vírus do mosaico do tabaco (TMV) e *Phytophthora capsica* (Yu et al., 2024).

### 1.2.3. Simbiose com modelos *in vivo*

Muitas das linhagens da espécie *B. paralicheniformis* são reconhecidas pelo *status* de Plant Growth-Promoting Rhizobacteria (PGPR), sendo isoladas das raízes e solo, e observadas em relação à simbiose com plantas (Sun et al., 2025). Nesse contexto, a linhagem 2R5 foi capaz de promover o melhor efeito pró-crescimento em canola, se comparado a outras rizobactérias como *Pseudomonas* sp. e *Pseudomonas atacamensis* (Świątczak et al., 2023). Esse efeito pode ser associado à modulação da microbiota do solo, que aumenta a quantidade de outros organismos benéficos como *Nitrospira*, *Ramlibacter*, *Solicoccozyma* entre outros (Świątczak et al., 2024). Com a linhagem FMCH001, foi observado que o crescimento de soja sob o estresse de seca foi 22,94% maior que o controle, e essa taxa foi relacionada à melhora na taxa de fotossíntese, condutância estomatal e taxa de transpiração, que melhoraram a eficiência da planta quanto ao uso de água (Liu et al., 2013). Outro estudo similar demonstrou o potencial da linhagem LBEndo1 em aumentar a taxa de crescimento da planta *Carya illinoensis* mesmo sob estresse salino, mostrando a habilidade da espécie *B. paralicheniformis* em aliviar possíveis estresses osmóticos (Palacio-Rodríguez et al., 2023).

Embora em menor quantidade, há estudos demonstrando a atuação da espécie em modelos animais. Com patos, a linhagem LN33 foi capaz de promover um efeito antioxidante e anti-inflamatório, modulação do sistema imune e da microbiota intestinal, promovendo aumento de linhagens probióticas como *Faecalibacterium* (Jiang et al., 2025). Em modelo murino de enterocolite induzida por *Clostridium perfringens*, a linhagem SN-6 foi capaz de modular a microbiota, aumentando a quantidade de bactérias benéficas como *Lactobacillus* e *Akkermansia*, além da redução da inflamação e danos a barreira intestinal causados pelo patógeno (Yang et al., 2025). Além disso, experimentos com um dos isolados do estudo, BAC220, mostraram o efeito benéfico dessa bactéria na redução da necrose de embriões de galinha desafiados com *Salmonella* spp. (Reis et al., 2022).

## 1.3 Fundamentos e aplicações em bioinformática e genômica

Com a crença da época de que proteínas eram portadoras de informações genéticas, o primeiro trabalho de bioinformática surgiu em 1962: COMPROTEIN, *software* que usava cartão perfurado contendo uma linha de código FORTRAN para montar corretamente os resíduos das sequências proteicas oriundas da reação de Edman, por Margaret Dayhoff e

Robert Ledley (Edman & Begg, 1966; Dayhoff & Ledley, 1962; Gauthier, 2018). Este trabalho foi tão impactante que permitiu estudos de alinhamento entre proteínas de diferentes espécies, com Zuckerkandl e Pauling (Morgan, 1998) que revelou a necessidade de julgar substituições entre aminoácidos, levando ao desenvolvimento de algoritmos de alinhamento por Needleman e Wunsch, em 1970 (Needleman & Wunsch, 1970); por Feng e Doolittle em 1987 (Feng & Doolittle, 1987) e, durante este período, também por Dayhoff, Schwartz e Orcutt (1978), trazendo a publicação da matriz PAM, algoritmo probabilístico de avaliação de mudança entre aminoácidos de uma proteína durante momentos da evolução (Yang et al., 2010).

Somente próximo da década de oitenta que o foco voltou-se para o Ácido Desoxirribonucleico (DNA). O bacteriófago  $\Phi$ X174, de 5386 pares de bases, foi o primeiro genoma a ser codificado, usando o sequenciamento de DNA “mais e menos”, de 1975, posteriormente aprimorado e chamado de Sanger (Sanger et al., 1977; Sanger & Coulson, 1977). Essa foi a primeira vez que se obteve a sequência completa do DNA de um organismo, o que se denomina genoma. O pacote Staden foi o primeiro software associado à leitura de sequenciamentos derivadas da técnica de Sanger, sendo capaz de analisar sobreposições de gel, unir *contigs* e manipular arquivos de sequência, ainda em 1979, sendo mais um marco da bioinformática (Staden, 1979).

Outros meios de sequenciamento foram apresentados neste período (Barnes, 1978; Maxam & Gilbert, 1977; McCombie et al, 2018), contudo, a técnica de Sanger foi a mais utilizada, sendo usada para revelar os fragmentos de DNA dos próximos genomas. Antes de 1995, todos os organismos com genoma completo sequenciado eram vírus, devido ao tamanho pequeno dos mesmos (Sanger et al., 1977; Fiers et al., 1978; Sanger et al., 1982; Baer et al., 1984). *Haemophilus influenzae Rd* foi o primeiro organismo de vida livre (não viral) com genoma completamente sequenciado, inaugurando a era da genômica bacteriana, bem como marcou o início do uso de algoritmos computacionais para montagem do genoma fragmentado pela técnica *shotgun sequencing*, através da abordagem *overlap-layout-consensus* (OLC) (Fleischmann et al., 1995). Somente no início dos anos 2000 que os genomas de organismos eucariotos multicelulares *Drosophila melanogaster*, *Homo sapiens* e *Mus musculus* foram apresentados (Adams et al., 2000; Venter et al., 2001; Nature, 2001; Dijk et al., 2023; Nature, 2002).

Após essa primeira leva de tecnologias de sequenciamento, no início dos anos 2000 passaram a ser descritos novos meios, apelidados de Sequenciamento de Segunda Geração

(SGS), produzindo vários fragmentos de sequências curtas (centenas de nucleotídeos), entre 50 - 300 pares de base (pb). Alguns exemplos incluem tecnologias como o Illumina HiSeq 300 e o Ion Torrent, que ainda possuem grande relevância no sequenciamento de genomas pequenos, como de procariotos (Meslier et al., 2022). Posteriormente, em 2010, surgiram os Sequenciamentos de Terceira Geração (TGS), os quais produzem menos fragmentos, mas de longos tamanhos (dezenas de milhares de nucleotídeos) (Heather & Chain, 2016; Shendure et al., 2017; Goodwin et al., 2016; Segerman, 2020; Zhang et al., 2021); Trofimova et al., 2023; Dijk, 2023). Alguns exemplos de sequenciadores de terceira geração incluem o Oxford Nanopore e o Pacific Biosciences (PacBio) (Athanasopoulou et al., 2021). Para a essencial e correta interpretação dos genomas, a montagem das sequências curtas e/ou longas derivadas dos meios de sequenciamento requer *softwares* de montagem, a fim de gerar a representação do genoma real, por meio de sobreposição e cobertura de cada região do genoma várias vezes (Simpson & Pop, 2015; Sohn & Nam, 2016; Segerman, 2020). A montagem resulta em “*contigs*” (do inglês, “contiguous sequences”), representando partes do genoma montados a partir das leituras; “*scaffold*”, quando uma única sequência nucleotídica contínua representa a união dos contigs conectados por “Ns” que representam lacunas entre eles e, por fim, genoma completo/genoma fechado (do inglês, “closed genome”) (WGS), apresentando todo o genoma sem lacunas (Segerman, 2020).

Desde então, a montagem híbrida se consolidou como meio de caracterizar o genoma procariótico, pois utiliza leituras longas e curtas para melhor estruturar os *contigs* consenso (Chen et al., 2020). O Unicycler é uma ferramenta de montagem de genomas de bactérias que utiliza essa abordagem e se sobressai perante outros algoritmos (Wick et al., 2017; Trisakul et al., 2024). Essa etapa deve passar por uma boa curadoria, já que bancos de dados, como o RefSeq (NCBI, 2025), requerem depósitos de genomas bacterianos com alta qualidade e pureza (NCBI, 2025). Assim, um genoma bem montado é indispensável para uma melhor precisão na etapa subsequente de anotação gênica (Kimbrel et al., 2021).

Nesse contexto, a relevância da anotação está em determinar as porções do genoma sequenciado que são responsáveis por abrigar sequências codificadoras (CDSs) (do inglês, “coding DNA sequence”), bem como genes de tRNA/rRNA (Lobb et al., 2020). Esta etapa rotineira é crucial para estudos comparativos e de identificação de genes, como os usados em pesquisas de bioprospecção e biotecnologia, devendo prezar por ferramentas de boa precisão e abrangência (Schwengers et al., 2021). Dentre os diferentes *pipelines* de bioinformática, Prokka e RAST (Rapid Annotations using Subsystems Technology), possuem bons

desempenhos para anotação de genomas bacterianos de maneira automática, rápida, consistente e de alta qualidade (Seemann, 2014; Aziz et al., 2008). Assim, com a bioprospecção, identificam-se genes e vias metabólicas associadas à produção de compostos de interesse biotecnológico, viabilizando a prospecção de microrganismos com potencial de aplicação em indústrias farmacêuticas e agrícolas (Saini & Mishra, 2024; Ashajyothi et al., 2024). Todavia, a ampla variedade de *pipelines*, aliada ao crescimento exponencial no número de genomas procarióticos depositados, vinha resultando em erros de anotação. Para mitigar esse problema, foi desenvolvido o projeto *Reference Sequence* (RefSeq), que utiliza o pipeline *Prokaryotic Genome Annotation Pipeline* (PGAP), do NCBI, responsável pela anotação automática das sequências depositadas. Os resultados já demonstram uma redução significativa nas anotações incorretas e têm ampliado o número de sequências com acurácia disponíveis, enriquecendo outras ferramentas bioinformáticas (Li et al., 2021).

No contexto do crescimento de genomas depositados e anotados, vários isolados do gênero *Bacillus*, como *B. amyloliquefaciens*, *B. subtilis* e *B. tequilensis* estão sendo identificados e caracterizados *in silico*. Considerando o potencial biotecnológico dessas espécies, as mesmas estão sendo identificadas como produtoras de surfactinas, iturinas, fengicinas, macrolactinas, bacilomicina-D e sideróforo bacilibactina e, por isso, possuem atividade antifúngica e potencial como biofertilizantes e agentes de biocontrole em plantações de trigo (Shahid et al., 2021). Da mesma forma, uma cepa marinha de *Bacillus siamensis* teve associação com 19 *clusters* de interesse em biossíntese (BGCs) de fengicinas, macrolactinas, surfactinas e bacilibactina, que teve suas produções confirmadas por LC-DAD-MS, mostrando produção de moléculas bioativas para uso farmacêutico (Pan et al., 2019). A espécie *B. paralicheniformis* também apresenta grande contribuição em bioprospecção e biotecnologia. Em 2024 foi demonstrado enriquecimento em genes de lipopeptídeos, como fengicina e surfactina, no genoma acessório do grupo *Bacillus subtilis* (Wang et al., 2024). Em 2018 já se reconhecia que *B. licheniformis* possuía potencial biossintético inferior a *B. paralicheniformis*, em relação ao enriquecimento em peptídeos não-ribossomais (NRPS) e *clusters* híbridos PKS/NRPS (Othoum et al., 2018). Em outro estudo focado em isolado de rizosfera, foram identificados BGCs que incluíam fengicina, liquenisina, bacitracina e sideróforos, todos associados ao biocontrole de fungos e bactérias (Du et al., 2019). Estes trabalhos demonstram a importância dos *pipelines* bioinformáticos de anotação em descobertas genômicas em espécies do gênero *Bacillus* ligadas à biotecnologia e biocontrole.

#### 1.4 Abordagens de identificação e caracterização bacteriana

Em um estudo avaliando *Bacillus paralicheniformis* derivado do microbioma da panícula do milho perolado, suas características fenotípicas mostraram colônias enrugadas, semelhantes a líquens, ásperas e brancas cremosas, com uma margem irregular e sendo bactérias gram-positivas móveis com atividade de celulase e hidrólise do amido, capazes de crescer à 37°C/24h (Ahire et al., 2020; Fatani et al., 2021; Ashajyothi et al., 2024). Por mais que há análises fenotípicas da espécie, estudos recentes vêm questionando a confiabilidade dos mesmos na identificação de espécies do gênero *Bacillus*, especialmente em contextos clínicos. A identificação é difícil devido ao compartilhamento do mesmo padrão morfológico, bioquímico e genético (Pereira et al., 2020). Esta preocupação muitas vezes ultrapassa e chega a métodos mais modernos (Caldeira et al., 2024). A tecnologia de Ionização por Dessorção a Laser Assistida por Matriz – Tempo de Voo/Espectrometria de Massas (MALDI-TOF MS) vem sendo usada, como uma boa opção, na identificação de microrganismos, bem como o método molecular de sequenciamento do gene 16S rRNA (Costa et al., 2022).

Embora eficiente para identificar os organismos, o método de filogenia do gene 16S rRNA tem mostrado dificuldades em diferenciar espécies pertencentes aos mesmos grupos, haja vista a semelhança entre as sequências nucleotídicas no RNA ribossomal (Priest et al., 1994; Wang et al., 2007; Fernández-No et al., 2013; Gupta et al., 2020). Esta dificuldade trouxe a proposta de seis novos gêneros de espécies de *Bacillus*, deixando no gênero *Bacillus* espécies pertencentes ao grupo *B. cereus sensu lato* intimamente relacionadas filogeneticamente sem diferenciação pelo gene 16S rRNA (Patel et al., 2020). Muitas vezes, o sequenciamento do gene 16S rRNA necessita de sequenciamento complementar de genes *housekeeping*, como *rpoB* ou *gurB* (Caldeira et al., 2024). O MALDI-TOF, por sua vez, consegue diferenciar espécies bacterianas pertencentes a um mesmo grupo (Fernández-No et al., 2013). Contudo, na distinção de espécies dos gêneros *Bacillus* e *Paenibacillus* em 2016, a tecnologia apresentou limitações na identificação correta de alguns isolados (Celandroni et al., 2016). Vale ressaltar que em 2022, o MALDI-TOF recebeu expansão no banco de dados usado para identificar *Bacillus* e gêneros relacionados (Costa et al., 2022).

Embora tais métodos tenham sido amplamente usados para identificação de *Bacillus*, estudos demonstraram que diversas linhagens inicialmente classificadas como *B. licheniformis* foram reclassificadas como *B. paralicheniformis* após análise de WGS (EFSA CEP Panel, 2024; Ashajyothi et al., 2024). A técnica de MALDI-TOF MS também não

consegue identificar a nível de espécie a maioria das bactérias aeróbicas formadoras de esporos (Caldeira et al., 2024). Contudo, análises indicam certa influência na precisão e eficiência do método em casos de diferentes métodos de preparação, bem como tempo de incubação (Janiszewska et al., 2023). Todos estes achados indicam que em grupos tão semelhantes, apenas o WGS oferece resolução suficiente para evitar erros taxonômicos, sendo, portanto, o padrão ouro de identificação para esta espécie (Mazhari et al., 2025; Olajide et al., 2021). Desta forma, apesar de úteis na triagem inicial, métodos fenotípicos, MALDI-TOF e filogenia por 16S rRNA devem ser complementados com abordagens genômicas para garantir maior precisão (Pereira et al., 2020; Caldeira et al., 2024; Ashajyothi et al., 2024).

A hibridização DNA-DNA (DDH), que mede a similaridade genômica entre bactérias por hibridização cruzada de DNA fragmentado, vinha sendo usada há quase 50 anos como o padrão-ouro para identificação de espécies procarióticas em nível genômico, tendo sido proposto por Socransky et al, em 1994 (Socransky et al., 1994; Alarcón-Sánchez et al., 2025). Com a ascensão da genômica em 2002, este método foi se tornando cada vez mais obsoleto, devido à falta de disponibilidade desses dados em bancos de dados e pela demora (Stackebrandt et al., 2002; Richter & Rosselló-Mora, 2009). Com os avanços da bioinformática, o método foi atualizado, dando origem a Hibridização DNA-DNA digital (dDDH), que com algoritmos, faz a análise *in silico* similar ao método original, podendo aproveitar os dados históricos advindos do mesmo, mas com uma acurácia e eficiência superiores (Meier-Kolthoff et al., 2013). Além dessa ferramenta, também foi estabelecido o método de identidade média de nucleotídeos (do inglês, “Average Nucleotide Identity”) (ANI) baseado em WGS, que é fundamental na delimitação de espécies baseada em sequências genômicas (Richter & Rosselló-Mora, 2009; Hernández-Salmerón et al., 2023). Este método se mostrou fundamental na distinção de cepas inicialmente caracterizadas como *B. licheniformis* por sequenciamento 16S rRNA que, na realidade, eram *B. paralicheniformis* (Ashajyothi et al., 2024). O cálculo realizado pelo ANI envolve fragmentação de sequências genômicas, seguida pela busca de sequências nucleotídicas, alinhamento e cálculo de identidade (Yoon et al, 2017). Outros métodos também são considerados relevantes, como a Tipagem por Sequenciamento Multilocus de Genes Ribossômicos (rMLST), que se baseia na análise de variações em 53 genes codificadores de subunidades de proteínas ribossômicas (genes *rps*), permitindo a identificação da posição filogenética de bactérias em diferentes níveis taxonômicos, desde o domínio até o nível de linhagem, integrando classificação

taxonômica e tipagem microbiana (Jolley et al., 2012). Além disso, o método do Índice de Correlação da Assinatura de Tetranucleotídeos (TETRA), que compara genomas com base na frequência relativa de tetranucleotídeos (sequências de quatro nucleotídeos), gerando assinaturas genômicas utilizadas para estimar a proximidade filogenética entre eles (Teeling et al., 2004). Todos esses métodos representam opções relevantes, que se corroboram para a identificação filogenômica correta de organismos.

A filogenômica, por sua vez, é caracterizada como a junção da filogenética e genômica para reconstrução de árvores filogenômicas confiáveis e propostas de novas espécies (Chan & Ragan, 2003; Ocaña & Dávila, 2011; Patel & Gupta, 2020; Kapli et al., 2020). Através dela, é possível reconstruir as histórias evolutivas dos organismos, considerando genomas inteiros ou grandes frações de genomas (Patané et al., 2017). Mais uma vez, esse cenário abraça a importância do WGS como padrão-ouro para identificação, taxonomia e inferência filogenética de *B. paralicheniformis*. Uma boa acurácia é necessária para melhor confiabilidade de análises subsequentes de genômica comparativa, como a análise de pangenoma. Em resumo, com ideias sendo aplicadas desde 2007, a genômica comparativa realiza comparações diretas entre o conteúdo genético de dois ou mais organismos, sejam da mesma espécie ou não (Sivashankari et al., 2007; Pudova et al., 2022). Essa abordagem é amplamente aplicada em estudos bioinformáticos em associação com filogenômica e pan-genômica (Benevides et al., 2017; Figueiredo et al., 2015; Zaid et al., 2022).

No contexto das análises comparativas, o pan-genoma pode ser caracterizado como o repertório completo não-redundante de genes da espécie ou gênero de interesse, sendo dividido em três categorias: *core*, acessório (também chamado de *core* estendido ou genoma dispensável) e *singletons*. O primeiro engloba todos os genes compartilhados entre todas as linhagens participantes do *n* amostral estudado; o segundo consiste em genes compartilhados por um subconjunto de linhagens e, finalmente, o terceiro indica genes exclusivos de cada cepa (Tettelin et al., 2005; Tettelin et al., 2008; Asif et al., 2023). Genes acessórios e/ou exclusivos podem estar relacionados com processos de adaptação (Medini, 2005). Tais genes, ainda, conferem a tendência aberta ou fechada dos pan-genomas, indicando inclusão de novos genes específicos a algumas cepas, como foi relatado anteriormente em cepas de *B. paralicheniformis* (Asif et al., 2023). Desta forma, o uso de WGS garante a acurácia necessária na identificação taxonômica e filogenética de espécies geneticamente próximas, cujas distinções por métodos convencionais é desafiador. Posteriormente, essa precisão é

crucial na categorização de diversidade intra e interespecífica, evitando distorções causadas por classificações incorretas.

## 2. JUSTIFICATIVA DA PESQUISA DA DISSERTAÇÃO

Linhagens da espécie *Bacillus paralicheniformis* já vinham sendo estudadas, porém reconhecidas como *B.licheniformis*. Apenas em 2015 houve distinção de ambas espécies, onde nota-se a escassez de estudos que abordem a espécie *B. paralicheniformis* como um todo e de forma isolada, com análises comparativas e filogenômicas que contribuam para o entendimento de seu repertório genético. Consideram-se também as divergências na identificação taxonômica de novas espécies de *Bacillus* em métodos como o MALDI-TOF, em relação aos métodos mais recentes, derivados do Sequenciamento de Nova Geração (NGS) como o *Whole Genome Sequencing* (WGS). Isso foi observado em um estudo prévio que utilizou um dos isolados do estudo atual, o qual foi erroneamente identificado como *Bacillus subtilis*, ressaltando a importância das análises dos genomas para identificações mais precisas. Ademais, a espécie *B. paralicheniformis* demonstra grande potencial biotecnológico, tendo em vista o conhecimento da espécie sobre a produção isolada de metabólitos, que incluem compostos antimicrobianos (AMCs), vitaminas do complexo B, com atuação comprovada no crescimento e biocontrole em plantas, além da atuação ainda escassa em modelos animais.

Em suma, é questionado como análises comparativas, filogenômicas e funcionais em linhagens sequenciadas por NGS poderiam contribuir para a caracterização taxonômica e biotecnológica da espécie *Bacillus paralicheniformis* em um âmbito geral.

### 3. OBJETIVO DA DISSERTAÇÃO

#### 3.1 Geral

Contribuir para a compreensão do repertório genético e do potencial biotecnológico da espécie *Bacillus paralicheniformis*, com ênfase em duas linhagens isoladas do solo da Bahia, por meio de análises comparativas, filogenômicas e funcionais com base em genomas sequenciados por NGS.

#### 3.2 Específicos

- Verificar a identidade taxonômica e caracterizar o perfil genômico dos isolados BAC30 e BAC220.
- Integrar os isolados a análises filogenômicas com outras linhagens de *Bacillus paralicheniformis* disponíveis em bancos de dados públicos.
- Executar análises comparativas entre os genomas dos isolados e demais linhagens da espécie, visando avaliar a diversidade intraespecífica.
- Realizar a anotação funcional dos isolados BAC30 e BAC220 para identificar genes associados a metabólitos bioativos e ao potencial biotecnológico da espécie.

# CAPÍTULO I

#### 4. CAPÍTULO 1

##### Artigo científico:

##### **Comparative and Functional Analyses of *Bacillus paralicheniformis* Strains BAC30 and BAC220 by WGS Uncover Species Homogeneity and Biotechnological Potential**

**Submetido em: World Journal of Microbiology and Biotechnology**

Baseado na escassez de estudos comparativos envolvendo a espécie *Bacillus paralicheniformis*, o artigo a seguir procurou realizar análises taxonômicas, filogenômicas e comparativas com 28 genomas completos da espécie, obtidos pelo NCBI, em conjunto com os isolados do solo da Bahia, BAC30 e BAC220, também identificados como *B. paralicheniformis*. Essas análises buscam entender melhor o repertório genético da espécie, além de ressaltar a importância do Sequenciamento de Genoma Completo (WGS) na identificação taxonômica de espécies do gênero *Bacillus*. Além disso, pela importância industrial da espécie, foram feitas anotações funcionais, buscando genes relacionados ao contexto biotecnológico, especificamente genes de resistência, biossíntese de vitaminas e de peptídeos antimicrobianos, a fim de avaliar, *in silico*, o potencial de uso desses organismos como bactérias de biocontrole.

**Comparative and Functional Analyses of *Bacillus paralicheniformis* Strains BAC30 and BAC220 by WGS Uncover Species Homogeneity and Biotechnological Potential**

Gabriel Camargos Gomes<sup>1</sup>, Eduarda Guimarães Sousa<sup>1</sup>, Ludmila Silva Quaresma<sup>1</sup>, Rhayane Cristina Viegas Santos, Gabriela Munis Campos<sup>1</sup>, Janaíne Aparecida de Paula<sup>1</sup>, Marcus Vinícius Canário Viana<sup>1</sup>, Rafael Junio de Oliveira<sup>1</sup>, Giovanna Karine Viana Silva<sup>1</sup>, Bertram Brenig<sup>2</sup>, Vasco Azevedo<sup>1\*</sup>

<sup>1</sup> Cellular and Molecular Genetics Laboratory, Department of General Biology, Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

<sup>2</sup> Institute of Veterinary Medicine, University of Göttingen, Göttingen, Germany

\*Correspondence: Vasco Azevedo, [vascoariston@gmail.com](mailto:vascoariston@gmail.com)

## Abstract

The *Bacillus* genus includes plant growth-promoting rhizobacteria (PGPR), and the discovery of new strains within this group is of great biotechnological interest due to their ability to produce antimicrobial compounds (AMCs), vitamins, enzymes, and heterologous proteins. Among these, *Bacillus paralicheniformis* is a recently described species whose phylogeny remains poorly resolved, highlighting the need for further investigation. This study aimed to identify and characterize the isolates BAC30 and BAC220 using whole-genome sequencing (WGS). Both were confirmed as *B. paralicheniformis* and included in phylogenomic and comparative analyses with 28 other strains to assess the species' genetic structure and inter-strain similarity. Functional annotation of BAC30 and BAC220 was also performed, focusing on biotechnological potential. Comparative analysis revealed high genomic similarity among strains, including the two isolates. Pangenome analysis showed a low proportion of core genes relative to accessory genes (shell and cloud), and the rarefaction curve suggested an open pangenome, indicating the species' ubiquity and co-evolution with other organisms. Functional analysis identified genes of defense mechanisms related to beta-lactam resistance. Regarding secondary metabolite production, genes involved in the biosynthesis of vitamins (e.g., riboflavin) and AMCs (e.g., bacitracin) were detected. Although further *in vitro* and *in vivo* assays are needed to confirm gene expression, the findings support the biotechnological relevance of these isolates as potential biocontrol agents and/or producers of industrially valuable compounds.

**Keywords:** *Bacillus*; Biocontrol; Agriculture; Industry; Metabolites

## Introduction

The *Bacillus* genus is highly diverse and extensively studied. Its main characteristics include rod-shaped bacteria capable of sporulating when exposed to environmental stress factors, primarily nutrient deprivation (Paredes-Sabja et al. 2011), but also sudden changes in pH, salinity, and temperature (Gauvry et al. 2021). Originally, the genus included species such as *B. subtilis* (the most extensively studied and a global reference among Gram-positive bacteria (Errington and Aart 2020)), *B. licheniformis*, *B. pumilus*, and *B. amyloliquefaciens*. Later, additional species were incorporated, such as pathogenic ones, like *B. cereus* and *B. anthracis*, as well as others of importance to agriculture and industry, such as *B. thuringiensis* (Blanco Crivelli et al. 2024). With the advancement of sequencing and genome annotation technologies, new species and subspecies have been identified, such as *B. paralicheniformis*, expanding the genetic landscape of the genus. However, as *B. paralicheniformis* was only recently established (many of its strains were previously classified as *B. licheniformis*), the number of available studies and deposited genomes remains limited compared to other species in the group. This highlights the need for further investigations to characterize its genomic profile, particularly regarding evolutionary and phylogenomic patterns.

This species, recently described and separated from *B. licheniformis* (Dunlap et al. 2015), already shows industrial application potential, with notable prominence in agriculture, where it is recognized as a plant growth-promoting rhizobacterium (PGPR) (Du et al. 2019). It is known to produce bacteriocins and other metabolites (Choyam et al. 2021) that help maintain the local microbiota by inhibiting pathogens such as fungi (Ramírez-Cariño et al. 2020), bacteria (Zhao et al. 2023), viruses (Yu et al. 2024), and nematodes (Chavarria-Quicaño et al. 2023). In addition, these bacteria promote the growth of plants like wheat and cotton (Xu et al. 2023), and may contribute to nitrogen fixation, phosphate solubilization, potassium release, and the production and release of phytohormones (Olanrewaju et al. 2017).

Although the induction of biosynthesis of molecules of interest and the establishment of optimal conditions for efficient production require the evaluation of multiple variables, studied under the scope of metabolic engineering (Kim et al. 2016), the use of individual bacterial strains that produce secondary metabolites, such as bacitracin, bacillibactin, fengycin, and surfactin, has proven valuable. These metabolites have potential applications as natural preservatives and surfactants, offering a promising alternative to chemical additives currently used in the food industry, which are associated with potential toxicological and

teratogenic effects harmful to human health. In this context, consumer demand for healthier and more natural biopreservatives has increased (Kumariya et al. 2019).

Moreover, previous studies have assessed the application of *B. paralicheniformis* strains in animal production, including strain BAC220, which is the focus of the present study (initially identified as *B. subtilis* via MALDI-TOF analysis) (Dos Reis et al. 2022). This strain was shown to reduce necrosis rates and inflammatory markers in chicken embryos challenged with *Salmonella Pullorum* (*Salmonella enterica* subsp. *enterica* serovar Gallinarum biovar Pullorum).

Given the biotechnological relevance of the *Bacillus* genus, this study aimed to explore the evolutionary profile of the *B. paralicheniformis* species, conducting phylogenomic and comparative analyses of strains obtained from the National Center for Biotechnology Information (NCBI) database, along with functional analyses of two isolated strains, BAC30 and BAC220. These analyses focused on functional genes related to defense mechanisms and the biosynthesis of secondary metabolites with potential for biotechnological application.

## Materials and methods

### Genome assembly, contamination check, and completeness evaluation

The strains, designated BAC30 and BAC220, were originally isolated in the 1990s from soil and rabbit feces, respectively, deposited at Fundação Oswaldo Cruz (FIOCRUZ), and donated by Dr. Leon Rabinovitch. Initially, both strains were identified as *Bacillus subtilis* using MALDI-TOF analysis (Dos Reis et al., 2022), and later as *Bacillus cereus* and “unknown”, respectively, which prompted further taxonomic identification through sequencing due to inconsistencies in the MALDI-TOF results. DNA extraction was performed using the Wizard® Genomic DNA Purification Kit (Promega), following the manufacturer's instructions. Next-generation sequencing was conducted on the HiSeq 2500 platform (2×150 bp) (Illumina®, United States), and paired-end libraries were constructed using the ThruPLEX DNA-Seq Kit (Takara).

For genome assembly, annotation, and taxonomic identification, initial trimming was performed using FastP (v0.24.0), a tool capable of detecting and removing adapters and low-quality reads, thereby improving the overall quality of the sequenced genome (Chen et al. 2018). This was followed by a quality control assessment using FastQC (v0.12.1), which generates a comprehensive report on several sequencing metrics (Babraham Bioinformatics, 2010). Genome assembly was carried out with Unicycler (v0.5.1), a tool optimized for short-read data generated by platforms such as Illumina (Wick et al. 2017).

Contamination and completeness assessments were performed using CheckM2 (v1.0.2), which applies a machine learning model independent of phylogeny (Chklovski et al. 2023). Additional quality control analyses were conducted using GUNC (v1.0.6), which detects possible chimeric contigs (fragments erroneously assembled from different genomic sources) (Orakov et al. 2021). Moreover, Barrnap (v0.9) was employed to predict conserved ribosomal genes such as 5S, 16S, and 23S rRNA, providing an additional metric for evaluating the completeness of the assembled genomes (Seemann, 2013)

### Taxonomic Identification

Taxonomic identification was performed using the Genome Taxonomy Database Toolkit (GTDB-Tk) (v2.4.1) and the Type (Strain) Genome Server (TYGS), to ensure high confidence in the classification results. GTDB-Tk infers taxonomic assignments based on the standardized taxonomy of the Genome Taxonomy Database (GTDB), applying the Relative Evolutionary Divergence (RED) criteria, followed by Average Nucleotide Identity (ANI)

validation for species-level resolution (Chaumeil et al. 2020). The commonly accepted ANI threshold for confident species-level identification is >95–96% (Elbir 2024). TYGS performs digital DNA-DNA hybridization (dDDH) analysis, which assesses phylogenomic relatedness between submitted genomes and database entries based on the degree of hybridization between complete genomes or genomic fragments, considering overlap, segment size, and similarity (Meier-Kolthoff and Göker 2019). For dDDH, the standard threshold for confident species-level identification is >70% (Meier-Kolthoff et al. 2013).

In addition, the pyANI tool was used to perform all-vs-all ANI comparisons, generating a pairwise similarity distance matrix (Pritchard et al. 2016). This analysis included 28 complete *B. paralicheniformis* genomes retrieved from the NCBI database (Table 1), along with the two isolates from this study (BAC30 and BAC220). A heatmap was generated using a custom script (Script S1) in RStudio (v2024.12.1) (Posit Team, 2025). Subsequently, using the pyANI software, the two isolates were compared against 14 genomes representing other *Bacillus* species (Table S3). All genomes were retrieved directly from the NCBI database (except for the isolates) and are designated as reference genomes for their respective species according to NCBI. They are not type strains, except for *B. licheniformis* ATCC 14580 and *B. siamensis* KCTC 13613.

### Compilation of the *Bacillus paralicheniformis* dataset

Table 1 presents 28 complete *Bacillus paralicheniformis* genomes (RefSeq), all obtained from the NCBI using the NCBI Datasets tool (v16.30.0) (O’Leary et al. 2024). This same dataset was used in all subsequent analyses, with the addition of the two isolates from this study, BAC30 and BAC220. The table includes information such as strain names, NCBI assembly accession numbers, approximate total genome size (in Megabases, Mb), as well as the source and country of isolation, in order to assess genomic similarities and differences among the strains.

**Table 1** Information on the 28 complete *Bacillus paralicheniformis* genomes retrieved via NCBI Datasets.

| Strain Name | NCBI RefSeq Assembly | Genome size (Mb) | Isolation source | Country of origin |
|-------------|----------------------|------------------|------------------|-------------------|
| Bac84       | GCF_002993925.1      | 4.38             | Red Sea Lagoon   | Saudi Arabia      |
| PRO109      | GCF_029536955.1      | 4.62             | Soy-based food   | Ghana             |

|            |                 |      |                           |              |
|------------|-----------------|------|---------------------------|--------------|
| A4-3       | GCF_008365255.1 | 4.58 | Tomato                    | South Korea  |
| CP47       | GCF_031326025.1 | 4.54 | Fermented soy-based food  | South Korea  |
| 14DA11     | GCF_002393225.1 | 4.54 | Fermented soy-based food  | South Korea  |
| J41TS8     | GCF_021654755.1 | 4.50 | Japanese honey            | Japan        |
| DY4        | GCF_030166695.1 | 4.55 | Pig feces                 | China        |
| CBMAI 1303 | GCF_003711025.1 | 4.48 | -                         | Brazil       |
| Bac48      | GCF_002993945.1 | 4.46 | Red Sea mangrove mud      | Saudi Arabia |
| A5         | GCF_035985545.1 | 4.46 | Marine sediment           | China        |
| CamBx3     | GCF_026210435.1 | 4.45 | Hot spring water          | -            |
| FA6        | GCF_009497935.1 | 4.45 | Grass carp                | China        |
| NCTC8721   | GCF_900635765.1 | 4.43 | -                         | -            |
| J36TS2     | GCF_021654735.1 | 4.60 | Japanese honey            | Japan        |
| DSM 28591  | GCF_029537035.1 | 4.39 | C57BL/6 mouse feces       | Germany      |
| J25TS1     | GCF_021654715.1 | 4.39 | Japanese honey            | Japan        |
| BL-09      | GCF_000876525.1 | 4.39 | Fermented Congee Soup     | China        |
| MDJK30     | GCF_002068155.1 | 4.35 | Peony rhizosphere         | China        |
| RP01       | GCF_029000405.1 | 4.34 | Soil                      | China        |
| CEW 1W     | GCF_030123025.1 | 4.34 | Marine sediment           | India        |
| 285-3      | GCF_026723705.1 | 4.32 | Soil                      | China        |
| RSC-1      | GCF_018324505.1 | 4.32 | Red sea water             | Saudi Arabia |
| RSC-2      | GCF_018324565.1 | 4.32 | Red sea water             | Saudi Arabia |
| SUBG0010   | GCF_003171815.2 | 4.32 | Rhizosphere               | India        |
| SYN-191    | GCF_029590455.1 | 4.29 | Soil                      | China        |
| ATCC 9945a | GCF_000408885.1 | 4.38 | -                         | -            |
| Baplich1   | GCF_038049955.1 | 4.38 | Tomato rhizosphere        | China        |
| FUA2150    | GCF_039619465.1 | 4.45 | Fermentation starter Daqu | China        |

### Phylogenomic analyses

Starting from the 28 complete genomes and the two isolated genomes (BAC30 and BAC220), one outgroup was included: *Bacillus sonorensis* ASM3405511v1 (GCF\_035804855.1). This strain was chosen as an outgroup due to its close taxonomic relationship, although it does not belong to the same species (Olajide et al. 2021). For phylogenomic tree construction, the programs Gegenees and SplitsTree4 (v4.14.4) were used. Gegenees generates a pairwise distance matrix (all vs all) between genomes by fragmenting them and assigning scores to gene regions, thus standardizing the comparison, which is performed using BLAST (Ågren et al. 2012). SplitsTree4 then uses this distance matrix to infer phylogenetic relationships through the Neighbor Joining (NJ) method (Huson and Bryant 2006).

Additionally, Orthofinder (v3.01b1) was used to construct a phylogenetic tree based on all-vs-all comparisons of protein sequences, leveraging orthologous and paralogous relationships. The tree was generated using the Maximum Likelihood (ML) method with 1000 bootstrap replicates (Emms and Kelly 2019). The resulting tree was visualized and edited using the Interactive Tree of Life (iTOL) platform (Letunic and Bork 2024).

### **Comparative analyses**

The 28 complete *Bacillus paralicheniformis* genomes obtained from NCBI were annotated using Prokka (v1.14.6), in order to standardize them with the two isolated genomes from this study, RNAmmer (Lagesen et al. 2007), and others (Seemann, 2014). A comparative circular genome map of the 30 genomes was then generated using the Proksee platform (Grant et al. 2023), with the NCBI reference genome *Bacillus paralicheniformis* Bac84 (GCF\_002993925.1) as the base. From this reference genome, a BLAST comparison was performed against each of the other genomes using BLAST+ (v2.16.0), integrated within the Proksee software, using the default settings provided by the tool.

### **Pangenomic Analyses**

The PPanGGolin software (v2.0.5) was used to perform the pangenome analysis. This tool integrates information from protein-coding genes and their genomic neighborhoods to construct a gene family graph, where each node represents a gene family, and edges between nodes reflect similarity relationships among families (Gautreau et al. 2020). The analysis included the 28 previously deposited genomes along with the two newly sequenced isolates. The program outputs several key parameters, including the number and percentage of gene families, categorized into accessory (cloud), shared (shell), and persistent (core) genes.

A rarefaction curve was also constructed based on Heap's Law, a mathematical model describing the increase in the number of gene families (or unique genes) as additional genomes are included in the analysis. This curve allows inference of whether the pangenome is open or closed, based on the  $\alpha$  (alpha) parameter. Alpha values below 1 indicate an open pangenome, suggesting that the inclusion of new genomes continues to yield novel genes (cloud). In contrast, alpha values equal to or greater than 1 indicate a closed pangenome, in which most genes have already been captured and the addition of new genomes has minimal impact on the total gene count (Felice et al. 2023).

### **Functional analysis**

For the functional analyses, the COG classifier tool was first used to perform functional annotation, classification, and analysis of all genes based on comparisons with the COG (Clusters of Orthologous Groups) database (Shimoyama 2022). This tool was used to compare the core genes of the isolated strains with those of the other genomes in the dataset, in order to evaluate the impact of the isolates on the overall core genome of the species and to assess differences in the gene proportions across COG functional categories.

Next, the BlastKOALA tool, integrated with the KEGG platform, was used to functionally annotate the sequences based on the KEGG database, assigning KEGG Orthology (KO) identifiers. This enabled the characterization of individual gene functions and the reconstruction of specific KEGG pathways and modules (Kanehisa et al. 2016), considered relevant to the biotechnological potential of the isolates, with emphasis on genes associated with defense mechanisms, secondary metabolites production, and genes related to nitrogen metabolism.

Finally, the antiSMASH software (v8.0) was used, a tool focused on the analysis of microbial genomes with an emphasis on the annotation of genes involved in biosynthetic gene clusters (BGCs), particularly those encoding nonribosomal peptide synthetases (NRPS), which are of considerable biotechnological interest (Blin et al. 2025).

## Results

### Assembly, quality assessment, and taxonomic identification of the isolated strains

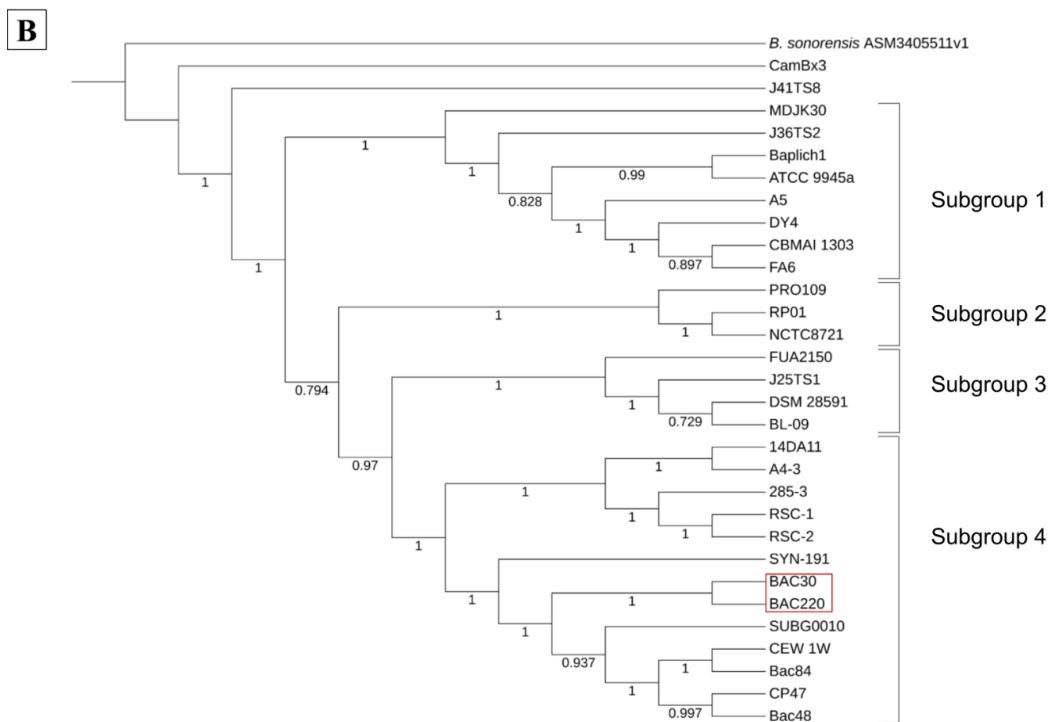
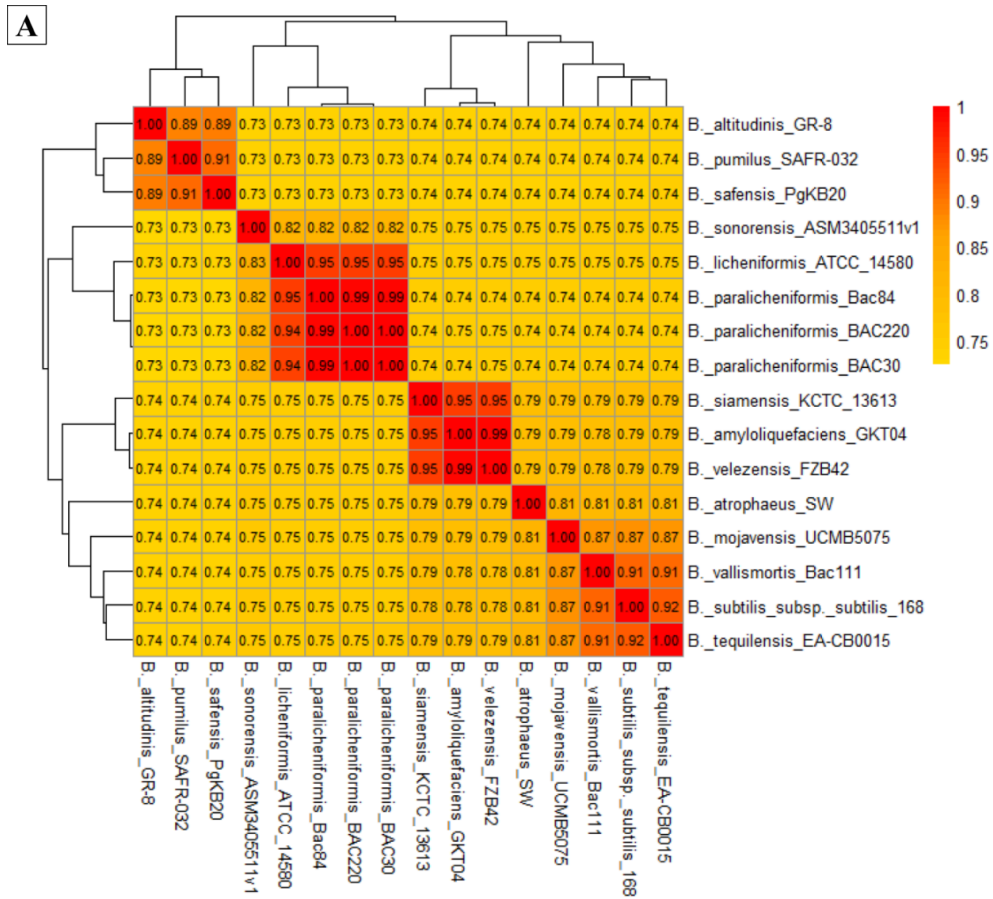
The genomes of the isolates were designated BAC30 and BAC220 and were subsequently identified as *Bacillus paralicheniformis* by both applied methods (GTDB-Tk and TYGS), showing an acceptable degree of similarity (with thresholds above 95% for ANI and 70% for dDDH) (Table S1). The GC content, ranging between 45% and 46%, is consistent with previously reported *B. paralicheniformis* genomes (Du et al. 2019; Asif et al. 2023). Quality assessments indicated 100% completeness, with negligible contamination (below 5%), and both samples were confirmed to be free of chimeric sequences. Furthermore, both isolates harbored the three conserved ribosomal RNA genes (5S, 16S, and 23S), further supporting genome completeness. These findings confirm that the isolation, genomic DNA extraction, and sequencing procedures were properly conducted, yielding high-quality and contamination-free samples (Table S2). The genomes have been deposited in the public GenBank database, with the strain BAC220 being designated as *Bacillus paralicheniformis* Bp Unicon 01 (SAMN37735107) and the strain BAC30 as *B. paralicheniformis* Bp Unicon 02 (SAMN43408015).

### Taxonomic and phylogenomic analysis

Using the pyANI software, all genomes of *B. paralicheniformis* showed high ANI values (0.97–1.00), including the publicly available genomes and the newly isolated ones. In contrast, the outgroup *B. sonorensis*, highlighted in blue, displayed considerably lower similarity scores (0.81–0.82) (Fig. S1). Subsequently, the isolates were compared to other members of the *Bacillus* genus (Fig. 1a), and showed a high level of similarity with strain Bac84 (0.99–1.00), followed by the reference genome of *B. licheniformis* (0.82), thereby supporting the previous taxonomic identifications.

The analysis of the full *B. paralicheniformis* dataset (30 genomes) using Gegenees and SplitsTree4, in a Neighbor Joining (NJ) phylogenetic tree, revealed high similarity scores among strains (81–100), while the outgroup exhibited significantly lower values (20–22). Notably, the strains CamBx3 and J41TS8 appeared more phylogenetically distant from the rest of the species (Fig. S2). This pattern was consistent with the maximum likelihood (ML) phylogenetic tree generated using OrthoFinder (Fig. 1b), which also positioned the outgroup distinctly apart, followed by the same two divergent strains. Additionally, the ML tree

suggested the formation of four subgroups, supported by satisfactory bootstrap values, all above 0.7 (minimum of 0.729).



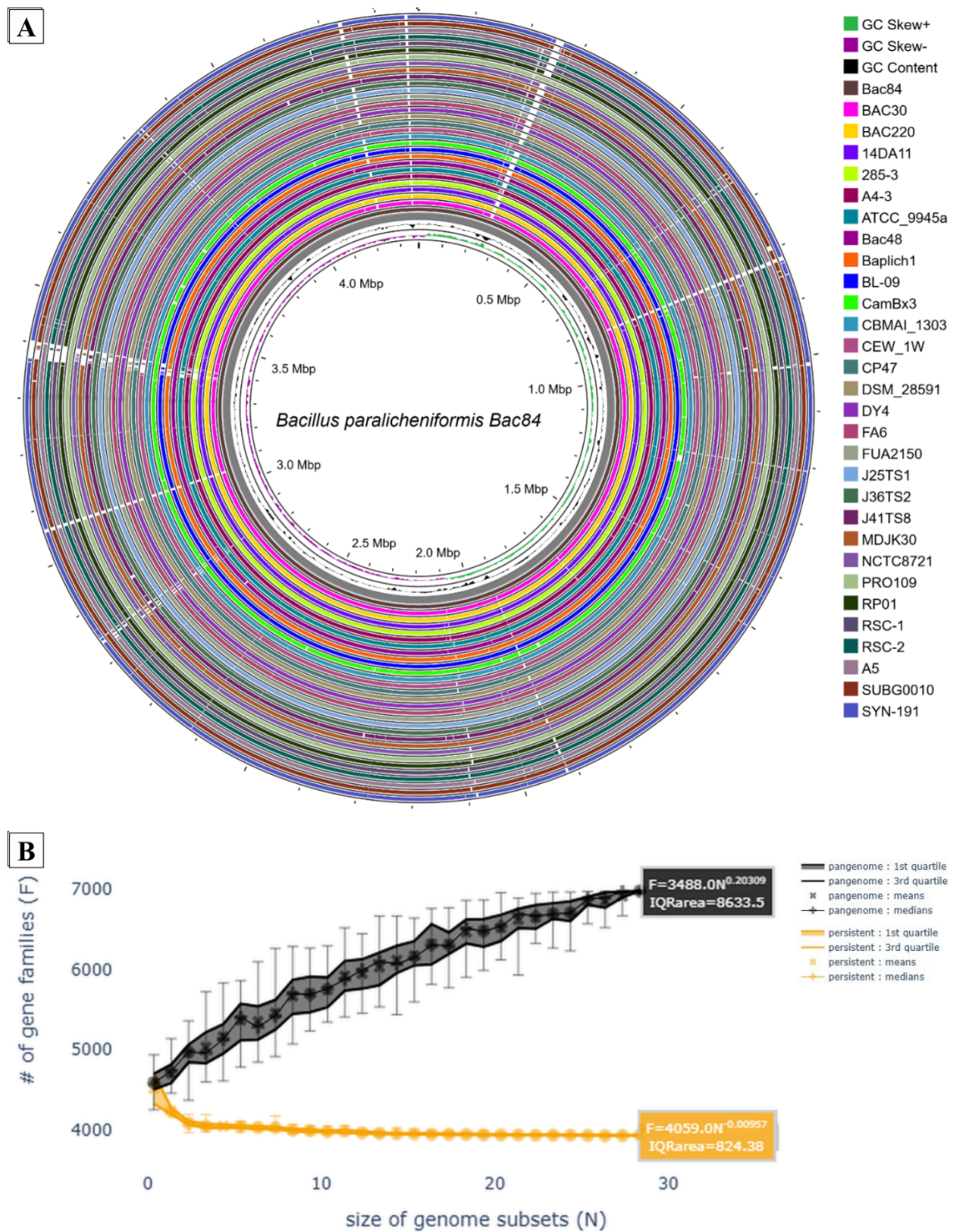
**Fig. 1** Taxonomic and phylogenomic analysis of the *B. paralicheniformis* species **(a)** ANI analysis of 16 *Bacillus* genomes using the pyANI software, including reference genomes for each species and the addition of the isolated strains BAC30 and BAC220 **(b)** Phylogenomic analysis of *B. paralicheniformis* strains performed using OrthoFinder based on the Maximum Likelihood (ML) method. The isolated strains BAC30 and BAC220 are highlighted in red.

## Comparative analysis

### Circular map of *Bacillus paralicheniformis* strains and pangenome analysis

The comparative circular map (Fig. 2a) displays the reference genome Bac84, retrieved from the NCBI database, aligned against the remaining 29 *B. paralicheniformis* genomes, including the isolated strains BAC30 and BAC220. As observed in the phylogenomic analyses, there is a high degree of similarity among the strains and the reference genome (Bac84), evidenced by the minimal presence of white gaps.

The pangenome analysis revealed a total of 129,406 genes grouped into 6,973 gene families. Of these, 3,935 (56.44%) were classified as core genes, 2,581 (37.01%) as cloud genes, and only 457 (6.55%) as shell genes (Fig. S3). In the rarefaction curve of *B. paralicheniformis* genomes (Fig. 2b), the calculated gamma ( $\gamma$ ) value was 0.20309. From this, the corresponding alpha ( $\alpha$ ) value, calculated as  $1 - \gamma$ , was 0.79691. Since  $\alpha < 1$ , the pangenome is considered open, meaning that as more genomes are added, the number of genes in the pangenome continues to increase. To assess the impact of the isolates on the overall species core,  $\alpha$  values were analyzed using only the 28 public genomes, with the isolates excluded. This analysis yielded a  $\gamma$  value of 0.18142, corresponding to an  $\alpha$  value of 0.81858 (Fig. S5).



**Fig. 2** Comparative analyses of the *B. paralicheniformis* dataset (a) Circular map generated in Proksee for the 30 *B. paralicheniformis* strains, with the reference genome Bac84 at the center. Comparisons with the remaining strains were performed using NCBI's BLAST+ (b) Rarefaction curve of the 30 *B. paralicheniformis* genomes, generated using PPanGGolin. The curve for persistent (core) genes is shown in orange, and the pangenome curve is shown in black. The gamma ( $\gamma$ ) value, derived from parameter F, is displayed on both curves.

### Gene categories comparison in COG Classifier

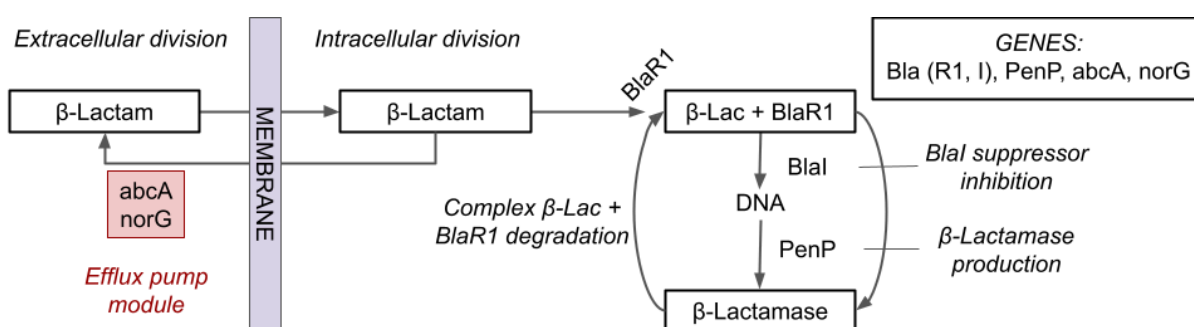
Using the COG Classifier software, the core genes of the two isolates BAC30 and BAC220 were first analyzed (Fig. S4a), followed by an assessment of their impact on the overall gene repertoire of the species (Fig. S4b & S4c). No significant difference in gene quantity was observed, highlighting the similarity between the isolates and the other 28 genomes available in the NCBI database. Additionally, in the core genome of the isolates (Fig. S4a), a proportionally higher number of genes classified under category C (Energy production and conversion) was noted, in contrast to category G (Carbohydrate transport and metabolism). This pattern is reversed when analyzing the core genome of the entire *B. paralicheniformis* dataset (Fig. S4a).

### Functional analysis

The genomes of the two isolated strains, BAC30 and BAC220, were analyzed using the KEGG database, focusing on genes of interest within categories considered relevant for the biotechnological application of these organisms (defense mechanisms, secondary metabolite production, and nitrogen metabolism).

### Defense mechanisms

A gene module related to beta-lactam resistance was identified, involving the Bla system (*BlaR1*, *BlaI*, *penP*), which operates through a negative feedback mechanism (Fig. 3). In this system, the presence of a beta-lactam compound in the intracellular environment binds to the *BlaR1* repressor, forming the B-lactam + *BlaR1* complex and inhibiting the repressor's activity. This inhibition activates the *penP* gene, which induces the production of beta-lactamase. The enzyme then degrades both the complex and the beta-lactam molecule, thereby releasing the *BlaR1* repressor. Moreover, a multidrug efflux system *AbcA* (*abcA*, *norG*) was also identified.



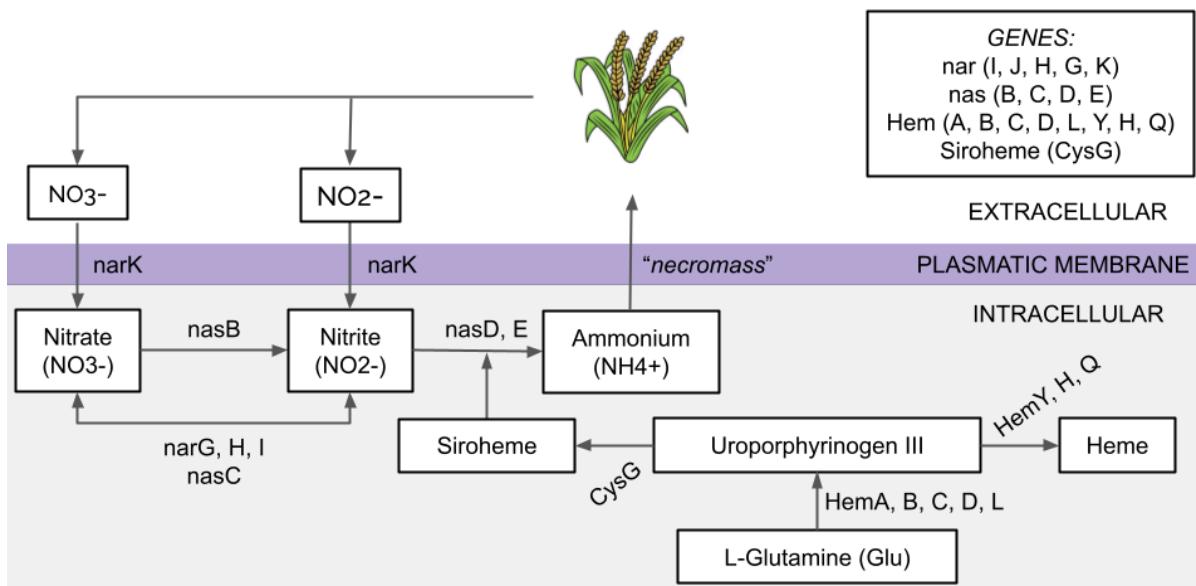
**Fig. 3** Metabolic map of beta-lactam resistance mechanisms based on genes from the BAC30 and BAC220 strains identified through KEGG analysis.

## Secondary metabolites production

Using KEGG, modules related to the biosynthesis of vitamins (B1, B2, B5, B6, B7, and B9) and the cofactors coenzyme A (CoA) and flavin adenine dinucleotide (FAD) were identified. In this context, the metabolic pathway responsible for the biosynthesis of the most relevant vitamin, riboflavin (vitamin B2), is explained. Riboflavin biosynthesis proceeds through the conversion of guanosine triphosphate (GTP) and ribulose 5-phosphate (Ru5P) which are converted into 5-Amino-6(ribityl-amino)uracil (ARU) and 3,4-Dihydroxy-2-butanone 4-phosphate (DHBP), respectively. Subsequently, both form the intermediate 6,7-dimethyl-8-ribityllumazine (DMRL) via the genes *ribBA*, *ribD*, *ribH*, *ybjI*, and *ycsE*, followed by conversion to riboflavin mediated by the *ribE* gene. Riboflavin can then be converted into the cofactor flavin adenine dinucleotide (FAD) through the action of the *ribF* gene.

## Nitrogen metabolism

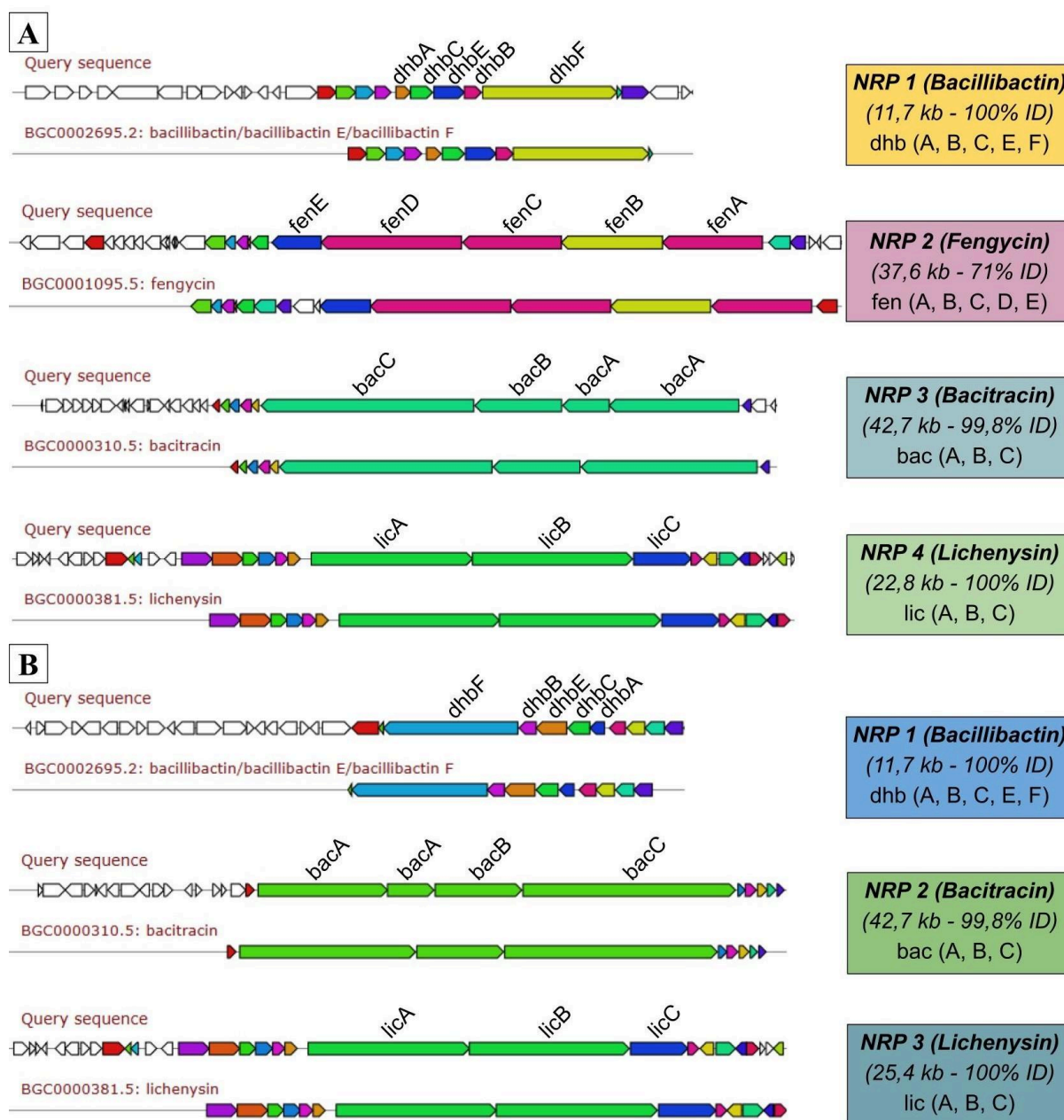
Genes related to nitrogen metabolism were analyzed using KEGG (Fig. 4), revealing a complete module associated with assimilatory nitrate reduction, in which nitrate is converted to nitrite and subsequently to ammonia through a set of *nas* genes (*nasB*, *nasC*, *nasD*, *nasE*). Additionally, the *nar* operon (*narG*, *narH*, *narI*, *narJ*), responsible for the interconversion of nitrate and nitrite, was identified, along with the *narK* gene involved in the transport of nitrate and nitrite from the extracellular environment to the intracellular space. Furthermore, a complete biosynthesis module for siroheme was detected, which is involved in the formation of the nitrite reductase enzyme (responsible for reducing nitrite to ammonium). This pathway includes *hem* genes (*hemA*, *hemB*, *hemC*, *hemD*, *hemL*), which convert L-glutamine (Glu) into uroporphyrinogen III, subsequently transformed into siroheme (*cysG*) and heme (*hemY*, *hemH*, *hemQ*).



**Fig. 4** Metabolic map related to nitrogen metabolism based on genes from the BAC30 and BAC220 strains identified through KEGG analysis.

### Clusters of Non-Ribosomal Peptide Synthesis (NRPS)

Using the AntiSMASH software, the genomes of the BAC30 and BAC220 strains were analyzed, identifying nonribosomal peptide synthesis (NRPS) regions associated with secondary metabolite production, exhibiting high identity (above 70%). In the BAC30 strain (Fig. 5a), regions linked to the biosynthesis of bacitracin, bacillibactin, lichenysin, and fengycin were detected, along with the biosynthetic genes involved in these processes. In contrast, the BAC220 strain exhibited only three regions associated with the production of bacillibactin, bacitracin, and lichenysin (Fig. 5b).



**Fig. 5** Identification of regions in the genomes of strains BAC30 and BAC220 associated with nonribosomal peptide synthesis (NRPS), specifically bacillibactin, fengycin, bacitracin, and lichenysin, using the AntiSMASH software. The figure also shows the NRPS region sizes in kilobases (kb), identity scores based on BLAST analysis (NCBI), and the core biosynthetic genes for each NRPS cluster. **(a)** BAC30 **(b)** BAC220

## Discussion

Due to the relatively recent description of the species *Bacillus paralicheniformis* and the limited number of comprehensive studies on its genomic and functional landscape, this study aimed to deepen the understanding of its phylogeny and behavior in a comparative context. Despite current knowledge limitations, *B. paralicheniformis* has been identified as a species capable of producing bacitracin, a metabolite of industrial interest, which has demonstrated

pathogen-inhibitory activity both *in vitro* and *in vivo*. In this context, the characterization of new strains is essential to broaden the biotechnological application potential of the species, as exemplified by the BAC30 and BAC220 isolates analyzed in this study.

In the context of isolates identification, both strains were successfully sequenced and subsequently processed through genome assembly and annotation. Additionally, they showed high similarity to each other in terms of genome size, GC content (%), and taxonomic classification, as determined by the TYGS and GTDB-Tk tools (Tables S1 & S2). When comparing *B. paralicheniformis* strains among themselves, a high degree of genomic similarity was observed across all genomes (Fig. S1). Furthermore, comparison of the BAC30 and BAC220 isolates with other *Bacillus* species (Fig. 1) revealed strong similarity to the reference genome of *B. paralicheniformis* (Bac84). These three *B. paralicheniformis* genomes also showed greater similarity to the *B. licheniformis* reference genome than to other species within the *Bacillus* genus. Additionally, the clustering of other *Bacillus* species aligned with previous studies that reported similar associations, supporting the reliability of our analysis (Nannan et al. 2021). These results further confirm the relevance and reliability of WGS analyses, considering that the initial identification of the isolates was incorrect, having been previously classified as *unknown* and *Bacillus cereus*, respectively.

In the phylogenomic analyses performed using Gegenees and SplitsTree4 (Fig. S2), the outgroup *Bacillus sonorensis* (ASM3405511v1) exhibited a relatively high similarity score, reflecting its close phylogenetic relationship due to its inclusion within the same genus as the other analyzed genomes. This similarity contributes to the robustness of the phylogenomic inferences, minimizing potential biases and branching errors that could arise from the use of an excessively divergent outgroup (Wilberg 2015). In the analysis conducted with Orthofinder (Fig. 1b), the formation of four distinct subgroups was observed, consistent with the results obtained through Gegenees. The structure of these groups remained stable, with only minor changes in the relative positioning of genomes across different approaches. However, to date, no studies have conducted specific and in-depth analyses focused solely on *Bacillus paralicheniformis* genomes, with most comparative studies involving closely related species within the same genus, such as *Bacillus licheniformis*, *Bacillus subtilis*, and others (Asif et al. 2023). In both analyses, the strains CamBx3 (Narsing Rao et al. 2024) and J41TS8 (Okumura et al. 2022) stood out by clustering farther from the remaining genomes, potentially indicating a relevant phylogenetic divergence. Although these strains share over 95% ANI identity, they may represent a sub-lineage or even a distinct lineage from the currently recognized *B.*

*paralicheniformis*, considering the limited number of available phylogenetic markers for this species. Furthermore, the circular map (Fig. 2a) generated for all *B. paralicheniformis* strains using Proksee once again revealed a high degree of genomic similarity. This finding corroborates previous results from the taxonomic analysis (Fig. S1), the minimal impact observed when comparing the gene repertoire of the isolates with that of the overall species (Fig. S4), and the negligible difference in the alpha value when comparing the pangenome with and without the isolates (Fig. S5).

The analysis performed using the PPanGGolin tool revealed that, despite the high genomic similarity among the *Bacillus paralicheniformis* strains in the dataset, the proportion of persistent genes (Fig. S3) is relatively low when compared to other *Bacillus* species, such as *B. amyloliquefaciens* (75.20%) and *B. anthracis* (74.01%) (Kim et al. 2017). The high number of unique genes and the limited number of shared genes among the analyzed genomes, suggests a pattern of ubiquity for the species. Such dispersion reflects the diversity of isolation sources and the broad geographic distribution of the strains, as shown in Table 1, and highlights the adaptive capacity of the species to acquire new genes. The rarefaction curve (Fig. 2b) also indicates a continuing trend of gene acquisition among the strains analyzed, characterizing the pangenome of the species as open, a feature previously observed in other studies on the species (Du et al. 2019; Asif et al. 2023). Although the strains are highly similar and the inclusion of the isolates did not significantly impact the species' overall gene repertoire (Fig. S4b & S4c) (Fig. S5), this pattern may indicate the species' adaptive capacity to different ecological niches, as well as its widespread environmental distribution, consistent with the distribution of gene families (Fig. S3). However, to confirm this genomic plasticity and adaptive potential, additional analyses are still required, particularly regarding gene synteny and the presence of mobile genetic elements.

Regarding the functional analyses, the results related to defense mechanisms obtained through KEGG revealed the presence of two distinct  $\beta$ -lactam resistance modules, which may act synergistically (Fig. 3). This resistance may be considered intrinsic to organisms of the *Bacillus* genus as, for example, a study analyzing 114 *Bacillus* genomes, including *B. paralicheniformis*, found that approximately 86% of the strains exhibited resistance to penicillin (Zhai et al. 2023). The low abundance of antibiotic resistance genes (ARGs) is crucial for the efficient disposal of these microorganisms during equipment cleaning and for mitigating the spread of ARGs from industrial waste (Li et al. 2022). It also allows for the introduction of classical vectors containing expression inducers or heterologous protein genes,

which frequently include resistance genes used for clonal culture selection (Dong and Zhang 2014). Regarding B-complex vitamin biosynthesis, recent studies have primarily focused on the development of food formulations using *Bacillus* strains capable of producing riboflavin (vitamin B2), an essential compound for monogastric organisms such as humans, poultry, and swine (Lambertz et al. 2020). In this context, several studies have characterized and assessed the safety of riboflavin produced by *B. subtilis* strains, considering it safe for consumption (EFSA Panel on Additives and Products or Substances used in Animal Feed (FEEDAP) et al. 2022). However, there is a potential risk of residual genes in the final product, which may include resistance or virulence genes with the potential for horizontal transfer to pathogenic bacteria (EFSA Panel on Additives and Products or Substances used in Animal Feed (FEEDAP) et al. 2018). Another limiting factor for the commercialization of isolated vitamin compounds is the high production cost, associated with low yield and the availability of more efficient alternatives, which currently makes the process economically unfeasible (Revuelta et al. 2016).

Regarding nitrogen metabolism (Fig. 4), the *nas* and *nar* genes, associated with nitrate reduction and conversion, were identified. These genes indicate the adaptability of the strains to different forms of nitrogen compounds (nitrate, nitrite, and ammonia), which vary in concentration depending on soil type, fertilizers, and animal-derived inputs (Ferraz-Almeida 2024), enabling the balancing of these compounds. Additionally, the conversion of nitrogen compounds may reduce  $\text{NO}_3^-$  (nitrate) levels, whose elevated concentrations contribute to soil acidification and nutrient depletion, ultimately hindering plant growth (Florio et al. 2025). It is also considered that these bacteria may increase nitrogen retention time in the soil, preventing its volatilization as nitrous oxide ( $\text{N}_2\text{O}$ ), a greenhouse gas that lowers fertilizer efficiency (Wu et al. 2018), as well as enabling a reduction in excessive fertilizer use, which is recognized as a major source of environmental pollution (Ahmed et al. 2017). Although nitrogenous compounds generated by these bacteria can be taken up by plants, no genes related to their extracellular transport were found. However, due to the rapid cell cycle of *Bacillus* strains, intracellular content, including metabolites and nitrogen compounds, can be released into the rhizosphere after bacterial lysis, making these compounds available to plants as necromass (Pausch et al. 2024). The identified heme biosynthesis pathway is also noteworthy, as microbial heme production has been explored in pharmaceutical and food industries as an economically viable and animal-free alternative to traditional production methods involving animal blood (Yang et al. 2023).

In the analysis of nonribosomal peptides (NRPs) associated with the production of AMCs (Fig. 5), bacitracin stands out as the most relevant, being primarily produced by strains of *Bacillus licheniformis* and *B. subtilis* (Cai et al. 2020). This antibiotic is widely used in topical formulations, such as ointments, often in combination with other antimicrobials like neomycin (Jones et al. 2006), contributing to the prevention of infections caused by pathogens and/or opportunistic microorganisms. Its efficacy has been demonstrated, for example, in a study indicating bacitracin as the main metabolite responsible for inhibitory activity against the pathogen *Staphylococcus aureus*, compared to other metabolites such as bacilysin, fengycin, and bacillibactin (Luo et al. 2023). Furthermore, the production of natural antibiotics and preservatives like bacitracin is highly relevant, considering that the consumption of artificial preservatives have been associated with gut diseases, obesity (Reardon 2015), lung and liver damage, and cancer (Aldabayan 2025). However, efficient bacitracin production requires, for example the genes *bcrABC*, which confers resistance to bacitracin itself (Podlesek et al. 2006), and were not found in the BAC30 and BAC220 strains. Additionally, limiting factors such as optical density, nitrate concentration, and NADH oxidation must be carefully considered in in vitro assays (Zhu et al. 2023). Nevertheless, the NRPs, combined with vitamin biosynthesis genes such as those for riboflavin, suggest a diversified potential for the biotechnological application of the isolated BAC30 and BAC220 strains.

## Conclusion

The isolated *Bacillus paralicheniformis* strains, BAC30 and BAC220, demonstrated high reliability in sequencing, genome assembly, and annotation. Phylogenomic analyses revealed high similarity between the genomes, with particular attention to the CamBx3 and J41TS8 strains, which may warrant further investigation due to their lower similarity relative to the others. Additionally, the formation of subgroups was proposed, although more specific phylogenomic analyses are still needed to support these subgroup distinctions. Regarding comparative analyses, the pangenomic approach provided insight into the species' tendency to adapt to different ecological niches and acquire new genes, helping to explain the speciation and divergence between *B. licheniformis* and *B. paralicheniformis*.

*In silico* analyses suggest that the isolated strains BAC30 and BAC220 hold biotechnological potential, as they only exhibited KEGG modules related to beta-lactam resistance. Moreover, modules involved in the biosynthesis of metabolites relevant to various industrial sectors were

identified, including those responsible for the production of vitamins, cofactors, and natural antibiotics, which can be incorporated into commercial products or produced independently.

Nevertheless, despite the observed potential, further *in vitro* and *in vivo* assays are required to confirm the expression of genes involved in metabolite biosynthesis and antibiotic resistance, in order to properly assess the actual effectiveness of these strains as producers of biotechnologically valuable compounds.

### **Acknowledgements**

The authors would like to acknowledge the Pró-Reitoria de Pesquisa—Universidade Federal de Minas Gerais, Rede de Ciências Ômicas (RECOM), the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) for their financial support and fellowships.

### **Statements & Declarations**

#### **Funding**

This research was funded by Fundação de Amparo à Pesquisa do Estado de Minas Gerais -FAPEMIG (5.18/2022).

#### **Competing Interests**

The authors declare no conflict of interest.

#### **Consent for Publication**

All authors consent to the publication of this article.

#### **Author Contributions**

Conceptualization: Gabriel Camargos Gomes, Eduarda Guimarães Sousa; Methodology: Gabriel Camargos Gomes, Eduarda Guimarães Sousa, Marcus Vinícius Canário Viana, Bertram Brenig; Formal analysis and investigation: Gabriel Camargos Gomes, Giovanna Karine Viana Silva, Rafael Junio de oliveira; Writing-original draft preparation: Gabriel Camargos Gomes, Eduarda Guimarães Sousa, Janaíne Aparecida de Paula; Writing-review and editing: Gabriel Camargos Gomes, Eduarda Guimarães Sousa, Ludmila Silva Quaresma, Rhayane Cristina Viegas Santos, Gabriela Munis Campos, Vasco Azevedo; Supervision:

Vasco Azevedo; Funding acquisition: Vasco Azevedo. All authors read and approved the final manuscript.

## Data Availability

The current study's data are available from the corresponding author upon reasonable request.

## References

- Ågren J, Sundström A, Håfström T, Segerman B (2012) Gegenees: Fragmented Alignment of Multiple Genomes for Determining Phylogenomic Distances and Genetic Signatures Unique for Specified Target Groups. *PLoS ONE* 7(6):e39107. <https://doi.org/10.1371/journal.pone.0039107>
- Ahmed M, Rauf M, Mukhtar Z, Saeed NA (2017) Excessive use of nitrogenous fertilizers: an unawareness causing serious threats to environment and human health. *Environ Sci Pollut Res* 24(35):26983–26987. <https://doi.org/10.1007/s11356-017-0589-7>
- Aldabayan YS (2025) Effect of Artificial Food Additives on Lung Health—An Overview. *Medicina* 61(4):684. <https://doi.org/10.3390/medicina61040684>
- Asif M, Li-Qun Z, Zeng Q, Atiq M, Ahmad K, Tariq A, Al-Ansari N, Blom J, Fenske L, Alodaini HA, Hatamleh AA (2023) Comprehensive genomic analysis of *Bacillus paralicheniformis* strain BP9, pan-genomic and genetic basis of biocontrol mechanism. *Computational and Structural Biotechnology Journal* 21:4647–4662. <https://doi.org/10.1016/j.csbj.2023.09.043>
- Babraham Bioinformatics (2010) FastQC: A quality control tool for high throughput sequence data. Available at: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Blanco Crivelli X, Cundon C, Bonino MP, Sanin MS, Bentancor A (2024) The Complex and Changing Genus *Bacillus*: A Diverse Bacterial Powerhouse for Many Applications. *Bacteria* 3(3):256–270. <https://doi.org/10.3390/bacteria3030017>
- Blin K, Shaw S, Vader L, Szenei J, Reitz ZL, Augustijn HE, Cediél-Becerra JDD, de Crécy-Lagard V, Koetsier RA, Williams SE, Cruz-Morales P, Wongwas S, Segurado Luchsinger AE, Biermann F, Korenskaia A, Zdouc MM, Meijer D, Terlouw BR, van der Hooft JJJ, Ziemert N, Helfrich EJM, Masschelein J, Corre C, Chevrette MG, van Wezel GP, Medema MH, Weber T (2025) antiSMASH 8.0: extended gene cluster detection capabilities and analyses of chemistry, enzymology, and regulation. *Nucleic Acids Research* :gkaf334. <https://doi.org/10.1093/nar/gkaf334>
- Cai D, Zhang B, Zhu J, Xu H, Liu P, Wang Z, Li J, Yang Z, Ma X, Chen S (2020) Enhanced Bacitracin Production by Systematically Engineering S-Adenosylmethionine Supply Modules in *Bacillus licheniformis*. *Front Bioeng Biotechnol* 8:305. <https://doi.org/10.3389/fbioe.2020.00305>
- Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH (2020) GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36(6):1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>
- Chavarria-Quicaño E, Contreras-Jáquez V, Carrillo-Fasio A, De La Torre-González F, Asaff-Torres A (2023) Native *Bacillus paralicheniformis* isolate as a potential agent for phytopathogenic nematodes control. *Front Microbiol* 14. <https://doi.org/10.3389/fmicb.2023.1213306>
- Chen S, Zhou Y, Chen Y, Gu J (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34(17):i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>

- Chklovski A, Parks DH, Woodcroft BJ, Tyson GW (2023) CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat Methods* 20(8):1203–1212. <https://doi.org/10.1038/s41592-023-01940-w>
- Choyam S, Jain PM, Kammara R (2021) Characterization of a Potent New-Generation Antimicrobial Peptide of *Bacillus*. *Front Microbiol* 12. <https://doi.org/10.3389/fmicb.2021.710741>
- Dong H, Zhang D (2014) Current development in genetic engineering strategies of *Bacillus* species. *Microb Cell Fact* 13(1):63. <https://doi.org/10.1186/1475-2859-13-63>
- Dos Reis TFM, Hoepers PG, Azevedo VADC, Da Silva GR, Notário FO, Soares MC, Schlemper AE, Costa ÍM, Coelho-Rocha ND, Da Costa MS, Sommerfeld S, Medeiros-Ronchi AA, Rossi DA, De Castro IP, Fonseca BB (2022) Chicken embryos are a valuable model for the selection of *Bacillus subtilis* for probiotic purposes. *Arch Microbiol* 204(12):715. <https://doi.org/10.1007/s00203-022-03307-9>
- Du Y, Ma J, Yin Z, Liu K, Yao G, Xu W, Fan L, Du B, Ding Y, Wang C (2019) Comparative genomic analysis of *Bacillus paralicheniformis* MDJK30 with its closely related species reveals an evolutionary relationship between *B. paralicheniformis* and *B. licheniformis*. *BMC Genomics* 20(1):283. <https://doi.org/10.1186/s12864-019-5646-9>
- Dunlap CA, Kwon S-W, Rooney AP, Kim S-J (2015) *Bacillus paralicheniformis* sp. nov., isolated from fermented soybean paste. *International Journal of Systematic and Evolutionary Microbiology* 65(Pt\_10):3487–3492. <https://doi.org/10.1099/ijsem.0.000441>
- EFSA Panel on Additives and Products or Substances used in Animal Feed (FEEDAP), Bampidis V, Azimonti G, Bastos M de L, Christensen H, Dusemund B, Fašmon Durjava M, Kouba M, López-Alonso M, López Puente S, Marcon F, Mayo B, Pechová A, Petkova M, Ramos F, Sanz Y, Villa RE, Woutersen R, Anguita M, Brozzi R, Galobart J, Pettenati E (2022) Safety of the feed additive consisting of vitamin B2/riboflavin (produced by *Bacillus subtilis* KCCM 10445) for all animal species (Hubei Guangji Pharmaceutical Co. Ltd.). *EFSA* 20(10). <https://doi.org/10.2903/j.efsa.2022.7607>
- EFSA Panel on Additives and Products or Substances used in Animal Feed (FEEDAP), Rychen G, Aquilina G, Azimonti G, Bampidis V, Bastos M de L, Bories G, Chesson A, Flachowsky G, Gropp J, Kolar B, Kouba M, López-Alonso M, López Puente S, Mantovani A, Mayo B, Ramos F, Saarela M, Villa RE, Wallace RJ, Wester P, Herman L, Glandorf B, Kärenlampi S, Aguilera J, Cocconcelli PS (2018) Safety of vitamin B2 (80%) as riboflavin produced by *Bacillus subtilis* KCCM-10445 for all animal species. *EFSA* 16(3). <https://doi.org/10.2903/j.efsa.2018.5223>
- Elbir H (2024) Updating the Relationship Between the Threshold Value of Average Nucleotide Identity and Digital DNA–DNA Hybridization for Reliable Taxonomy of *Corynebacterium*. *Veterinary Sciences* 11(12):661. <https://doi.org/10.3390/vetsci11120661>
- Emms DM, Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 20(1). <https://doi.org/10.1186/s13059-019-1832-y>
- Errington J, Aart LTVD (2020) Microbe Profile: *Bacillus subtilis*: model organism for cellular development, and industrial workhorse: This article is part of the Microbe Profiles collection. *Microbiology* 166(5):425–427. <https://doi.org/10.1099/mic.0.000922>
- Felice AG, Sousa EG, Dominici FV, Azevedo VADC, Soares SDC (2023) Pangenome Analysis Reveals a High Degree of Genetic Diversity in *Gardnerella vaginalis*: An In Silico Approach. *Venereology* 2(4):132–146. <https://doi.org/10.3390/venereology2040012>

- Ferraz-Almeida R (2024) Balance of Nitrate and Ammonium in Tropical Soil Conditions: Soil Factors Analyzed by Machine Learning. *Nitrogen* 5(3):732–745. <https://doi.org/10.3390/nitrogen5030048>
- Florio A, Legout A, Marechal M, Clesse M, Delort A, Des Chatelliers CC, Gervais J, Shi Y, Van Der Heijden G, Zeller B, Le Roux X (2025) Nitrate leaching from soil under different forest tree species is related to the vertical distribution of *Nitrobacter* abundance. *Science of The Total Environment* 967:178776. <https://doi.org/10.1016/j.scitotenv.2025.178776>
- Gautreau G, Bazin A, Gachet M, Planel R, Burlot L, Dubois M, Perrin A, Médigue C, Calteau A, Cruveiller S, Matias C, Ambroise C, Rocha EPC, Vallenet D (2020) PPanGGOLiN: Depicting microbial diversity via a partitioned pangenome graph. *PLoS Comput Biol* 16(3):e1007732. <https://doi.org/10.1371/journal.pcbi.1007732>
- Gauvry E, Mathot A-G, Couvert O, Leguérinel I, Coroller L (2021) Effects of temperature, pH and water activity on the growth and the sporulation abilities of *Bacillus subtilis* BSB1. *International Journal of Food Microbiology* 337:108915. <https://doi.org/10.1016/j.ijfoodmicro.2020.108915>
- Grant JR, Enns E, Marinier E, Mandal A, Herman EK, Chen C, Graham M, Van Domselaar G, Stothard P (2023) Proksee: in-depth characterization and visualization of bacterial genomes. *Nucleic Acids Research* 51(W1):W484–W492. <https://doi.org/10.1093/nar/gkad326>
- Huson DH, Bryant D (2006) Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution* 23(2):254–267. <https://doi.org/10.1093/molbev/msj030>
- Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11(1):119. <https://doi.org/10.1186/1471-2105-11-119>
- Jones RN, Li Q, Kohut B, Biedenbach DJ, Bell J, Turnidge JD (2006) Contemporary antimicrobial activity of triple antibiotic ointment: a multiphased study of recent clinical isolates in the United States and Australia. *Diagnostic Microbiology and Infectious Disease* 54(1):63–71. <https://doi.org/10.1016/j.diagmicrobio.2005.08.009>
- Kanehisa M, Sato Y, Morishima K (2016) BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology* 428(4):726–731. <https://doi.org/10.1016/j.jmb.2015.11.006>
- Kim HU, Charusanti P, Lee SY, Weber T (2016) Metabolic engineering with systems biology tools to optimize production of prokaryotic secondary metabolites. *Nat Prod Rep* 33(8):933–941. <https://doi.org/10.1039/C6NP00019C>
- Kim Y, Koh I, Young Lim M, Chung W-H, Rho M (2017) Pan-genome analysis of *Bacillus* for microbiome profiling. *Sci Rep* 7(1). <https://doi.org/10.1038/s41598-017-11385-9>
- Kumariya R, Garsa AK, Rajput YS, Sood SK, Akhtar N, Patel S (2019) Bacteriocins: Classification, synthesis, mechanism of action and resistance development in food spoilage causing bacteria. *Microbial Pathogenesis* 128:171–177. <https://doi.org/10.1016/j.micpath.2019.01.002>
- Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW (2007) RNAMmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research* 35(9):3100–3108. <https://doi.org/10.1093/nar/gkm160>
- Lambertz C, Leopold J, Damme K, Vogt-Kaute W, Ammer S, Leiber F (2020) Effects of a riboflavin source suitable for use in organic broiler diets on performance traits and health indicators. *Animal* 14(4):716–724. <https://doi.org/10.1017/S175173111900243X>

- Letunic I, Bork P (2024) Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Research* 52(W1):W78–W82. <https://doi.org/10.1093/nar/gkae268>
- Li S, Ondon BS, Ho S-H, Jiang J, Li F (2022) Antibiotic resistant bacteria and genes in wastewater treatment plants: From occurrence to treatment strategies. *Science of The Total Environment* 838:156544. <https://doi.org/10.1016/j.scitotenv.2022.156544>
- Luo C, Chen M, Luo K, Yin X, Onchari MM, Wang X, Zhang J, Zhong H, Tian B (2023) Genome Sequencing and Genetic Engineering Reveal the Contribution of Bacitracin Produced by *Bacillus paralicheniformis* CPL618 to Anti-*Staphylococcus aureus* Activity. *Curr Microbiol* 80(5):135. <https://doi.org/10.1007/s00284-023-03196-1>
- Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14(1):60. <https://doi.org/10.1186/1471-2105-14-60>
- Meier-Kolthoff JP, Göker M (2019) TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat Commun* 10(1). <https://doi.org/10.1038/s41467-019-10210-3>. Available at <https://tygs.dsmz.de/>.
- Nannan C, Vu HQ, Gillis A, Caulier S, Nguyen TTT, Mahillon J (2021) Bacilysin within the *Bacillus subtilis* group: gene prevalence versus antagonistic activity against Gram-negative foodborne pathogens. *Journal of Biotechnology* 327:28–35. <https://doi.org/10.1016/j.jbiotec.2020.12.017>
- Narsing Rao MP, Singh RN, Sani RK, Banerjee A (2024) Genome-based approach to evaluate the metabolic potentials and exopolysaccharides production of *Bacillus paralicheniformis* CamBx3 isolated from a Chilean hot spring. *Front Microbiol* 15:1377965. <https://doi.org/10.3389/fmicb.2024.1377965>
- Okumura K, Okamoto M, Takamatsu D (2022) Whole-Genome Sequences of *Bacillus* and *Paenibacillus* sp. Strains Isolated from Honey in Japan. *Microbiol Resour Announc* 11(11):e00842-22. <https://doi.org/10.1128/mra.00842-22>
- Olajide AM, Chen S, LaPointe G (2021) Markers to Rapidly Distinguish *Bacillus paralicheniformis* From the Very Close Relative, *Bacillus licheniformis*. *Front Microbiol* 11. <https://doi.org/10.3389/fmicb.2020.596828>
- Olanrewaju OS, Glick BR, Babalola OO (2017) Mechanisms of action of plant growth promoting bacteria. *World J Microbiol Biotechnol* 33(11). <https://doi.org/10.1007/s11274-017-2364-9>
- O’Leary NA, Cox E, Holmes JB, Anderson WR, Falk R, Hem V, Tsuchiya MTN, Schuler GD, Zhang X, Torcivia J, Ketter A, Breen L, Cothran J, Bajwa H, Tinne J, Meric PA, Hlavina W, Schneider VA (2024) Exploring and retrieving sequence and metadata for species across the tree of life with NCBI Datasets. *Sci Data* 11(1). <https://doi.org/10.1038/s41597-024-03571-y>
- Orakov A, Fullam A, Coelho LP, Khedkar S, Szklarczyk D, Mende DR, Schmidt TSB, Bork P (2021) GUNC: detection of chimerism and contamination in prokaryotic genomes. *Genome Biol* 22(1). <https://doi.org/10.1186/s13059-021-02393-0>
- Paredes-Sabja D, Setlow P, Sarker MR (2011) Germination of spores of Bacillales and Clostridiales species: mechanisms and proteins involved. *Trends in Microbiology* 19(2):85–94. <https://doi.org/10.1016/j.tim.2010.10.004>
- Pausch J, Holz M, Zhu B, Cheng W (2024) Rhizosphere priming promotes plant nitrogen acquisition by microbial necromass recycling. *Plant Cell & Environment* 47(6):1987–1996. <https://doi.org/10.1111/pce.14858>

- Podlesek Z, Herzog B, Comino A (2006) Amplification of bacitracin transporter genes in the bacitracin producing *Bacillus licheniformis*. *FEMS Microbiology Letters* 157(1):201–205. <https://doi.org/10.1111/j.1574-6968.1997.tb12774.x>
- Posit Team (2025) RStudio: Integrated Development Environment for R. Boston, MA: Posit Software, PBC. Available at: <http://www.posit.co/>
- Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK (2016) Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal Methods* 8(1):12–24. <https://doi.org/10.1039/C5AY02550H>
- Ramírez-Cariño HF, Guadarrama-Mendoza PC, Sánchez-López V, Cuervo-Parra JA, Ramírez-Reyes T, Dunlap CA, Valadez-Blanco R (2020) Biocontrol of *Alternaria alternata* and *Fusarium oxysporum* by *Trichoderma asperelloides* and *Bacillus paralicheniformis* in tomato plants. *Antonie van Leeuwenhoek* 113(9):1247–1261. <https://doi.org/10.1007/s10482-020-01433-2>
- Reardon S (2015) Food preservatives linked to obesity and gut disease. *Nature* :nature.2015.16984. <https://doi.org/10.1038/nature.2015.16984>
- Revuelta JL, Buey RM, Ledesma-Amaro R, Vandamme EJ (2016) Microbial biotechnology for the synthesis of (pro)vitamins, biopigments and antioxidants: challenges and opportunities. *Microbial Biotechnology* 9(5):564–567. <https://doi.org/10.1111/1751-7915.12379>
- Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9(7):671–675. <https://doi.org/10.1038/nmeth.2089>
- Seeman T (2013) Barrnap 0.9: rapid ribosomal RNA prediction. Available at: <https://github.com/tseemann/barrnap>.
- Seemann T (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30(14):2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>
- Shimoyama Y (2022) COGclassifier: A tool for classifying prokaryote protein sequences into COG functional category [Computador software]. Available at: <https://github.com/moshi4/COGclassifier>.
- Wick RR, Judd LM, Gorrie CL, Holt KE (2017) Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13(6):e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>
- Wilberg EW (2015) What's in an Outgroup? The Impact of Outgroup Choice on the Phylogenetic Position of *Thalattosuchia* (Crocodylomorpha) and the Origin of Crocodyliformes. *Systematic Biology* 64(4):621–637. <https://doi.org/10.1093/sysbio/syv020>
- Wu S, Zhuang G, Bai Z, Cen Y, Xu S, Sun H, Han X, Zhuang X (2018) Mitigation of nitrous oxide emissions from acidic soils by *Bacillus amyloliquefaciens*, a plant growth-promoting bacterium. *Global Change Biology* 24(6):2352–2365. <https://doi.org/10.1111/gcb.14025>
- Xu J, Qin L, Xu X, Shen H, Yang X (2023) *Bacillus paralicheniformis* RP01 Enhances the Expression of Growth-Related Genes in Cotton and Promotes Plant Growth by Altering Microbiota inside and outside the Root. *IJMS* 24(8):7227. <https://doi.org/10.3390/ijms24087227>
- Yang S, Wang A, Li J, Shao Y, Sun F, Li S, Cao K, Liu H, Xiong P, Gao Z (2023) Improved biosynthesis of heme in *Bacillus subtilis* through metabolic engineering assisted fed-batch fermentation. *Microb Cell Fact* 22(1):102. <https://doi.org/10.1186/s12934-023-02077-3>

Yu C, Liang X, Song Y, Ali Q, Yang X, Zhu L, Gu Q, Kuptsov V, Kolomiets E, Wu H, Gao X (2024) A glycoside hydrolase 30 protein BpXynC of *Bacillus paralicheniformis* NMSW12 recognized as A MAMP triggers plant immunity response. *International Journal of Biological Macromolecules* 261:129750. <https://doi.org/10.1016/j.ijbiomac.2024.129750>

Zhai Z, Cui C, Li X, Yan J, Sun E, Wang C, Guo H, Hao Y (2023) Prevalence, antimicrobial susceptibility, and antibiotic resistance gene transfer of *Bacillus* strains isolated from pasteurized milk. *Journal of Dairy Science* 106(1):75–83. <https://doi.org/10.3168/jds.2022-22199>

Zhao X, Zhu D, Tan J, Wang R, Qi G (2023) Cooperative Action of Fulvic Acid and *Bacillus paralicheniformis* Ferment in Regulating Soil Microbiota and Improving Soil Fertility and Plant Resistance to Bacterial Wilt Disease. *Microbiol Spectr* 11(2). <https://doi.org/10.1128/spectrum.04079-22>

Zhu J, Wang S, Wang C, Wang Z, Luo G, Li J, Zhan Y, Cai D, Chen S (2023) Microbial synthesis of bacitracin: Recent progress, challenges, and prospects. *Synthetic and Systems Biotechnology* 8(2):314–322. <https://doi.org/10.1016/j.synbio.2023.03.009>

## MATERIAL SUPLEMENTAR - CAPÍTULO I

Material Suplementar do artigo de pesquisa intitulado: Comparative and Functional Analyses of *Bacillus paralicheniformis* Strains BAC30 and BAC220 by WGS Uncover Species Homogeneity and Biotechnological Potential.

**Tabela suplementar 1:** Informações referentes à classificação taxonômica dos 2 genomas montados.

| Amostra | Classificação GTDB-Tk  | ANI (%) | Closest genome reference | TYGS classification                     | dDDH (d0, in %) |
|---------|--|---------|--------------------------|---|-----------------|
| BAC30   | f__Bacillaceae;g__Bacillus;s__ <i>Bacillus paralicheniformis</i> | 98.91   | GCF_001042485.2          | <i>Bacillus paralicheniformis</i> KJ-16 | 97,5            |
| BAC220  | f__Bacillaceae;g__Bacillus;s__ <i>Bacillus paralicheniformis</i> | 98.8    | GCF_001042485.2          | <i>Bacillus paralicheniformis</i> KJ-16 | 92,7            |

**Tabela suplementar 2:** Avaliação da qualidade do sequenciamento e montagem e informações sobre o tamanho dos genomas

| Amostra | CheckM2 Completeness | CheckM2 Contamination | Genome size (pb) | % GC | Pass. GUNC | Barnap 5S | Barnap 16S | Barnap 23S |
|---------|----------------------|-----------------------|------------------|------|------------|-----------|------------|------------|
| BAC30   | 100.0                | 0.03                  | 4471444          | 46   | True       | Complete  | Partial    | Partial    |
| BAC220  | 100.0                | 3.06                  | 4785386          | 45   | True       | Complete  | Partial    | Partial    |

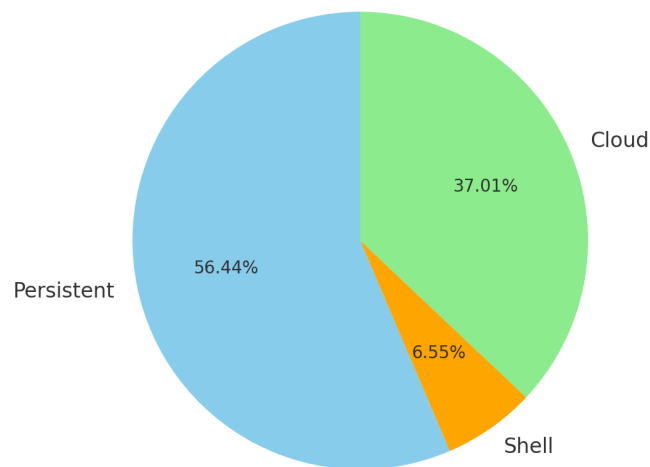
**Tabela suplementar 3:** Informações sobre os 14 genomas de *Bacillus* utilizados na **Figura 1A**

| Species name                       | Strain name | NCBI RefSeq Assembly |
|------------------------------------|-------------|----------------------|
| <i>B. altitudinis</i>              | GR-8        | GCF_001191605.1      |
| <i>B. pumilus</i>                  | SAFR-032    | GCF_000017885.4      |
| <i>B. safensis</i>                 | PgKB20      | GCF_008244765.1      |
| <i>B. sonorensis</i>               | PMC204      | GCF_034055115.1      |
| <i>B. licheniformis</i>            | ATCC 14580  | GCF_034478925.1      |
| <i>B. paralicheniformis</i>        | Bac84       | GCF_002993925.1      |
| <i>B. siamensis</i>                | KCTC 13613  | GCF_000262045.1      |
| <i>B. amyloliquefaciens</i>        | GKT04       | GCF_019396925.1      |
| <i>B. velezensis</i>               | FZB42       | GCF_000015785.2      |
| <i>B. atrophaeus</i>               | SW          | GCF_039519175.1      |
| <i>B. mojavensis</i>               | UCMB5075    | GCF_012648005.1      |
| <i>B. valismortis</i>              | Bac111      | GCF_003667885.1      |
| <i>B. subtilis subsp. subtilis</i> | 168         | GCF_000009045.1      |
| <i>B. tequilensis</i>              | EA-CB0015   | GCF_012225885.1      |

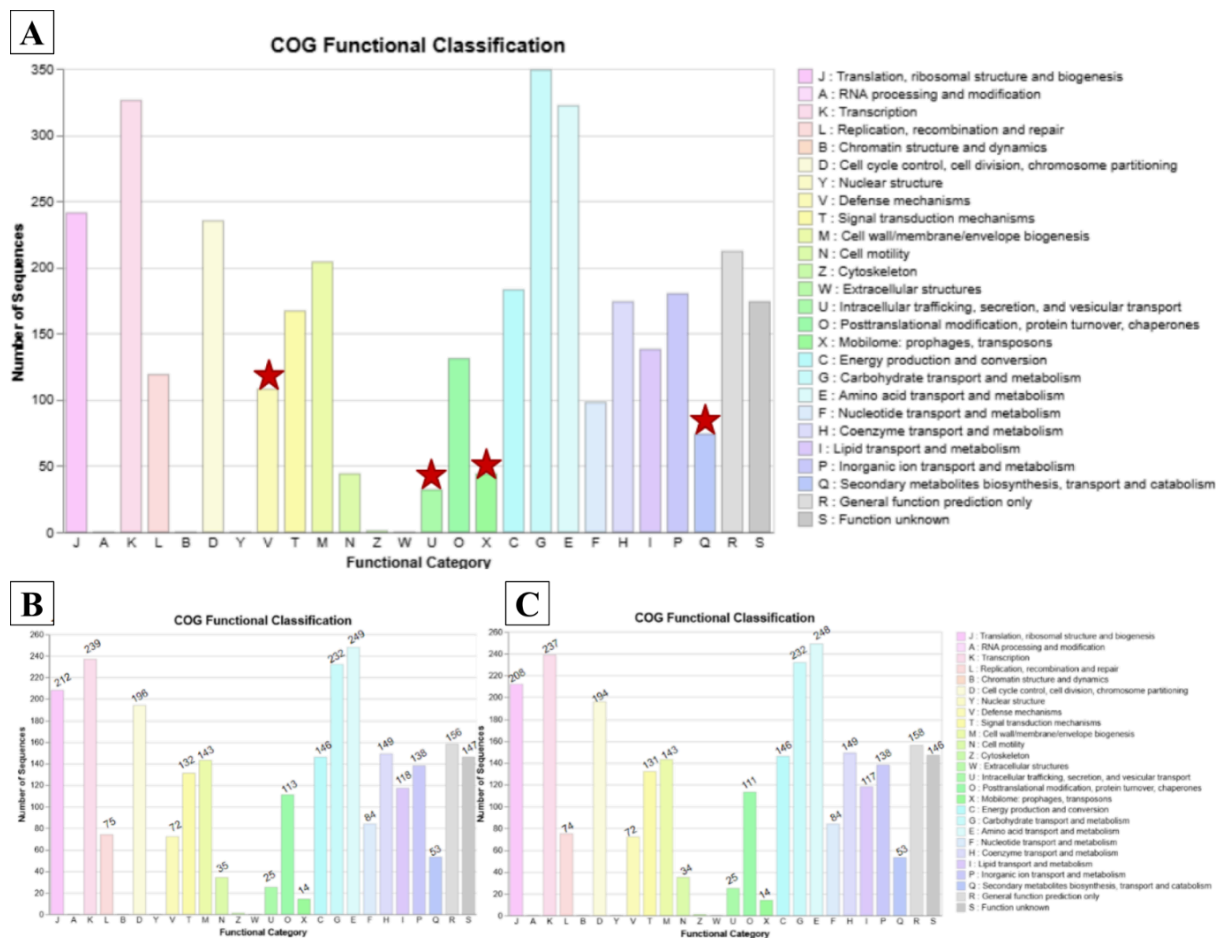




**Figura Suplementar 3:** Distribuição das famílias gênicas do pangenoma, obtida pelo programa PPanGGolin. Em azul, genes da categoria *Persistent* (core); em laranja, *Shell* (compartilhados); e em verde, *Cloud* (singulares).



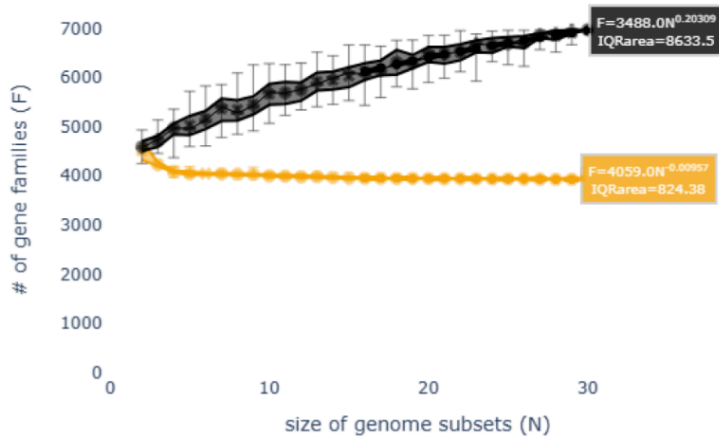
**Figura Suplementar 4:** Classificação funcional dos genes *core* dos isolados BAC30 e BAC220 a partir do COG Classifier; **3b:** Comparação entre os genes *core* das 28 linhagens de *Bacillus paralicheniformis* **sem** os isolados; **3c:** Comparação entre os genes *core* das 28 linhagens de *Bacillus paralicheniformis* **com** os isolados



**Figura Suplementar 5:** Curvas de rarefação comparando: (a) o pangenoma de 30 linhagens, incluindo os isolados BAC30 e BAC220; e (b) o pangenoma de 28 linhagens, excluindo os isolados.

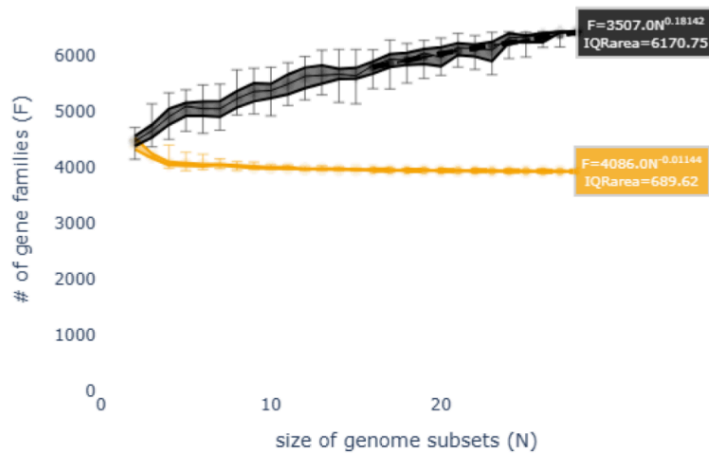
**A**

Rarefaction curve



**B**

Rarefaction curve



- pangenome : 1st quartile
- pangenome : 3rd quartile
- \* pangenome : means
- pangenome : medians
- persistent : 1st quartile
- persistent : 3rd quartile
- \* persistent : means
- persistent : medians

## Scripts utilizados no artigo:

**Script S1:** Script R customizado, utilizado para editar outputs do software pyANI.

```

install.packages("pheatmap")
# Load the ANI identity matrix from the .tab file
df <- read.table("ANIm_percentage_identity.tab", header = TRUE, row.names = 1, check.names = FALSE)

# Convert the ANI identity matrix (similarity) into a distance matrix (1 - ANI)
dist_matrix <- as.dist(1 - df)

# Perform hierarchical clustering using the UPGMA method (average linkage)
hc <- hclust(dist_matrix, method = "average")

# Plot the dendrogram with title and font size adjustment
plot(hc, main = "ANI Dendrogram", cex = 0.8)

# Load the required library to generate the heatmap
library(pheatmap)

# Generate a heatmap with the ANI matrix and the phylogenetic tree derived from clustering
pheatmap(df,
  clustering_distance_rows = dist_matrix, # Use the distance matrix to cluster rows
  clustering_distance_cols = dist_matrix) # Use the distance matrix to cluster columns

# Define a custom color gradient (yellow for lower ANI, red for higher ANI)
color_palette <- colorRampPalette(c("beige", "gold", "orange", "darkred"))(50)

# Generate the heatmap with the new color scale
pheatmap(df,
  clustering_distance_rows = dist_matrix, # Use the distance matrix to cluster rows
  clustering_distance_cols = dist_matrix, # Use the distance matrix to cluster columns
  color = color_palette,                # Apply the custom color gradient
  display_numbers = TRUE,              # Show values in the cells
  number_format = "%.2f",              # Format with 2 decimal places
  number_color = "black",              # Text color
  fontsize_number = 8,                 # Font size of numbers
  main = "ANI Heatmap with Values")

```

## 5. DISCUSSÃO GERAL

A alta similaridade observada entre as linhagens da espécie, somado às análises taxonômicas dos isolados *B. paralicheniformis* BAC30 e BAC220, que não mostraram contaminação, quimerismos ou erros de montagem, fornecem evidências sólidas da classificação correta a partir do sequenciamento genômico completo (WGS). Com o barateamento e a maior eficiência das técnicas de sequenciamento, além do crescimento exponencial de genomas depositados, este estudo destaca a relevância do sequenciamento de nova geração (NGS) tanto para a descoberta de novas espécies quanto para o estabelecimento de relações filogenômicas, seja entre membros da mesma espécie (como observado em *B. paralicheniformis*), e/ou entre espécies distintas, como no caso do gênero *Bacillus*. As análises de pangenoma corroboram com a ubiquidade da espécie, evidenciada pela diversidade geográfica das amostras. Essa diversidade genética resulta da tendência dessas linhagens em adquirir novos genes, como demonstrado pela alta proporção de genes acessórios e pelo pangenoma aberto, observado na curva de rarefação.

Quanto às linhagens isoladas, as análises funcionais via KEGG indicaram resistência apenas a  $\beta$ -lactâmicos, corroborando os resultados dos testes de disco-difusão, embora também tenha sido observada resistência à eritromicina. No entanto, a presença de halo frente à eritromicina, em contraste com a ausência total frente à ampicilina, sugere uma resistência intermediária, não necessariamente associada a genes específicos de resistência à eritromicina. Ademais, a capacidade de conversão de compostos nitrogenados, aliada à produção de peptídeos antimicrobianos (AMPs), indica o potencial uso dos isolados em simbiose com plantas. No entanto, aplicações em modelos animais, como mucosite e colite ulcerativa, também devem ser consideradas, dado que linhagens de *B. paralicheniformis* já demonstraram atividade moduladora da microbiota, além da produção de metabólitos que podem auxiliar na resposta imune dos hospedeiros, reduzindo as citocinas inflamatórias.

## 6. CONCLUSÃO GERAL

Partindo das análises taxonômicas, comparativas e filogenômicas, o estudo ajudou a entender melhor o comportamento genético da espécie *B. paralicheniformis*, fornecendo evidências, como a alta taxa de genes acessórios, assim como o pangenoma aberto, obtido na curva de rarefação, que correlacionam com a ubiquidade e tendência desta espécie a obter novos genes. Esses achados se relacionam com o fato de muitas dessas linhagens estarem em simbiose com hospedeiros, sejam plantas, animais e até humanos, onde são buscados genes que auxiliem na adaptação e sobrevivência desses organismos em diferentes nichos.

Quanto aos isolados, as baixas taxas de contaminação, somadas às análises taxonômicas evidenciam a relevância do NGS e da bioinformática no geral, no âmbito de identificação e caracterização de novas espécies, partindo de análises *in silico*. Nas análises funcionais, as linhagens BAC30 e BAC220 mostraram poucos genes de resistência, com expressão corroborada em testes *in vitro*, além da presença de genes de biossíntese de compostos fundamentais para a atuação dessas bactérias na indústria, como produtoras de metabólitos isolados, ou como agentes de biocontrole em produtos alimentícios, plantas e animais, possivelmente substituindo o uso de antibióticos artificiais.

No entanto, para que os isolados sejam utilizados no contexto biotecnológico, ainda é necessário comprovar a expressão dos genes observados *in silico*, em testes *in vitro*, e posteriormente testes *in vivo*, que avaliem as relações dessas linhagens com outros organismos, seja na modulação da biota do hospedeiro quanto a indução do sistema imune e também possíveis fatores de virulência que ainda não foram detectados.

## 7. PERSPECTIVAS

Esse estudo foi fundamental em estabelecer ideias sobre como aprofundar o entendimento sobre a filogenia da espécie *B. paralicheniformis*. O dataset formado para o estudo foi feito apenas com genomas completos, então considera-se relevante a utilização de todos os 253 genomas do banco de dados do NCBI, que incluem os organizados a nível cromossômico, scaffold e contig, em análises mais aprofundadas. Nesse contexto, deve-se analisar o resistoma completo da espécie, incluindo genes de resistência e elementos móveis (plasmídeos, integrons, transposases e ilhas genômicas), assim como os genes de virulência, visto que a espécie já é utilizada em simbiose com plantas e animais. Além disso, deve-se identificar mais marcadores filogenômicos que contribuam para a distinção entre *B. licheniformis* e *B. paralicheniformis*. Quanto às linhagens isoladas BAC30 e BAC220, as mesmas precisam ser submetidas à testes *in vitro*, como os de atividade inibitória contra patógenos de interesse, assim como as análises de antibiograma. Posteriormente, testes *in vivo* que avaliem o potencial como bactéria de biocontrole em plantas. Também devem ser considerados os modelos animais, principalmente devido à escassez de estudos nesse contexto, onde devem ser avaliados fatores de interação com o hospedeiro, como a modulação da microbiota, a indução do sistema imune e também possíveis fatores de virulência que ainda não foram detectados nos testes *in silico*.

## 8. REFERÊNCIAS BIBLIOGRÁFICAS

ADAMS, Mark D. *et al.* The Genome Sequence of *Drosophila melanogaster*. *Science*, v. 287, n. 5461, p. 2185–2195, 24 mar. 2000.

AHIRE, J. J. *et al.* Identification and characterization of antimicrobial peptide produced by indigenously isolated *Bacillus paralicheniformis* UBBLi30 strain. *3 Biotech*, v. 10, n. 3, p. 112, mar. 2020.

AKHTAR, Saqib Saleem *et al.* *Bacillus licheniformis* FMCH001 Increases Water Use Efficiency via Growth Stimulation in Both Normal and Drought Conditions. *Frontiers in Plant Science*, v. 11, 7 abr. 2020.

ALARCÓN-SÁNCHEZ, Mario Alberto *et al.* Detection of *Enterococcus faecalis* and the red complex bacteria analyzed by the Checkerboard technique for DNA-DNA hybridization in endodontic infections: A systematic review and meta-analysis. *Diagnostic Microbiology and Infectious Disease*, v. 111, n. 3, p. 116654, mar. 2025.

ALBA-TERCEDOR, Javier; VILCHEZ, Susana. Anatomical damage caused by *Bacillus thuringiensis* variety *israelensis* in yellow fever mosquito *Aedes aegypti* (L.) larvae revealed by micro-computed tomography. *Scientific Reports*, v. 13, n. 1, p. 8759, 30 maio 2023.

ASHAJYOTHI, Mushineni *et al.* Comprehensive genomic analysis of *Bacillus subtilis* and *Bacillus paralicheniformis* associated with the pearl millet panicle reveals their antimicrobial potential against important plant pathogens. *BMC Plant Biology*, v. 24, n. 1, p. 197, 18 mar. 2024.

ASHOK, Anup *et al.* Microbes Producing L-Asparaginase free of Glutaminase and Urease isolated from Extreme Locations of Antarctic Soil and Moss. *Scientific Reports*, v. 9, n. 1, p. 1423, 5 fev. 2019.

ASIF, Muhammad *et al.* Comprehensive genomic analysis of *Bacillus paralicheniformis* strain BP9, pan-genomic and genetic basis of biocontrol mechanism. *Computational and Structural Biotechnology Journal*, v. 21, p. 4647–4662, 2023.

ATHANASOPOULOU, Konstantina *et al.* Third-Generation Sequencing: The Spearhead towards the Radical Transformation of Modern Genomics. *Life*, v. 12, n. 1, p. 30, 26 dez. 2021.

AZIZ, Ramy K. *et al.* The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics*, v. 9, n. 1, p. 75, dez. 2008.

BAEK, Beomyeol *et al.* Optimizing MALDI-TOF Mass Spectrometry for the Identification of *Bacillus cereus*: The Impact of Sporulation and Cultivation Time. *International Journal of Molecular Sciences*, v. 26, n. 9, p. 4355, 4 maio 2025.

BAER, R. *et al.* DNA sequence and expression of the B95-8 Epstein—Barr virus genome. *Nature*, v. 310, n. 5974, p. 207–211, jul. 1984.

BARBA-VIDAL, E. *et al.* Response to a *Salmonella* Typhimurium challenge in piglets supplemented with protected sodium butyrate or *Bacillus licheniformis*: effects on performance, intestinal health and behavior<sup>1,2</sup>. *Translational Animal Science*, v. 1, n. 2, p. 186–200, 1 abr. 2017.

BARNES, Wayne M. DNA sequencing by partial ribosubstitution. *Journal of Molecular Biology*, v. 119, n. 1, p. 83–99, fev. 1978.

BELOUSOVA, Maria E. *et al.* Dissecting the Environmental Consequences of *Bacillus thuringiensis* Application for Natural Ecosystems. *Toxins*, v. 13, n. 5, p. 355, 16 maio 2021.

BENEVIDES, Leandro *et al.* New Insights into the Diversity of the Genus *Faecalibacterium*. *Frontiers in Microbiology*, v. 8, p. 1790, 22 set. 2017.

BHANDARI, Vaibhav *et al.* Molecular signatures for *Bacillus* species: demarcation of the *Bacillus subtilis* and *Bacillus cereus* clades in molecular terms and proposal to limit the placement of new species into the genus *Bacillus*. *International Journal of Systematic and Evolutionary Microbiology*, v. 63, n. Pt\_7, p. 2712–2726, 1 jul. 2013.

BLANCO CRIVELLI, Ximena *et al.* The Complex and Changing Genus *Bacillus*: A Diverse Bacterial Powerhouse for Many Applications. *Bacteria*, v. 3, n. 3, p. 256–270, 2 set. 2024.

CAI, Guolin *et al.* Levan from *Bacillus amyloliquefaciens* JN4 acts as a prebiotic for enhancing the intestinal adhesion capacity of *Lactobacillus reuteri* JN101. *International Journal of Biological Macromolecules*, v. 146, p. 482–487, mar. 2020.

CALDEIRA, Nathalia Gonçalves Santos *et al.* Characterization by MALDI-TOF MS and 16S rRNA Gene Sequencing of Aerobic Endospore-Forming Bacteria Isolated from Pharmaceutical Facility in Rio de Janeiro, Brazil. *Microorganisms*, v. 12, n. 4, p. 724, 3 abr. 2024.

CARROLL, Laura M. *et al.* Genomic Sequencing of *Bacillus cereus* Sensu Lato Strains Isolated from Meat and Poultry Products in South Africa Enables Inter- and Intranational

Surveillance and Source Tracking. *Microbiology Spectrum*, v. 10, n. 3, p. e00700-22, 29 jun. 2022.

CAULIER, Simon *et al.* Overview of the Antimicrobial Compounds Produced by Members of the *Bacillus subtilis* Group. *Frontiers in Microbiology*, v. 10, p. 302, 26 fev. 2019.

CELANDRONI, Francesco *et al.* Identification and Pathogenic Potential of Clinical *Bacillus* and *Paenibacillus* Isolates. *PLOS ONE*, v. 11, n. 3, p. e0152831, 31 mar. 2016.

CHAN, Cheong Xin; RAGAN, Mark A. Next-generation phylogenomics. *Biology Direct*, v. 8, n. 1, p. 3, dez. 2013.

CHEN, Lin *et al.* Induced maize salt tolerance by rhizosphere inoculation of *Bacillus amyloliquefaciens* SQR9. *Physiologia Plantarum*, v. 158, n. 1, p. 34–44, set. 2016.

CHEN, Zhao; ERICKSON, David L.; MENG, Jianghong. Benchmarking hybrid assembly approaches for genomic analyses of bacterial pathogens using Illumina and Oxford Nanopore sequencing. *BMC Genomics*, v. 21, n. 1, p. 631, dez. 2020.

COSTA, Luciana Veloso Da *et al.* Assessment of VITEK® 2, MALDI-TOF MS and full gene 16S rRNA sequencing for aerobic endospore-forming bacteria isolated from a pharmaceutical facility. *Journal of Microbiological Methods*, v. 194, p. 106419, mar. 2022a.

COSTA, Luciana Veloso Da *et al.* MALDI-TOF MS database expansion for identification of *Bacillus* and related genera isolated from a pharmaceutical facility. *Journal of Microbiological Methods*, v. 203, p. 106625, dez. 2022b.

DAYHOFF, Margaret Oakley; LEDLEY, Robert S. Comproteín: a computer program to aid primary protein structure determination. *In: THE DECEMBER 4-6, 1962, FALL JOINT COMPUTER CONFERENCE. Proceedings of the December 4-6, 1962, fall joint computer conference on - AFIPS '62 (Fall)*. Philadelphia, Pennsylvania: ACM Press, 1962. Disponível em: <<http://portal.acm.org/citation.cfm?doid=1461518.1461546>>. Acesso em: 24 jun. 2025

DOS REIS, Thais Fernanda Martins *et al.* Chicken embryos are a valuable model for the selection of *Bacillus subtilis* for probiotic purposes. *Archives of Microbiology*, v. 204, n. 12, p. 715, dez. 2022.

DU, Yuhui *et al.* Comparative genomic analysis of *Bacillus paralicheniformis* MDJK30 with its closely related species reveals an evolutionary relationship between *B. paralicheniformis* and *B. licheniformis*. *BMC Genomics*, v. 20, n. 1, p. 283, dez. 2019.

DUNLAP, Christopher A. *et al.* *Bacillus paralicheniformis* sp. nov., isolated from fermented soybean paste. *International Journal of Systematic and Evolutionary Microbiology*, v. 65, n. Pt\_10, p. 3487–3492, 1 out. 2015.

EDMAN, P.; BEGG, G. A Protein Sequenator. *In: LIÉBECQ, Claude (Org.). European Journal of Biochemistry*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1967. p. 80–91.

EFSA PANEL ON FOOD CONTACT MATERIALS, ENZYMES AND PROCESSING AIDS (CEP) *et al.* Taxonomic identity of the *Bacillus licheniformis* strains used to produce food enzymes evaluated in published EFSA opinions. *EFSA Journal*, v. 22, n. 5, maio 2024.

EFSA PANEL ON FOOD CONTACT MATERIALS, ENZYMES AND PROCESSING AIDS (EFSA CEP PANEL) *et al.* Scientific Guidance for the submission of dossiers on Food Enzymes. *EFSA Journal*, v. 19, n. 10, out. 2021.

EGLER, Rachel A.; AHUJA, Sanjay P.; MATLOUB, Yousif. L-asparaginase in the treatment of patients with acute lymphoblastic leukemia. *Journal of Pharmacology and Pharmacotherapeutics*, v. 7, n. 2, p. 62–71, jun. 2016.

EHLING-SCHULZ, Monika; FRICKER, Martina; SCHERER, Siegfried. *Bacillus cereus*, the causative agent of an emetic type of food-borne illness. *Molecular Nutrition & Food Research*, v. 48, n. 7, p. 479–487, dez. 2004.

FATANI, Siham *et al.* Genome sequencing and identification of cellulase genes in *Bacillus paralicheniformis* strains from the Red Sea. *BMC Microbiology*, v. 21, n. 1, p. 254, dez. 2021.

FENG, Da-Fei; DOOLITTLE, Russell F. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *Journal of Molecular Evolution*, v. 25, n. 4, p. 351–360, ago. 1987.

FERNÁNDEZ-NO, I. C. *et al.* Characterisation and profiling of *Bacillus subtilis*, *Bacillus cereus* and *Bacillus licheniformis* by MALDI-TOF mass fingerprinting. *Food Microbiology*, v. 33, n. 2, p. 235–242, abr. 2013.

FIERS, W. *et al.* Complete nucleotide sequence of SV40 DNA. *Nature*, v. 273, n. 5658, p. 113–120, 11 maio 1978.

FIGUEIREDO, Henrique C. P. *et al.* Comparative genome analysis of *Weissella ceti*, an emerging pathogen of farm-raised rainbow trout. *BMC Genomics*, v. 16, n. 1, p. 1095, dez. 2015.

FLEISCHMANN, Robert D. *et al.* Whole-Genome Random Sequencing and Assembly of *Haemophilus influenzae* Rd. *Science*, v. 269, n. 5223, p. 496–512, 28 jul. 1995.

GARCÍAS, Josinete Torres *et al.* Bacillus Strains Isolated from Agroforestry Systems in the Amazon Promote Açai Palm Seedling Growth When Inoculated as a Consortium. *Agronomy*, v. 15, n. 3, p. 569, 25 fev. 2025.

GAUTHIER, Jeff *et al.* A brief history of bioinformatics. *Briefings in Bioinformatics*, v. 20, n. 6, p. 1981–1996, 27 nov. 2019.

GOODWIN, Sara; MCPHERSON, John D.; MCCOMBIE, W. Richard. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, v. 17, n. 6, p. 333–351, jun. 2016.

GUICHARD, Annabel; NIZET, Victor; BIER, Ethan. New insights into the biological effects of anthrax toxins: linking cellular to organismal responses. *Microbes and Infection*, v. 14, n. 2, p. 97–118, fev. 2012.

GUPTA, Mamta; KUMAR, Harish; KAUR, Sarvjeet. Vegetative Insecticidal Protein (Vip): A Potential Contender From *Bacillus thuringiensis* for Efficient Management of Various Detrimental Agricultural Pests. *Frontiers in Microbiology*, v. 12, p. 659736, 13 maio 2021.

GUPTA, Radhey S. *et al.* Robust demarcation of 17 distinct *Bacillus* species clades, proposed as novel Bacillaceae genera, by phylogenomics and comparative genomic analyses: description of *Robertmurraya kyonggiensis* sp. nov. and proposal for an emended genus *Bacillus* limiting it only to the members of the Subtilis and Cereus clades of species. *International Journal of Systematic and Evolutionary Microbiology*, v. 70, n. 11, p. 5753–5798, 1 nov. 2020.

HARWOOD, Colin R. *et al.* Secondary metabolite production and the safety of industrially important members of the *Bacillus subtilis* group. *FEMS Microbiology Reviews*, v. 42, n. 6, p. 721–738, 1 nov. 2018.

HE, Hehe *et al.* Biotechnological and food synthetic biology potential of platform strain: *Bacillus licheniformis*. *Synthetic and Systems Biotechnology*, v. 8, n. 2, p. 281–291, jun. 2023.

HEATHER, James M.; CHAIN, Benjamin. The sequence of sequencers: The history of sequencing DNA. *Genomics*, v. 107, n. 1, p. 1–8, jan. 2016.

HERNÁNDEZ-SALMERÓN, Julie E.; IRANI, Tanya; MORENO-HAGELSIEB, Gabriel. Fast genome-based delimitation of Enterobacterales species. *PLOS ONE*, v. 18, n. 9, p. e0291492, 14 set. 2023.

INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM *et al.* Initial sequencing and analysis of the human genome. *Nature*, v. 409, n. 6822, p. 860–921, 15 fev. 2001.

IQBAL, Sajid *et al.* Genome mining, antimicrobial and plant growth-promoting potentials of halotolerant *Bacillus paralicheniformis* ES-1 isolated from salt mine. *Molecular Genetics and Genomics*, v. 298, n. 1, p. 79–93, jan. 2023.

JANDA, J. Michael; ABBOTT, Sharon L. 16S rRNA Gene Sequencing for Bacterial Identification in the Diagnostic Laboratory: Pluses, Perils, and Pitfalls. *Journal of Clinical Microbiology*, v. 45, n. 9, p. 2761–2764, set. 2007.

JANISZEWSKA, Daria *et al.* Implications of Sample Preparation Methods on the MALDI-TOF MS Identification of Spore-Forming *Bacillus* Species from Food Samples: A Closer Look at *Bacillus licheniformis*, *Peribacillus simplex*, *Lysinibacillus fusiformis*, *Bacillus flexus*, and *Bacillus marisflavi*. *ACS Omega*, v. 8, n. 38, p. 34982–34994, 26 set. 2023.

JOLLEY, Keith A. *et al.* Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology*, v. 158, n. 4, p. 1005–1015, 1 abr. 2012.

KANG, Sang-Mo *et al.* Integrated phytohormone production by the plant growth-promoting rhizobacterium *Bacillus tequilensis* SSB07 induced thermotolerance in soybean. *Journal of Plant Interactions*, v. 14, n. 1, p. 416–423, 1 jan. 2019.

KAPLI, Paschalia; YANG, Ziheng; TELFORD, Maximilian J. Phylogenetic tree building in the genomic age. *Nature Reviews Genetics*, v. 21, n. 7, p. 428–444, jul. 2020.

KIMBREL, Jeffrey A.; JEFFREY, Brendan M.; WARD, Christopher S. Prokaryotic Genome Annotation. *In: NAVID, Ali (Org.). Microbial Systems Biology. Methods in Molecular Biology*. New York, NY: Springer US, 2022. v. 2349 p. 193–214.

KOVÁCS, Ákos T. *Bacillus subtilis*. *Trends in Microbiology*, v. 27, n. 8, p. 724–725, ago. 2019.

KUMAR, Pankaj; DUBEY, R. C.; MAHESHWARI, D. K. *Bacillus* strains isolated from rhizosphere showed plant growth promoting and antagonistic activity against phytopathogens. *Microbiological Research*, v. 167, n. 8, p. 493–499, set. 2012.

LAND, Miriam *et al.* Insights from 20 years of bacterial genome sequencing. *Functional & Integrative Genomics*, v. 15, n. 2, p. 141–161, mar. 2015.

- LI, Guannan *et al.* Probiotic *Bacillus subtilis* contributes to the modulation of gut microbiota and blood metabolic profile of hosts. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, v. 272, p. 109712, out. 2023.
- LI, Wenjun *et al.* RefSeq: expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation. *Nucleic Acids Research*, v. 49, n. D1, p. D1020–D1028, 8 jan. 2021.
- LIU, Jie *et al.* Effect of *Bacillus paralicheniformis* on soybean (*Glycine max*) roots colonization, nutrient uptake and water use efficiency under drought stress. *Journal of Agronomy and Crop Science*, v. 209, n. 4, p. 547–565, ago. 2023.
- LOBB, Briallen *et al.* An assessment of genome annotation coverage across the bacterial tree of life. *Microbial Genomics*, v. 6, n. 3, 1 mar. 2020.
- MAUGHAN, Heather; VAN DER AUWERA, Geraldine. *Bacillus* taxonomy in the genomic era finds phenotypes to be essential though often misleading. *Infection, Genetics and Evolution*, v. 11, n. 5, p. 789–797, jul. 2011.
- MAXAM, A. M.; GILBERT, W. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences*, v. 74, n. 2, p. 560–564, fev. 1977.
- MAZHARI, Farnaz *et al.* Resolution of MALDI-TOF compared to whole genome sequencing for identification of *Bacillus* species isolated from cleanrooms at NASA Johnson Space Center. *Frontiers in Microbiology*, v. 16, p. 1499516, 9 abr. 2025.
- MCCOMBIE, W. Richard; MCPHERSON, John D.; MARDIS, Elaine R. Next-Generation Sequencing Technologies. *Cold Spring Harbor Perspectives in Medicine*, v. 9, n. 11, p. a036798, nov. 2019.
- MEDINI, Duccio *et al.* The microbial pan-genome. *Current Opinion in Genetics & Development*, v. 15, n. 6, p. 589–594, dez. 2005.
- MEIER-KOLTHOFF, Jan P. *et al.* Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics*, v. 14, n. 1, p. 60, dez. 2013.
- MESLIER, Victoria *et al.* Benchmarking second and third-generation sequencing platforms for microbial metagenomics. *Scientific Data*, v. 9, n. 1, p. 694, 11 nov. 2022.
- MOAYERI, Mahtab *et al.* Anthrax Pathogenesis. *Annual Review of Microbiology*, v. 69, n. 1, p. 185–208, 15 out. 2015.

MOUSE GENOME SEQUENCING CONSORTIUM. Initial sequencing and comparative analysis of the mouse genome. *Nature*, v. 420, n. 6915, p. 520–562, dez. 2002.

MURSALIN, Md Huzzatul; LIVINGSTON, Erin T.; CALLEGAN, Michelle C. The cereus matter of *Bacillus endophthalmitis*. *Experimental Eye Research*, v. 193, p. 107959, abr. 2020.

NAIMOV, Samir *et al.* *Bacillus thuringiensis* Delta-Endotoxin Cry1 Hybrid Proteins with Increased Activity against the Colorado Potato Beetle. *Applied and Environmental Microbiology*, v. 67, n. 11, p. 5328–5330, nov. 2001.

NARSING RAO, Manik Prabhu *et al.* Genome-based approach to evaluate the metabolic potentials and exopolysaccharides production of *Bacillus paralicheniformis* CamBx3 isolated from a Chilean hot spring. *Frontiers in Microbiology*, v. 15, p. 1377965, 2 abr. 2024.

NEEDLEMAN, Saul B.; WUNSCH, Christian D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, v. 48, n. 3, p. 443–453, mar. 1970.

OCAÑA, Kary A. C. S.; DÁVILA, Alberto M. R. Phylogenomics-Based Reconstruction of Protozoan Species Tree. *Evolutionary Bioinformatics*, v. 7, p. EBO.S6861, jan. 2011.

OLAJIDE, Atinuke M.; CHEN, Shu; LAPOINTE, Gisèle. Markers to Rapidly Distinguish *Bacillus paralicheniformis* From the Very Close Relative, *Bacillus licheniformis*. *Frontiers in Microbiology*, v. 11, p. 596828, 11 jan. 2021.

OTHOUM, Ghofran *et al.* In silico exploration of Red Sea *Bacillus* genomes for natural product biosynthetic gene clusters. *BMC Genomics*, v. 19, n. 1, p. 382, dez. 2018.

PALACIO-RODRÍGUEZ, Rubén *et al.* Halotolerant Rhizobacteria Promote Plant Growth and Decrease Salt Stress in *Carya illinoensis* (Wangenh.) K. Koch. *Agronomy*, v. 13, n. 12, p. 3045, 13 dez. 2023.

PALKOVICSNÉ PÉZSA, Nikolett *et al.* Effects of *Bacillus licheniformis* and *Bacillus subtilis* on Gut Barrier Function, Proinflammatory Response, ROS Production and Pathogen Inhibition Properties in IPEC-J2—*Escherichia coli*/*Salmonella* Typhimurium Co-Culture. *Microorganisms*, v. 10, n. 5, p. 936, 29 abr. 2022.

PAN, Huaqi *et al.* Genome mining and metabolic profiling illuminate the chemistry driving diverse biological activities of *Bacillus siamensis* SCSIO 05746. *Applied Microbiology and Biotechnology*, v. 103, n. 10, p. 4153–4165, maio 2019.

PATANÉ, José S. L.; MARTINS, Joaquim; SETUBAL, João C. Phylogenomics. *In*: SETUBAL, João C.; STOYE, Jens; STADLER, Peter F. (Orgs.). *Comparative Genomics. Methods in Molecular Biology*. New York, NY: Springer New York, 2018. v. 1704 p. 103–187.

PATEL, Sudip; GUPTA, Radhey S. A phylogenomic and comparative genomic framework for resolving the polyphyly of the genus *Bacillus*: Proposal for six new genera of *Bacillus* species, *Peribacillus* gen. nov., *Cytobacillus* gen. nov., *Mesobacillus* gen. nov., *Neobacillus* gen. nov., *Metabacillus* gen. nov. and *Alkalihalobacillus* gen. nov. *International Journal of Systematic and Evolutionary Microbiology*, v. 70, n. 1, p. 406–438, 1 jan. 2020.

PEREIRA, Paola P. *et al.* Polyphasic characterization and identification of the bioremediation agent *Bacillus* sp. SFC 500-1E. *Genomics*, v. 112, n. 6, p. 4525–4535, nov. 2020.

PETKOVA, Mariana *et al.* Plant Growth-Promoting and Biocontrol Characteristics of Four *Bacillus* Strains and Evaluation of Their Effects on Wheat (*Tr. aestivum* L.). *International Journal of Plant Biology*, v. 16, n. 1, p. 1, 27 dez. 2024.

PLATA-RUEDA, Angelica *et al.* Insecticidal Activity of *Bacillus thuringiensis* Strains on the Nettle Caterpillar, *Euprosterina elaeasa* (Lepidoptera: Limacodidae). *Insects*, v. 11, n. 5, p. 310, 15 maio 2020.

PRIEST, F. G. *et al.* Characterization of *Bacillus thuringiensis* and related bacteria by ribosomal RNA gene restriction fragment length polymorphisms. *Microbiology*, v. 140, n. 5, p. 1015–1022, 1 maio 1994.

PUDOVA, Daria S. *et al.* Comparative Genome Analysis of Two *Bacillus pumilus* Strains Producing High Level of Extracellular Hydrolases. *Genes*, v. 13, n. 3, p. 409, 24 fev. 2022.

REY, Michael W. *et al.* Complete genome sequence of the industrial bacterium *Bacillus licheniformis* and comparisons with closely related *Bacillus* species. *Genome Biology*, v. 5, n. 10, 13 set. 2004.

RICHTER, Michael; ROSSELLÓ-MÓRA, Ramon. Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences*, v. 106, n. 45, p. 19126–19131, 10 nov. 2009.

RUSSI, Alessandra; ALMANÇA, Marcus André K.; SCHWAMBACH, Joséli. *Bacillus subtilis* strain F62 against *Fusarium oxysporum* and promoting plant growth in the grapevine rootstock SO4. *Anais da Academia Brasileira de Ciências*, v. 94, n. suppl 3, p. e20210860, 2022.

SAINI, Pinki; MISHRA, Pragya. Revolution in Microbial Bioprospecting via the Development of Omics-Based Technologies. *In*: SAINI, Pinki; MISHRA, Pragya (Eds.). *Bioprospecting of Microbial Resources for Agriculture, Environment and Bio-chemical Industry*. Cham: Springer Nature Switzerland, 2024. p. 171–191.

SALEH, Abdullah Abobakr *et al.* Enhanced production and purification of L-asparaginase from *Bacillus paralicheniformis* AUMC B-516 with potent cytotoxicity against MCF-7 cell lines. *AMB Express*, v. 15, n. 1, p. 80, 22 maio 2025.

SANGER, F. *et al.* Nucleotide sequence of bacteriophage  $\phi$ X174 DNA. *Nature*, v. 265, n. 5596, p. 687–695, fev. 1977.

SANGER, F. *et al.* Nucleotide sequence of bacteriophage  $\lambda$  DNA. *Journal of Molecular Biology*, v. 162, n. 4, p. 729–773, dez. 1982.

SANGER, F.; NICKLEN, S.; COULSON, A. R. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, v. 74, n. 12, p. 5463–5467, dez. 1977.

SCHWENGER, Oliver *et al.* Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification: Find out more about Bakta, the motivation, challenges and applications, here. *Microbial Genomics*, v. 7, n. 11, 30 nov. 2021.

SEEMANN, Torsten. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, v. 30, n. 14, p. 2068–2069, 15 jul. 2014.

SEGERMAN, Bo. The Most Frequently Used Sequencing Technologies and Assembly Methods in Different Time Segments of the Bacterial Surveillance and RefSeq Genome Databases. *Frontiers in Cellular and Infection Microbiology*, v. 10, p. 527102, 19 out. 2020.

SERMKAEW, Namfa *et al.* A Novel Bacitracin-like Peptide from Mangrove-Isolated *Bacillus paralicheniformis* NNS4-3 against MRSA and Its Genomic Insights. *Antibiotics*, v. 13, n. 8, p. 716, 30 jul. 2024.

SHAHID, Izzah *et al.* Profiling of Metabolites of *Bacillus* spp. and Their Application in Sustainable Plant Growth Promotion and Biocontrol. *Frontiers in Sustainable Food Systems*, v. 5, p. 605195, 22 fev. 2021.

SHENDURE, Jay *et al.* DNA sequencing at 40: past, present and future. *Nature*, v. 550, n. 7676, p. 345–353, 19 out. 2017.

SHU, Lin-Jie; YANG, Yu-Liang. Bacillus Classification Based on Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry—Effects of Culture Conditions. *Scientific Reports*, v. 7, n. 1, p. 15546, 14 nov. 2017.

SIMPSON, Jared T.; POP, Mihai. The Theory and Practice of Genome Sequence Assembly. *Annual Review of Genomics and Human Genetics*, v. 16, n. 1, p. 153–172, 24 ago. 2015.

SIVASHANKARI, Selvarajan; SHANMUGHAVEL, Piramanayagam. Comparative genomics - A perspective. *Bioinformation*, v. 1, n. 9, p. 376–378, 2 jan. 2007.

SOCRANSKY, S. S. *et al.* “Checkerboard” DNA-DNA hybridization. *BioTechniques*, v. 17, n. 4, p. 788–792, out. 1994.

SOHN, Jang-il; NAM, Jin-Wu. The present and future of *de novo* whole-genome assembly. *Briefings in Bioinformatics*, p. bbw096, 14 out. 2016.

STACKEBRANDT, Erko *et al.* Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *International Journal of Systematic and Evolutionary Microbiology*, v. 52, n. 3, p. 1043–1047, 1 maio 2002.

STADEN, R. A strategy of DNA sequencing employing computer programs. *Nucleic Acids Research*, v. 6, n. 7, p. 2601–2610, 1979.

STÜLKE, Jörg *et al.* *Bacillus subtilis*, a Swiss Army Knife in Science and Biotechnology. *Journal of Bacteriology*, v. 205, n. 5, p. e00102-23, 25 maio 2023.

SUN, Yanan *et al.* *Bacillus paralicheniformis* SYN-191 isolated from ginger rhizosphere soil and its growth-promoting effects in ginger farming. *BMC Microbiology*, v. 25, n. 1, p. 75, 14 fev. 2025.

ŚWIĄTCZAK, Joanna *et al.* *Bacillus paralicheniformis* 2R5 and its impact on canola growth and N-cycle genes in the rhizosphere. *FEMS Microbiology Ecology*, v. 99, n. 9, p. fiad093, 22 ago. 2023a.

ŚWIĄTCZAK, Joanna *et al.* The effect of seed bacterization with *Bacillus paralicheniformis* 2R5 on bacterial and fungal communities in the canola rhizosphere. *Microbiological Research*, v. 275, p. 127448, out. 2023b.

TAN, Irene S.; RAMAMURTHI, Kumaran S. Spore formation in *B. acillus subtilis*. *Environmental Microbiology Reports*, v. 6, n. 3, p. 212–225, jun. 2014.

TEELING, Hanno *et al.* TETRA: a web-service and a stand-alone program for the analysis and comparison of tetranucleotide usage patterns in DNA sequences. *BMC Bioinformatics*, v. 5, n. 1, p. 163, 26 out. 2004.

TETTELIN, Hervé *et al.* Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial “pan-genome”. *Proceedings of the National Academy of Sciences*, v. 102, n. 39, p. 13950–13955, 27 set. 2005.

TETTELIN, Hervé *et al.* Comparative genomics: the bacterial pan-genome. *Current Opinion in Microbiology*, v. 11, n. 5, p. 472–477, out. 2008.

TRISAKUL, Kanwara *et al.* Comparisons of genome assembly tools for characterization of *Mycobacterium tuberculosis* genomes using hybrid sequencing technologies. *PeerJ*, v. 12, p. e17964, 29 ago. 2024.

TROFIMOVA, Ellina *et al.* A bacterial genome assembly and annotation laboratory using a virtual machine. *Biochemistry and Molecular Biology Education*, v. 51, n. 3, p. 276–285, maio 2023.

VALENZUELA-RUIZ, Valeria *et al.* Potential biological control Mechanisms of *Bacillus paralicheniformis* TRQ65 against phytopathogenic fungi. *Revista Mexicana de Fitopatología, Mexican Journal of Phytopathology*, v. 42, n. 4, 20 dez. 2024.

VAN DIJK, Erwin L. *et al.* Genomics in the long-read sequencing era. *Trends in Genetics*, v. 39, n. 9, p. 649–671, set. 2023.

VENTER, J. Craig *et al.* The Sequence of the Human Genome. *Science*, v. 291, n. 5507, p. 1304–1351, 16 fev. 2001.

WANG, Li-Ting *et al.* Comparison of *gyrB* gene sequences, 16S rRNA gene sequences and DNA–DNA hybridization in the *Bacillus subtilis* group. *International Journal of Systematic and Evolutionary Microbiology*, v. 57, n. 8, p. 1846–1850, 1 ago. 2007.

WANG, Taiquan *et al.* Comparative Genomics Unveils Functional Diversity, Pangenome Openness, and Underlying Biological Drivers among *Bacillus subtilis* Group. *Microorganisms*, v. 12, n. 5, p. 986, 14 maio 2024.

WICK, Ryan R. *et al.* Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Computational Biology*, v. 13, n. 6, p. e1005595, 8 jun. 2017.

XIAO, Xiao *et al.* *Bacillus licheniformis* suppresses *Clostridium perfringens* infection via modulating inflammatory response, antioxidant status, inflammasome activation and microbial homeostasis in broilers. *Poultry Science*, v. 103, n. 11, p. 104222, nov. 2024.

YANG, Ming; SOLIDAR, Ada; J. WYCKOFF, Gerald. Novel method for discerning the action of selection during evolution. *Journal of Biomedical Science and Engineering*, v. 03, n. 02, p. 109–113, 2010.

YANG, Shumin *et al.* *Bacillus paralicheniformis*-mediated gut microbiota promotes M2 macrophage polarization by inhibiting P38 MAPK signaling to alleviate necrotizing enterocolitis and apoptosis in mice. *Microbiological Research*, v. 296, p. 128136, jul. 2025.

YOON, Seok-Hwan *et al.* A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie van Leeuwenhoek*, v. 110, n. 10, p. 1281–1286, out. 2017.

YOSSA, Nadine *et al.* Genomic characterization of *Bacillus cereus sensu stricto* 3A ES isolated from eye shadow cosmetic products. *BMC Microbiology*, v. 22, n. 1, p. 240, 5 out. 2022.

YU, Chenjie *et al.* A glycoside hydrolase 30 protein BpXynC of *Bacillus paralicheniformis* NMSW12 recognized as A MAMP triggers plant immunity response. *International Journal of Biological Macromolecules*, v. 261, p. 129750, mar. 2024.

ZAID, Doaa 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*, v. 10, n. 1, p. e02169-21, 23 fev. 2022.

ZHANG, Pengfei *et al.* Comparison of De Novo Assembly Strategies for Bacterial Genomes. *International Journal of Molecular Sciences*, v. 22, n. 14, p. 7668, 17 jul. 2021.

ZHU, Jiang *et al.* Microbial synthesis of bacitracin: Recent progress, challenges, and prospects. *Synthetic and Systems Biotechnology*, v. 8, n. 2, p. 314–322, jun. 2023.

## 9. APÊNDICE

### 9.1. Análises em andamento

**Tabela suplementar 3:** Tamanhos dos halos de inibição, em milímetros (mm), obtidos nos testes de antibiograma dos isolados BAC30 e BAC220, referentes aos seguintes antibióticos (na ordem da tabela): estreptomicina, eritromicina, gentamicina, tetraciclina, clindamicina, cloranfenicol, ampicilina e vancomicina. Os testes foram realizados em duplicata, com os valores em mm correspondentes.

|        | EST     | ERY     | GEN     | TET     | CLI     | CLO     | AMP   | VAN     |
|--------|---------|---------|---------|---------|---------|---------|-------|---------|
| BAC30  | 21 / 20 | 14 / 13 | 23 / 22 | 22 / 21 | 23 / 22 | 15 / 16 | 0 / 0 | 13 / 14 |
| BAC220 | 20 / 21 | 19 / 18 | 25 / 26 | 24 / 23 | 24 / 24 | 18 / 18 | 0 / 0 | 15 / 14 |

**Tabela suplementar 4:** Tabela de susceptibilidade das linhagens BAC30 e BAC220, baseada nos breakpoints estabelecidos pela EUCAST e pelo CLSI. Na tabela, a susceptibilidade (S) a antibióticos está representada pela cor verde, enquanto a resistência (R) está representada pela cor vermelha.

|        | EST | ERY | GEN | TET | CLI | CLO | AMP | VAN |
|--------|-----|-----|-----|-----|-----|-----|-----|-----|
| BAC30  | S   | R   | S   | S   | S   | S   | R   | S   |
| BAC220 | S   | R   | S   | S   | S   | S   | R   | S   |

### 9.2. Produção científica no mestrado

Produção científica referente às atividades de pesquisa no período de agosto de 2023 a julho de 2025, a qual estive no mestrado. Nesse período, foram publicados 4 artigos e 1 capítulo de livro, todos como colaborador, conforme pode ser observado abaixo no currículo lattes resumido e comprovantes. Além disso, participei do Curso de Férias em Bioinformática da UFMG em 2024, como ouvinte, e posteriormente como palestrante e organizador, em 2025.

#### Curriculum Lattes (Resumido)

**Gabriel Camargos Gomes (<http://lattes.cnpq.br/5073835604323924>)**

#### Curriculum Vitae

---

#### Formação acadêmica/titulação

2023

Mestrado em andamento em Bioinformática.

Universidade Federal de Minas Gerais, UFMG, Brasil.

Título: Bioprospecção in silico de Bactérias Lácticas probióticas.

Orientador:  Vasco Ariston de Carvalho Azevedo.

#### **2018 - 2023**

Graduação em Abi - Ciências Biológicas.

Universidade Federal de Minas Gerais, UFMG, Brasil.

---

## **Formação Complementar**

#### **2023 - 2023**

Workshop on Phenomics. (Carga horária: 15h).

Universidade Federal de Minas Gerais, UFMG, Brasil.

#### **2021 - 2021**

Extensão universitária em Princípios Éticos e Manejo de Animais em Pesquisa. (Carga horária: 60h).

Universidade de São Paulo, USP, Brasil.

---

## **Prêmios e títulos**

#### **2022**

Viabilidade Celular da *Lactococcus lactis* NCDO2118 após a técnica de spray-drying utilizando leite desnatado em pó, Pró-Reitoria de Pesquisa.

#### **2022**

HSP65 protein DNA vaccine and heterologous protein administration for the treatment of DSS-induced colitis, Pró-Reitoria de Pesquisa.

#### **2021**

EFEITO TERAPÊUTICO DO LEITE FERMENTADO POR *Lactobacillus delbrueckii* subsp. *lactis* CIDCA 133 (pExu:hsp65) EM MODELO MURINO DE MUCOSITE INTESTINAL, Pró-Reitoria de Pesquisa.

---

## **Artigos completos publicados em periódicos**

#### **1.**

SOUSA, EDUARDA GUIMARÃES ; CAMPOS, GABRIELA MUNIS ; VIANA, MARCUS VINÍCIUS CANÁRIO ; **GOMES, GABRIEL CAMARGOS** ; RODRIGUES, DIEGO LUCAS NERES ; ABURJAILE, FLAVIA FIGUEIRA ; FONSECA,

BELCHIOLINA BEATRIZ ; DE ARAÚJO, MAX ROBERTO BATISTA ; DA COSTA, MATEUS MATIUZZI ; GUEDON, ERIC ; BRENIG, BERTRAM ; SOARES, SIOMAR ; AZEVEDO, VASCO . The research on the identification, taxonomy, and comparative genomics analysis of nine *Bacillus velezensis* strains significantly contributes to microbiology, genetics, bioinformatics, and biotechnology. *Frontiers in Microbiology* **JCR** , v. 16, p. 1544934, 2025.

2.

★ SOUSA, EDUARDA GUIMARÃES ; CAMPOS, GABRIELA MUNIS ; QUARESMA, LUDMILA SILVA ; MOTA, THAÍS FERNANDES MENDONÇA ; DE CASTILHOS GHISI, NÉDIA ; **GOMES, GABRIEL CAMARGOS** ; SANTOS, RHAYANE CRISTINA VIEGAS ; DE SOUZA, BEATRIZ GAMARANO ROCHA ; GUÉDON, ÉRIC ; DE CASTRO SOARES, SIOMAR ; DA CRUZ FERRAZ DUTRA, JOYCE ; DE CARVALHO AZEVEDO, VASCO ARISTON . Exploring *Bacillus velezensis* in a biomedical context: a systematic review. *Academia Molecular Biology and Genomics*, v. 2, p. 7598, 2025.

3.

★ QUARESMA, LUDMILA SILVA ; SANTOS, RHAYANE CRISTINA VIEGAS ; **GOMES, GABRIEL CAMARGOS** ; AMÉRICO, MONIQUE FERRARY ; CAMPOS, GABRIELA MUNIS ; LAGUNA, JULIANA GUIMARÃES ; BARROSO, FERNANDA ALVARENGA LIMA ; AZEVEDO, VASCO ; DE JESUS, LUÍS CLÁUDIO LIMA . Multidrug resistance profile in *Lactobacillus delbrueckii*: a food industry species with probiotic properties. *WORLD JOURNAL OF MICROBIOLOGY & BIOTECHNOLOGY* **JCR** , v. 40, p. 1-11, 2024. **Citações:** [WEB OF SCIENCE](#) <sup>1</sup>

4.

FREITAS, ANDRIA DOS SANTOS ; BARROSO, FERNANDA ALVARENGA LIMA ; CAMPOS, GABRIELA MUNIS ; AMÉRICO, MONIQUE FERRARY ; VIEGAS, RHAYANE CRISTINA DOS SANTOS ; **GOMES, GABRIEL CAMARGOS** ; VITAL, KÁTIA DUARTE ; FERNANDES, SIMONE ODÍLIA ANTUNES ; CARVALHO, RODRIGO DIAS DE OLIVEIRA ; JARDIN, JULIEN ; MIRANDA, ANA PAULA GOMES DOS SANTOS ; FERREIRA, ENIO ; MARTINS, FLAVIANO SANTOS ; LAGUNA, JULIANA GUIMARÃES ; JAN, GWÉNAËL ; AZEVEDO, VASCO ; DE JESUS, LUÍS CLÁUDIO LIMA . Exploring the anti-inflammatory effects of postbiotic proteins from *Lactobacillus delbrueckii* CIDCA 133 on inflammatory bowel disease model. *INTERNATIONAL JOURNAL OF BIOLOGICAL MACROMOLECULES* **JCR** , v. 277, p. 134216, 2024. **Citações:** [WEB OF SCIENCE](#) <sup>2</sup> | [SCOPUS](#) <sup>1</sup>

5.

★ AMÉRICO, MONIQUE FERRARY ; FREITAS, ANDRIA DOS SANTOS ; DA SILVA, TALES FERNANDO ; DE JESUS, LUÍS CLÁUDIO LIMA ; BARROSO, FERNANDA ALVARENGA LIMA ; CAMPOS, GABRIELA MUNIS ; SANTOS, RHAYANE CRISTINA VIEGAS ; **GOMES, GABRIEL CAMARGOS** ; ASSIS, RAFAEL ; FERREIRA, ÊNIO ; MANCHA-AGRESTI, PAMELA ; LAGUNA, JULIANA GUIMARÃES ; CHATEL, JEAN-MARC ; CARVALHO, RODRIGO DIAS DE OLIVEIRA ; AZEVEDO, VASCO . Growth differentiation factor 11 delivered by dairy *Lactococcus lactis* strains modulates inflammation and prevents mucosal damage in a mice model of intestinal mucositis. *Frontiers in Microbiology* **JCR** , v. 14, p. 1-1, 2023. **Citações:** [WEB OF SCIENCE](#) <sup>5</sup> | [SCOPUS](#) <sup>6</sup>

## Capítulos de livros publicados

1.

SOUZA, E. G. ; GOMES, L. G. R. ; PRATES, F. D. ; GOMES, T. P. ; **GOMES, G. C.** ; PAULA, J. A. ; VINHAL, A. L. O. ; MENDONÇA, B. B. A. ; PEDROSA, M. L. C. ; REIS, L. P. ; OLIVEIRA, A. F. M. ; VIANA, M. V. C. ; JAISWAL, A. K. ; SOARES, S. C. ; AZEVEDO, V. . State of the Art of Immunoinformatics. In: Paulo Fazendeiro, Carmelina Leite. (Org.). *Effective Techniques for Bioinformatic Exploration*. 1ed.: IGI Global, 2024, v. 3, p. 69-106.

2.

★ DE JESUS, L. C. L. ; SILVA, T. F. ; GLORIA, R. A. ; FREITAS, A. S. ; AMÉRICO, M. F. ; FERNANDES, L. J. S. ; CAMPOS, G. M. ; **GOMES, G. C.** ; SANTOS, R. C. V. ; CARVALHO, R. D. O. ; BARH, D. ; AZEVEDO, V. . Lactic acid bacteria?based beverages in the promotion of gastrointestinal tract health. In: Debasis Bagchi. ; Bernard William Downs.. (Org.). *Microbiome, Immunity, Digestive Health and Nutrition*. 1ed.: Academic Press, 2022, v. 26, p. 373-385.

## 9.3. Totais de produção

| <b>Produção bibliográfica</b>             |           |
|---|-----------|
| Artigos completos publicados em periódico | <b>5</b>  |
| Capítulos de livros publicados            | <b>2</b>  |
| Apresentações de trabalhos                | <b>5</b>  |
| Demais produções bibliográficas           | <b>12</b> |

| <b>Eventos</b>                    |          |
|-----------------------------------|----------|
| Participação em eventos           | <b>3</b> |
| Organização de evento (congresso) | <b>2</b> |

## 9.4. Artigos científicos

> [Front Microbiol.](#) 2025 Mar 19;16:1544934. doi: 10.3389/fmicb.2025.1544934. eCollection 2025.

## The research on the identification, taxonomy, and comparative genomics analysis of nine *Bacillus velezensis* strains significantly contributes to microbiology, genetics, bioinformatics, and biotechnology


Eduarda Guimarães Sousa<sup>1</sup>, Gabriela Munis Campos<sup>1</sup>, Marcus Vinícius Canário Viana<sup>1</sup>, Gabriel Camargos Gomes<sup>1</sup>, Diego Lucas Neres Rodrigues<sup>2</sup>, Flavia Figueira Aburjaile<sup>2</sup>, Belchiolina Beatriz Fonseca<sup>3</sup>, Max Roberto Batista de Araújo<sup>1 4</sup>, Mateus Matiuzzi da Costa<sup>5</sup>, Eric Guedon<sup>6</sup>, Bertram Brenig<sup>7</sup>, Siomar Soares<sup>8</sup>, Vasco Azevedo<sup>1</sup>

Affiliations + expand

PMID: 40177483 PMID: PMC11962042 DOI: 10.3389/fmicb.2025.1544934

Acesso: <https://doi.org/10.3389/fmicb.2025.1544934>

**Contribuição:** Participação na metodologia, escrita, revisão e edição

 Systematic review

March 26, 2025

## Exploring *Bacillus velezensis* in a biomedical context: a systematic review

Eduarda Guimarães Sousa\* [1], Gabriela Munis Campos [1], Ludmila Silva Quaresma [1], Thaís Fernandes Mendonça Mota [2], Nédia de Castilhos Ghisi [2], Gabriel Camargos Gomes [1], Rhayane Cristina Viegas Santos [1], Beatriz Gamarano Rocha de Souza [1], Éric Guédon [3], Siomar de Castro Soares [4], Joyce da Cruz Ferraz Dutra [1], Vasco Ariston de Carvalho Azevedo\* [1]

Acesso: <https://www.doi.org/10.20935/AcadMolBioGen7598>

**Contribuição:** Participação na escrita do tópico de produção de metabólitos secundários e revisão bibliográfica.

Review > World J Microbiol Biotechnol. 2024 Jun 8;40(8):235. doi: 10.1007/s11274-024-04046-3.

## Multidrug resistance profile in *Lactobacillus delbrueckii*: a food industry species with probiotic properties

Ludmila Silva Quaresma <sup># 1</sup>, Rhayane Cristina Viegas Santos <sup># 1</sup>, Gabriel Camargos Gomes <sup>1</sup>, Monique Ferrary Américo <sup>1</sup>, Gabriela Munis Campos <sup>1</sup>, Juliana Guimarães Laguna <sup>1</sup>, Fernanda Alvarenga Lima Barroso <sup>1</sup>, Vasco Azevedo <sup># 2</sup>, Luís Cláudio Lima de Jesus <sup># 3</sup>

Affiliations + expand

PMID: 38850338 DOI: 10.1007/s11274-024-04046-3

Acesso: <https://doi.org/10.1007/s11274-024-04046-3>

**Contribuição:** Participação na extração e visualização de dados, escrita da revisão e revisão do texto.

> Int J Biol Macromol. 2024 Oct;277(Pt 2):134216. doi: 10.1016/j.ijbiomac.2024.134216.  
Epub 2024 Jul 26.

## Exploring the anti-inflammatory effects of postbiotic proteins from *Lactobacillus delbrueckii* CIDCA 133 on inflammatory bowel disease model

Andria Dos Santos Freitas <sup>1</sup>, Fernanda Alvarenga Lima Barroso <sup>1</sup>, Gabriela Munis Campos <sup>1</sup>, Monique Ferrary Américo <sup>1</sup>, Rhayane Cristina Dos Santos Viegas <sup>1</sup>, Gabriel Camargos Gomes <sup>1</sup>, Kátia Duarte Vital <sup>2</sup>, Simone Odília Antunes Fernandes <sup>2</sup>, Rodrigo Dias de Oliveira Carvalho <sup>2</sup>, Julien Jardin <sup>3</sup>, Ana Paula Gomes Dos Santos Miranda <sup>4</sup>, Enio Ferreira <sup>4</sup>, Flaviano Santos Martins <sup>5</sup>, Juliana Guimarães Laguna <sup>1</sup>, Gwénaél Jan <sup>3</sup>, Vasco Azevedo <sup>6</sup>, Luís Cláudio Lima de Jesus <sup>7</sup>

Affiliations + expand

PMID: 39069058 DOI: 10.1016/j.ijbiomac.2024.134216

Acesso: <https://doi.org/10.1016/j.ijbiomac.2024.134216>

**Contribuição:** Escrita, revisão e edição, e metodologia

## 9.5. Capítulo de livro

## State of the Art of Immunoinformatics

[Eduarda Guimarães Sousa](#) (Federal University of Minas Gerais, Brazil), [Lucas Gabriel Rodrigues Gomes](#) (Federal University of Minas Gerais, Brazil), [Fernanda Diniz Prates](#) (Federal University of Minas Gerais, Brazil), [Talita Pereira Gomes](#) (Federal University of Minas Gerais, Brazil), [Gabriel Camargos Gomes](#) (Federal University of Minas Gerais, Brazil), [Janaína Aparecida de Paula](#) (Federal University of Minas Gerais, Brazil), [Ana Lua de Oliveira Vinhal](#) (Federal University of Minas Gerais, Brazil), Bernardo Buhr Alves Mendonça (Federal University of Minas Gerais, Brazil), [Mariana Letícia Costa Pedrosa](#) (Federal University of Minas Gerais, Brazil), [Luiza Pereira Reis](#) (Federal University of Minas Gerais, Brazil), Aline Ferreira Maciel de Oliveira (Federal University of Minas Gerais, Brazil), [Marcus Vinicius Canário Viana](#) (Federal University of Minas Gerais, Brazil), Arun Kumar Jaiswal (Federal University of Minas Gerais, Brazil), [Siomar de Castro Soares](#) (Federal University of Triângulo Mineiro, Brazil), and [Vasco Ariston de Carvalho Azevedo](#) (Federal University of Minas Gerais, Brazil)

Source Title: [Effective Techniques for Bioinformatic Exploration](#)

Copyright: © 2025

Pages: 38

ISBN13: 9798369331927 ISBN13 Softcover: 9798369366493 EISBN13: 9798369331934

DOI: 10.4018/979-8-3693-3192-7.ch003

**Acesso:** <http://dx.doi.org/10.4018/979-8-3693-3192-7.ch003>

**Contribuição:** Escrita do tópico “Future Horizons and Emerging Trends”.