



UNIVERSIDADE DE BRASÍLIA
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA MICROBIANA

**ANÁLISE *IN SILICO* DO POTENCIAL GENÔMICO
PARA PRODUÇÃO DE TERPENOS EM
BACTÉRIAS AERÓBIAS FORMADORAS DE ENDÓSPOROS**

FELIPE DE ARAÚJO MESQUITA

Brasília
2023



Universidade de Brasília
Instituto de Ciências Biológicas
Programa de pós-graduação em Biologia
Microbiana



FELIPE DE ARAÚJO MESQUITA

**ANÁLISE *IN SILICO* DO POTENCIAL GENÔMICO PARA PRODUÇÃO DE
TERPENOS EM BACTÉRIAS AERÓBIAS FORMADORAS DE ENDÓSPOROS**

Dissertação apresentada ao Programa de Pós-graduação em Biologia Microbiana, do Departamento de Biologia Celular, do Instituto de Ciências Biológicas da Universidade de Brasília como pré-requisito parcial para obtenção de título de Mestre em Biologia Microbiana.

Prof^a. Dr^a. Marlene Teixeira De-Souza
Orientadora

Prof. Dr. Waldeyr Mendes Cordeiro da Silva
Coorientador

Brasília

2023

AGRADECIMENTOS

Agradeço primeiramente à minha família. À minha mãe Shirlei, meu pai José Airton e minhas irmãs Letícia e Fernanda. Obrigado pelo apoio dado nestes últimos anos, os incentivos, as conversas e os momentos de lazer. Agradeço meus pais que nunca mediram esforços para proporcionarem dentro de suas possibilidades educação de qualidade para mim e minhas irmãs. Obrigado por me ensinarem que é através do estudo que nos tornamos pessoas de caráter, justas, honestas e cidadãs, em uma sociedade tão injusta e desigual.

Agradeço à Universidade de Brasília, pela formação profissional e ética que obtive ao longo destes últimos sete anos. Sentirei saudades desta instituição, das amizades construídas e dos momentos de frustrações, alegrias e conquistas. Obrigado pelo aprendizado e pela pessoa que me transformei nesta jornada. Apesar da despedida sinto que a Universidade sempre estará comigo ao longo dessas novas portas que se abrirão. Agradeço também à Capes pela bolsa de mestrado incluindo o reajuste, tão importante, para muitos estudantes em todo o país.

Agradeço ao meu coorientador, professor Waldeyr, pela disponibilidade em me ajudar na realização deste estudo, bem como pelo aprendizado na área de bioinformática. Obrigado pela paciência nos momentos de preocupação e por me guiar em uma área inexplorada por minha pessoa até então. Agradeço também aos professores Gabriel Costa Alves e Thaís Amaral, que de bom grado, concordaram em participar como membros da banca de defesa deste estudo.

Por fim, agradeço minha orientadora, professora Marlene. Obrigado pela orientação dada não somente neste trabalho, mas em todos os projetos desenvolvidos ao longo dos últimos cinco anos. Agradeço o aprendizado profissional, ético e humano que aprendi em momentos compartilhados com a senhora. Obrigado pela amizade e pelos conselhos que levarei para a vida. Agradeço por me mostrar que o mundo é enorme e que tenho todo o potencial para conquistá-lo.

RESUMO

Metabólitos secundários ou especializados são moléculas não essenciais ao crescimento celular, porém, desempenham grande importância adaptativa para os organismos produtores, como bactérias, fungos e plantas. Bactérias isoladas de solos são conhecidas por sintetizarem uma ampla diversidade de metabólitos especializados com diferentes propriedades físico-químicas. Espécies de *Bacillus* e gêneros relacionados, coletivamente designadas bactérias aeróbias formadoras de endósporos (Bafes), são promissoras para a busca desses compostos. Porém, relatos sobre a síntese de terpenos por Bafes são deveras escassos na literatura científica. O objetivo deste estudo foi analisar, *in silico*, o potencial genômico para produção de terpenos em dez linhagens de Bafes isoladas do solo do Distrito Federal (linhagens SDF). Utilizando técnicas e ferramentas de Bioinformática, foi possível detectar prováveis *biosynthetic gene clusters* (BGC) envolvidos na síntese de metabólitos especializados nestas linhagens. Adicionalmente, foi realizada a reconstrução *in silico* das respectivas redes metabólicas para detectar a presença das enzimas da via metil-eritrol fosfato (MEP) responsáveis pela síntese de terpeno em procariotos. Os resultados obtidos apontaram que parte das linhagens SDF sondadas possui o aparato genético completo para produção dos terpenos esqualeno, hopanóide e licopeno. Em complemento, análises filogenéticas baseadas em sequências de aminoácidos de terpeno sintases detectadas, revelaram que estes catalisadores são conservados evolutivamente, sugerindo que o potencial de produção de terpenos é amplamente distribuído em Bafes. Além de terpenos, os resultados obtidos também revelaram que as linhagens SDF são fontes promissoras de policetídeos e peptídeos não ribossomais. Portanto, estas bactérias são boas candidatas para prospecção de novos bioprodutos, notadamente antimicrobianos.

Palavras-chave: Metabólitos especializados, bioinformática, isoprenóides, *Bacillales*.

SUMMARY

Secondary or specialized metabolites are non-essential molecules for cell growth. However, they play an important adaptive role for the producing organisms, such as bacteria, fungi, and plants. Bacteria isolated from soils are known for synthesizing a wide range of specialized metabolites with different physicochemical properties. *Bacillus* species and from related genera, collectively quoted as aerobic endospore-forming bacteria (AEFB), are promising for searching these compounds. Nevertheless, reports on the synthesis of terpenes by AEFB are very scarce in the scientific literature. This study aimed to analyse the genomic potential for terpene production in ten AEFB strains isolated from soil in the Federal District (SDF strains). Using Bioinformatics techniques and tools, it was possible to detect probable biosynthetic gene clusters (BGC) involved in the syntheses of specialized metabolites in these strains. Additionally, *in silico* reconstruction of the respective metabolic networks was carried out to detect the presence of enzymes in the methylerythrol phosphate (MEP) pathway, responsible for terpene synthesis in prokaryotes. The results indicate that part of the probed SDF strains has the complete genetic apparatus to produce the terpenes squalene, hopanoid, and lycopene. Phylogenetic analyses based on amino acid sequences of the detected terpene synthases revealed that these catalysts are well conserved, indicating that the potential for terpene production is widely distributed among AEFB. In addition to terpenes, the results showed that AEFB are promising sources of polyketides and non-ribosomal peptides. Therefore, these bacteria are good candidates for prospecting new bioproducts, notably antimicrobials.

Keywords: Specialized metabolites, bioinformatics, isoprenoids, *Bacillales*.

LISTA DE FIGURAS

- Figura 1. Esquema simplificado das principais vias bioquímicas sintetizadoras de metabólitos especializados.** A glicose é o precursor fundamental para síntese de diferentes classes de metabólitos especializados pela via do ácido chiquímico, via do ácido malônico, via do mevalonato (MEV) ou pela via metileritrotol-fosfato (MEP).....18
- Figura 2. BGC de talionol que dirige a síntese de talianina em *A. thaliana*.** Os genes THAS (seta vermelha), THAH (seta azul claro), THAO (seta azul escuro), THAA1 e THAA2 (seta rosa) codificam as enzimas responsáveis pela via biossintética de talianina em *A. thaliana*.....19
- Figura 3. Rotas bioquímicas das vias do mevalonato (MVA) e metileritrotol-fosfato (MEP) para síntese de terpenos.** As enzimas DXS, DXR, MCT, CMK, MDS, HDS e HDR catalisam reações sequenciais e indispensáveis para produção final de isopentenil pirofosfato (IPP) a partir de glicose na via MEP.....21
- Figura 4. Os sete estágios para formação de um endósporo em Bafes (modelo *B. subtilis*).** (1) Formação do filamento axial, (2) Formação do septo assimétrico, (3) Engolfamento do pré-esporo pelo envelope celular, Síntese do córtex (4) e capa (5) nas camadas externas no esporo, (6) Maturação e (7) Liberação do esporo após lise celular.....24
- Figura 5. Microscopia de contraste de fase de células de Bafes em esporulação.** Endósporos (seta vermelha) no citoplasma do esporângio de uma espécie de Bafes.....25
- Figura 6. Relações filogenéticas de 192 linhagens SDF depositadas na CBafes inferidas a partir de sequências de genes de rRNA 16S.** O resultado aponta pelo menos oito agrupamentos filogenéticos. As letras indicam (A) Complexo do *B. subtilis*, (B) Grupo do *B. cereus*, (C) *B. simplex*, (D) Grupo do *B. megaterium*, (E) *Lysinibacillus*, (F) *B. clausii*, (G) *Paenibacillus* e (H) *Brevibacillus*.....29
- Figura 7. Fluxograma resumido com as principais atividades efetuadas para realização deste estudo**.....38
- Figura 8. Total de BGC relacionados à síntese dos metabólitos especializados identificados neste estudo pelo antiSMASH.** A classe representativa com o maior número de BGC identificados foi aquela que codifica a síntese de NRPS, com 47 BGC reconhecidos, seguidos de trans ATPKS (17), terpenos (16), T3PKS (12), Betalactonas (10), RiPP-like (7), T1PKS (6), Sideróforos (6), LAP (5), RRE (4), PKS like (4), NRPS like (4), Bacilisina (4), Tiopeptídeos (3), Lantipeptídeos (2), Autoindutor de LC (2), Sactipeptídeos (1), Rantipeptídeos (1), Lasso-peptídeos (1), Fosfanato (1).....39
- Figura 9. Classes de metabólito especializado identificados nas linhagens SDF utilizadas neste estudo.** As cores que representam as classes estão identificadas na parte inferior da figura. A linhagem *Paenibacillus* sp. SDF0028 teve 38 BGC reconhecidos, seguido de *B. velezensis* SDF0150 (22), *B. velezensis* SDF0141 (21), *Lysinibacillus* sp. SDF0063 (19), *B. pumilus* SDF011 (16), *B. simplex* SDF0024 (10), *Bacillus* sp. SDF0016 (10), *Lysinibacillus* sp. SDF0037 (9), *L. fusiformis* SDF0005 (5) e *B. oleronius* SDF0015 (3)40

- Figura 10. BGC de bacillibactina.** BGC que codifica atividades responsáveis pela síntese de bacillibactina detectado nas linhagens *B. velezensis* (SDF0141 e SDF0150) com índice de similaridade de 100%.....42
- Figura 11. BGC de bacilisina.** BGC que codifica atividades responsáveis pela síntese de bacilisina detectado nas duas linhagens *B. velezensis* (SDF0141 e SDF0150) com índice de similaridade de 100%.....43
- Figura 12. BGC de macrolactina H.** BGC que codifica atividades responsáveis pela síntese de macrolactina detectado nas duas linhagens *B. velezensis* (SDF0141 e SDF0150) com índice de similaridade de 100%.....43
- Figura 13. BGC de bacillaeno.** BGC que codifica atividades responsáveis pela síntese de bacillaeno detectado nas duas linhagens *B. velezensis* (SDF0141 e SDF0150) com índice de similaridade de 100%.....43
- Figura 14. BGC de paenibacterina.** BGC que codifica atividades responsáveis pela síntese de paenibacterina detectado na linhagem *Paenibacillus* sp. SDF0028 com índice de similaridade de 100%.....44
- Figura 15. BGC de macrobrevina.** BGC que codifica atividades responsáveis pela síntese de macrobrevina detectado na linhagem *Lysinibacillus* sp. SDF0063 com índice de similaridade de 100%.....44
- Figura 16. Gene *sqhC* identificado pelo antiSMASH dentro dos respectivos BGC, que codifica as atividades responsáveis pela síntese de terpenos nas linhagens SDF avaliadas.** As cores indicam o tipo de atividade codificada por cada sequência.....45
- Figura 17. Localização do gene codificador de FES.** A posição do gene codificador de FES adjacente ou não ao gene *CrtI* identificados dentro dos respectivos BGC, que codificam as atividades responsáveis pela síntese de terpenos nas linhagens SDF avaliadas. As cores indicam o tipo de atividade codificada por cada sequência.....46
- Figura 18. BGC de carotenoide.** BGC que codificam as atividades responsáveis pela síntese de carotenoide nas linhagens *B. pumilus* SDF0011 e *Bacillus* sp. SDF0016 com índice de similaridade de 50%.....47
- Figura 19. Via metileritrotol-fosfato (MEP).** Diagrama da reconstrução metabólica *in silico* da via MEP detectado pelo aplicativo *Pathways tools* para a linhagem *B. velezensis* SDF0150 e demais linhagens analisadas (canto superior direito). As cores indicam a presença ou ausência do catalisador para a respectiva reação.....48
- Figura 20. Relações filogenéticas entre linhagens SDF baseado na sequência de aminoácidos da enzima SHC.** As linhagens *B. velezensis* SDF0141 e SDF0150 são mais relacionadas com as linhagens *B. pumilus* SDF0011 e *Bacillus* sp. SDF0016. As sequências de aminoácidos de SHC de *B. simplex* SDF0024 e *B. oleronius* SDF0015 são mais relacionadas

quando comparadas às de *Lysinibacillus* sp. SDF0063. A linhagem *Paenibacillus* sp. SDF0028 possui menor grau de parentesco com as demais linhagens.....50

Figura 21. Relações filogenéticas entre linhagens SDF baseado na sequência de aminoácidos da enzima FES. As linhagens *B. velezensis* SDF0141 e SDF0150 são mais relacionadas com as linhagens *B. pumilus* SDF0011 e *Bacillus* sp. SDF0016. As sequências de aminoácidos de FES de *L. fusiformis* SDF0005 e *Lysinibacillus* sp. SDF0037 são mais relacionadas quando comparadas às de *B. simplex* SDF0024.....51

Figura 22. Similaridade entre o conjunto enzimático envolvidos na produção de terpenos nas linhagens SDF avaliadas. A cor azul representa a presença de uma enzima de interesse para produção de terpeno, enquanto ausência está representado pela cor preta. As linhagens *B. pumilus* SDF0011 e *Bacillus* sp. SDF0016 apresentaram mesma similaridade do conjunto enzimático, assim como as linhagens *B. velezensis* (SDF0141 e SDF0150) e *B. simplex* SDF0024. As linhagens *Paenibacillus* sp. SDF0028 e *Lysinibacillus* sp. SDF0063 também apresentaram o mesmo conjunto enzimático.52

Figura 23. Domínios usuais de PKS – cetoacila sintase (KS), acetil-transferase (AT) e carreadora de acila (ACP). Usualmente, os domínios de PKS são adjacentes na estrutura enzimática (Cis-AT). Em trans AT-PKS o domínio AT não está posicionado de modo subsequente aos demais.....57

Figura 24. Mapa das reações catalíticas detectadas na biossíntese de terpenos nas linhagens SDF sondadas. As cores indicam a linhagem SDF que apresenta a enzima que catalisa a respectiva reação. Todas as linhagens apresentaram as enzima da via MEP. A enzima da família poliprenil sintase foi detectada nas linhagens SDF0005, SDF0015, SDF0024, SDF0141 e SDF0150. A enzima FES foi detectada nas linhagens SDF0005, SDF0011, SDF0016, SDF0024, SDF0037, SDF0141 e SDF0150. A enzima SHC foi detectado nas linhagens SDF0011, SDF0015, SDF0016, SDF0024, SDF0028, SDF0063, SDF0141 e SDF0150.....68

Figura 25. Alinhamento de sequências de aminoácidos para detecção de enzima da família poliprenil sintase em *L. fusiformis* SDF0005.....84

Figura 26. Alinhamento de sequências de aminoácidos para detecção de enzima da família poliprenil sintase em *B. oleronius* SDF0015.....84

Figura 27. Alinhamento de sequências de aminoácidos para detecção de enzima da família poliprenil sintase em *B. simplex* SDF0024.....85

Figura 28. Alinhamento de sequências de aminoácidos para detecção de enzima da família poliprenil sintase em *B. velezensis* SDF014.....85

Figura 29. Alinhamento de sequências de aminoácidos para detecção de enzima da família poliprenil sintase em *B. velezensis* SDF0150.....86

LISTA DE TABELAS

Tabela 1. Enzimas necessárias para produção de IPP pela via MEP para síntese de terpenos em bactérias.....	22
Tabela 2. Linhagens SDF utilizadas para análise de produção de metabólitos especializados neste estudo.....	35
Tabela 3. Metabólitos especializados identificados pelo antiSMASH nos dez genomas analisados neste estudo.....	41
Tabela 4. Presença de TS identificadas pelo antiSMASH nas linhagens SDF investigadas	47
Tabela 5. Detecção de enzimas da família poliprenil sintase nas linhagens SDF pelo BLAST. Os níveis mais altos de identidade resultante dos alinhamentos das respectivas sequências de cada SDF estão descritos.....	49

LISTA DE SIGLAS E ABREVIATURAS

- ACP:** proteína carreadora de acila
- antiSMASH:** (do inglês, *antibiotics and secondary metabolite analysis shell*)
- Bafes:** bactérias aeróbias formadoras de endósporos
- BGC:** (do inglês, *biosynthetic gene clusters*)
- CBafes:** Coleção de bactérias aeróbias formadoras de endósporos
- CD-ME:** 4-difosfocitidil-2-C-metil-D-eritritol
- CD-MEP:** 4-difosfocitidil-2-C-metil-D-eritritol 2-fosfato
- CMK:** 4-difosfocitidil-2-C-metil-D-eritritol quinase
- DMAPP:** dimetilalil pirofosfato
- DNA:** ácido desoxirribonucleico
- DXP:** 1-Desoxi-D-xilulose 5-fosfato
- DXR:** 1-Desoxi-D-xilulose 5-fosfato redutoisomerase
- DXS:** 1-Desoxi-D-xilulose 5-fosfato sintase
- FES:** fitoeno/esqualeno sintase
- FPP:** farnesil pirofosfato
- FPPS:** farnesil pirofosfato sintase
- GGPP:** geranil geranil pirofosfato
- GGPPS:** geranil geranil pirofosfato sintase
- GPP:** geranil pirofosfato
- GPPS:** geranil pirofosfato sintase
- HDR:** 4-hidroxi-3-metilbut-2-enil difosfato redutase
- HDS:** 4-hidroxi-3-metilbut-2-enil difosfato sintase
- HMBPP:** 4-hidroxi-3-metilbut-2-enil difosfato
- IDI:** isopentenil pirofosfato isomerase
- IPP:** isopentenil pirofosfato
- KS:** cetoacila sintase
- LAP:** (do inglês, *linear azole-containing peptides*)
- LC:** lactona cíclica
- MCT:** 2-C-metil-D-eritritol 4-fosfato citidiltransferase
- MDS:** 2-C-metil-D-eritritol-2,4-ciclodifosfato sintase
- MEC:** 2-C-metil-D-eritritol-2,4-ciclodifosfato
- MEGA:** (do inglês, *Molecular Evolutionary Genetics Analysis*)

MEP: via metileritrotol-fosfato

MIBiG: (do inglês, *minimum information about a biosynthetic gene cluster*)

MVA: via do mevalonato

NRP: peptídeo não ribossomal

NRPS: peptídeo não ribossomal sintetase

OMS: Organização Mundial da Saúde

PDS: fitoeno desaturase

pH: potencial hidrogeniônico

PHS: fitoeno sintase

PK: policetídeo

PKS: policetídeo sintase

PPAS: pequenas proteínas ácido solúveis

RiPP: (do inglês, *ribosomally synthesized and post-translationally modified peptides*)

RNA: ácido ribonucleico

rRNA: ácido ribonucleico ribossomal

RRE: (do inglês, *RiPP recognition element*)

SDF: solo do Distrito Federal

SHC: esqualeno hopeno ciclase

SQS: esqualeno sintase

T1PKS: policetídeo sintase tipo I

T2PKS: policetídeo sintase tipo II

T3PKS: policetídeo sintase tipo III

trans AT-PKS: policetídeo sintase acil-transferase

TS: terpeno sintase

UnB: Universidade de Brasília

SUMÁRIO

1. Introdução.....	13
2. Revisão de Literatura.....	16
2.1 Metabólitos especializados.....	16
2.2 Terpenos.....	20
2.3 Filo Firmicutes.....	22
2.4 Bactérias aeróbias formadoras de endósporos.....	26
2.5 Taxonomia e filogenia de Bafes.....	27
2.6 Coleção de Bactérias aeróbias formadoras de endósporos.....	30
2.7 Principais ferramentas computacionais utilizadas neste estudo.....	30
2.7.1 Plataforma antibiotics and secondary metabolite analysis shell	30
2.7.2 Molecular evolutionary genetics analysis (MEGA)	32
2.7.3 Pathways tools.....	32
3. Justificativa.....	33
4. Objetivos.....	34
4.1 Objetivo geral.....	34
4.2 Objetivos específicos para as linhagens SDF avaliadas.....	34
5. Metodologia.....	35
5.1 Linhagens SDF.....	35
5.2 Identificação de <i>BGC</i> envolvidos na síntese de metabólitos especializados.....	35
5.3 Reconstrução metabólica para detecção dos genes de enzimas da via MEP.....	36
5.4 Detecção de enzimas da família poliprenil sintase.....	37
5.5 Construção de árvores filogenéticas.....	37
5.6 Similaridade do conjunto enzimático para produção de terpenos.....	38
6. Resultados.....	39
7. Discussão.....	53
7.1 Potencial genômico global para síntese de metabólitos especializados em Bafes.....	53
7.2 Potencial genômico para produção de metabólitos especializados nas linhagens SDF...60	
7.3 Perspectiva para identificação de novos metabólitos especializados em Bafes.....	61
7.4 Potencial genômico para produção de terpenos nas linhagens SDF selecionadas.....	63
7.5 Caráter evolutivo para produção de terpenos em Bafes.....	68
8. Considerações Finais.....	71
9. Referências bibliográficas.....	73
10. Apêndice.....	84

1. INTRODUÇÃO

Metabólitos secundários ou especializados são moléculas de vital importância biológica para os organismos produtores como bactérias, fungos e plantas (BILLS; GLOER, 2017). Estes compostos não são fundamentais ao crescimento celular, porém, apresentam papel ecológico relevante ao garantir disponibilidade de nutrientes em ambientes competitivos (SINGH et al., 2019). Os metabólitos especializados são sintetizados pela combinação, ou não, de vias biossintéticas de poliketídeos, do mevalonato ou de aminoácidos (KELLER et al., 2005).

Bactérias isoladas de solos sintetizam grande quantidade de metabólitos especializados com diferentes propriedades físico-químicas, e que são explorados em diversas áreas da saúde, agricultura e produção de alimentos (SINGH et al., 2019). Notadamente, espécies do gênero *Bacillus* e outros gêneros relacionados, coletivamente designadas bactérias aeróbias formadoras de endósporos (Bafes), possuem o solo como principal reservatório, e produzem metabólitos especializados com alta variabilidade funcional (MONDOL et al., 2013). A importância ecológica e socioeconômica destes metabólitos de Bafes é reconhecida pela capacidade de promoção de crescimento vegetal, ação imunossupressora, antimicrobiana, antitumoral, além de controle de insetos praga e vetores de doenças (VAN DER AUWERA et al., 2007; MONDOL et al., 2013; FALQUETO et al., 2021).

Os bioprodutos de interesse englobam diversas classes como bacteriocinas, poliketídeos (PK) e peptídeos não ribossomais (NRP) (KELLER et al., 2005; TYC et al., 2017). Atualmente essas classes de metabólitos são investigadas e exploradas em Bafes como fontes promissoras na busca de novos agentes antimicrobianos e demais aplicações biotecnológicas (SUMI et al., 2015; HEILBRONNER et al., 2021). Dessa forma, estudos para prospecção de metabólitos especializados são relevantes para sedimentar o conhecimento sobre metabolismo especializado em Bafes e ampliar o arcabouço para exploração de novos bioprodutos.

Terpenos revelam-se como uma importante classe de metabólitos especializados para sondagem destas novas moléculas. Estes compostos são amplamente caracterizados e estudados em vegetais, porém, a produção é reconhecida também em procariotos e fungos (QUIN et al., 2014; YAMADA et al., 2015). Embora a síntese de terpenos de origem bacteriana seja relatada na literatura científica, as funções biológicas e ecológicas destes metabólitos neste domínio procariótico ainda são amplamente desconhecidas (TYC et al., 2017). Enquanto o gênero *Streptomyces* é explorado como modelo na investigação da produção de terpenos em bactérias (TYC et al., 2017), estudos para produção destes metabólitos em Bafes ainda são escassos. Logo, análises de identificação do potencial de produção de terpenos em Bafes configuram-se

importantes ferramentas para a compreensão do papel biológico destas moléculas neste grupo, além de ampliar o potencial biotecnológico desses microrganismos.

Os terpenos são sintetizados a partir de correspondentes intermediários originários do isopentenil pirofosfato (IPP) provenientes da via do mevalonato (MVA) ou da via do metileritritol-fosfato (MEP) (ZHANG et al., 2016). Condensações contínuas de IPP, submetidos posteriormente a ciclização e pós-modificações bioquímicas geram uma ampla diversidade de terpenos encontrados na natureza (LIANG et al., 2021). Em procariotos a produção de terpenos ocorre exclusivamente via MEP e pelo menos sete enzimas estão envolvidas na produção de IPP (LIANG et al., 2021). Desta maneira, a avaliação do potencial de produção de terpenos em Bafes apenas se viabiliza com a investigação dos genes envolvidos na produção de enzimas catalíticas presentes na via MEP nestas bactérias.

Os precursores para síntese de terpenos são essenciais, entretanto, a capacidade de um microrganismo produzir estes metabólitos sujeita-se à presença de terpeno sintases (TS). Estudos recentes revelaram que as enzimas TS estão amplamente distribuídas em bactérias da família *Streptomycetaceae*, isoladas de solos (YAMADA et al., 2015). Contudo, algumas TS não estão distribuídas uniformemente no gênero *Streptomyces* levando a formação de diferentes grupos filogenéticos (SÁNCHEZ et al., 2019). Essa perspectiva pode ser extrapolada para Bafes. Ao analisar a presença de genes de TS distribuídos especificamente, ou de modo randômico em diferentes linhagens, a natureza e a importância dos terpenos em Bafes pode ser expandida, além de possibilitar o aprofundamento da discussão de relações evolutivas nesses microrganismos.

Baseando-se nos elementos destacados acima, esse estudo se propôs a analisar *in silico* o potencial para produção de terpenos em dez genomas de linhagens de Bafes isoladas do solo do Distrito Federal (linhagens SDF) da Coleção de Bactérias aeróbias formadoras de endósporos (CBafes) da Universidade de Brasília (UnB). A CBafes está localizada no Laboratório de Microbiologia do Instituto de Ciências Biológicas, Departamento de Biologia Celular da UnB e abriga 312 linhagens designadas SDF0001-SDF0312. Para fins taxonômicos e biotecnológicos, a CBafes tem sido inspecionada por uma abordagem polifásica, onde estão sendo reunidos dados fenotípicos e genotípicos. A Coleção abrange desde espécies bem conhecidas na literatura até aquelas com descrição escassa (OREM et al., 2019; CAVALCANTE et al., 2019; MARTINS et al., 2020; MARTINS et al., 2023).

O genoma de parte das linhagens SDF foi sequenciado, montado e anotado. Análises *in silico* permitem buscar nestes genomas agrupamentos de genes que configurem vias bioquímicas produtoras de metabólitos especializados conhecidos como *biosynthetic gene*

clusters (*BGC*). Utilizando ferramentas e técnicas de Bioinformática, este estudo se propôs a identificar possíveis *BGC* envolvidos na síntese global de metabólitos especializados e, em especial de terpenos, em dez linhagens SDF selecionadas. Para confirmar o efetivo potencial de produção de terpenos foi avaliada a presença dos genes que sintetizam os precursores fundamentais de terpenos pela via MEP. Posteriormente, as sequências de aminoácidos de TS identificadas foram utilizadas para construção de árvores filogenéticas, a fim de caracterizar a natureza evolutiva de terpenos em Bafes e ampliar a discussão sobre as relações filogenéticas dentro deste grupo.

2. REVISÃO DE LITERATURA

2.1. *Metabólitos especializados*

Metabólitos especializados ou secundários são moléculas não essenciais ao crescimento celular, porém, desempenham papel complementar ao desenvolvimento biológico pois promovem vantagens adaptativas ao organismo produtor (SINGH et al., 2019). Diversas espécies de bactérias, fungos e plantas são amplamente conhecidas por produzirem alta variedade de metabólitos especializados, entretanto, a síntese desses compostos também já foi detectada em outros organismos como esponjas, tunicados, corais e caramujos (VAISHNAV; DEMAINS, 2010).

O potencial biotecnológico de metabólitos especializados é extensivamente explorado em diversas áreas da saúde, agricultura e produção de alimentos (SINGH et al., 2019). Os metabólitos especializados são inestimáveis agentes antibacterianos, antifúngicos, herbicidas, antiparasitas, antitumorais, imunossupressores e agentes redutores de colesterol (MONDOL et al., 2013; VAISHNAV; DEMAINS, 2010; SINGH et al., 2019; FALQUETO et al., 2021). Alguns desses compostos exercem papel em tratamentos ou na prevenção de distúrbios biológicos, até então, sem cura conhecida. Estudos recentes, por exemplo, identificaram metabólitos originários de plantas com potencial para o tratamento da Covid-19 (WINK, 2020).

Os metabólitos especializados apresentam alta variabilidade estrutural e podem ser classificados como peptídeos não ribossomais (NRP), policetídeos (PK), alcaloides, bacteriocinas (KELLER et al., 2005; TYC et al., 2017), entre outras classes.

Os NRP são derivados de modificações de aminoácidos proteínogênicos ou sintetizados *de novo*. Esses peptídeos podem ser dímeros, trímeros ou estruturas macrocíclicas, ramificadas ou não (NÚÑEZ; LÓPEZ, 2016). Normalmente, os NRP são sintetizados por grandes complexos enzimáticos de peptídeos não ribossomais sintetases (NRPS), caracterizados como enzimas modulares de múltiplos domínios (TYC et al., 2017). As NRPS estão presentes em eucariotos e *archaeas*, entretanto, são encontradas sobretudo em bactérias alocadas nos filos *Actinobacteria*, *Proteobacteria*, *Cyanobacteria* e *Firmicutes* (NÚÑEZ; LÓPEZ, 2016). Os NRP possuem várias aplicações, desde ação antimicrobiana, imunossupressora, citostática, entre outros (SCHWARZER et al., 2003).

Os PK são resultantes da condensação de unidades de acetila ou malonil-CoA mediada por enzimas designadas policetídeo sintases (PKS), também relacionadas com a síntese de ácidos graxos em eucariotos (BILLS; GLOER, 2017). De acordo com o conjunto de domínios proteicos presentes nas enzimas, as PKS podem ser classificadas em tipo I, II ou III. As PKS do tipo I são enzimas multidomínios modulares, em que cada domínio é utilizado apenas uma

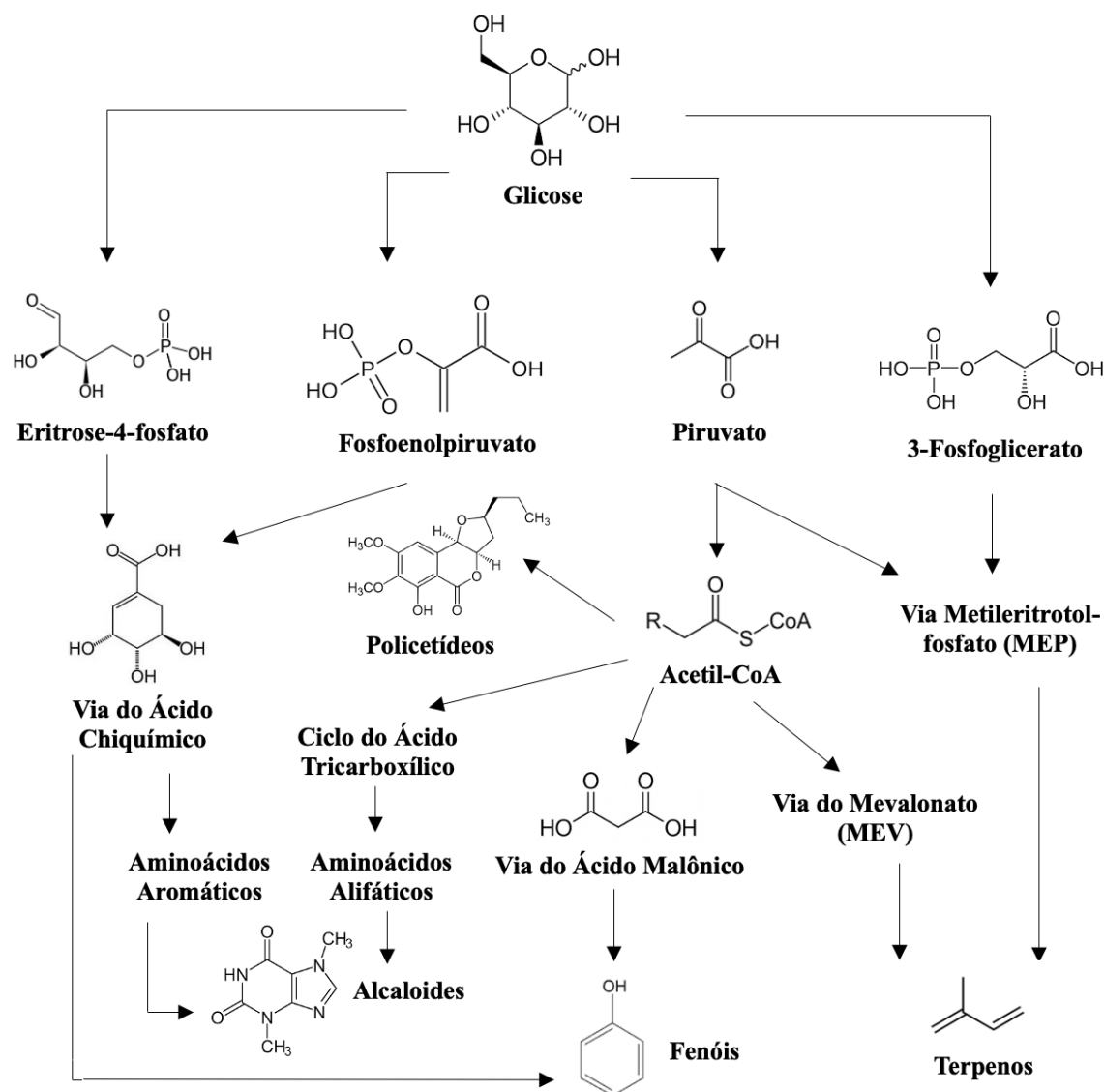
vez até o processamento do PK, ou iterativas com os domínios utilizados repetidas vezes durante o processamento do metabólito (CUADRAT, 2010). Em PKS do tipo II, a atividade catalítica para a produção do PK é realizada por um complexo de proteínas dissociáveis, e cada domínio é utilizado de maneira iterativa. Por outro lado, as PKS do tipo III apresentam estrutura homodimérica, ação interativa e são independentes do domínio denominado proteína carreadora de acila (ACP) para a síntese da cadeia policetídica em formação (CUADRAT, 2010; PAULO; SIGRIST; OLIVEIRA, 2019). Os PK são produzidos por bactérias, fungos e plantas e apresentam alta diversidade estrutural e funcional. Os fármacos derivados dessa classe de metabólitos representam fração significativa de produtos comercializados na indústria farmacêutica, tais como a eritromicina e tetraciclina (WALSH; FISHBACH, 2010).

Os alcaloides constituem uma classe heterogênea de estruturas nitrogenadas de caráter básico. São encontrados predominantemente na forma combinada com ácidos orgânicos e, em concentração menor, na forma livre (KELLER, 2005). Os alcaloides são sintetizados a partir de aminoácidos aromáticos decorrentes da via do ácido chiquímico (CABRAL; PITA, 2015). Porém, podem ser derivados de terpenos e fenóis, o que amplia a diversidade estrutural e funcional destes metabólitos. Os alcaloides apresentam diversas propriedades biológicas como ação sinalizadora, antimicrobiana, além de serem utilizados para proteção nas relações de predação e defesa de animais e vegetais. Os alcaloides são produzidos por bactérias, fungos, protozoários e animais (CABRAL; PITA, 2015). Entretanto, as plantas continuam sendo a principal fonte conhecida dessas moléculas.

As bacteriocinas consistem em peptídeos ribossomais bacterianos, que podem sofrer modificações pós-traducionais, adquirindo propriedades específicas ou diversificadas (HEILBRONNER, 2021). As bacteriocinas produzidas por uma célula exibem atividade contra outros microrganismos de uma mesma espécie ou gêneros diferentes. Atualmente, acredita-se que a produção dessas moléculas é uma estratégia utilizada por bactérias para controlar organismos competidores por nutrientes em um mesmo nicho ambiental (TYC et al., 2017). Consequentemente, as bacteriocinas, influenciam a dinâmica populacional periférica tanto em nível de populações, quanto de comunidade. Esse entendimento permitiu o emprego desses peptídeos como importantes reguladores da microbiota humana (HEILBRONNER, 2021).

Os metabólitos especializados são produzidos por diferentes rotas bioquímicas de modo independente ou integrado (Figura 1). Estas vias caracterizam-se pela alta complexidade, o que impossibilita em muitos casos a síntese destes compostos em laboratório (TWAIJ; HASAN, 2022). Nas principais rotas biossintéticas para produção de metabólitos especializados, a acetil-

CoA é uma molécula precursora fundamental e integrativa para produção de estruturas variadas (Figura 1).

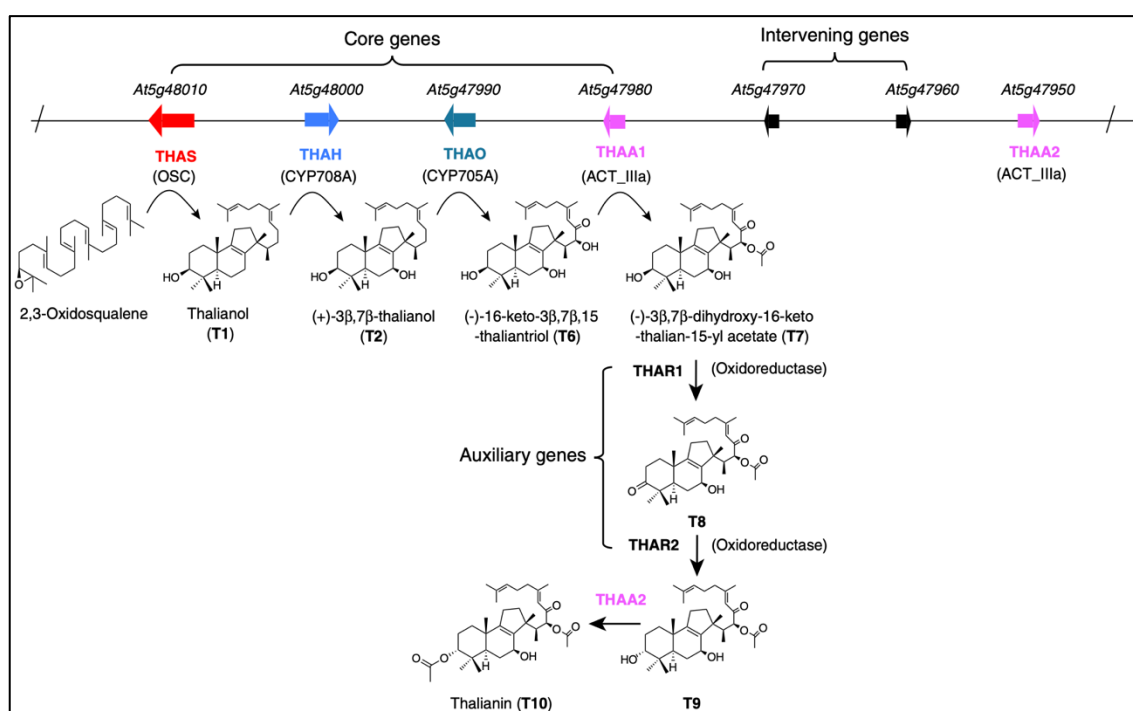


Fonte: TWAI; HASSAN, 2022, com modificações.

Figura 1. Esquema simplificado das principais vias bioquímicas sintetizadoras de metabólitos especializados. A glicose é o precursor fundamental para síntese de diferentes classes de metabólitos especializados pela via do ácido chiquímico, via do ácido malônico, via do mevalonato (MEV) ou pela via metileritrotol-fosfato (MEP).

Os genes que codificam as atividades envolvidas na síntese de vários metabólitos especializados encontram-se em agrupamentos denominados *biosynthetic gene clusters* (BGC). Geralmente, os BGC contêm toda a informação para biossíntese do precursor, a montagem, modificação e regulação do produto final (LEE et al., 2020). A expressão destes *clusters* é

controlada por redes regulatórias complexas, influenciadas por fatores celulares e do hábitat natural circundante do organismo produtor (SILVA, 2014; LEE et al., 2020). O *cluster* de thalianol, por exemplo, é altamente conservado em espécies de planta do gênero *Arabidopsis*. Este *cluster* possui quatro genes que são expressos simultaneamente (*THAS*, *THAH*, *THAO* e *THAA1*) e um *linked gene* (*THAA2*) cujos produtos participam da biossíntese de thalianina em *A. thaliana* (LIU et al., 2020) (Figura 2). Este metabólito foi associado ao controle bioquímico que as plantas exercem sobre a microbiota da raiz (DEANE, 2019).



Fonte: LIU et al., 2020, com modificações.

Figura 2. *BGC* de thalianol que dirige a síntese de thalianina em *A. thaliana*. Os genes *THAS* (seta vermelha), *THAH* (seta azul claro), *THAO* (seta azul escuro), *THAA1* e *THAA2* (seta rosa) codificam as enzimas responsáveis pela via biossintética de thalianina em *A. thaliana*.

Os *BGC* envolvidos na produção de metabólitos especializados são amplamente descritos em diferentes grupos taxonômicos de bactérias (ALETI et al., 2019). Porém, estudos de metagenômica revelam ainda que diversos *BGC* identificados não estão associados a produção de nenhum metabólito conhecido (SUMI et al., 2015; ALETI et al., 2019). Os metabólitos especializados constituem fonte promissora para uma nova geração de antimicrobianos. A prospecção da existência de vias bioquímicas para produção de novos compostos são essenciais para auxiliar no combate à histórica e crescente onda de resistência

atual verificada na área clínica (URBAN-CHMIEL et al., 2022). Ademais, outras aplicações biotecnológicas podem ser desenvolvidas a partir da identificação de novos compostos.

2.2. Terpenos

Os terpenos configuram-se como uma classe destaque de metabólitos especializados formados por cadeias de unidades isoprênicas com tamanhos definidos (KELLER et al., 2005). Quando possuem oxigênio na estrutura são nomeadas terpenoides. Embora tenham sido reconhecidos inicialmente como metabólitos de origem vegetal, atualmente a produção também é descrita em diversos organismos, com destaque para bactérias, fungos e protozoários (QUIN et al., 2014; YAMADA et al., 2015). Em plantas os terpenos estão distribuídos em todos os tecidos, inclusive em frutos, onde são fontes de pigmentos e odores que protegem contra predação, além de atrair predadores e parasitas de herbívoros. Estes metabólitos possuem ainda ação sinalizadora e atuam como hormônios vegetais (YAMADA et al., 2015).

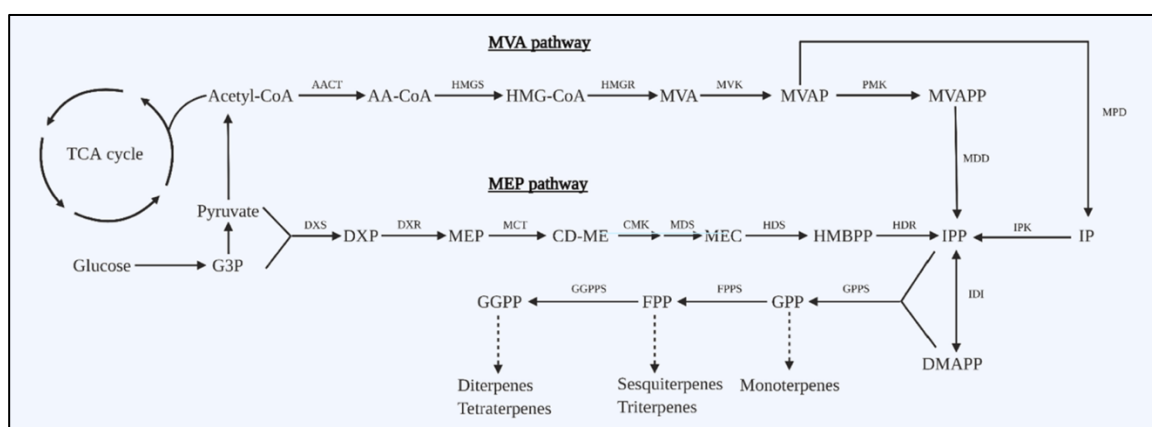
Em bactérias, terpenos como a albaflavenona e o germacreno D, ambos com atividade antimicrobiana, foram caracterizados (TYC et al., 2017). As funções biológicas e ecológicas de terpenos, porém, permanecem amplamente desconhecidas para procariotos, incluindo a geosmina, composto emitido por bactérias de solos responsável pelo cheiro de terra molhada. O gênero *Streptomyces* é certamente o mais bem investigado para produção de terpenos em bactérias, entretanto, outras espécies de solos têm sido caracterizadas como produtoras destes metabólitos especializados (SONG et al., 2015). Análises comparativas em genomas de linhagens de *Collimonas*, por exemplo, revela que algumas linhagens abrigam genes de terpeno sintases (TS) para produção de sesquiterpenos.

Em espécies de *Bacillus* e gêneros relacionados, ou Bafes, a produção de TS é conhecida. Alinhamentos entre sequências gênicas que codificam TS com enzimas homólogas encontradas nestas bactérias revelam altos níveis de conservação, o que indica enorme potencial para produção de terpenos (SATO et al., 2011). No entanto, em muitos casos os mecanismos catalíticos para produção dessas moléculas ainda são pouco elucidados, o que ainda torna a identificação de novos compostos um desafio (SATO et al., 2011).

A classificação de terpenos pode ser fundamentada no número de unidades de isopreno (C_5H_8) presentes na cadeia. Dessa maneira, encontramos na natureza: monoterpenos (com dez C ou C_{10}), sesquiterpenos (C_{15}), diterpenos (C_{20}), triterpenos (C_{30}) e tetraterpenos (C_{40}). Quando associadas à oxigênio, as estruturas são nomeadas monoterpenoides, sesquiterpenoides, diterpenoides, triterpenoides e tetraterpenoides, respectivamente. Alguns estudos ainda

apontam classes mais raras como os sesquiterpenos (C₃₅), identificados também em *Bacillus* spp. (SATO, 2013).

Os terpenos são derivados, principalmente, do isopentenil pirofosfato (IPP) e do isômero dimetilalil pirofosfato (DMAPP), decorrentes da via do mevalonato (MVA) ou da via metileritritol-fosfato (MEP) (ZHANG et al., 2016). Estas vias estão distribuídas de modo conjunto ou particular nos organismos produtores. Plantas, por exemplo, produzem terpenos pelas duas vias, fungos produzem pela via MVA, apenas, enquanto bactérias sintetizam terpenos unicamente pela via MEP (LIANG et al., 2021). Os diferentes tipos de terpenos são produzidos a partir de correspondentes intermediários gerados pela condensação de IPP e DMAPP. Estes produtos intermediários são o geranyl difosfato (GPP), farnesil difosfato (FPP) ou geranylgeranyl difosfato (GGPP) (Figura 3). Estes produtos são os substratos utilizados por TS para sintetizar a ampla variedade de terpenos encontrados na natureza (GUTBORD; ROMER; DORMANN, 2019; LIANG et al., 2021).



Fonte: LIANG et al., 2021.

Figura 3. Rotas bioquímicas das vias do mevalonato (MVA) e metileritritol-fosfato (MEP) para síntese de terpenos. As enzimas DXS, DXR, MCT, CMK, MDS, HDS e HDR catalisam reações sequenciais e indispensáveis para produção final de isopentenil pirofosfato (IPP) a partir de glicose na via MEP.

Pelo menos sete enzimas são necessárias para produção de IPP a partir do piruvato ou 3-fosfoglicerato (G3P) na via MEP (Figura 3). A ausência de qualquer uma dessas enzimas em bactérias pode comprometer a biossíntese de terpenos. As enzimas DXS, DXR, MCT, CMK, MDS, HDS e HDR catalisam reações sequenciais e indispensáveis para produção final de IPP. As informações sobre as enzimas, os substratos necessários e os produtos de cada uma delas estão destacados na tabela 1.

A enzima IPP isomerase (IDI) catalisa a conversão de IPP no isômero DMAPP. Unidades de IPP e DMAPP são condensadas por enzimas da família poliprenil sintase (GPPS, FPPS e GGPPS) e dão origem a GPP, FPP e GGPP (Figura 3). As enzimas dessa família são conhecidas pela capacidade de catalisar diferentes reações para síntese dos três precursores intermediários de terpenos (VATTEKKATTE et al., 2018). A FPPS, por exemplo, possui mecanismos básicos para sintetizar tanto FPP como GPP (SUN et al., 2011), o que torna possível a síntese de uma mesma molécula por catalisadores diferentes em vias bioquímicas produtoras de terpenos.

Tabela 1. Enzimas necessárias para produção de IPP pela via MEP para síntese de terpenos em bactérias

Reação*	Enzima	Substratos	Produto
2.2.1.7	1-Desoxi-D-xilulose 5-fosfato sintase (DXS)	Piruvato e G3P	1-Desoxi-D-xilulose 5-fosfato (DXP)
1.1.1.267	DXP redutoisomerase (DXR)	DXP e NADPH	2-C-metil-D-eritritol 4-fosfato (MEP)
2.7.7.60	MEP citidiltransferase (MCT)	MEP	4-difosfocitidil-2-C-metil-Deritritol (CD-ME)
2.7.1.148	CD-ME quinase (CMK)	CD-ME e ATP	4-difosfocitidil-2-C-metil-Deritritol 2-fosfato (CD-MEP)
4.6.1.12	2-C-metil-D-eritritol-2,4-ciclodifosfato sintase (MDS)	CD-MEP	2-C-metil-D-eritritol-2,4-ciclodifosfato (MEC)
1.17.7.3	4-hidroxi-3-metilbut-2-enil difosfato sintase (HDS)	MEC e NADPH	4-hidroxi-3-metilbut-2-enil difosfato (HMBPP)
1.17.7.4	HMBPP redutase (HDR)	HMBPP e NADPH	IPP

* Código da reação disponível no banco de dados KEGG (*Kyoto Encyclopedia of Genes and Genomes*).

2.3. Filo *Firmicutes*

Dentro do domínio *Bacteria*, *Firmicutes* compreende um dos filos com maior número de espécies. Quando proposto, o filo abrigava todas as bactérias Gram positivas (GIBBONS; MURRAY, 1978), alocadas em três classes denominadas *Bacilli*, *Clostridia* e *Mollicutes* (GARRITY et al., 2005). Baseado em análises de sequências de gene de rRNA 16S e características morfológicas como formação de endósporo (GALPERIN, 2013; DE VOS et al., 2009), o filo *Firmicutes* abriga atualmente sete classes – *Bacilli*, *Clostridia*, *Culicoidibacteria*, *Erysipelotrichia*, *Limnochordia*, *Negativicutes* e *Thermolithobacteria* – com 15 ordens e 54 famílias (www.bacterio.net; acessado em 01 de março de 2023).

Em virtude de características fenotípicas e filogenéticas distintas, a antiga classe *Mollicutes* foi promovida a filo. A família *Erysipelotrichaceae*, inicialmente alocada em

Mollicutes, originou uma nova classe do filo *Firmicutes*, denominada *Erysipelotrichia* (GALPERIN, 2013; DE VOS et al., 2009). A classe *Negativicutes*, por sua vez, abriga os membros Gram negativos, até então ausentes na descrição original de *Firmicutes* (GIBBONS; MURRAY, 1978; GALPERIN, 2013). Apesar da resposta negativa à coloração de Gram, os genomas de linhagens alocadas na classe *Negativicutes* apresentam sequências gênicas de marcadores moleculares do filo com níveis elevados de conservação (DE VOS et al., 2011; LUDWIG et al., 2011; GALPERIN, 2013).

O filo *Firmicutes* é fenotipicamente diverso, com células curvas, bastões, esféricas, helicoidais ou retas, podendo apresentar flagelo ou não. Quanto a utilização de oxigênio, podem ser aeróbias, facultativas e anaeróbias (DE VOS et al., 2009; GALPERIN, 2013). A maioria das espécies é mesófila, porém, o crescimento em ampla faixa de temperatura pode ser observado em algumas espécies, classificadas em psicotolerantes e/ou termotolerantes. Algumas espécies são psicrófilas ou termófilas. O crescimento ótimo *in vitro* da maior parte dos organismos ocorre em pH neutro, encontrando-se, porém, espécies de caráter acidófilo ou alcalifílico. Em geral, o genoma apresenta conteúdo de G+C < 50% (FRITZE, 2004; DE VOS et al., 2009; GALPERIN, 2013).

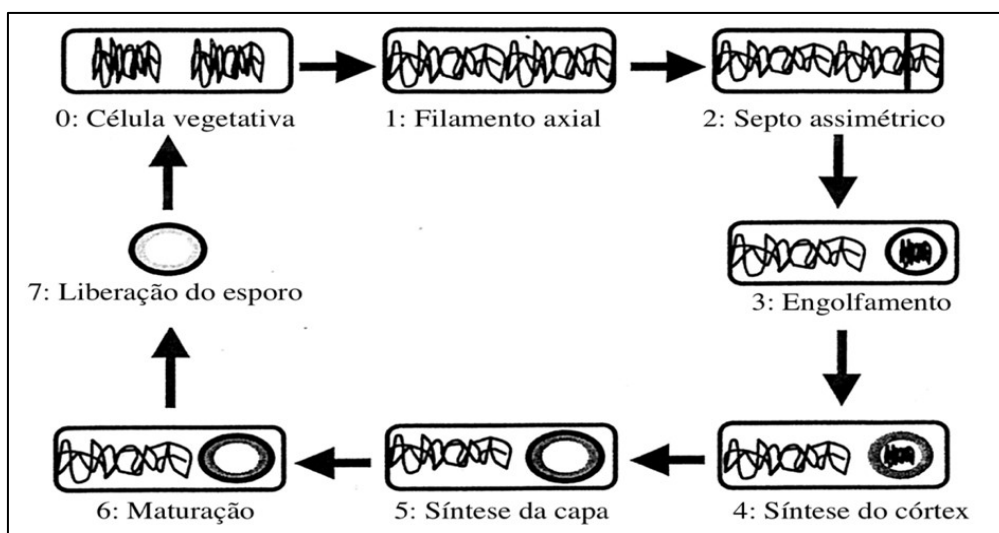
A habilidade de formar endósporos é uma característica marcante de membros do filo *Firmicutes*. Entretanto, apesar de observada em várias espécies, tanto aeróbias quanto anaeróbias, a endoesporulação não constitui uma característica universal do filo. Linhagens da classe *Thermolithobacteria* são asporogênicas. Em outras classes, espécies formadoras de endósporos coexistem com espécies não formadoras, sendo que um mesmo gênero pode incluir linhagens esporulantes ou não (FRITZE, 2004; DE VOS et al., 2009; LOGAN, 2009).

As espécies formadoras de endósporos compartilham um conjunto mínimo de genes homólogos envolvidos na esporulação (FRITZE, 2004; DE VOS et al., 2011; GALPERIN, 2013). Mutações nesses genes podem interromper o evento de endoesporulação em diversos estágios, diminuir o número de células capazes de esporular ou resultar em mutantes asporogênicos. É importante destacar que inexiste um gene ubíquo em genomas de espécies formadoras de endósporos e ausente em linhagens asporogênicas, que possa ser utilizado como marcador universal da esporulação (FRITZE, 2004; GALPERIN, 2013). Aparentemente, a habilidade de formar esporos foi perdida por algumas linhagens ao longo da evolução, sugerindo que a sobrevivência ambiental baseada na produção de um esporo viável tenha um custo elevado.

O esporo maduro e livre, célula resultante do processo de endoesporulação, apresenta metabolismo indetectável, sendo resistente a agentes químicos e físicos, tipicamente letais às

células vegetativas (DRIKS, 2004; GIORNO et al., 2007). A indução à endoesporulação acontece na transição da fase logarítmica para a fase estacionária, induzida por estresses ambientais como alta densidade populacional e, especialmente, depleção de nutrientes (MOIR, 2006; GIORNO et al., 2007).

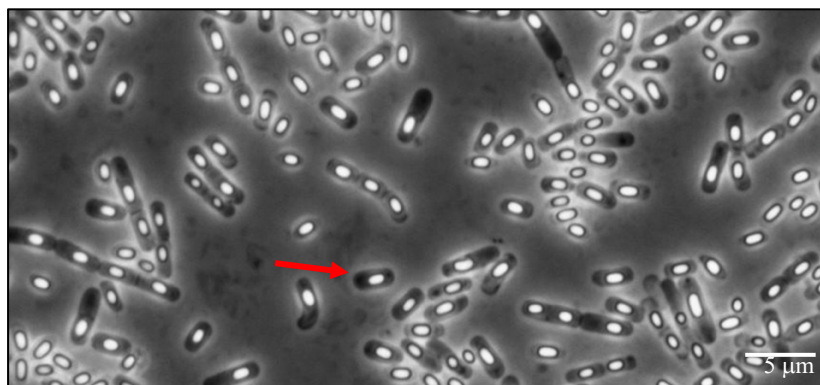
Didaticamente, a endoesporulação pode ser dividida em sete estágios, resumidos na figura 4 e fundamentado na espécie modelo *B. subtilis* (HILBERT; PIGGOT, 2004). Inicialmente ocorre a duplicação do genoma do esporângio ou célula-mãe (célula em diferenciação). O genoma duplicado estende-se pelo citoplasma com a formação do filamento axial (estágio I). No estágio II ocorre a formação do septo assimétrico, que dá origem a dois compartimentos celulares: célula-mãe (maior) e pré-esporo (menor). Neste estágio o evento torna-se irreversível. O estágio III é caracterizado pelo engolfamento do pré-esporo pelo envelope da célula-mãe. No estágio IV ocorre a síntese do córtex, seguido da formação de camadas internas e externas da capa (estágio V). No estágio V também ocorre incorporação de alta concentração de cálcio pelo endósporo e produção de dipicolinato de cálcio, que promovem desidratação da célula. A intensa produção de pequenas proteínas ácido solúveis (PPAS) que protegem o DNA e atuam como fonte de carbono durante a germinação, também é um evento deste estágio. Nos estágios VI e VII ocorre a maturação do endósporo, a degradação do envelope celular do esporângio e liberação do esporo maduro ou livre para o meio (HILBERT; PIGGOT, 2004).



Fonte: HILBERT; PIGGOT, 2004.

Figura 4. Os sete estágios para formação de um endósporo em Bafes (modelo *B. subtilis*). (1) Formação do filamento axial, (2) Formação do septo assimétrico, (3) Engolfamento do pré-esporo pelo envelope celular, Síntese do córtex (4) e capa (5) nas camadas externas no esporo, (6) Maturação e (7) Liberação do esporo após lise celular.

O esporo maduro livre, decorrente do evento de diferenciação celular, representa um mecanismo evolutivo de adaptação e sobrevivência. O esporo é uma estrutura refratável e visto na microscopia de contraste de fase com um brilho destacado (Figura 5). Os aspectos fenotípicos dos esporos e a localização no interior da célula-mãe são utilizados ainda como parâmetros para taxonomia em Bafes (DE VOS et al., 2009).



Fonte: INRAE, 2018.

Figura 5. Microscopia de contraste de fase de células de Bafes em esporulação. Endósporos (seta vermelha) no citoplasma do esporângio de uma espécie de Bafes.

Os esporos monitoram as condições ambientais por meio de receptores, e quando encontram uma situação favorável, a germinação se inicia (MOIR, 2006; CHRISTIE, 2012). Moléculas de baixa massa molecular tais como aminoácidos, açúcares, íons, dentre outras, denominadas agentes germinantes, desencadeiam uma cascata de reações bioquímicas e hidrolíticas levando a formação de uma nova célula vegetativa. No processo de germinação ocorre hidrólise de estruturas externas, aumento de pH citoplasmático e reidratação do citoplasma, que possibilita a retomada das atividades metabólicas da nova célula (SETLOW, 2008).

Estudos recentes de taxonomia propuseram a renomeação do filo *Firmicutes* para *Bacillota* (OREN; GARRITY, 2021; www.bacterio.net; acessado em 02 de março de 2023). Entretanto, em razão da ampla utilização na literatura científica e reconhecimento ao longo de décadas, no presente estudo manteremos o termo *Firmicutes* para se referir ao filo.

2.4. *Bactérias aeróbias formadoras de endósporos*

Espécies do gênero *Bacillus* e gêneros relacionados são coletivamente designadas bactérias aeróbias formadoras de endósporos ou Bafes (FRITZE, 2004; GALPERIN, 2013). As Bafes são ubíquas, sendo o solo o principal reservatório, onde desempenham papel essencial no funcionamento de ecossistemas e ciclagem de nutrientes (MANDIC-MULLER; PROSSER, 2011). O caráter heterotrófico promove importante papel no ciclo geoquímico do carbono, e as Bafes atuam também como desnitrificadoras, fixadoras e degradadoras de nitrogênio inorgânico e orgânico, respectivamente (CHENEBY et al., 2000). Outras propriedades bioquímicas de Bafes, incluem oxidação de enxofre, redução de manganês e metilação de mercúrio (CHENEBY et al., 2000).

As Bafes podem ser isoladas por métodos simples, como seleção de esporos por choque térmico ou agentes químicos (FRITZE, 2004; LOGAN et al., 2009). A diversidade metabólica de Bafes permite isolamento desses microrganismos a partir de substratos variados de origem terrestre e aquática, e, inclusive, de ambientes com características físico-químicas extremas, com temperaturas elevadas, alta salinidade, condições ácidas ou alcalinas, entre outros (LOGAN; DE VOS, 2009).

Os gêneros de destaque em biodiversidade e importância industrial de Bafes são *Bacillus* e *Lysinibacillus*, alocados na família *Bacillaceae*; e *Paenibacillus* e *Brevibacillus*, alocados na família *Paenibacillaceae* (FRITZE, 2004; DE VOS et al., 2011; GALPERIN, 2013). A endoesporulação é uma característica marcante desse grupo de microrganismos. As Bafes são de particular interesse em razão do alto potencial biotecnológico que apresentam, como produção de enzimas e outros bioprodutos de importância industrial, toxinas para controle de insetos pragas e veículos de doenças e nematoides, biorremediação, promoção de crescimento de plantas, entre outras aplicações (MANDIC-MULLER; PROSSER, 2011). A amplitude metabólica presente em Bafes também tem sido aproveitada para a produção de moléculas como riboflavina, estreptavidina e betalactamases (ZEIGLER; PERKINS, 2021). As Bafes também são produtoras de grande variedade de metabólitos especializados como antimicrobianos peptídicos de síntese ribossomal ou não (LARSSON, 2014; ZHAO; KUIPERS, 2016).

Embora a maioria das Bafes não seja patogênica, algumas espécies como *B. cereus* e *B. anthracis* representam riscos à saúde pública, por causarem quadros de intoxicação alimentar e o antraz, respectivamente (EHLING-SCHULZ; MESSELHAUSSER, 2013). A alta capacidade de dispersão e adesão de esporos, promove a presença de *B. cereus* em vários ambientes dedicados a produção de alimentos. Toxinas produzidas pelo *B. cereus* já foram encontradas em arroz, molhos, vegetais, carnes em natura e produtos lácteos (OSMAN et al., 2018; PARK et

al., 2018). Esse quadro de relevância clínica, causada por algumas espécies, amplia as propriedades e a importância de Bafes.

2.5. Taxonomia e filogenia de Bafes

Dentro do filo *Firmicutes*, as Bafes estão alocadas na classe *Bacilli*, ordem *Bacillales*, constituída de pelo menos 16 famílias (www.bacterio.net; acessado em 02 de março de 2023), sendo sete formadas por gêneros produtores de endósporos – *Alicyclobacillaceae*, *Bacillaceae*, *Paenibacillaceae*, *Pasteuriaceae*, *Planococcaceae*, *Sporolactobacillaceae* e *Thermoactinomycetacea* (FRITZE, 2004; GALPERIN, 2013). Incluso na família *Bacillaceae*, o gênero *Bacillus* abriga a maior quantidade de espécies de Bafes descritas, apresentando membros de alta diversidade estrutural e fisiológica (www.bacterio.net; acessado em 01 de março de 2023). O gênero foi descrito pela primeira vez em 1872 pelo botânico e microbiologista alemão Ferdinand Cohn (DE VOS et al., 2009). Desde então, o número de espécies do gênero sempre sofre alterações nos levantamentos realizados por pesquisadores (FRITZE, 2004; GALPERIN, 2013, EHLING-SCHULZ; LERECLUS; KOEHLER, 2019; www.bacterio.net; acessado em 01 de março de 2023).

A taxonomia dentro do gênero *Bacillus* é complexa e suscetível a divergências, o que torna desafio em muitas situações a discriminação de espécies relacionadas. Algumas espécies apresentam quase 100% de identidade nas sequências do gene rRNA 16S como *B. megaterium* e *B. aryabhatai* (Figura 6) (SHIVAJI et al., 2009). Além disso, linhagens de uma mesma espécie podem apresentar características bioquímicas distintas (MARTINS et al., 2023). Apesar destes desafios é reconhecido dentro do gênero a formação de, pelo menos, dois clados filogenéticos importantes: o grupo do *B. cereus* e complexo do *B. subtilis* (EHLING-SCHULZ; LERECLUS; KOEHLER, 2019).

O grupo do *B. cereus* ou *B. cereus sensu lato* (*s.l.*) compreende as espécies *B. cereus sensu stricto* (*s.s.*), *B. anthracis*, *B. thuringiensis*, *B. mycoides*, *B. pseudomycoides*, *B. toyonensis*, *B. weihenstephanensis* e *B. cytotoxicus* (EHLING-SCHULZ; LERECLUS; KOEHLER, 2019). Estas bactérias são classificadas com base em características fenotípicas como o potencial patogênico. Entretanto, este grupo constitui ainda uma subdivisão homogênea dentro do gênero como um todo (EHLING-SCHULZ; LERECLUS; KOEHLER, 2019). O complexo do *B. subtilis*, por sua vez, abriga o *B. subtilis s.s.*, modelo de estudo em bactérias Gram positivas para a compreensão do processo de endoesporulação (HILBERT; PIGGOT, 2004). O complexo do *B. subtilis* pode ser dividido em quatro clados. O clado I abriga *B. subtilis*

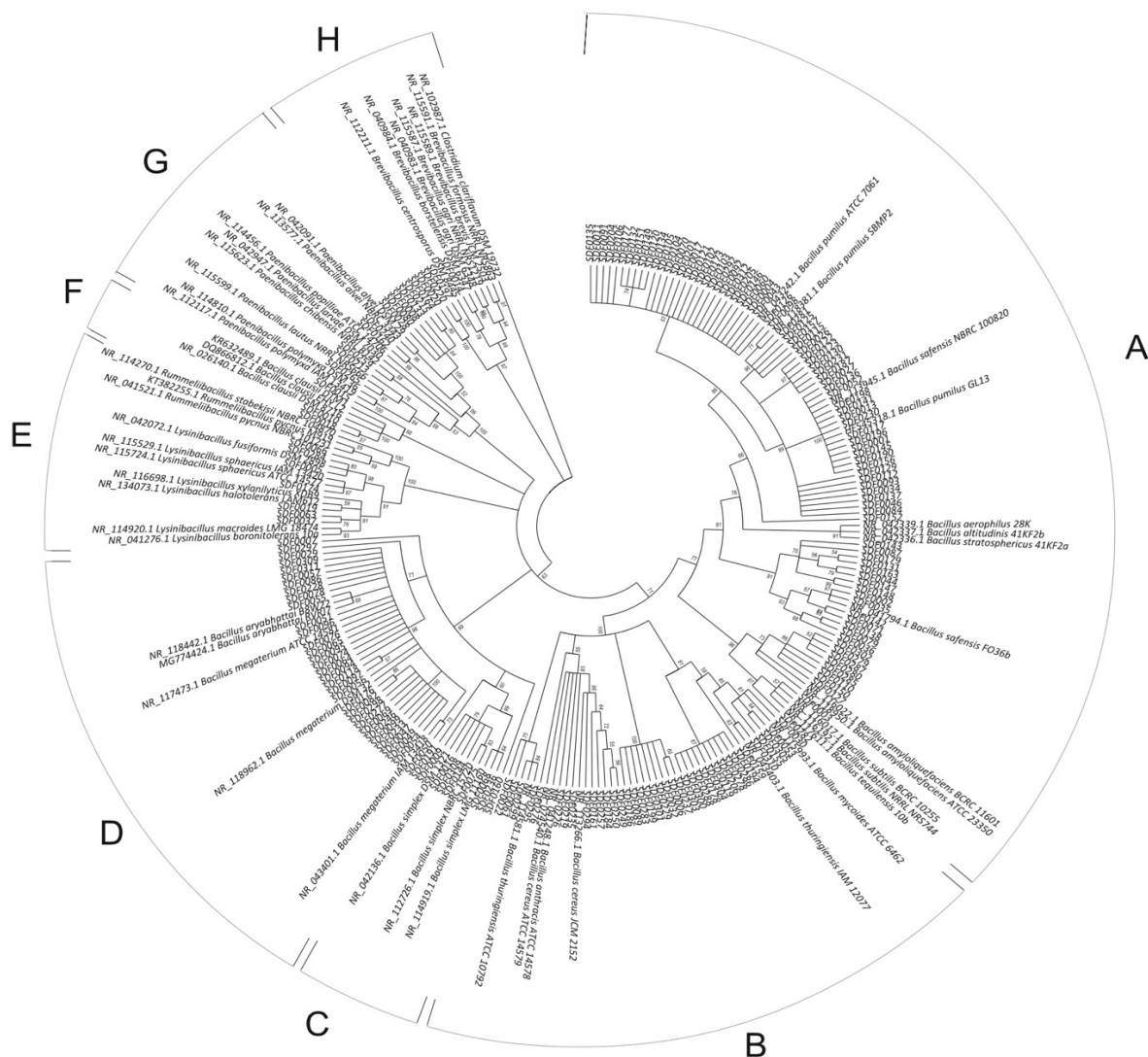
s.s, *B. tequilensis*, *B. vallismortis*, *B. mojavensis* e *B. atrophaeus*. O clado II compreende as espécies *B. siamensis*, *B. amyloliquefaciens*, *B. methylotrophicus* e *B. velezensis*. O clado III abriga *B. licheniformis*, *B. sonorensis* e espécies relacionadas. O clado IV inclui *B. pumilus*, *B. safensis*, *B. altitudinis*, *B. xiamenensis* e *B. aerophilus* (FRITZE, 2004; BHANDARI et al., 2013).

Embora as subcategorizações do complexo do *B. subtilis* representem padrões para as relações filogenéticas entre as espécies destacadas, as semelhanças genotípicas e fenotípicas entre os membros não permitem determinar sempre com ampla precisão as relações evolutivas. Esse fato possibilita sob certas circunstâncias acomodação de espécies em clados diferentes (EHLING-SCHULZ; LERECLUS; KOEHLER, 2019). Anteriormente, qualquer espécie de bactéria com a morfologia de bacilo, aeróbia ou anaeróbia facultativa e formadora de endósporos, era classificadas dentro do gênero *Bacillus* (ZEIGLER, 2013). Entretanto, essas características nem sempre são adequadas para agrupar espécies em um mesmo gênero. Com o avanço da Bioinformática e Biologia Molecular, desde a década de 1990, frequentemente, as Bafes são realocadas em famílias e gêneros diferentes.

Em 2007, estudos sobre a composição da parede celular, por exemplo, estabeleceram o gênero *Lysinibacillus* (AHMED et al., 2007), alocada inicialmente dentro de *Bacillus*. Estes microrganismos possuem lisina como terceiro resíduo de aminoácido na ligação entre as cadeias de peptideoglicano. Geralmente, o aminoácido encontrado nessa posição é o ácido diaminopimélico (AHMED et al., 2007). O gênero *Lysinibacillus* criado também pertence à família *Bacillaceae*, porém, outros estudos baseados em análises de sequências de genes rRNA 16S também realocaram membros desta família para novas famílias criadas para comportar as relações moleculares estabelecidas (PRIEST et al., 1988; ASH et al., 1991).

Assim, em 1993 ocorreu formalmente a criação da família *Paenibacillaceae* cujo gênero característico é o *Paenibacillus* (ASH et al., 1993). O advérbio *paene*, proveniente do latim, significa quase, ou seja, quase *Bacillus*. Os membros deste gênero são isolados de ambientes diversos e possuem grande relevância biotecnológica. A maioria é encontrado no solo, muitas vezes associada às raízes de plantas, o que os tornam suscetíveis para exploração na agricultura (GRADY et al., 2016). Até o presente, o número de *Paenibacillus* spp. é alvo de intensas alterações à medida que novas espécies são isoladas e descritas (GRADY et al., 2016). Dessa maneira, espera-se que o gênero sofra divisões taxonômicas no futuro (PRIEST et al., 1988; GRADY et al., 2016). Desde a criação da família *Paenibacillaceae*, novos gêneros foram criados para acomodar novas espécies ou realocar espécies conhecidas, como o *Brevibacillus* e outros seis gêneros (SHIDA et al., 1997; GRADY et al., 2016).

A fim de ampliar a compreensão sobre taxonomia e filogenia de Bafes, estudos realizados por nosso grupo de pesquisa (OREM et al., 2019), a partir de sequências de genes rRNA 16S de 192 linhagens SDF aponta oito grupos representativos para as relações filogenéticas das linhagens avaliadas (Figura 6).



Fonte: OREM et al., 2019.

Figura 6. Relações filogenéticas de 192 linhagens SDF depositadas na CBafes inferidas a partir de sequências de genes de rRNA 16S. O resultado aponta pelo menos oito agrupamentos filogenéticos. As letram indicam (A) Complexo do *B. subtilis*, (B) Grupo do *B. cereus*, (C) *B. simplex*, (D) Grupo do *B. megaterium*, (E) *Lysinibacillus*, (F) *B. clausii*, (G) *Paenibacillus* e (H) *Brevibacillus*.

O grupo A representa o complexo do *B. subtilis*, enquanto o grupo B aloca os membros do grupo do *B. cereus*. Ambos são próximos evolutivamente como citado anteriormente. Outros gêneros de *Bacillus* compreendem mais grupos relacionados como o formado por *B. simplex*

(grupo C) e *B. megaterium* / *B. aryabhatai* (grupo D) (SHIVAJI et al., 2009). O *B. clausii* (grupo F) neste estudo é mais próximo do gênero *Lysinibacillus* (Grupo E), sugerindo que *Bacillus* seja parafilético. O gênero *Paenibacillus* (grupo G) e *Brevibacillus* (grupo H), ambos pertencentes a família *Paenibacillaceae*, formam os grupos restantes.

2.6. Coleção de Bactérias aeróbias formadoras de endósporos

A Coleção de Bactérias aeróbias formadoras de endósporos (CBafes) da UnB/Instituto de Ciências Biológicas/Departamento de Biologia Celular está localizada no Laboratório de Microbiologia/LaBafes e é especializada em linhagens de interesse científico e com potencial biotecnológico para áreas da saúde, agricultura e produção de alimentos. Atualmente a Coleção possui um acervo de 312 linhagens isoladas de solo do DF, designadas SDF0001-SDF0312. As linhagens SDF depositadas na CBafes compreendem espécies bem conhecidas na literatura, como aquelas com descrição escassa. As linhagens SDF tem sido inspecionadas por uma abordagem polifásica, onde estão sendo reunidos dados fenotípicos e genotípicos (OREM et al., 2019; CAVALCANTE et al., 2019; MARTINS et al., 2020; MARTINS et al., 2023).

A importância da CBafes manifesta-se pela crescente demanda de armazenamento de amostras biológicas e disponibilização para a pesquisa científica. Coleções biológicas permitem que isolados sejam preservados adequadamente e estejam disponíveis para análises futuras, reanálise de dados e comparação entre resultados. As amostras depositadas na CBafes estão armazenadas em tiras de papel filtro em tubos de preservação a temperatura ambiente (CAVALCANTE et al., 2019; OREM et al., 2019). Quando requisitado, os esporos das linhagens de interesse são manipulados cuidadosamente para a elaboração de testes em estudos com as mais variadas finalidades.

2.7. Principais ferramentas computacionais utilizadas neste estudo

2.7.1. Plataforma antibiotics and secondary metabolite analysis shell

Tradicionalmente, metabólitos especializados têm sido identificados por abordagens clássicas de isolamento e análise. Recentemente, a capacidade de identificação desses metabólitos de origem natural tem sido expandida graças ao advento e à otimização de ferramentas computacionais. Assim, durante as últimas décadas, os avanços nas áreas de Biologia Molecular e Bioinformática abriram margem para estender a exploração de novas moléculas a partir da identificação e caracterização de vias biossintéticas de produtos naturais

em dados de genomas e metagenomas (BLIN, 2019). Dentre as plataformas públicas, a *antibiotics and secondary metabolite analysis shell* (antiSMASH) se destaca. Disponível a partir de 2011, essa ferramenta é usada por cientistas em todo o mundo para identificação de *BGC* codificadores de atividades envolvidas na biossíntese de metabólitos especializados em genomas de interesse (BLIN, 2019).

O antiSMASH detecta os *BGC* baseado em regras previamente definidas para identificar 45 tipos diferentes de vias biossintéticas de metabólitos especializados a partir da similaridade e da anotação das sequências biológicas (BLIN, 2019). A identificação desses *clusters* ocorre pela detecção de sequências genéticas que codificam as principais enzimas envolvidas na produção dessas moléculas. Para NRPS, PKS do tipo I, terpenos, lantipeptídeos, thiopeptídeos e outros grupos de peptídeos, o antiSMASH pode indicar o provável composto produzido pelo agrupamento de *BGC*. (BLIN, 2019). Os *clusters* são comparados frente a um banco de dados da própria plataforma, usando o algoritmo *Cluster Blast* integrado. Outro algoritmo denominado *Known Cluster Blast* compara o *cluster* de interesse com um conjunto de *BGC* selecionados manualmente do banco de dados MIBiG, do inglês *Minimum information about a biosynthetic gene cluster* (MEDEMA et al., 2015).

O antiSMASH fornece os resultados obtidos com base em sequências de interesse originárias de outros genomas e depositados na plataforma com acesso público, o que amplia a disponibilidade de análises comparativas. A versão atual do *software* (6.0) é resultado da atualização constante na última década (BLIN, 2021). As regras de detecção de *BGC* foram otimizadas ao longo do tempo, tornando a identificação da informação que codifica a estrutura modular de enzimas mais acessível e outras ferramentas de anotação de *clusters* mais aprimoradas.

As sequências gênicas acessadas pelo algoritmo *Cluster Blast* são baseadas nos registros contidos no banco de dados do antiSMASH (BLIN, 2021). Esse conjunto de dados é atualizado constantemente e atualmente inclui mais de 140.000 sequências de *BGC* encontrados em 388 genomas de espécies do domínio *Archaea*, 25.236 genomas de *Bacteria* e 177 do reino *Fungi*. Graças a estas habilidades, o antiSMASH fornece resultados relevantes em entradas de sequências genéticas desses microrganismos e eleva a capacidade de identificação, realçando a importância dos metabólitos especializados dessas espécies (BLIN, 2021).

2.7.2. *Molecular Evolutionary Genetics Analysis*

O *software Molecular Evolutionary Genetics Analysis* (MEGA) é projetado para realizar análises comparativas de sequências de genes homólogos ou famílias multigênicas de organismos de espécies similares bem como não relacionadas. A plataforma possibilita inferências de relações evolutivas a partir de sequências de nucleotídeos ou aminoácidos (KUMAR et al., 2018). Análises estatísticas e montagem de conjuntos de dados a partir de arquivos ou repositórios da *internet* ampliam o arcabouço de ferramentas com utilidade prática deste *software*. Além disso, as relações de parentesco entre os organismos são apresentadas visualmente pelo MEGA em forma de árvores filogenéticas interativas e matrizes de distância evolutiva (KUMAR et al., 2008).

2.7.3. *Pathways tools*

O aplicativo *Pathways tools* fornece dispositivos integrados para modelagem metabólica e análise de dados de alto rendimento. Esta ferramenta permite criação e o gerenciamento de bancos de dados de genes, proteínas, vias e reações metabólicas (KARP et al., 2021). Desde a década de 1990, o *Pathways tools* sofre atualizações, que permitem inferências em genomas anotados, incluindo reconstrução metabólica, criação de modelos de vias bioquímicas, análise de dados de expressão gênica, metabolômica e comparação de dados entre organismos de espécies diferentes (PALEY; KARP, 2006). Porém, a reconstrução metabólica continua sendo o recurso mais requisitado em estudos desenvolvidos com esta ferramenta, com mais de 4.000 bancos de dados gerados (KARP et al., 2021).

Os dados gerados pelo *Pathways tools* são executados e processados a partir de uma entrada de um genoma de interesse no formato GenBank ou GFF. As análises podem ser realizadas em formato *desktop* ou em servidores de *internet* (PALEY; KARP, 2006; KARP et al., 2021).

A reconstrução de vias metabólicas ocorre empregando o algoritmo *PathoLogic*, cujos resultados preveem as reações metabólicas catalisadas por enzimas presentes no genoma. As vias bioquímicas detectadas são visualizadas globalmente ou individualizadas (KARP et al., 2021). Este sistema de análise de dados é constantemente submetido a curadorias. Os genomas dos organismos modelos *Saccharomyces cerevisiae* e *Synechococcus elongatus*, por exemplo, são utilizados para atualizar periodicamente as informações experimentais do *software* e otimizar o uso do algoritmo de reconstrução metabólica (KARP et al., 2021).

3. JUSTIFICATIVA

Bactérias isoladas de solos sintetizam grande variedade de metabólitos especializados com amplo potencial biotecnológico. As Bafes, em particular, são conhecidas por produzirem metabólitos com alta variabilidade funcional e estrutural. Desse modo, estudos para prospectar o potencial de produção de metabólitos especializados em Bafes auxilia na exploração de novos compostos, principalmente com o avanço da Bioinformática. O metabolismo especializado permite também compreender a importância funcional e biológica destas moléculas para estes microrganismos. Os terpenos constituem um importante grupo de metabólitos especializados para ampliar o arsenal de novas estruturas identificadas. Embora a produção de terpenos seja reportada em bactérias, as funções biológicas e o caráter evolutivo destas moléculas ainda são amplamente desconhecidas para os domínios procariotos em geral, o que torna a investigação do potencial de produção de terpenos em Bafes pertinente. Os genomas das linhagens SDF aqui avaliadas estão depositadas na CBafes da UnB. Este estudo permite ampliar o conhecimento das linhagens escolhidas e ampliar o potencial biotecnológico de nossa Coleção para os mais variados fins.

4. OBJETIVOS

4.1. *Objetivo geral*

Analisar *in silico* o potencial para produção de terpenos em dez genomas de linhagens de Bafes (linhagens SDF) depositadas na CBafes da UnB.

4.2. *Objetivos específicos para as linhagens SDF avaliadas*

- Verificar o potencial genômico global para produção de metabólitos especializados;
- Comparar o potencial genômico de produção de metabólitos entre as linhagens;
- Identificar *clusters* gênicos possivelmente envolvidos na biossíntese de terpenos;
- Investigar a presença de enzimas precursoras para produção de terpenos da via MEP e da família poliprenil sintase;
- Avaliar a produção de moléculas de terpenos a partir de TS identificadas;
- Estabelecer as relações filogenéticas a partir da sequência de aminoácidos das enzimas de TS identificadas;
- Comparar a similaridade do conjunto enzimático detectado envolvidos na produção de terpenos entre as linhagens.

5. METODOLOGIA

5.1. Linhagens SDF

As permissões específicas requeridas para a coleta e isolamento de linhagens SDF usadas nesse estudo foram endossadas pelo CNPq por meio da Autorização de Acesso e Amostra do Patrimônio Genético número **010439/2015-3**. O isolamento não envolveu espécies ameaçadas ou protegidas. As linhagens SDF utilizadas (Tabela 2) neste estudo foram previamente isoladas do solo do Distrito Federal e estocadas como descrito em Orem et al., (2019) e Cavalcante et al., (2019). O DNA total das linhagens SDF utilizados neste estudo foram sequenciados na plataforma IlluminaMiseq e os genomas depositados na plataforma NCBI (<https://www.ncbi.nlm.nih.gov>) (Tabela 2).

Tabela 2. Linhagens SDF utilizadas para análise de produção de metabólitos especializados neste estudo

Linhagem SDF	Número de acesso NCBI	Nível de acesso
<i>Lysinibacillus fusiformis</i> SDF0005	SAMN12262387	restrito
<i>Bacillus pumilus</i> SDF0011	SAMN12262389	restrito
<i>Bacillus oleronius</i> SDF0015	SAMN12262388	restrito
<i>Bacillus simplex</i> SDF0024	SAMN12262386	restrito
<i>Bacillus velezensis</i> SDF0141	SAMN12262409	restrito
<i>Bacillus velezensis</i> SDF0150	SAMN12262428	restrito
<i>Bacillus</i> sp. SDF0016	SAMN06921004	público
<i>Paenibacillus</i> sp. SDF0028	SAMN06917517	público
<i>Lysinibacillus</i> sp. SDF0037	SAMN06921005	público
<i>Lysinibacillus</i> sp. SDF0063	SAMN06917521	público

5.2. Identificação de BGC envolvidos na síntese de metabólitos especializados

Para identificar potenciais BGC envolvidos na síntese de metabólitos especializados foi instalado o *software* antiSMASH 6.0 *bacterial standalone version*, otimizada para uso de sequência de procariotos (<https://antismash.secondarymetabolites.org/#!/start>). O *software* foi configurado dentro do sistema operacional Ubuntu, distribuição Linux na versão 22.04.2 LTS (<https://ubuntu.com/download/desktop>). O *software* e as dependências auxiliares da ferramenta necessárias para identificação de BGC foram instaladas utilizando os *scripts* a seguir:

```
sudo apt-get update
sudo apt-get install -y apt-transport-https
sudo wget http://dl.secondarymetabolites.org/antismash-stretch.list -O /etc/apt/sources.list.d/antismash.list
sudo wget -q -O- http://dl.secondarymetabolites.org/antismash.asc | sudo apt-key add -
sudo apt-get update
```

```
sudo apt-get install hmmer2 hmmer diamond-aligner fasttree prodigal ncbi-blast+ muscle glimmerhmm
```

As sequências genômicas das dez linhagens de interesse (Tabela 2) foram baixados diretamente da plataforma NCBI para as análises. Foram utilizados os arquivos em formato GenBank (.gbk) como *input*. O parâmetro utilizado para a acurácia de detecção dos *BGC* no antiSMASH foi o completo relaxado (*Full featured-run*), com os algoritmos disponíveis ativados (*KnownClusterBlast*, *ClusterBlast*, *ClusterPfam*, *ActiveSiteFinder* e *Pfam-based GO term annotation*). A análise foi efetuada no Laboratório de Biologia Molecular do Departamento de Biologia Celular da UnB, via SSH – ferramenta para conexão de servidores remotos. O *script* a seguir foi utilizado para obtenção dos *BGC* de cada linhagem, onde X é o número da SDF testada.

```
antismash SDF(X).gbk | --output-dir SDF(X)_results --taxon bacteria --genefinding-tool prodigal --cb-general --knownclusters --cb-subclusters --asf --pfam2go --smcog-trees
```

Os *BGC* identificados pelo antiSMASH foram mensurados e classificados quanto aos tipos (<https://docs.antismash.secondarymetabolites.org/glossary/#clustertypes>; acessado em 04 de Julho de 2023 (Quadro 1 – Apêndice)), possíveis metabólitos produzidos por *cluster* e índice de similaridade (0-100%). Os dados encontrados foram tabulados por cada linhagem. Potencias *BGC* envolvidos na biossíntese de terpenos foram curados e os genes de TS presentes em cada *cluster* identificados.

5.3. Reconstrução metabólica para detecção dos genes de enzimas da via MEP

A detecção dos genes das enzimas da via MEP (DXS, DXR, MCT, CMK, MDS, HDS e HDR) foi realizada com o uso do *software Pathways tools*. A instalação da ferramenta ocorreu após obtenção de licença acadêmica para *download*. As análises foram realizadas utilizando o *Pathways Tools* versão 26.5 (<http://bioinformatics.ai.sri.com/ptools/>) em formato *desktop*. Foi utilizado o algoritmo *PathoLogic* para reconstrução metabólica global *in silico* de cada linhagem SDF. Os genomas das dez linhagens de interesse (Tabela 2) foram baixados diretamente da plataforma NCBI para as análises. Foram utilizados os arquivos em formato GenBank (.gbk) como *input*. O banco de dados gerado das vias bioquímicas de cada linhagem foi feito com o *score cutoff* de 0.15 e os parâmetros *Transport Inference Parser*, *Pathway Hole Filler*, *Operon Predictor*, *Protein Complex Predictor* ativados. Ao final da análise foi gerado o diagrama de visualização das vias bioquímicas identificadas.

5.4. Detecção de enzimas da família poliprenil sintase

A análise para detectar a presença das enzimas GPPS, FPPS e GGPPS nas linhagens SDF de interesse foi realizada na plataforma BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) em servidores de *internet*. A detecção das atividades desempenhadas pelas respectivas enzimas foi realizada perante a construção de um banco de dados com 6.146 arquivos de sequências de aminoácidos em formato FASTA, obtido a partir de sequências depositadas no NCBI, para síntese de poliprenil sintases isolados da família *Bacillaceae*. A extração das sequências de proteínas de cada genoma (.gbk) foi realizado utilizando o *script* abaixo (BOGDANOVE et al., 2011), onde X é o número de cada linhagem SDF testada. O alinhamento para comparação entre as sequências de aminoácidos foi realizado empregando o pBLAST e as enzimas detectadas com maior percentual de similaridade foram mensurados.

```
perl 01_retrieveProteinSeqFromGB.pl -g SDFX.gbk -o SDFX.fasta
```

5.5. Construção de árvores filogenéticas

As sequências de aminoácidos de uma mesma TS identificada nas linhagens SDF foram extraídas do arquivo *regions.js* (Java Script), disponibilizado como *output* dos resultados da busca de *BGC* no antiSMASH. A extração ocorreu via *script in-house* em linguagem de programação *Biopython* (<http://biopython.org/DIST/docs/tutorial/Tutorial.html>) descrito abaixo:

```
def getFile(cluster_name, lista_interesse):
    with open(f"{cluster_name}.fasta", "w") as sdf_file:
        for path in Path().resolve().glob("*.gbk"):
            for sequencia in SeqIO.parse(path.as_posix(), "genbank"):
                for feature in sequencia.features:
                    if feature.type == "CDS":
                        if feature.qualifiers['locus_tag'][0] in lista_interesse:
                            sdf_id = f">{path.stem}|{feature.qualifiers['protein_id'][0]}|{feature.qualifiers['product'][0]}"
                            sdf_sq = str(feature.qualifiers['translation'][0])
                            sdf_file.write(sdf_id)
                            sdf_file.write("\n")
                            sdf_file.write(sdf_sq)
                            sdf_file.write("\n")
                    sdf_file.close()

lista_interesse_terpene = ["FOA25_16900", "FOA26_17055", "C7Y46_10735", "FOA23_09685", "FOA20_12850",
"FOA22_00740", "C7Y47_13875", "C7Y45_10095"] getFile("terpene", lista_interesse_terpene_SHC)

lista_interesse_terpene = ["C7Y46_00835", "FOA23_17590", "FOA25_09515", "FOA26_04920", "FOA20_02380",
"C7Y47_07685", "FOA21_15700"] getFile("terpene", lista_interesse_terpene_FES)
```

As sequências de aminoácidos obtidas em formato FASTA foram alinhadas utilizando o *software* MEGA versão 11 (<https://www.megasoftware.net>), previamente instalada e configurada em formato *desktop*. O algoritmo de alinhamento empregado foi o MUSCLE com os parâmetros padrões. O arquivo gerado (.mas) foi utilizado para a construção das árvores filogenéticas pelo método estatístico de máxima verossimilhança.

5.6. Similaridade do conjunto enzimático para produção de terpenos

As enzimas responsáveis pela biossíntese de terpenos detectadas foram organizadas em *heatmaps* com auxílio do *software* R para revelar visualmente a similaridade do potencial de promoção de terpenos entre as linhagens SDF. Foi utilizado os valores dicotômicos 0 (negativo) e 1 (positivo) como variáveis binárias para representar a associação entre as linhagens a partir da detecção ou ausência das enzimas. Usando a correlação de Pearson, as linhagens foram agrupadas pela similaridade de resultados. O fluxograma abaixo resume todas as etapas realizadas no desenvolvimento deste estudo (Figura 7).

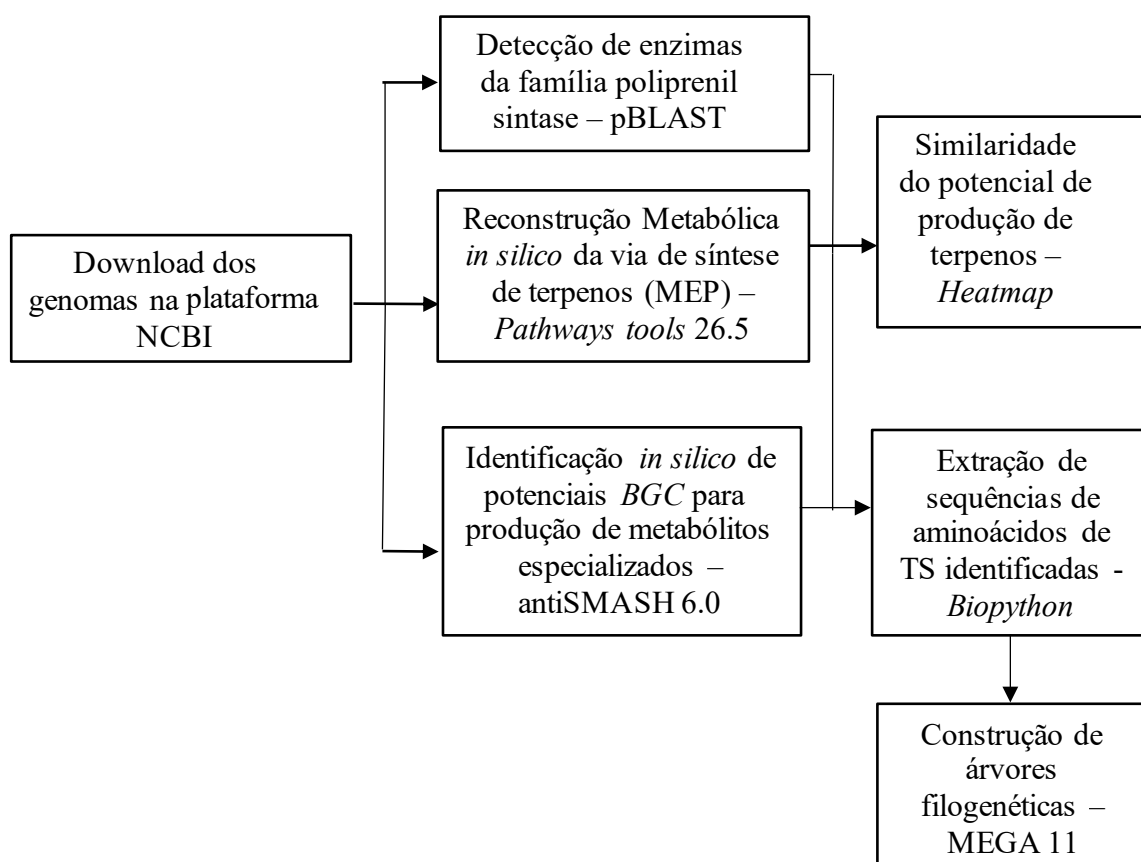


Figura 7. Resumo das principais atividades efetuadas para realização deste estudo

6. RESULTADOS

A análise efetivada no genoma das dez linhagens SDF sondadas (Tabela 2) utilizando o antiSMASH detectou prováveis 153 *BGC* envolvidos na síntese de, pelo menos, 20 classes ou subclasses de metabólitos especializados (Figura 8). A classe representativa com o maior número de *BGC* identificados foi aquela que codifica a síntese de NRPS, com 47 (30,7%) *clusters* reconhecidos (Figura 8). Outras classes de metabólitos especializados detectados com número de *BGC* significativos compreenderam a informação para expressão de policetídeo sintase acil-transferase (trans AT-PKS) sendo 17 ou 11,1%; terpenos (16; 10,5%), policetídeo sintase tipo III (T3PKS tipo III) (12; 7,8%) e betalactonas (10; 6,5%) (Figura 8). A demais quantidades e tipos de *BGC* envolvidos na síntese de metabólitos especializados encontradas nas linhagens SDF de interesse (Tabela 2) também estão descritas na figura 8.

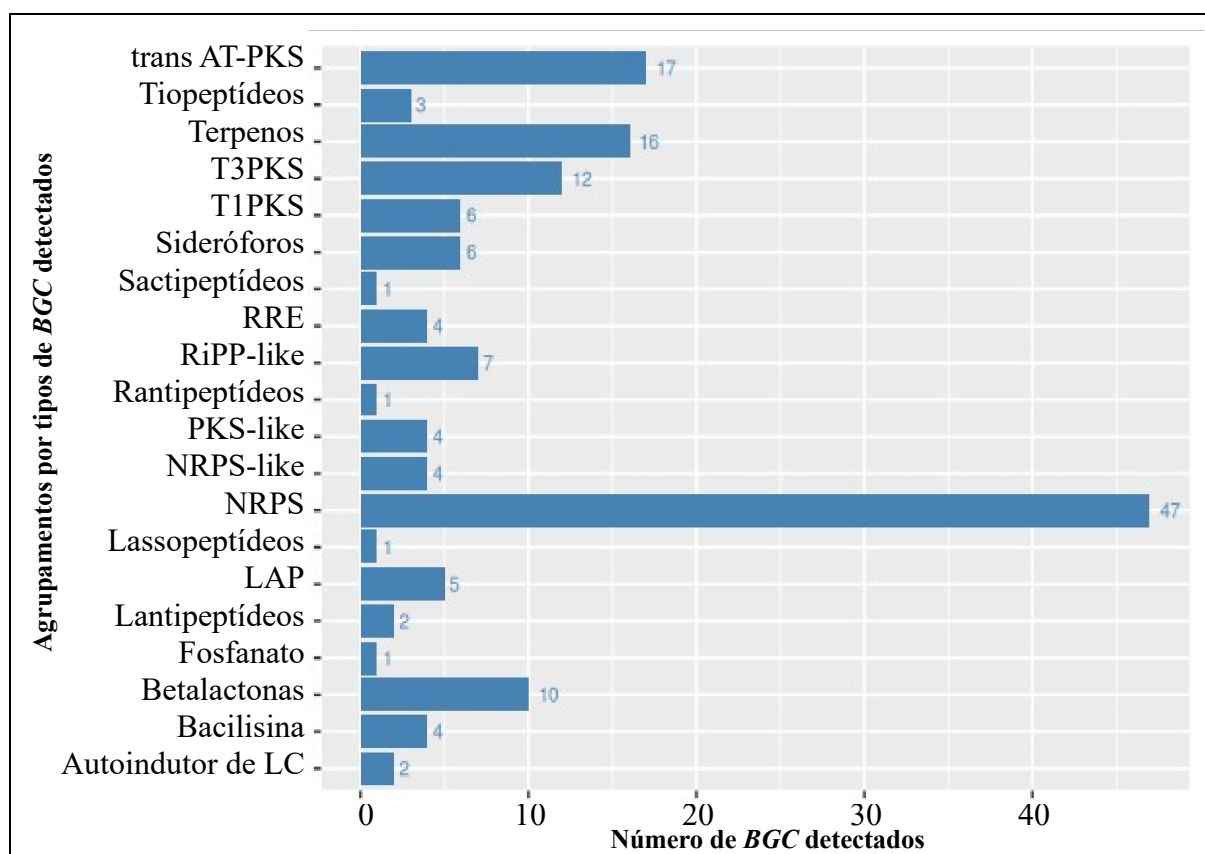


Figura 8. Total de *BGC* relacionados à síntese dos metabólitos especializados identificados neste estudo pelo antiSMASH. A classe representativa com o maior número de *BGC* identificados foi aquela que codifica a síntese de NRPS, com 47 *BGC* reconhecidos, seguidos de trans ATPKS (17), terpenos (16), T3PKS (12), Betalactonas (10), RiPP-like (7), T1PKS (6), Sideróforos (6), LAP (5), RRE (4), PKS like (4), NRPS like (4), Bacilisina (4), Tiopeptídeos (3), Lantipeptídeos (2), Autoindutor de LC (2), Sactipeptídeos (1), Rantipeptídeos (1), Lasso-peptídeos (1), Fosfanato (1).

A linhagem SDF que apresentou a maior quantidade de prováveis *BGC* detectados foi *Paenibacillus* sp. SDF0028 com 38 (24,8%) *clusters* reconhecidos, seguida pelas duas linhagens da espécie *B. velezensis* (SDF0150 e SDF0141) com 22 (14,3%) e 21 (13,7%) *clusters*, respectivamente. Foram detectados 19 (12,4%) *BGC* no genoma da linhagem *Lysinibacillus* sp. SDF0063 e 16 (10,4%) em *B. pumilus* SDF011. As demais linhagens SDF apresentaram: *B. simplex* SDF0024 (10; 6,5%), *Bacillus* sp. SDF0016 (10; 6,5%), *Lysinibacillus* sp. SDF0037 (9; 5,8%) e *L. fusiformis* SDF0005 (5; 3,2%). A linhagem *B. oleronius* SDF0015 apresentou a menor quantidade de *BGC* entre as dez linhagens avaliadas, com 3 (1,9%) *clusters* detectados. Constatou-se grande variedade de tipos de *BGC* envolvidos na síntese de metabólitos especializados, inclusive, quando observado em uma mesma linhagem investigada (Figura 9). Prováveis *BGC* relacionados com a síntese de PK por enzimas PKS do tipo III e terpenos foram encontradas em todas as linhagens SDF (Figura 9).

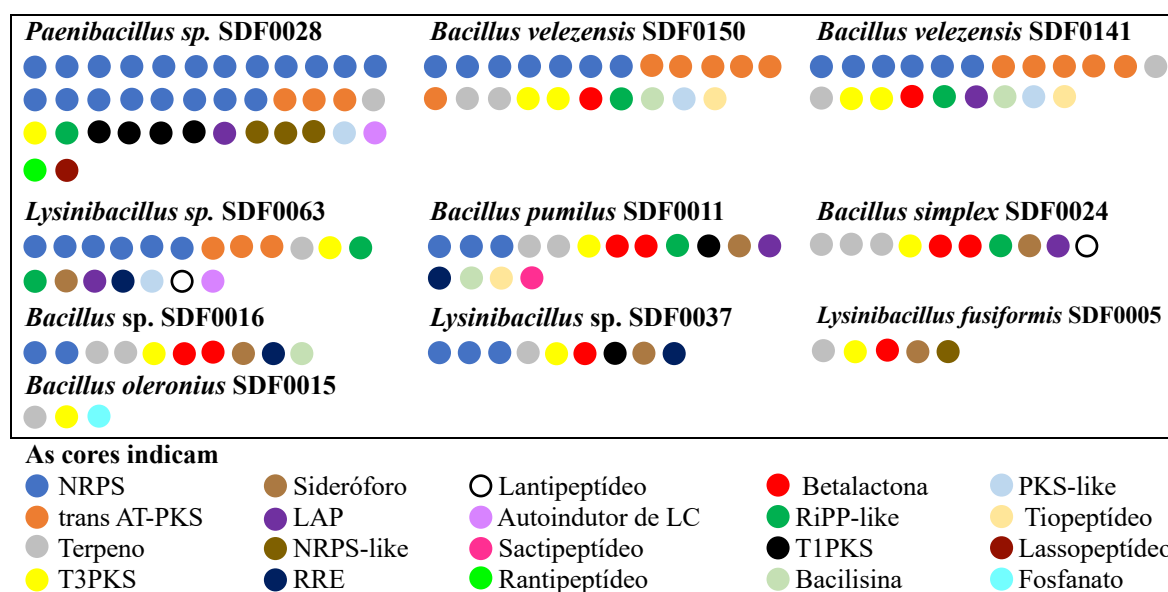


Figura 9. Classes de metabólito especializado identificados nas linhagens SDF utilizadas neste estudo. As cores que representam as classes estão identificadas na parte inferior da figura. A linhagem *Paenibacillus* sp. SDF0028 teve 38 *BGC* reconhecidos, seguido de *B. velezensis* SDF0150 (22), *B. velezensis* SDF0141 (21), *Lysinibacillus* sp. SDF0063 (19), *B. pumilus* SDF011 (16), *B. simplex* SDF0024 (10), *Bacillus* sp. SDF0016 (10), *Lysinibacillus* sp. SDF0037 (9), *L. fusiformis* SDF0005 (5) e *B. oleronius* SDF0015 (3).

Partindo dos prováveis 153 *BGC* reconhecidos, o antiSMASH identificou em 63 (41,1%) ocorrências o potencial metabólito especializado produzido por um *cluster* em particular. Frente a sequência do *BGC* envolvida na produção do respectivo metabólito depositada no banco de

dados MIBiG, o índice de similaridade entre as sequências que dirigem a síntese dos compostos reconhecidos variou de 0-100% (Tabela 3). O índice de similaridade foi baseado na quantidade de genes presentes no genoma avaliado, como revelado pelo algoritmo *KnwonClusterBlast*. Entre os 63 metabólitos reconhecidos, em 10 (6,5%) situações, o índice de similaridade foi de 100% e os demais 53 (34,6%), apresentaram índice <90% (Tabela 3).

Tabela 3. Metabólitos especializados identificados pelo antiSMASH nos dez genomas analisados neste estudo¹

Linagem	Metabólito identificado	I.S (%) ²	Classe BGC	MIBiG ³
<i>Lysinibacillus fusiformis</i> SDF0005	Fengicina	46	Betalactona	BGC0001095
	Petrobactina	33	Sideróforo	BGC0000942
	Canamicina	4	NRPS-Like	BGC0000082
<i>Bacillus pumilus</i> SDF0011	Lichenicina	85	NRPS	BGC0000381
	<i>Sporulation killing factor</i>	85	Sactipeptídeo	BGC0000601
	Bacilisina	85	Outro	BGC0001184
	Fengicina	53	Betalactona	BGC0001095
	Bacillibactina	53	NRPS	BGC0000309
	Zwittermicina	18	NRPS	BGC0001059
<i>Bacillus oleronius</i> SDF0015	-	-	-	-
<i>Bacillus</i> sp. SDF0016	Bacilisina	85	Outro	BGC0001184
	Lichenisina	85	NRPS	BGC0000381
	Bacillibactina	53	NRPS	BGC0000309
	Fengicina	53	Betalactona	BGC0001095
	Botromicina A2	6	Betalactona	BGC0000469
<i>Bacillus simplex</i> SDF0024	Cerecidina	64	Lantipeptídeo	BGC0000502
	Fengicina	40	Betalactona	BGC0001095
<i>Paenibacillus</i> sp. SDF0028	Paenibacterina	100	NRPS	BGC0000400
	Nostamida A	85	NRPS	BGC0001479
	Bacillibactina	53	NRPS	BGC0000309
	Pelgipeptina	37	NRPS	BGC0000403
	Locillomicina	14	NRPS	BGC0001005
	Zwittermicina A	14	NRPS	BGC0001059
	Paenilipoheptim	11	NRPS	BGC0001728
	Auritinin	9	-	BGC0001520
	Butirosina	7	-	BGC0000693
	Zwittermicina A	7	NRPS	BGC0001059
Iersiniabactina	5	NRPS	BGC0001055	
<i>Lysinibacillus</i> sp. SDF0037	Fengicina	46	Betalactona	BGC0001095
	Pellasoren	33	NRPS	BGC0001034
	Petrobactina	33	Sideróforo	BGC0000942
<i>Lysinibacillus</i> sp. SDF0063	Macrobrevina	100	TransAT-PKS	BGC0001470
	Gramicidina	90	NRPS	BGC0000367
	Petrobactina	83	Sideróforo	BGC0000942
	Tirocidina	81	NRPS	BGC0000452
	Aurantina	25	TransAT-PKS	BGC0001520
	Pacidamicina	13	NRPS	BGC0000951
	Octapeptina C4	11	-	BGC0001715
	Zwittermicina A	11	NRPS	BGC0001059

<i>Bacillus velezensis</i> SDF0141	Bacillibactina	100	NRPS	BGC0000309	
	Bacilisina	100	Outro	BGC0001184	
	Macrolactina H	100	TransAT-PKS	BGC0000181	
	Bacillaeno	100	PKS/NRPS	BGC0001089	
	Fengicina	86	Betalactona	BGC0001095	
	Surfactina	82	NRPS	BGC0000433	
	Difficidina	66	TransAT-PKS	BGC0000176	
	Difficidina	46	TransAT-PKS	BGC0000176	
	Pliplastatin	30	NRPS	BGC0000407	
	Pliplastatin	23	NRPS	BGC0000407	
	Butirosina A/B	7	PKS-like	BGC0000693	
	<i>Bacillus velezensis</i> SDF0150	Bacillibactina	100	NRPS	BGC0000309
		Bacilisina	100	Outro	BGC0001184
Macrolactina H		100	TransAT-PKS	BGC0000181	
Bacillaeno		100	PKS/NRPS	BGC0001089	
Difficidina		53	TransAT-PKS	BGC0000176	
Surfactina		52	NRPS	BGC0000433	
Surfactina		47	NRPS	BGC0000433	
Difficidina		46	TransAT-PKS	BGC0000176	
Plipastatin		30	NRPS	BGC0000407	
Difficidina		26	TransAT-PKS	BGC0000176	
Fengicina		26	Betalactonas	BGC0001095	
Butirosina A/B		7	PKS-like	BGC0000693	

¹Metabólitos especializados identificados como terpenos não estão contabilizados. ²Índice de similaridade. ³Número de acesso do banco de dado para o respectivo metabólito.

As sequências de *clusters* envolvidos na síntese dos metabólitos especializados que apresentaram índice de similaridade de 100% com base na biblioteca do banco de dados MIBiG foram a bacillibactina (Figura 10), bacilisina (Figura 11), macrolactina H (Figura 12) e bacillaeno (Figura 13), presentes nas duas linhagens *B. velezensis* SDF0141 e *B. velezensis* SDF0150. A paenibacterina (Figura 14), presente na linhagem *Paenibacillus* sp. SDF0028, assim como a macrobrevina (Figura 15), produzida pela linhagem *Lysinibacillus* sp. SDF0063 também obtiveram 100% de similaridade na análise efetuada pelo antiSMASH.

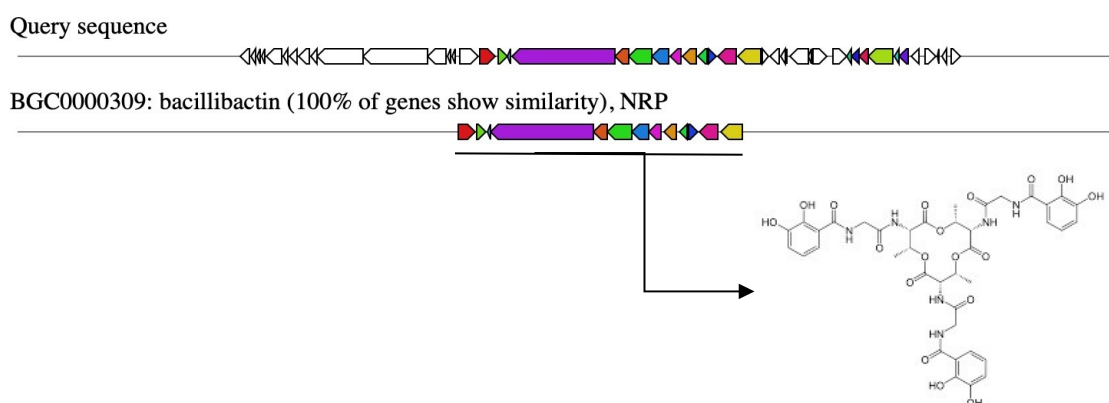


Figura 10. BGC de bacillibactina. BGC que codifica atividades responsáveis pela síntese de bacillibactina detectado nas linhagens *B. velezensis* (SDF0141 e SDF0150) com índice de similaridade de 100%.

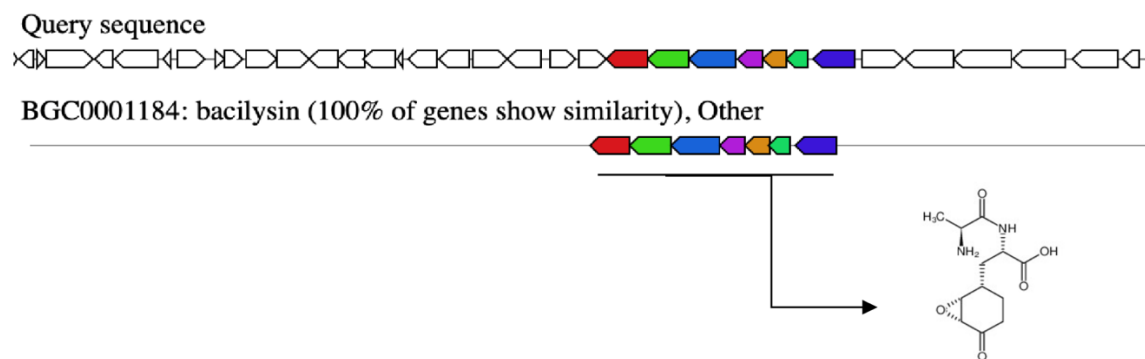


Figura 11. BGC de bacilisina. BGC que codifica atividades responsáveis pela síntese de bacilisina detectado nas duas linhagens *B. velezensis* (SDF0141 e SDF0150) com índice de similaridade de 100%.

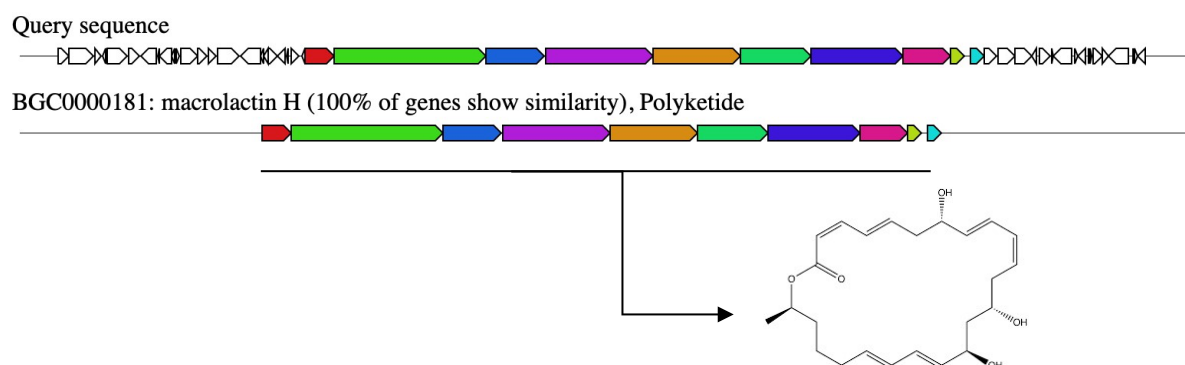


Figura 12. BGC de macrolactina H. BGC que codifica atividades responsáveis pela síntese de macrolactina detectado nas duas linhagens *B. velezensis* (SDF0141 e SDF0150) com índice de similaridade de 100%.

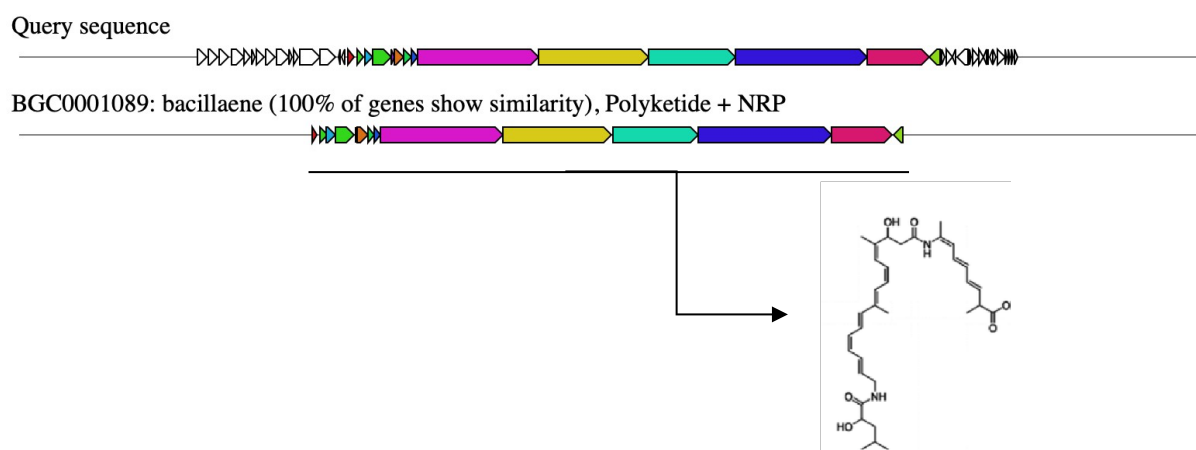


Figura 13. BGC de bacillaeno. BGC que codifica atividades responsáveis pela síntese de bacillaeno detectado nas duas linhagens *B. velezensis* (SDF0141 e SDF0150) com índice de similaridade de 100%.

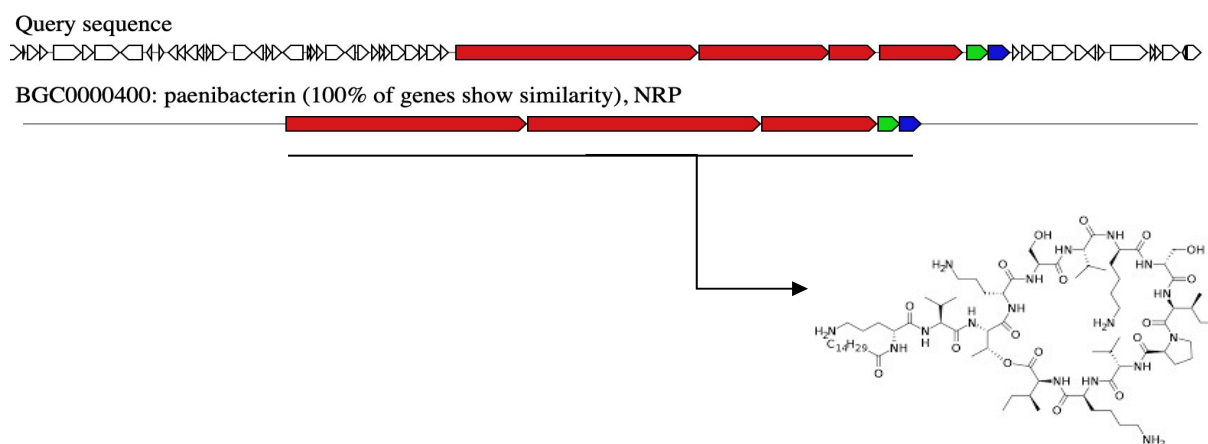


Figura 14. BGC de paenibacterina. BGC que codifica atividades responsáveis pela síntese de paenibacterina detectado na linhagem *Paenibacillus* sp. SDF0028 com índice de similaridade de 100%.

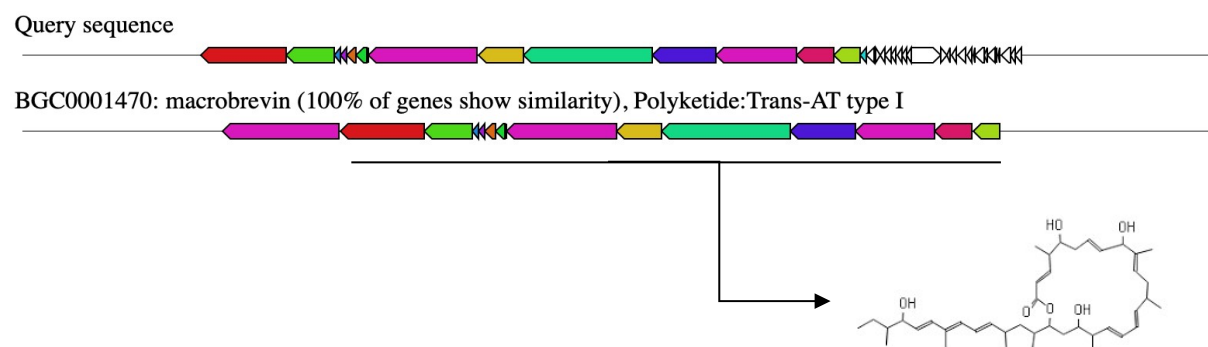


Figura 15. BGC de macrobrevin. BGC que codifica atividades responsáveis pela síntese de macrobrevin detectado na linhagem *Lysinibacillus* sp. SDF0063 com índice de similaridade de 100%.

Embora o índice de similaridade obtido tenha sido menor que 100%, os BGC envolvidos na biossíntese de fengicina foram detectados em pelo menos sete linhagens (Tabela 3). Neste caso, o maior índice (85%) foi encontrado em *B. velezensis* SDF0141. A bacillibactina com 100% de similaridade, detectado em duas linhagens de *B. velezensis* (SDF0141 e SDF0150), também foram encontrados em outras três linhagens SDF (Tabela 3). Ambas as moléculas estavam representadas pelo maior número de BGC que dirigem as respectivas sínteses nas linhagens avaliadas.

Entre os prováveis 16 BGC envolvidos na síntese de terpenos, foram detectados pelo menos três sequências de genes característicos, que codificam a expressão de TS nas linhagens

avaliadas: (i) *sqhC*, gene codificador de esqualeno hopeno ciclase (SHC); (ii) *CrtI*, codificador de fitoeno desaturase (PDS) e um gene que codifica atividade não determinada e relacionado com a produção de enzimas da família fitoeno/esqualeno sintase (FES).

Os *BGC* que continham a sequência do gene *sqhC*, que codifica a SHC foram detectados pelo antiSMASH em oito linhagens SDF (Figura 16): *B. pumilus* SDF0011; *B. oleronius* SDF0015; *Bacillus* sp. SDF0016; *B. simplex* SDF0024; *Paenibacillus* sp. SDF0028; *Lysinibacillus* sp. SDF0063; *B. velezensis* SDF0141 e *B. velezensis* SDF0150. Também foi observado sequências de outros genes biossintéticos, reguladores ou de transporte localizadas adjacentes à sequência do gene *sqhC*.

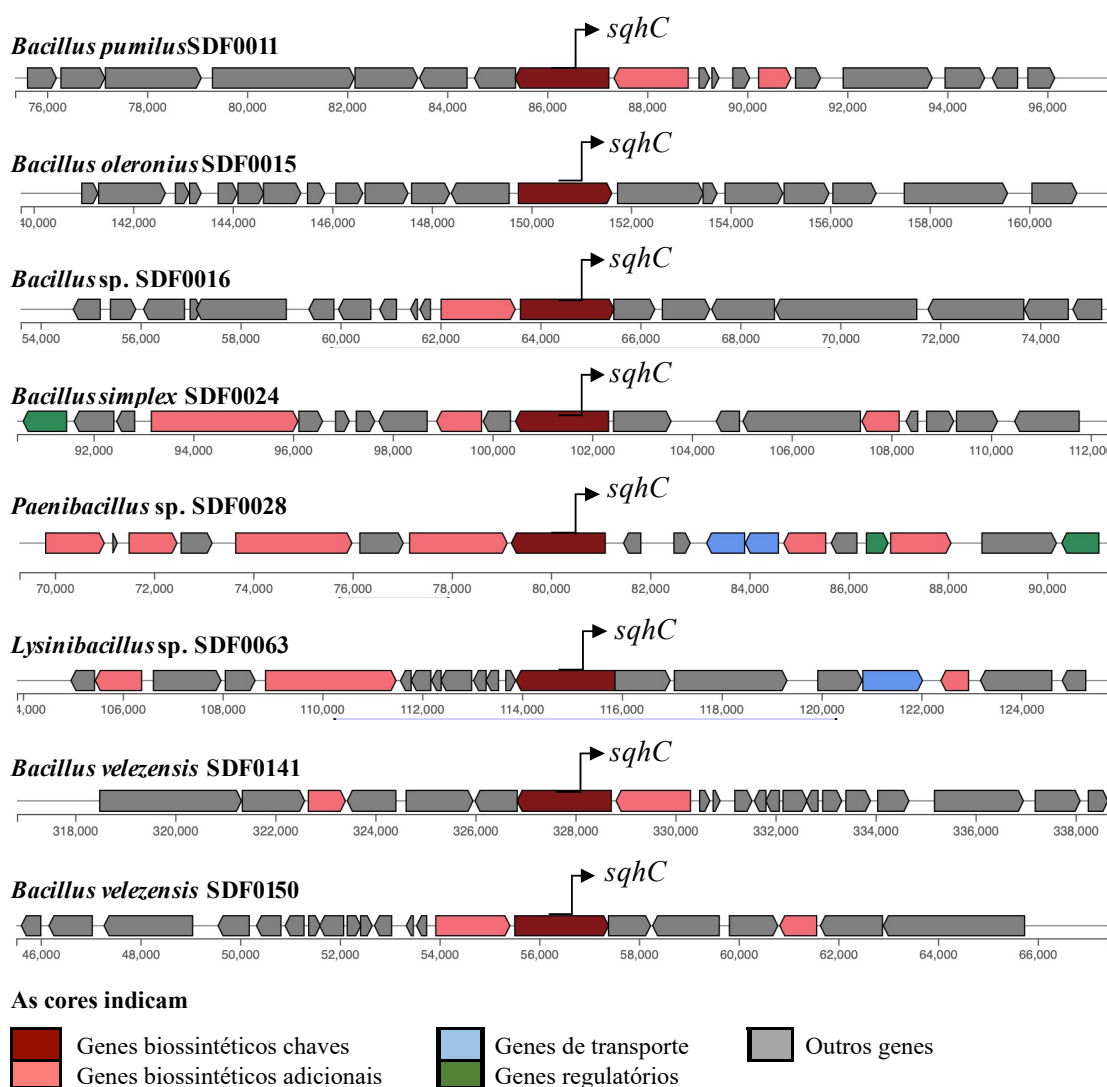


Figura 16. Gene *sqhC* identificado pelo antiSMASH dentro dos respectivos *BGC*, que codifica as atividades responsáveis pela síntese de terpenos nas linhagens SDF avaliadas. As cores indicam o tipo de atividade codificada por cada sequência.

As sequências de genes que codificam a síntese de FES estavam localizadas em *BGC* apontados como responsáveis por dirigir a síntese de terpeno nas linhagens *L. fusiformis* SDF0005; *B. simplex* SDF0024; *Lysinibacillus* sp. SDF0037; *B. velezensis* SDF0141 e *B. velezensis* SDF0150 (Figura 17). Entretanto, apesar de estarem presentes nas linhagens *B. pumilus* SDF0011 e *Bacillus* sp. SDF0016, os genes que codifica FES divergiram em termos de localização e estavam adjacentes à duas cópias do gene *CrtI*, codificador de PDS (Figura 17). Nestas duas linhagens, estes *BGC* identificados foram reportados como informação para biossíntese de carotenoides, com índice de similaridade de 50% (Figura 18).

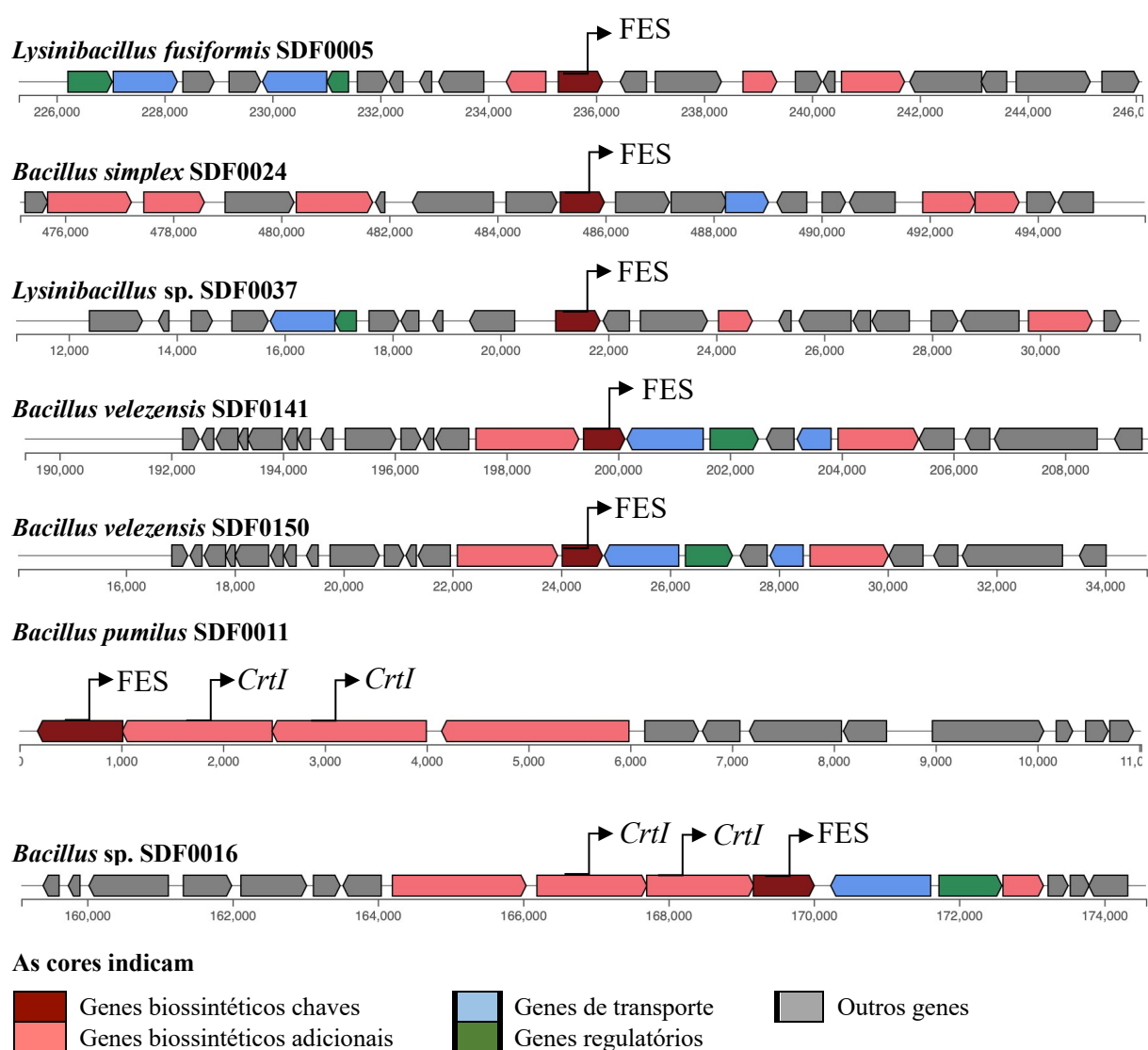
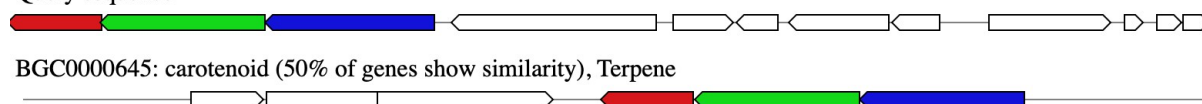


Figura 17. Localização do gene codificador de FES. A posição do gene codificador de FES adjacente ou não ao gene *CrtI* identificados dentro dos respectivos *BGC*, que codificam as atividades responsáveis pela síntese de terpenos nas linhagens SDF avaliadas. As cores indicam o tipo de atividade codificada por cada sequência.

***Bacillus pumilus* SDF0011**

Query sequence

***Bacillus* sp. SDF0016**

Query sequence

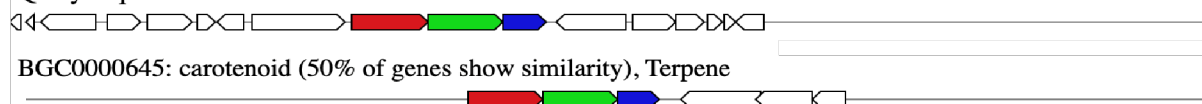


Figura 18. BGC de carotenoide. BGC que codificam as atividades responsáveis pela síntese de carotenoide nas linhagens *B. pumilus* SDF0011 e *Bacillus* sp. SDF0016 com índice de similaridade de 50%.

O resumo das TS detectadas pelo antiSMASH para as linhagens SDF neste estudo estão descritas na tabela 4.

Tabela 4. Presença de TS identificadas pelo antiSMASH nas linhagens SDF investigadas

Linhagem SDF	Enzimas		
	SHC	PDS	FES
<i>Lysinibacillus fusiformis</i> SDF0005	-	-	+
<i>Bacillus pumilus</i> SDF0011	+	+	+
<i>Bacillus oleronius</i> SDF0015	+	-	-
<i>Bacillus</i> sp. SDF0016	+	+	+
<i>Bacillus simplex</i> SDF0024	+	-	+
<i>Paenibacillus</i> sp. SDF0028	+	-	-
<i>Lysinibacillus</i> sp. SDF0037	-	-	+
<i>Lysinibacillus</i> sp. SDF0063	+	-	-
<i>Bacillus velezensis</i> SDF0141	+	-	+
<i>Bacillus velezensis</i> SDF0150	+	-	+

(+) Gene para enzima detectado; (-) Gene para enzima não detectado.

O algoritmo *PathoLogic* disponibilizado pelo *software Pathways tools* identificou em todas as linhagens SDF analisadas a presença das enzimas da via MEP (DXS, DXR, MCT, CMK, MDS, HDS e HDR), descritas na figura 19. A via bioquímica obtida pela reconstrução metabólica indica as reações descritas na Tabela 1, desde a conversão de glicose ou G3P até a produção de IPP, precursor indispensável para síntese de terpenos. Esta via bioquímica foi obtida para todas as linhagens SDF investigadas neste estudo. A linhagem *L. fusiformis* SDF0005 não apresentou o gene codificador da enzima IDI, responsável pela isomerização de IPP em DMAPP (Figura 19).

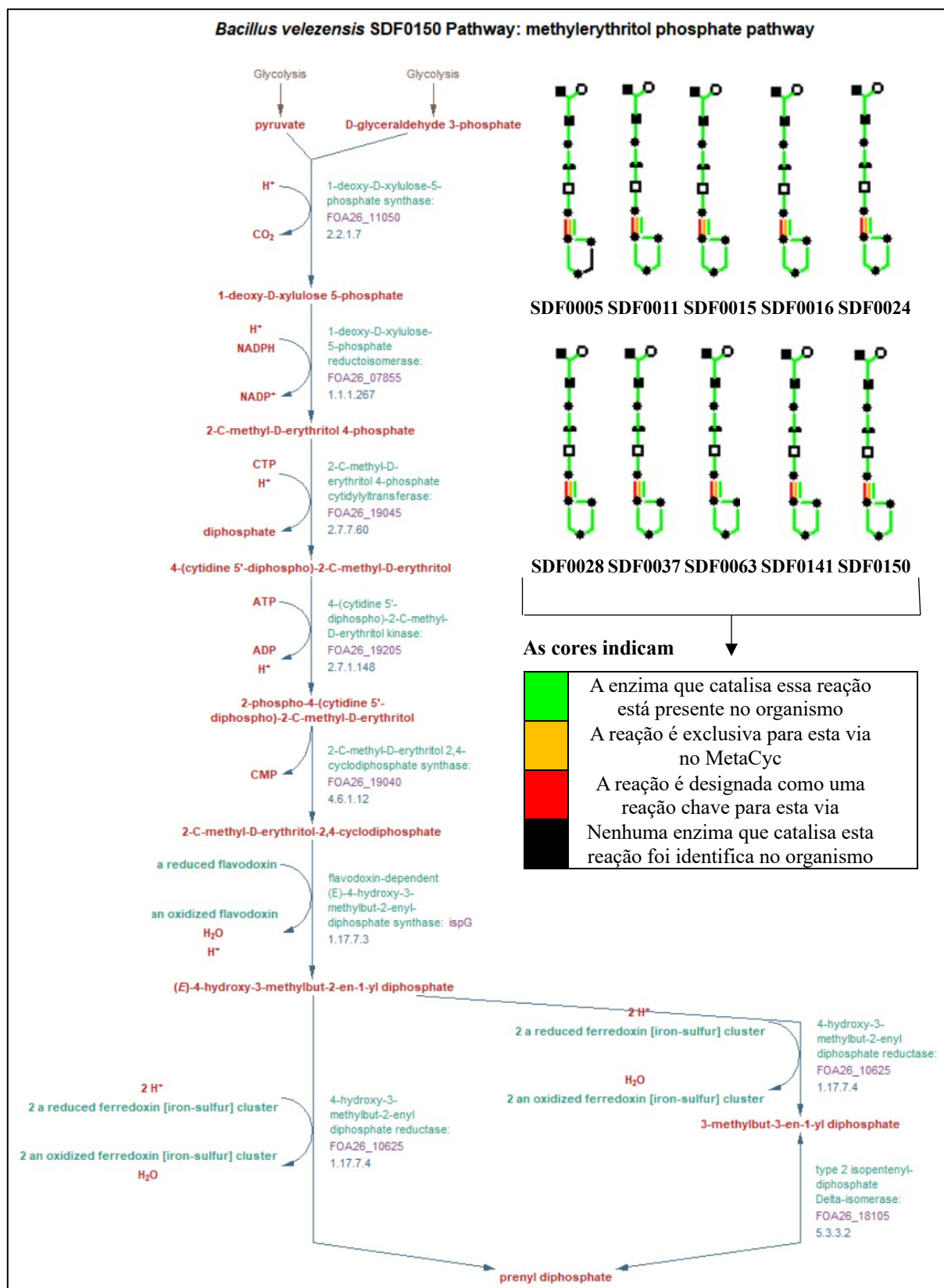


Figura 19. Via metileritrotol-fosfato (MEP). Diagrama da reconstrução metabólica *in silico* da via MEP detectado pelo aplicativo *Pathways tools* para a linhagem *B. velezensis* SDF0150 e demais linhagens analisadas (canto superior direito). As cores indicam a presença ou ausência do catalisador para a respectiva reação.

A análise efetivada pelo pBLAST detectou a presença de sequências de aminoácidos das enzimas da família poliprenil sintase (Tabela 5), responsáveis pela conversão de IPP em GPP, FPP e GGPP nas linhagens *L. fusiformis* SDF0005; *B. oleronius* SDF0015; *B. simplex* SDF0024; *B. velezensis* SDF0141 e *B. velezensis* SDF0150. Em todos os casos, o melhor percentual de identidade encontrado foi >98%, exceto para linhagem SDF0015, com percentual de 67,86% (Tabela 5). Os resultados obtidos do diagrama esquemático do alinhamento de cada linhagem em que foi detectado a enzima da família poliprenil sintase estão descritos nas figuras 24 a 28 do apêndice deste estudo.

Tabela 5. Detecção de enzimas da família poliprenil sintase nas linhagens SDF pelo BLAST. Os níveis mais altos de identidade resultante dos alinhamentos das respectivas sequências de cada SDF estão descritos.

Linhagem SDF	Resultado	Percentual de identidade (%)	Sequência de referência NCBI	Espécie de referência
<i>Lysinibacillus fusiformis</i> SDF0005	+	99,66	KAB0443654.1	<i>L. fusiformis</i>
<i>Bacillus pumilus</i> SDF0011	-	-	-	-
<i>Bacillus oleronius</i> SDF0015	+	67,86	WP_268443628.1	<i>B. pumilus</i>
<i>Bacillus</i> sp. SDF0016	-	-	-	-
<i>Bacillus simplex</i> SDF0024	+	98,65	WP_241589686.1	<i>Peribacillus</i> sp.
<i>Paenibacillus</i> sp. SDF0028	-	-	-	-
<i>Lysinibacillus</i> sp. SDF0037	-	-	-	-
<i>Lysinibacillus</i> sp. SDF0063	-	-	-	-
<i>Bacillus velezensis</i> SDF0141	+	100	ASK59031.1	<i>B. velezensis</i>
<i>Bacillus velezensis</i> SDF0150	+	99,65	QWC45887.1	<i>B. velezensis</i>

A árvore filogenética gerada a partir tradução *in silico* da sequência do gene *sqhC*, que dirige a síntese de SHC, em oito linhagens SDF apontou a formação de dois clados principais, o primeiro com membros apenas do gênero *Bacillus* (Figura 20). As respectivas enzimas das duas linhagens de *B. velezensis* SDF0141 e SDF0150 eram mais próximas evolutivamente, enquanto a sequência primária de SHC da linhagem *B. pumilus* SDF0011 se mostrou mais relacionada com a da linhagem *Bacillus* sp. SDF0016 (Figura 20). As sequências relacionadas da mesma enzima de outras linhagens do gênero, *B. simplex* SDF0024 e *B. oleronius* SDF0015, eram mais próximas evolutivamente, e foram localizadas no segundo clado gerado. A relação molecular entre a enzima SHC das duas espécies conjuntas eram mais próxima evolutivamente da sequência encontrada na linhagem *Lysinibacillus* sp. SDF0063. Por fim, a enzima da

linhagem SDF0028, classificada como *Paenibacillus* sp., foi a que apresentou menor grau de relação com as demais encontradas no segundo clado principal (Figura 20).

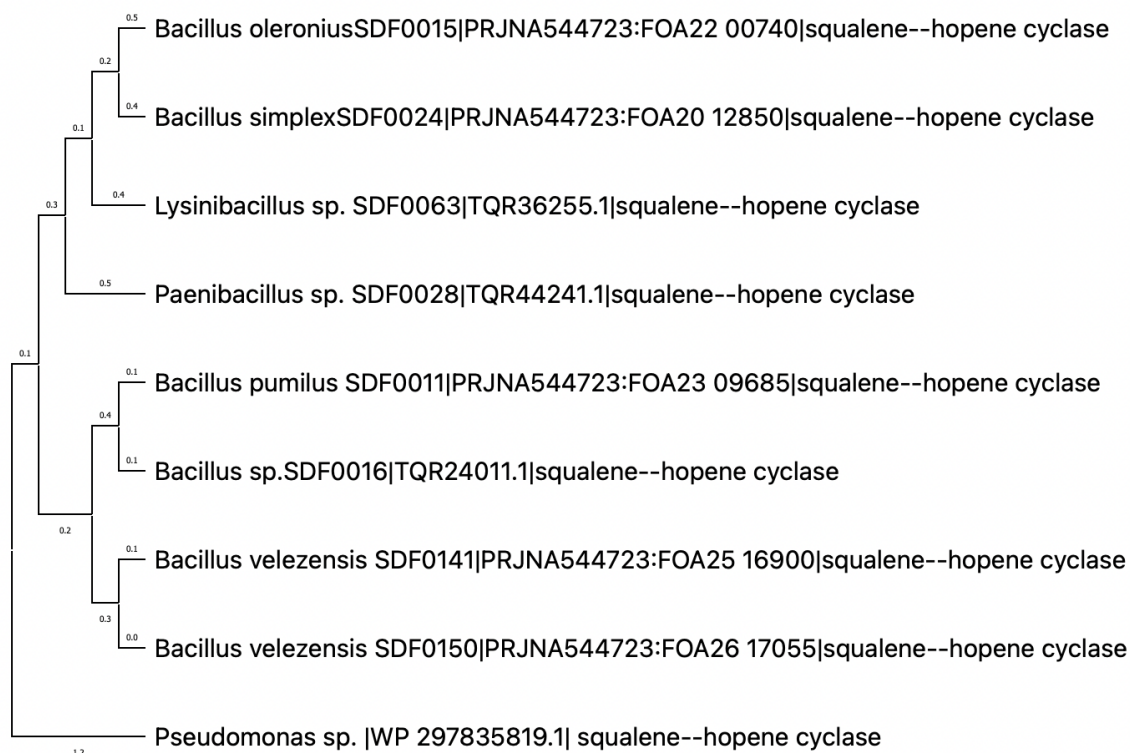


Figura 20. Relações filogenéticas entre linhagens SDF baseado na sequência de aminoácidos da enzima SHC. As linhagens *B. velezensis* SDF0141 e SDF0150 são mais relacionadas com as linhagens *B. pumilus* SDF0011 e *Bacillus* sp. SDF0016. As sequências de aminoácidos de SHC de *B. simplex* SDF0024 e *B. oleronius* SDF0015 são mais relacionadas quando comparadas às de *Lysinibacillus* sp. SDF0063. A linhagem *Paenibacillus* sp. SDF0028 possui menor grau de parentesco com as demais linhagens. A sequência de aminoácidos de SHC de *Pseudomonas* sp. foi utilizado como controle para a análise. Os valores de *bootstrap* estão indicados nos ramos.

Por outro lado, as sequências de aminoácidos da enzima FES detectadas em sete linhagens SDF e utilizadas na construção da árvore filogenética resultaram na deformação de dois clados principais (Figura 21). As enzimas das linhagens *B. velezensis* SDF0141 e SDF0150 eram mais próximas evolutivamente, enquanto a FES da linhagem *B. pumilus* SDF0011 mostrou-se mais relacionada com a linhagem *Bacillus* sp. SDF0016 dentro do primeiro clado (Figura 21). O outro clado apontou que as linhagens de mesmo gênero *L. fusiformis* SDF0005 e *Lysinibacillus* sp. SDF0037 podem expressar FES contendo nível superior de similaridade. A linhagem *B. simplex* SDF0024 apresentou FES mais característica e menos relacionada com as demais dentro do clado gerado (Figura 21).

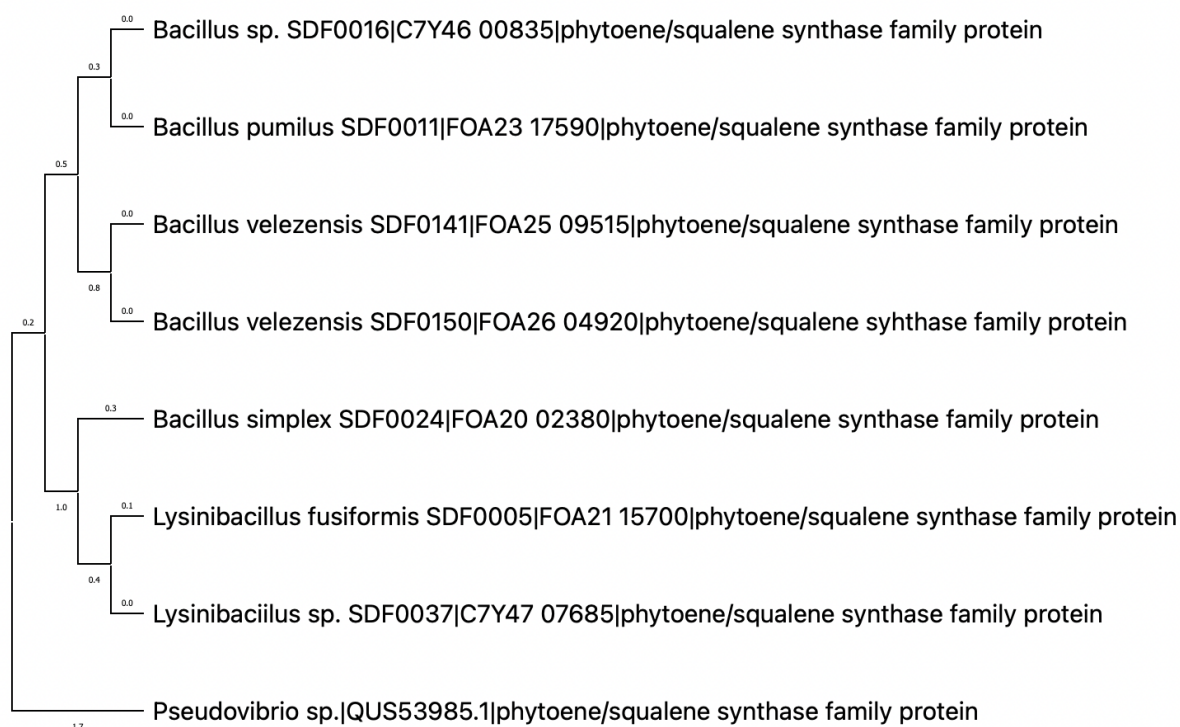
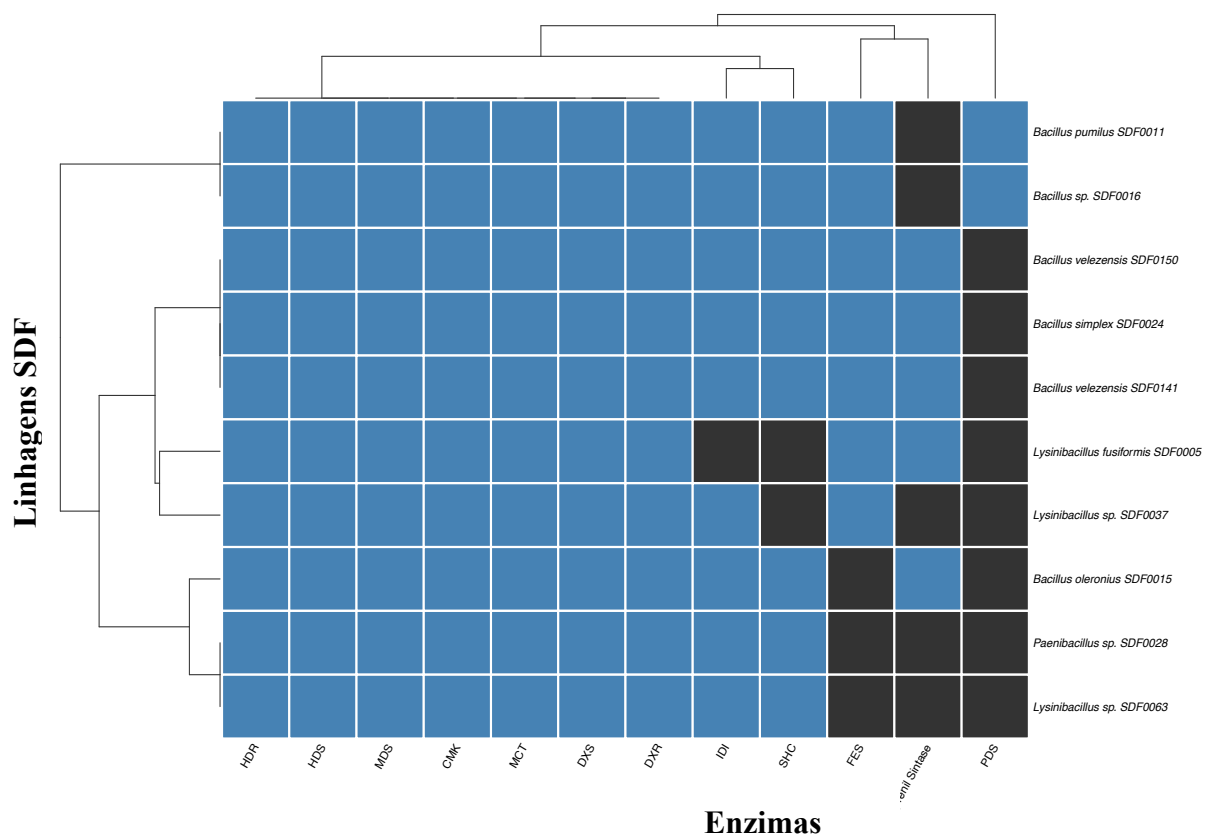


Figura 21. Relações filogenéticas entre linhagens SDF baseado na sequência de aminoácidos da enzima FES. As linhagens *B. velezensis* SDF0141 e SDF0150 são mais relacionadas com as linhagens *B. pumilus* SDF0011 e *Bacillus* sp. SDF0016. As sequências de aminoácidos de FES de *L. fusiformis* SDF0005 e *Lysinibacillus* sp. SDF0037 são mais relacionadas quando comparadas às de *B. simplex* SDF0024. A sequência de aminoácidos de FES de *Pseudovibrio* sp. foi utilizado como controle para a análise. Os valores de *bootstrap* estão indicados nos ramos.

O *heatmap* que descreve o nível de similaridade do conjunto enzimático envolvido na produção de terpenos apontou que as linhagens *B. pumilus* SDF0011 e *Bacillus* sp. SDF0016 possuem o mesmo nível de similaridade para produção de terpenos (Figura 22). As demais linhagens foram alocadas dentro do segundo clado principal gerado. A linhagem *B. simplex* SDF0024 e as duas de *B. velezensis* (SDF0141 e SDF0150) apresentaram o mesmo nível de similaridade encontrados entre as enzimas detectadas (Figura 22). Próximo a essas linhagens, *Lysinibacillus* sp. SDF0037 e *L. fusiformis* SDF0005 apresentaram maior nível de relação molecular. Para as linhagens SDF restantes, *B. oleronius* SDF0015 apresentou o conjunto enzimático do potencial de produção de terpenos mais característico, quando comparada com as linhagens *Paenibacillus* sp. SDF0028 e *Lysinibacillus* sp. SDF0063. Neste caso, ambas as linhagens apresentaram as sequências que representam o mesmo conjunto de enzimas identificadas (Figura 22).



Presença da enzima: **positivo**; negativo.

Figura 22. Similaridade entre o conjunto enzimático envolvidos na produção de terpenos nas linhagens SDF avaliadas. A cor azul representa a presença de uma enzima de interesse para produção de terpeno, enquanto ausência está representado pela cor preta. As linhagens *B. pumilus* SDF0011 e *Bacillus* sp. SDF0016 apresentaram mesma similaridade do conjunto enzimático, assim como as linhagens *B. velezensis* (SDF0141 e SDF0150) e *B. simplex* SDF0024. As linhagens *Paenibacillus* sp. SDF0028 e *Lysinibacillus* sp. SDF0063 também apresentaram o mesmo conjunto enzimático.

7. DISCUSSÃO

7.1. *Potencial genômico global para produção de metabólitos especializados em linhagens SDF*

A investigação da presença de *BGC* envolvidos na síntese de metabólitos especializados no genoma de dez linhagens SDF selecionadas da CBafes (Tabela 2) pelo antiSMASH definiu 153 *clusters*, de pelo menos 20 classes ou subclasses diferentes (Figura 8). Potencialmente, este resultado indica que, em geral, as Bafes apresentam um arsenal genético amplo devotado à síntese de diferentes grupos de metabólitos especializados. De fato, espécies do gênero *Bacillus* e gêneros relacionados (Bafes) são constantemente relatadas como fontes promissoras na busca de novos produtos originados do metabolismo especializado (MONDOL; SHIM; ISLAM, 2013; TYC et al., 2017; KIESEWALTER et al., 2021). Estes metabólitos são subdivididos em classes principais como bacteriocinas, NRP, terpenos, PK ou alcaloides (KELLER et al., 2005). Em razão, da diversidade estrutural e topológica identificada, esses catalisadores são capazes de utilizar uma grande variedade de substratos. (GUSMÁN-CHAVES et al., 2018). Esta habilidade explica as 20 classes ou subclasses diferentes de metabólitos especializados detectados pelo antiSMASH (Figura 8) durante o desenvolvimento deste trabalho. Os resultados obtidos são decorrentes da capacidade do *software* em identificar essas variações topológicas das enzimas e, conseqüente, o emprego de diferentes substratos (BLIN, 2021).

Neste estudo, os *BGC* que continham a informação envolvida na síntese de NRPS (47 de 153) foram os mais numerosos nos genomas avaliados (Figura 8). O potencial genômico para a síntese de NRP se ampliou com a detecção de quatro *clusters* adicionais para *BGC* envolvidos na síntese de NRPS-like (Figura 8). AS NRPS são enzimas modulares com os vários domínios ligados covalentemente. Existem ao menos três domínios presentes nas enzimas que produzem NRP: (i) o sítio de adenilação (A); (ii) tiolação (T) e (iii) condensação (C) (YEH et al., 2012). O domínio A catalisa a transferência de um grupo acila, proveniente de um ácido carboxílico, para o substrato ligado ao domínio T, formando uma ligação tioéster. O sítio C, por fim, promove a formação das ligações peptídicas entre os resíduos de aminoácidos por reação de condensação (SHI et al., 2021). A diferença existente entre as enzimas NRPS e NRPS-like é a ausência do sítio C na última forma. Assim, a alongação dos peptídeos, nesse caso, ocorre por outros mecanismos enzimáticos (SHI et al., 2021).

A grande quantidade de *BGC* dedicada à síntese de NRPS detectada aqui corrobora os resultados obtidos para *B. licheniformis*. Nesta espécie, os *BGC* envolvidos na síntese de NRPS representaram até 50% de *clusters* envolvidos na síntese de metabólitos especializados em linhagens de *B. licheniformis* avaliadas por OTHOUM et al., (2018). Esses peptídeos possuem

desde ação antimicrobiana até antitumoral (HUR; VICKERY; BURKART, 2012). Entretanto, segundo Sumi et al., (2015), em espécies do gênero *Bacillus*, é notória a atividade de NRP com ação antimicrobiana. Entre os metabólitos especializados com índice de similaridade de 100% preditos pelo antiSMASH com base no banco de dados MIBiG, a bacilisina (Figura 11), bacilaeno (Figura 13) e paenibacterina (Figura 14) estão dentro desta conjuntura.

A bacilisina é um dos dipeptídeos com atividade antimicrobiana estruturalmente mais simples, inicialmente caracterizado em *B. subtilis* (OZCENGIZ; OGULUR, 2015). Este peptídeo é composto por um resíduo de alanina ligado a um resíduo derivado de tirosina. A bacilisina pode inibir o crescimento de *Staphylococcus aureus* e *Candida albicans*, por interferir na síntese da parede celular (OZCENGIZ; OGULUR, 2015). Além de *B. subtilis*, a síntese deste metabólito foi descrita em *B. amyloliquefaciens*, *B. velezensis* e *B. pumilus* (NANNAN et al., 2021). Os *BGC* envolvidos na biossíntese de bacilisina que exibiram 100% de similaridade neste estudo estavam presentes nos genomas das duas linhagens de *B. velezensis* (SDF0141 e SDF0150) (Tabela 3), convergindo com os resultados encontrados por Nannan et al., (2021). Adicionalmente, os mesmos *BGC* foram detectados em *B. pumilus* SDF0011 e *Bacillus sp.* SDF0016 (Tabela 3). Em ambos os casos, o índice de similaridade obtido foi menor que 85%, o que pode ser explicado pela diferença entre os mecanismos de biossíntese de bacilisina observados para os membros do gênero (ISLAM et al., 2022). Entretanto, esta variação pode não ter sido detectada pelo antiSMASH, visto que o mesmo *BGC* de referência no banco MIBiG foi apontado em todas as situações (Tabela 3).

A paenibacterina é um peptídeo cíclico com uma cadeia lateral de ácido graxo, descrito inicialmente em *P. thiaminolyticus* (HUANG; YOUSEF, 2014). A atividade antimicrobiana da paenibacterina está relacionada com a atuação na membrana externa de bactérias Gram negativas, bem como na membrana citoplasmática de Gram positivas e Gram negativas. Estas estruturas do envelope celular são rompidas como consequência do vazamento de íons intracelulares e, conseqüente despolarização das membranas (HUANG; YOUSEF, 2014). O *BGC* envolvido na biossíntese desta molécula foi detectado pelo antiSMASH no genoma da linhagem SDF0028 de espécie não determinada pertencente ao gênero *Paenibacillus* (Tabela 3). Este resultado concorda com demais estudos desenvolvidos na área, que apontam a paenibacterina como um metabólito encontrado em diferentes espécies desse gênero, e, em casos excepcionais no gênero relacionado *Brevibacillus* (ALETTI; SESSITSCH; BRADER, 2015; GLARE et al., 2020).

Como destacado anteriormente, assim como a bacilisina e a paenibacterina, o bacillaeno possui ação antimicrobiana, sendo descrito inicialmente em *B. subtilis* (PATEL et al., 1995). A

estrutura molecular do bacillaeno, entretanto, foi elucidada apenas recentemente, uma vez que o *cluster* biossintético deste metabólito codifica um catalisador com características tanto de PKS como de NRPS (REDDICK; ANTOLAK; RANER, 2007). Atualmente, o bacillaeno é caracterizado como uma molécula linear com duas ligações amida (Figura 13). Este metabólito atua inibindo a síntese de proteínas em procariotos, embora o mecanismo de ação ainda não tenha sido elucidado precisamente (BUTCHER et al., 2007). Os *BGC* envolvidos na produção de bacillaeno são detectados em várias espécies do complexo do *B. subtilis*, como *B. amyloliquefaciens*, *B. velezensis*, *B. methylotrophicus*, entre outros (CHEN et al., 2018; RABEE et al., 2019). O *BGC* que dirige a síntese de bacillaeno foi identificado nas duas linhagens *B. velezensis* (SDF0141 e SDF0150) (Tabela 3), convergindo com os resultados obtidos nos estudos destes autores.

Em adição à atividade antimicrobiana, os NRP também estão envolvidos na captação de ferro no ambiente circundante do organismo produtor. De fato, a bacilibactina é um sideróforo encontrado em espécies de Bafes (MIETHKE et al., 2006). Os *BGC* envolvidos na produção de bacilibactina apresentaram 100% de similaridade nos genomas de ambas as linhagens de *B. velezensis* (SDF0141 e SDF0150) inspecionadas. Contudo, em outras linhagens como *B. pumilus* SDF0011; *Bacillus* sp. SDF0016 e *Paenibacillus* sp. SDF0028 também foram detectados *clusters* com informações dedicadas à síntese deste metabólito, apesar de índices de similaridade inferiores (Tabela 3). Este resultado sugere que a biossíntese de bacilibactina é compartilhada por várias espécies de Bafes e possui papel fisiológico e ecológico relevante para esses procariotos.

O ferro é um nutriente essencial porque participa no crescimento celular microbiano. Este nutriente é associado a estrutura de diversas proteínas, inclusive atuando como cofator de enzimas de diversos processos bioquímicos (RABEE et al., 2019). A alta capacidade de captação de ferro por um organismo produtor de sideróforos possui, ainda, atividade antimicrobiana de forma complementar. A habilidade de incorporar ferro com alta eficiência inibe o crescimento de outros organismos circundantes, porque dificulta a aquisição desse elemento em quantidades suficientes para suprir as necessidades fisiológicas de outras populações (RABEE et al., 2019).

A segunda família de sideróforos detectada neste estudo foi a petrobactina (Tabela 3). Coincidentemente, os *BGC* envolvidos na síntese desta estrutura foram encontrados em apenas membros do gênero *Lysinibacillus*: *L. fusiformis* SDF0005 e duas linhagens com espécie não identificada *Lysinibacillus* sp. (SDF0037 e SDF0063). Entretanto, outros estudos apontam que a petrobactina também é produzida por outras Bafes, como espécies de *Bacillus* e *Paenibacillus*

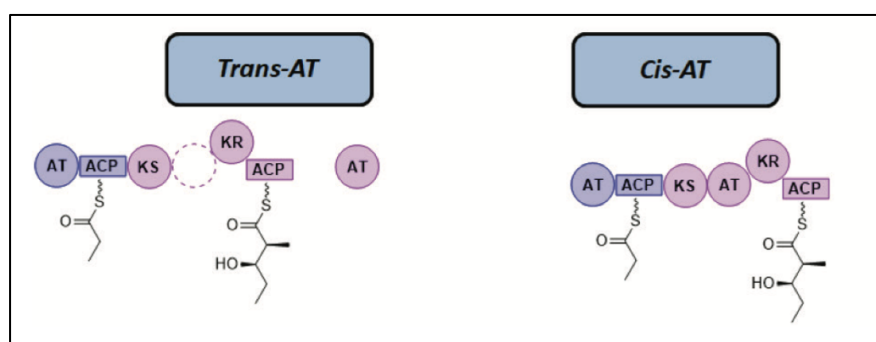
(WILSON et al., 2006; GLARE et al., 2020). O *B. anthracis*, por exemplo, é noticiado como produtor tanto de petrobactina como de bacilibactina, assim como outros membros do grupo do *B. cereus* (WILSON et al., 2006; LEE et al., 2011). Embora a petrobactina e a bacilibactina sejam produzidas por Bafes em geral, é tentador especular se membros de *Lysinibacillus* produzem apenas a petrobactina em detrimento da bacilibactina, o que torna este tópico alvo de futuras investigações.

Além de NRP, outra classe relevante de metabólitos especializados produzida por Bafes são os PK. Dentre os 153 *BGC* detectados pelo antiSMASH neste estudo, ao menos 40 estavam envolvidos na produção de diferentes PKS (Figura 8). Esses catalisadores foram classificados em quatro tipos principais pelo *software*: T1PKS, T3PKS, trans AT-PKS e PKS-like (Figura 8). Como destacado anteriormente, T1PKS são enzimas multidomínios modulares (cada domínio é utilizado apenas uma vez até o processamento do PK) ou iterativas (domínios utilizados repetidas vezes) (CUADRAT, 2010). T3PKS apresentam estrutura homodimérica e são independentes da atividade catalítica do domínio ACP (PAULO; SISTRIT; OLIVEIRA, 2019). Aparentemente, em Bafes, T2PKS não constituem um grupo enzimático relevante na biossíntese de metabólitos especializados, visto que nenhum *BGC* com informações para a síntese deste tipo enzimático em particular foi detectado nos dez genomas das linhagens SDF examinadas (Tabela 2).

Inversamente, as trans AT-PKS podem ser uma família de enzimas primordiais para síntese de PK em Bafes, pois compreende a maior quantidade de *BGC* envolvidos na produção de PKS detectados neste estudo (Figura 8). As trans AT-PKS, porém, integram um tipo específico de T1PKS (PAULO; SISTRIT, OLIVEIRA, 2019). Um módulo de T1PKS apresenta tipicamente três domínios: cetoacila sintase (KS), acetil-transferase (AT) e o domínio ACP. Usualmente, esses domínios são adjacentes na estrutura de PKS. Em trans AT-PKS o domínio AT não está posicionado de modo subsequente aos demais (Figura 23) (PAULO; SISTRIT; OLIVEIRA, 2019). Atualmente, as trans AT-PKS são conhecidas pela ampla distribuição em diferentes espécies de bactérias e fungos. Efetivamente, estudos em genomas bacterianos apontam que até 38% dos genes que codificam as PKS são do tipo trans AT (HELFRICH; PIEL, 2016).

A macrolactina (Figura 12) e a macrobrevina (Figura 15), identificados com 100% de nível de similaridade entre os *BGC* detectado pelo antiSMASH e o respectivo metabólito especializado previsto pelo banco de dados MIBiG são PK sintetizados por trans AT-PKS (Tabela 3). Macrolactinas, como outros macrolídeos, são moléculas que tipicamente contêm lactonas macrocíclicas na estrutura. Além da atividade antimicrobiana, geralmente exibem ação

citotóxica e são reportadas como agentes de controle biológico (HELFRICH; PIEL, 2016). Inicialmente, essas moléculas foram isoladas de bactérias marinhas (GUSTAFSON et al., 1989). Entretanto, macrolactinas são amplamente sintetizadas por microrganismos isolados de solos, notadamente *Bacillus* spp. (ORTIZ; SANSININEA, 2020). Desde a primeira descrição, mais de outros vinte derivados de macrolactinas foram caracterizados. Segundo estes autores, as novas estruturas são classificadas utilizando letras de A a W. As variações são decorrentes de diferentes agrupamentos químicos encontrados dentro das moléculas. Em alguns casos, as macrolactinas podem ser ainda conjugadas com monômeros de carboidratos ou álcoois (ORTIZ; SANSININEA, 2020).



Fonte: PAULO; SISTRIT; OLIVEIRA, 2019

Figura 23. Domínios usuais de PKS – cetoacila sintase (KS), acetil-transferase (AT) e carreadora de acila (ACP). Usualmente, os domínios de PKS são adjacentes na estrutura enzimática (Cis-AT). Em *trans* AT-PKS o domínio AT não está posicionado de modo subsequente aos demais.

Segundo Nagao et al., (2001), a macrolactina do tipo H, também detectada neste estudo em duas linhagens de *B. velezensis* (SDF0141 e SDF0150) exibe ação antimicrobiana contra *S. aureus* e, até mesmo, *B. subtilis*. Porém, o mecanismo de ação das macrolactinas no organismo alvo permanece desconhecido. Possivelmente, mais de um mecanismo esteja envolvido na atividade antimicrobiana, consequência da alta variabilidade estrutural química encontrada nessas moléculas (HUU-DAT et al., 2021). Estudos que dispõem sobre a distribuição dos diferentes tipos de macrolactinas em *Bacillus* spp. são escassos. A presença de macrolactina H em *B. velezensis* detectada neste estudo, entretanto, conflui com o mesmo resultado de Yan et al., (2022). Estes autores detectaram *BGC* envolvidos na síntese deste tipo de macrolactina a partir do sequenciamento de um genoma de *B. velezensis*.

Por outro lado, a atividade antimicrobiana da macrobrevina possivelmente consiste na estratégia usual de macrolídeos, atuando como inibidor da tradução em organismos susceptíveis (CHAKRABORTY et al., 2021). Neste estudo o *BGC* com informação para a síntese desse metabólito foi identificado apenas no genoma de *Lysinibacillus* sp. SDF0063 (Figura 15). O termo macrobrevina foi cunhado porque inicialmente este metabólito foi caracterizado em *Brevibacillus* spp isoladas de rizosfera de plantas (HELFRICH et al., 2018). Atualmente, graças aos avanços nas ferramentas para bioprospecção *in silico*, genes ou grupos de genes envolvidos na produção de metabólitos especializados, a produção de macrobrevina têm sido apontadas em espécies dos gêneros *Bacillus* e *Lysinibacillus* (família *Bacillaceae*) (JAHNE et al., 2023), além de *Paenibacillus*, alocados na família *Paenibacillaceae* (ELTOKHY et al., 2021). A abundante distribuição de *BGC* em diferentes famílias taxonômicas aponta que a produção da macrobrevina pode ser uma característica pleiomórfica de Bafes.

As betalactonas também constitui uma classe relevante de metabólitos especializados detectados pelas análises realizadas utilizando o antiSMASH (Figura 8). Porém, os *clusters* contendo sequências aparentemente relacionadas com a síntese destas estruturas apontaram a produção de fengicina (Tabela 3), que na realidade são sintetizados por NRPS (VILLEGAS-ESCOBAR et al., 2013). As fengicinas são lipopolipeptídeos cíclicos conjugados com ácidos graxos. Atualmente são descritos diferentes tipos de fengicina, sendo que em alguns casos são encontrados anéis lactona na porção cíclica da molécula (VILLEGAS-ESCOBAR et al., 2013). O antiSMASH realiza a predição dos metabólitos especializados baseado no tipo de enzima expressa pela sequência detectada pelo *software*, o que explica porque ocorreu a classificação de tais *BGC* como uma classe específica de betalactonas (BLIN, 2021).

As fengicinas foram pioneiramente caracterizadas como antifúngicos em *B. subtilis*, com ação, notadamente, contra fungos filamentosos (VANITTANAKOM et al., 1986). Nenhum *BGC* envolvido na síntese de fengicina apresentou índice de similaridade de 100% nos genomas investigados (Tabela 3). Como a mesma sequência do *cluster* de referência do banco de dados MIBiG foi a mesma em todas as situações apontadas (Tabela 3), os diferentes mecanismos bioquímicos envolvidos na síntese deste composto (VILLEGAS-ESCOBAR et al., 2013), possivelmente não foram detectados pelo antiSMASH. Este fato pode explicar os baixos índices de similaridade obtidos. Da mesma maneira, *BGC* contendo informações para a síntese de fengicina tiveram a maior distribuição nos genomas investigados, e se mostraram presentes em sete linhagens SDF (Tabela 3). Esse dado exprime a importância desse metabólito para as espécies de Bafes, visto que as mesmas, assim como fungos filamentosos, compartilham o solo como principal reservatório. (KANG; PARK; JUNG, 2020).

Peptídeos sintetizados tradicionalmente pela maquinaria ribossômica também são uma classe importante de metabólitos especializados em Bafes. Pelo menos 15 *BGC* de diferentes grupos foram detectados nos genomas avaliados (Figura 8). Entre esses estão RiPP (do inglês *ribosomally synthesized and post-translationally modified peptides*) e as variações como o RRE (do inglês *RiPP recognition element*), baseado em sítios de reconhecimento específicos, e LAP (do inglês *linear azole-containing peptides*) (TRAVIN et al., 2019). Esses peptídeos são genericamente conhecidos como bacteriocinas e possuem atividade antimicrobiana reportada em várias espécies relacionadas (HEILBRONNER et al., 2021). Após a tradução as bacteriocinas são sujeitas a modificações estruturais, o que gera uma ampla variedade de moléculas e diversidade de mecanismos de inibição. Tiopeptídeos, lantipeptídeos, rantipeptídeos, lassopeptídeos e sactipeptídeos são exemplos dessa biodiversidade bioquímica, cujos *BGC* envolvidos na respectiva síntese foram detectados nos genomas das linhagens SDF investigadas (Figura 8).

Embora esses peptídeos sintetizados por ribossomos configurem uma importante classe de metabólitos especializado nas espécies de Bafes analisadas (Tabela 2), observou-se maior abundância de *BGC* das classes de NRP e PK (Figura 8). Os resultados obtidos neste estudo apontam, dessa maneira, um alto potencial genômico para produção de uma ampla variedade de metabólitos especializados das linhagens SDF depositadas na CBafes (Tabela 3), notadamente, peptídeos sintetizados independentemente da maquinaria ribossômica e dos produtos derivados do malonil-CoA. *Bacillus* e gêneros relacionados são de fato relatados na literatura como fontes importantes de NRP e PK com atividade antimicrobiana (SUMI et al., 2015; HAMDACHE et al., 2013). Os metabólitos especializados identificados com 100 % de similaridade neste estudo estão inseridos neste contexto, visto que as informações genéticas que permitem as atividades relacionadas com a produção destas moléculas foram reportadas esta atividade.

O potencial biotecnológico da CBafes, dessa forma, é generosamente promissor. Porém, os genomas aqui utilizados representam apenas uma fração da biodiversidade de toda a Coleção, o que torna necessário estudos complementares para prospecção *in silico* de *BGC* envolvidos na síntese de metabólitos especializados. Além disso, futuras análises para detectar a produção *in vitro* destes metabólitos é de extrema importância, pois a expressão gênica para a síntese de um bioproduto é regulada por fatores genéticos e ambientais. (TYC et al., 2017; SHARRAR et al., 2020). Dessa maneira, a constatação *in vitro* comprovaria se o que foi identificado *in silico* no genoma do organismo produtor, é realmente convertido em um produto funcional.

7.2. Comparação do potencial genômico para produção de metabólitos especializados entre as linhagens SDF

Dentre as linhagens SDF sondadas, o genoma de *Paenibacillus* sp. SDF0028 foi o que apresentou a maior número de *BGC* envolvidos na síntese de metabólitos especializados (Figura 9). Este gênero tem sido relatado como promissor na busca de novos compostos, principalmente com ação antimicrobiana (PRANAV et al., 2021). Em conjunto com *Streptomyces*, *Paenibacillus* constitui os gêneros de bactérias que apresentam o maior número de *BGC* reportados com o potencial de produção desta atividade (COCHRANE; VEDERAS, 2016). O número mais elevado de *clusters* dedicados à síntese de NRPS neste estudo foi identificado na linhagem *Paenibacillus* sp. SDF0028 (Figura 9), o que corrobora os resultados obtidos, visto que NRP são constantemente apontados como agentes de ação antimicrobiana (SUMI et al., 2015). No estudo de Pranav et al., (2021), 61% dos *BGC* detectados envolvidos na síntese de metabólitos especializados em *P. alvei* é para síntese de NRPS. Os resultados observados em *Paenibacillus* sp. SDF0028 também são similares, o que torna esta linhagem e outras relacionadas, grandes candidatas para o rastreamento de novos antibióticos em Bafes.

O número de *BGC* detectados nas duas linhagens (SDF0141 e SDF0150) de mesma espécie (*B. velezensis*), seguiu como destaque e apresentou o maior número após *Paenibacillus* sp. SDF0028 (Figura 9). O *B. velezensis* foi isolado inicialmente na Espanha, e desde então, vem sendo utilizado como importante agente promotor de crescimento vegetal (RABEE et al., 2021). O *B. velezensis* exibe amplo potencial genômico para promover a síntese de lipopetídeos como fengicina e bacilibactina, além de PK como macrolactinas e bacillaeno (RABEE et al., 2021). Coincidentemente, neste estudo, com exceção da fengicina, os demais foram identificados com índice de similaridade de 100% nos resultados obtidos pelo antiSMASH (Tabela 3). Atualmente, o genoma de *B. velezensis* é bem caracterizado. Linhagens de *B. amyloliquefaciens*, por exemplo, têm sido reclassificadas como *B. velezensis* em razão do alto grau de parentesco entre ambas as espécies alocadas no complexo do *B. subtilis* (BHANDARI et al., 2013; DUNLAP et al., 2016). Este fato ajuda a explicar a maior quantidade de moléculas reconhecidas com ampla precisão a partir dos *BGC* detectados nas duas linhagens de *B. velezensis* (SDF0141 e SDF0150).

Em contrapartida, a linhagem *B. oleronius* SDF0015 teve apenas três prováveis *BGC* identificados pelo antiSMASH neste estudo (Figura 9). Além disso, na biblioteca de sequências do banco de dados MIBiG não foi reconhecido nenhum metabólito especializado (Tabela 3). Essa espécie foi isolada a partir do intestino de cupins da espécie *Reticulitermes santonensis* (KUHNIK et al., 1995) e, segundo nosso conhecimento, estudos sobre o potencial

biotecnológico associados à metabólitos especializados nessa espécie não está disponível. Os estudos que envolvam esta espécie são direcionados para exploração de proteínas de *B. oleronius* para o uso em tratamentos de rosácea (JARMUDA et al., 2014; MAHER; STAUNTON; KAVANAGH, 2018). Apesar de ser uma espécie aeróbia e formadora de endósporo, *B. oleronius* responde atipicamente como Gram negativa ao teste de coloração (HEYNDRICKX et al., 2012). Ademais, a relação filogenética conhecida atualmente levou a reclassificação e criação de um novo gênero designado *Heyndrickxia* para alocar esta espécie, em conjunto com *H. sporothermodurans*, também originalmente classificada como *Bacillus* (GUPTA et al., 2020).

As demais linhagens SDF variaram, em maior ou menor grau, quanto ao número de *BGC* identificados, inclusive intra gênero. Enquanto *L. fusiformis* SDF0005 e *Lysinibacillus* sp. SDF0037 apresentaram menor quantidade de *clusters* detectados, *Lysinibacillus* sp. SDF0063, por exemplo, foi a quarta linhagem SDF em números de *BGC* apontados pelo antiSMASH (Figura 9). Estes resultados indicam um potencial genômico envolvido na produção de metabólitos especializados diversificados nas linhagens avaliadas. Esses resultados não são surpreendentes frente a biodiversidade encontrada em Bafes (FRITZE, 2004; DE VOS et al., 2009). Entretanto, é importante ressaltar que o potencial genômico que permite a produção de metabólitos especializados em cada uma das dez linhagens deste estudo é apenas uma predição, podendo ou não, corresponder ao potencial real de cada amostra. Fatores adicionais como extração e purificação do DNA inadequados, bem como eventual anotação e sequenciamento incompleto dos genomas de cada linhagem, deve ser levado em consideração, frente a possibilidade de supressão de resultados.

7.3. Perspectiva para identificação de novos metabólitos especializados em Bafes

Os metabólitos especializados identificados neste estudo são previstos a partir de uma comparação entre sequências de um *BGC* provável detectado pelo antiSMASH frente ao banco de dados MIBiG, com índice de similaridade entre 0 e 100%. Esse índice reflete a quantidade de genes em comum presente no *BGC* de uma sequência de interesse contraposto com aquelas depositadas no banco de dados. Dessa maneira, uma linhagem SDF que apresenta 100% de similaridade de determinantes genéticos envolvidos na síntese de um determinado composto revela que esta possui todos os genes que codificam as enzimas necessárias para a via catalítica do metabólito previsto (BLIN, 2021).

Porém, índices de similaridade inferiores ao máximo não significa que a molécula prevista não possa ser sintetizada. Os metabólitos especializados produzidos por bactérias apresentam uma ampla diversidade bioquímica e estrutural, inclusive compostos com as mesmas funções (VILLEGAS-ESCOBAR et al., 2013). A macrolactina e a fengicina, como discutido, são exemplos de metabólitos que sofrem pequenas variações estruturais e por isso são reconhecidos pela ampla diversidade de estruturas conhecidas (VILLEGAS-ESCOBAR et al., 2013; ORTIZ; SANSININEA, 2020). Os metabólitos especializados sintetizados a partir de sequências de *BGC* que apresentaram um índice de similaridade considerável, pode ser um indicativo de novas variações na estrutura do respectivo metabólito, pois em muitas situações a linhagem de referência do banco de dados MIBiG para o respectivo metabólito foi a mesma em diferentes linhagens SDF (Tabela 3). Dessa forma, metabólitos previstos que apresentaram altos valores de similaridade são potenciais alvos de estudos futuros para possível detecção de variações ainda desconhecidas.

Para os *BGC* acessados e sem identificação de metabólitos especializados potenciais, novas análises são ainda mais promissoras, pois representam possíveis combinações enzimáticas com amplo potencial de produção de metabólitos ainda não caracterizados. Neste estudo, de 153 *BGC* reconhecidos, 90 não foram associados à síntese de metabólitos reconhecidos. Somado aos que apresentaram baixos índices de similaridades, estas sequências indicam um arsenal amplo e importante a ser investigado na busca de novos metabólitos especializados produzidos por Bafes. Embora seja possível reconhecer que esses *BGC* possam ao mesmo tempo não serem responsáveis pela síntese de um produto funcional, a grande quantidade de *clusters* detectados não podem ser ignorada frente ao potencial biotecnológico promitente.

Como mencionado anteriormente, os metabólitos especializados de Bafes são reconhecidos pela capacidade de promoção de crescimento vegetal, ação citotóxica, imunossupressora, antitumoral, além de controle biológico de insetos e vetores de doenças (MONDOL et al., 2013; FALQUETO et al., 2021). Entretanto, a atividade antimicrobiana é a mais frequentemente caracterizada para os metabólitos de Bafes (SUMI et al., 2015; PRANAV et al., 2021). Considerando-se as altas taxas de morbidade e mortalidade, somado aos impactos econômicos que geram, a OMS em conjunto a outras agências de saúde e especialistas, concordam que a resistência antimicrobiana, atualmente, é uma das mais graves ameaças mundiais para a saúde humana e o meio ambiente (<https://www.who.int/en/news-room/fact-sheets/detail/antimicrobial-resistance>). Estudos para dissecar a existência de genes de metabólitos especializados são, dessa maneira, essenciais para explorar novos candidatos a

agentes antimicrobianos. Os resultados obtidos por este estudo mostram que as linhagens da CBafes possuem potencial promissor na busca destes antimicrobianos a fim de auxiliar no combate à histórica e crescente onda de resistência.

7.4. Potencial genômico para produção de terpenos nas linhagens SDF selecionadas

Em bactérias os terpenos são sintetizados a partir da via MEP. Essa via é responsável pela produção de IPP, precursor fundamental para biossíntese desses isoprenoides (LIANG et al., 2021). As enzimas que participam nesta via (DXS, DXR, MCT, CMK, MDS, HDS e HDR) foram todas detectadas pelo aplicativo *Pathways tools* nas linhagens SDF investigadas neste estudo (Figuras 20 e 21). A detecção de determinantes genéticos relacionados com esta via esteve altamente presente nos resultados encontrado para as linhagens SDF, indicando que Bafes em geral possuem o conjunto enzimático elementar para biossíntese de terpenos.

As enzimas da família poliprenil sintase (GPPS, FPPS e GGPPS) sintetizam as moléculas moldes subsequentes geradas a partir do IPP, que são utilizadas pelas TS para produzir terpenos (ZEVALLOS et al., 2013). A enzima GPPS catalisa a conversão de IPP e DMAPP em GPP, precursor dos monoterpenos. O DMAPP é um isômero de IPP, sintetizado pela enzima IDI (Figura 3). A detecção de IDI neste estudo foi também realizada utilizando o *Pathways tools*, que identificou em todas as dez linhagens SDF a presença de sequências que codificam este catalisador, exceto na linhagem *L. fusiformis* SDF0005 (Figura 22). Este resultado indica que a respectiva linhagem possui certo nível de comprometimento na biossíntese de terpenos decorrente da ausência de IDI, o que pode afetar os próximos passos de reações da via.

A enzima FPPS, por sua vez, catalisa a reação de formação de FPP a partir de GPP e uma molécula de IPP. O FPP gerado é utilizado como substrato para síntese de GGPP pela enzima GGPPS com participação de outra molécula de IPP. O FPP dá origem aos triterpenos e sesquiterpenos, enquanto GGPP origina diterpenos e tetraterpenos (SONG et al., 2020) (Figura 3). A análise feita utilizando o pBLAST detectou a presença das enzimas da família poliprenil sintase em cinco linhagens SDF (*L. fusiformis* SDF0005, *B. oleronius* SDF0015, *B. simplex* SDF0024, *B. velezensis* SDF0141 e *B. velezensis* SDF0150) (Tabela 5). O alto índice de similaridade obtido na análise aponta de fato para presença de sequências codificadoras desta enzima nas respectivas linhagens, com exceção de *B. oleronius* SDF0015, que apresentou índice de 66% (Tabela 5). Essa constatação indica que essa linhagem não possui informação genética para realizar a síntese desta enzima ou apresenta um catalisador estruturalmente diferenciado

quando comparado aquelas presentes nas Bafes em geral, decorrente das características singulares reportadas para o *Heyndrickxia oleronius*, anteriormente classificado como *B. oleronius* (HEYNDRICKX et al., 2012; GUPTA et al., 2020).

Qualquer enzima da família poliprenil sintase possui a capacidade de sintetizar a formação de GPP, FPP ou GGPP a partir de IPP (Figura 3). Esta habilidade é alcançável em razão dos diferentes níveis de homologia funcional observados nestas enzimas, para além da utilização de IPP como substrato comum às reações (SUN et al., 2011; VATTEKKATTE et al., 2018). Dessa forma, é possível imaginar que as linhagens SDF que não apresentaram pelo menos uma enzima da família poliprenil sintase, possuam algum tipo de deficiência na produção de terpenos, visto que não portam o aparato enzimático para sintetizar os precursores intermediários dos isoprenoides.

O GPP, FPP ou GGPP gerados são utilizados por diferentes TS como substratos da síntese dos mais variados terpenos encontrados na natureza. A sequência que codifica a enzima FES foi identificada pelo antiSMASH em, pelo menos, sete linhagens SDF (Tabela 4). O esqualeno é sintetizado a partir de duas moléculas de FPP, enquanto o fitoeno é sintetizado por duas moléculas de GGPP (NAKASHIMA et al., 1995; TANSEY; SHECHTER, 2000). A literatura científica aponta que esqualenos sintases (SQS) e fitoeno sintases (PHS) são, na realidade, enzimas bastante relacionadas (TANSEY; SHECHTER, 2000). Inclusive, são noticiadas situações em que SQS catalisam a reação de formação de fitoeno (NAKASHIMA et al., 1995). Esse fato explica por que a enzima detectada pelo antiSMASH foi relacionada tanto na síntese de esqualeno como de fitoeno. Dessa maneira, as linhagens SDF que apresentaram informação que codifica a enzima da família poliprenil sintase e o gene para expressão de FES possuem todo o arcabouço genético para síntese de esqualeno e fitoeno, situação reportada para *B. simplex* SDF0024, *B. velezensis* SDF0141 e *B. velezensis* SDF0150 (Tabelas 4 e 5).

O esqualeno é utilizado como agente de atividade antioxidante, antitumoral e pela capacidade de fortalecer o sistema imunológico em humanos (BATISTA et al., 2014). As propriedades deste terpeno permitem a utilização como aditivo ou suplemento na indústria de alimentos e cuidados pessoais (SONG et al., 2020). Por esta razão, a otimização da produção industrial de esqualeno em larga escala é buscada constantemente (BATISTA et al., 2014). A síntese deste terpeno é reportada em vários organismos, incluindo em Bafes (SONG et al., 2020), o que corrobora os resultados obtidos neste estudo. Além das propriedades destacadas, o esqualeno é também um intermediário na via de biossíntese de esteróis (SIEDENBURG; JENDROSSEK, 2011).

Hopanóides são sintetizados a partir do esqualeno na presença da enzima SHC (SIEDENBURG; JENDROSSEK, 2011). Esta enzima é codificada pela sequência do gene *sqhC*, detectada pelo antiSMASH em oito linhagens SDF neste estudo (Tabela 4). Portanto, as linhagens que podem sintetizar as enzimas FES e SHC possuem o aparato responsável pelas reações enzimáticas finais da síntese de hopanóides, casos detectados para *B. pumilus* SDF0011; *Bacillus* sp. SDF0016; *B. simplex* SDF0024; *B. velezensis* SDF0141 e *B. velezensis* SDF0150 (Tabela 4). Os hopanóides são terpenos que se integram às membranas biológicas da célula produtora e regulam a fluidez e a permeabilidade. Por conseguinte, são responsáveis pela estabilização da membrana citoplasmática em bactérias, fato relevante frente a ausência de colesterol na membrana destes procariotos (SIEDENBURG; JENDROSSEK, 2011). A ausência de hopanóides não influencia o crescimento bacteriano, porém, afeta a tolerância às condições de estresse como temperaturas elevadas, ambientes anaeróbicos ou extremamente ácidos (NEVES, 2015).

A distribuição de hopanóides nas linhagens SDF sugeriu um fato considerável, visto que apenas os membros pertencentes ao gênero *Bacillus* apresentaram informação genética tanto para a síntese de FES quanto para SHC (Tabela 4). Efetivamente, estudos apontam que determinantes genéticos para a produção de hopanóides estão presentes em membros deste gênero, assim como em membros do gênero *Paenibacillus*, embora a linhagem SDF0028 não foi apontada como detetora da respectiva informação envolvida na síntese de hopanóide (DEINHARD et al., 1987; SADAF et al., 2016; TILBURG et al., 2022). Do nosso conhecimento, para o gênero *Lysinibacillus*, não há estudos aprofundados disponíveis envolvendo a detecção de hopanóides. Investigar a presença ou não desses terpenos na composição da membrana citoplasmática em *Lysinibacillus* spp. constituem alvo de possíveis investigações futuras, a fim de elucidar a natureza biológica desta estrutura celular neste gênero de Bafes.

Os *clusters* biossintéticos que foram detectados os genes envolvidos na síntese de FES e SHC diferiram nas linhagens SDF avaliadas (Figuras 16 e 17), indicando que os genes envolvidos na produção de hopanóides em Bafes estão localizados em regiões distintas do genoma e associados à genes diferentes de acordo com a espécie. Desta forma, a detecção de um *BGC* de terpeno pelo antiSMASH não explicita que a respectiva sequência seja de fato um *cluster* envolvido na produção desses isoprenóides. De fato, a informação genética que codifica hopanóides não se apresentou em formato de *BGC* nas linhagens SDF investigadas. Os hopanóides são classificados como terpenos do ponto de vista estrutural, o que explica a detecção destas moléculas como metabólito especializado pelo *software*. Contudo, este tópico

pode ser alvo de discussão, pois hopanóides embora não sejam essenciais para o organismo produtor, não possuem os aparatos genéticos estruturais típicos para síntese de produtos do metabolismo especializado.

De maneira oposta, o gene que codifica FES nas linhagens *B. pumilus* SDF0011 e *Bacillus* sp. SDF0016 está adjacente ao gene *CrtI* que codifica a enzima PDS (Figura 18), indicando, neste caso, que o provável *BGC* tem informação que pode promover a síntese de carotenoide para as respectivas linhagens (Figura 19; GARCIA-ASUA et al., 1998). As respectivas sequências dos três genes localizados de maneira subsequente indicam que o *cluster* detectado pelo antiSMASH é de fato um *BGC* típico com informação que permite a síntese de metabólitos especializados. Nesta situação, a enzima FES correspondente detectada pode ser de fato uma PHS, que sintetiza fitoeno a partir de duas moléculas de GGPP (NAKASHIMA et al., 1995). O gene *CrtI* codifica a enzima PDS, responsável por remover dois átomos de hidrogênio da molécula de fitoeno a cada reação catalítica, promovendo a formação inicial de ligações duplas entre dois carbonos até a síntese do licopeno (GARCIA-ASUA et al., 1998). A presença de duas sequências de *CrtI* pode ser resultado de uma duplicação gênica, comum em procariotos (MADIGAN et al., 2016).

O licopeno é o precursor direto da síntese de betacaroteno, o que explica 50% de similaridade (Figura 19) obtido entre as respectivas sequências encontradas nas linhagens avaliadas, visto a ausência do gene *CrtY* em *B. pumilus* SDF0011 e *Bacillus* sp. SDF0016. Este gene codifica a enzima que realiza a conversão dos dois carotenoides (QIANG et al., 2019). Portanto, as duas linhagens em questão possuem o aparato genético completo para as reações enzimáticas finais implicadas na síntese de licopeno, que possui propriedades bastantes similares ao do betacaroteno (SABIO et al., 2003). O licopeno é amplamente produzido por plantas, sendo responsável pela pigmentação avermelhada característica de tomates e melancias, entre outros vegetais (LI et al., 2020). Este terpeno é reportado na literatura como antioxidante, anti-inflamatório e antitumoral, portanto, possui aplicação diversificada na indústria farmacêutica e alimentícia (LI et al., 2020).

Em Bafes, a síntese de licopeno é reportada apenas em casos de expressão heteróloga em *B. amyloliquefaciens* e *B. subtilis* (ZOU et al., 2022; LUO; BAO; ZHU, 2023). O resultado obtido neste estudo potencializa as propriedades biotecnológicas destas duas linhagens SDF e abre margem para investigações futuras sobre produção de licopeno em *B. pumilus* e espécies relacionadas. A presença das enzimas dedicadas à síntese do licopeno problematiza o caráter evolutivo desta molécula em Bafes, seja como produto resultante de uma característica apomórfica em determinadas espécies ou como produto gerado a partir de aquisição de

informação genética por transferência horizontal de genes (MADIGAN et al., 2016). Ademais, novos estudos podem elucidar a importância funcional que o licopeno desempenha para o metabolismo de Bafes ou o papel que esta molécula possui para os organismos produtores em ecossistemas naturais.

Em síntese, o potencial genômico para produção de terpenos nas dez linhagens SDF sondadas apontou informação relacionadas com síntese das enzimas da via MEP, responsável pela síntese de IPP, em todos os genomas analisados (Figuras 20 e 21), o que indica que as linhagens SDF possuem os determinantes genéticos para a produção do aparato enzimático básico para síntese de terpenos. O DMAPP, que é gerado no processo catalítico pela enzima IDI esteve ausente na linhagem *L. fusiformis* SDF0005 (Figura 22). As enzimas da família poliprenil sintase, que convertem a formação de GPP, FPP e GGPP a partir de IPP e DMAPP foram encontradas em cinco linhagens avaliadas (Tabela 5). A ausência nas demais linhagens pode comprometer as reações subsequentes da síntese de terpenos. As linhagens SDF neste estudo alocadas no gênero *Bacillus* contêm a informação que codifica o aparato enzimático para a síntese do esqualeno, que é utilizado na síntese de hopanóides pela enzima SHC. Por outro lado, as linhagens *B. pumilus* SDF0011 e *Bacillus* sp. SDF0016 apresentaram as informações genéticas para a expressão das enzimas necessárias (FES e PDS) para síntese de licopeno a partir do fitoeno (Figura 18 e 19). A figura 23 resume as reações e as enzimas detectadas para biossíntese destes produtos presentes nos genomas das linhagens SDF avaliadas.

Embora os resultados obtidos para detecção das enzimas envolvidas na síntese de terpenos apontem que determinada linhagem SDF não possua alguma enzima necessária para a sequência das reações de síntese de um produto mencionado acima, não indica necessariamente que a respectiva linhagem não produza o terpeno em questão. Como discutido anteriormente, o presente estudo foi pautado em uma análise do potencial genômico para produção de metabólitos especializados. Dessa maneira, possíveis fatores adicionais como purificação, extração, sequenciamento e anotação dos genomas inadequados, pode levar a supressão de resultados. Além disso, as rotas biossintéticas de produção de terpenos são caracterizadas por enzimas promíscuas, que catalisam mais de uma reação (SUN et al., 2011; VATTEKKATTE et al., 2018), o que pode levar as linhagens SDF avaliadas a sintetizar o terpeno previsto, mesmo na ausência da informação que permite a produção de uma enzima não detectada da via em nossos resultados.

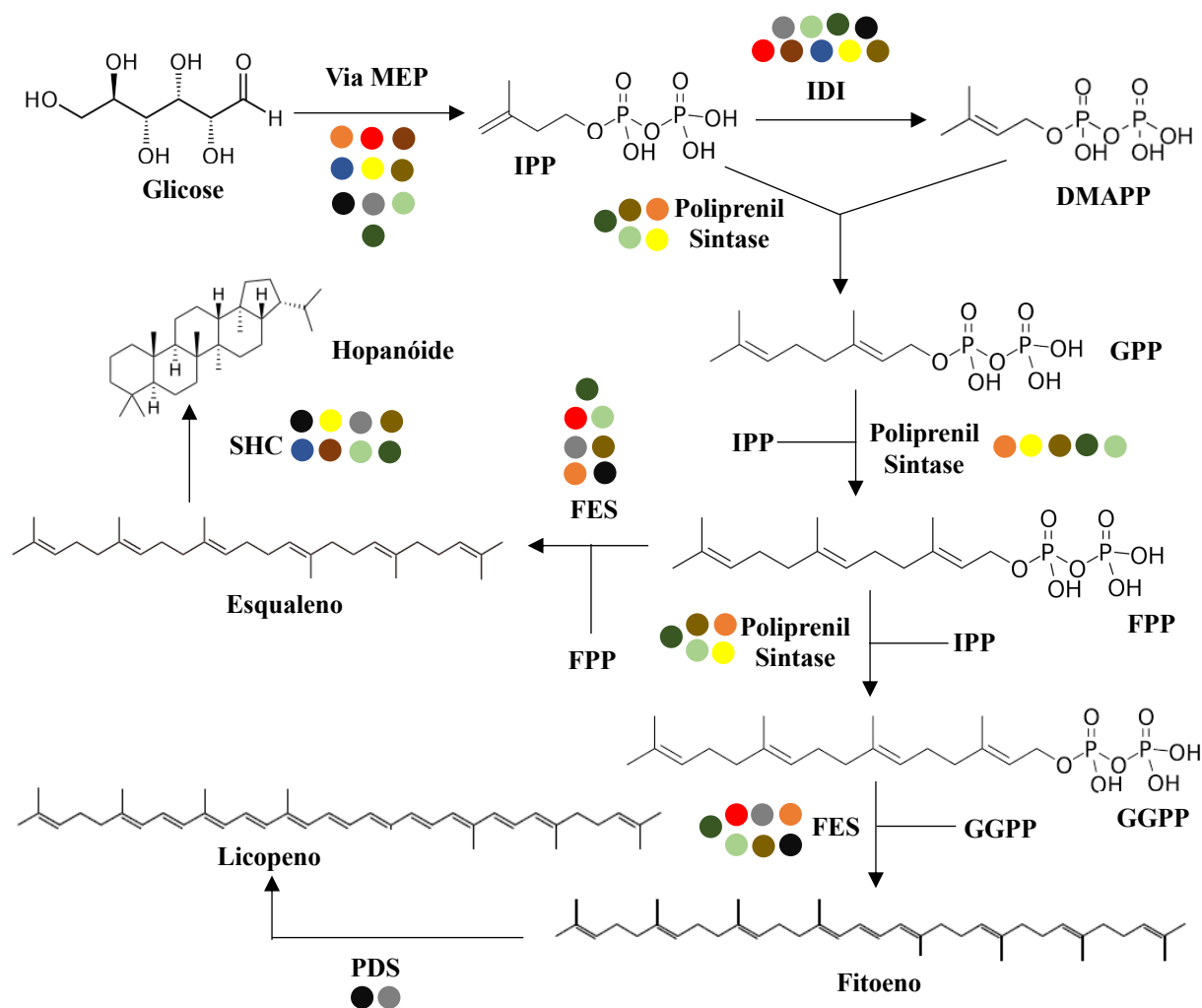


Figura 24. Mapa das reações catalíticas detectadas na biossíntese de terpenos nas linhagens SDF sondadas. Todas as linhagens apresentaram as enzima da via MEP. A enzima da família poliprenil sintase foi detectada nas linhagens SDF0005, SDF0015, SDF0024, SDF0141 e SDF0150. A enzima FES foi detectada nas linhagens SDF0005, SDF0011, SDF0016, SDF0024, SDF0037, SDF0141 e SDF0150. A enzima SHC foi detectado nas linhagens SDF0011, SDF0015, SDF0016, SDF0024, SDF0028, SDF0063, SDF0141 e SDF0150. As cores indicam a enzima detectada nas linhagens SDF de interesse

- *L. fusiformis* SDF0005 ● *B. pumilus* SDF0011 ● *B. oleronius* SDF0015 ● *Bacillus* sp. SDF0016
- *B. simplex* SDF0024 ● *Paenibacillus* sp. SDF0028 ● *Lysinibacillus* sp. SDF0037 ● *B. velezensis* SDF0141
- *B. velezensis* SDF0150 ● *Lysinibacillus* sp. SDF0063

7.5. Caráter evolutivo da produção de terpenos em Bafes

A sequência de aminoácidos de enzimas de TS detectadas neste estudo foram alinhadas para construção de árvores filogenéticas, com o propósito de elucidar o caráter evolutivo de enzimas envolvidas na produção de terpenos em Bafes. Foram alinhadas as sequências que dirigem a expressão das enzimas FES e SHC (Figuras 20 e 21), enquanto o alinhamento de

sequências codificadoras da enzima PDS não foi efetuado, em razão da presença em apenas duas linhagens SDF (Tabela 4). As enzimas SHC de *B. velezensis* SDF0140 e SDF0150 se apresentaram mais relacionadas, como esperado, visto que ambas as linhagens pertencem a mesma espécie (Figura 20). A SHC das linhagens *B. pumilus* SDF0011 e *Bacillus* sp. SDF0016, em seguida, foram as que apresentaram maior grau de similaridade com a enzima de *B. velezensis* (Figura 20). Este resultado está de acordo com a literatura, visto que *B. velezensis* e *B. pumilus* pertencem ao complexo do *B. subtilis* (BHANDARI et al., 2013), portanto espécies molecularmente relacionadas. Este resultado indica também que a respectiva linhagem de espécie não determinada é uma espécie de Bafes mais próxima evolutivamente de *B. pumilus* como membros do clado IV do complexo do *B. subtilis*, que engloba *B. safensis*, *B. altitudinis*, *B. xiamenensis*, *B. aerophilus* ou de outros clados do complexo (FRITZE, 2004; BHANDARI et al., 2013).

No segundo clado principal gerado na árvore filogenética, as linhagens *B. simplex* SDF0024 e *B. oleronius* SDF0015 apresentaram as sequências envolvidas na síntese da enzima SHC mais relacionadas. As duas espécies, de fato, não pertencem ao complexo do *B. subtilis*, o que explica a acomodação das espécies em clados distintos (Figura 20). O estudo de Darko et al., (2020), aponta que a espécie *B. oleronius* é próxima evolutivamente de *L. sphaericus*, o que converge com os resultados obtidos neste estudo, visto que a SHC de *B. oleronius* SDF0015 é mais relacionada com a SHC de *Lysinibacillus* sp. SDF0063 do que da linhagem *Paenibacillus* sp. SDF0028 (Figura 20). Além disso, *Paenibacillus* sp. SDF0028 é a única linhagem neste estudo que não pertence a família *Bacillaceae*, o que pode explicar a SHC com estrutura menos característica.

A árvore filogenética obtida também converge com outros estudos realizados pelo nosso grupo de pesquisa corroborando as relações filogenéticas entre linhagens SDF a partir de sequências de genes de rRNA 16S (Figura 6; OREM et al., 2020). Neste estudo, as espécies de Bafes apresentaram o mesmo padrão de distribuição filogenética reportada para as oito linhagens SDF produtoras de SHC (Figura 20), o que pode indicar que o potencial para a produção de terpenos nestas linhagens de Bafes é influenciado por fatores filogenéticos. De modo similar, a árvore filogenética gerada utilizando a enzima FES como marcador molecular (Figura 21) revelou o mesmo padrão de distribuição filogenética entre as linhagens SDF. Neste caso, as linhagens SDF0005 e SDF0037 alocadas no mesmo gênero (*Lysinibacillus*) possuem a enzima FES mais relacionadas, como esperado.

A similaridade do conjunto enzimático detectado para produção de terpenos, por sua vez, reforçou a especulação sobre uma relação mais próxima entre as linhagens *Bacillus* sp.

SDF0016 e *B. pumilus* SDF0011 (Figura 22). A enzima PDS foi identificada apenas nessas duas linhagens, o que explica o mesmo nível de similaridade com enzimas detectadas em ambas as espécies. O aparato genético de síntese para licopeno detectado indica que, possivelmente, a produção desta molécula pode ser uma característica compartilhada entre *B. pumilus* e as espécies mais relacionadas dentro do complexo do *B. subtilis* (FRITZE, 2004; BHANDARI et al., 2013; MARTINS et al., 2023). Embora as linhagens *B. velezensis* SDF0141 e SDF0150, assim como as linhagens *L. fusiformis* SDF0005 e *Lysinibacillus* sp. SDF0037 apresentaram o mesmo conjunto enzimático ou similar, convergindo com os resultados esperados para 16S (Figura 22), esta análise aponta que a capacidade de produção dos terpenos identificados pode variar entre as Bafes e não é totalmente determinado por fatores filogenéticos. Este fato é consumado, pois a similaridade do conjunto enzimático entre as linhagens SDF não seguiu o mesmo padrão encontrado nas árvores filogenéticas baseado nas sequências de aminoácidos de TS detectadas.

As sequências gênicas envolvidas na síntese das enzimas da via MEP foram todas identificadas nas dez linhagens avaliadas neste estudo (Figura 19), o que sugere que essa via é compartilhada por vários membros de Bafes. Os determinantes genéticos que dirigem a síntese das enzimas FES e SHC se mostraram presentes em grande parte das linhagens SDF sondadas (Tabela 4). A grande distribuição desses catalisadores demonstradas no *heatmap* (Figura 22) sugere que a via biossintética para produção de hopanóides é amplamente distribuída nas linhagens SDF e, potencialmente, para Bafes em geral. Este achado demonstra que esses tipos de terpenos desempenham função indispensável para estes microrganismos na estabilidade de membranas biológicas (SIEDENBURG; JENDROSSEK, 2011).

Em síntese, os resultados obtidos neste estudo mostram que as enzimas TS responsáveis pela produção de terpenos nas linhagens investigadas são conservadas evolutivamente e são influenciadas por aspectos filogenéticos. No entanto, a capacidade de produção ou não de diferentes terpenos por estes microrganismos é dinâmica e independe da filogenia global de Bafes.

8. CONSIDERAÇÕES FINAIS

Os terpenos constituem uma classe relevante de metabólitos especializados. Embora a síntese destas moléculas seja demonstrada em procariotos, pesquisas que dispõem sobre a produção destes compostos em Bafes ainda são escassos. Este estudo mostrou por análise *in silico*, que as linhagens SDF avaliadas possuem o aparato genético básico para codificar as enzimas da via MEP, responsável pela produção de IPP, precursor essencial na biossíntese de terpenos. As enzimas da família proliprenil sintase foram reveladas em cinco das dez linhagens sondadas, o que permite este subgrupo sintetizar GPP, FPP ou GGPP a partir de IPI. Apesar da ausência de informação genética detectada nas demais linhagens, não se pode descartar o potencial para a produção de terpenos nestas situações, seja pelo caráter catalítico plural de enzimas envolvidas nas vias de síntese destas moléculas, ou pelas limitações instrumentais metodológicas.

As sequências codificadoras de TS detectadas pelo *software* antiSMASH dentre os *BGC* identificados apontaram que membros da CBafes possuem o potencial genômico para sintetizar terpenos como esqualeno, hopanóides e licopeno. Os hopanóides são reconhecidos pela capacidade de estabilização das membranas biológicas, enquanto o esqualeno e o licopeno são reportados na indústria para atividades de ação antioxidante, anti-inflamatória e antitumoral. Estes resultados ampliam o potencial biotecnológico de Bafes e motivam estudos futuros para ampliar estas análises para demais linhagens SDF depositadas na nossa Coleção. Além disso, possibilitam investigações futuras para verificar a produção *in vitro* dos compostos detectados, visto os fatores plurais que regulam a expressão gênica.

As sequências de aminoácidos de TS identificadas das linhagens sondadas revelou ainda que estas enzimas são conservadas evolutivamente quando comparadas às filogenias baseadas em sequências de genes de rRNA 16S, o que indica que o potencial de produção de terpenos em Bafes pode ser compartilhada evolutivamente por estes microrganismos. Apesar desta constatação, a similaridade do conjunto enzimático para produção de terpenos nas linhagens SDF apontou parcialmente o mesmo padrão esperado para as relações moleculares conhecida de Bafes, o que mostra que o potencial de síntese de terpenos varia entre os membros a partir de enzimas produzidas ou não pelas linhagens estudadas. Dessa maneira, o potencial de produção de terpeno por membros de Bafes é dinâmica, e não é determinada necessariamente apenas por fatores filogenéticos.

Além de terpenos, os resultados deste estudo mostraram que Bafes possuem potencial genômico promissor para sintetizar metabólitos especializados das classes de NRP e PK. A análise efetivada pelo antiSMASH apontou em linhagens investigadas, o arcabouço genético

completo para biossíntese de bacillibactina, bacilisina, bacillaeno, macrobrevina, macrolactinas e paenibacterina, além de outros metabólitos especializados potenciais. Em maioria estes metabólitos são reportados como agentes de ação antimicrobiana, o que torna as Bafes promissoras na investigação destes novos agentes de combate a infecções, indispensáveis na área clínica. Os resultados obtidos complementam outros estudos desenvolvidos pelo nosso grupo de pesquisa e ampliam o conhecimento fenotípico e genotípico das linhagens depositadas na CBafes, estendendo a capacidade de exploração do potencial biotecnológico de linhagens SDF para os mais variados fins.

9. REFERÊNCIAS BIBLIOGRÁFICAS

- AHMED, Iftikhar et al. Proposal of *Lysinibacillus boronitolerans* gen. nov. sp. nov., and transfer of *Bacillus fusiformis* to *Lysinibacillus fusiformis* comb. nov. and *Bacillus sphaericus* to *Lysinibacillus sphaericus* comb. nov. **International Journal of Systematic and Evolutionary Microbiology**, v. 57, n. 5, p. 1117-1125, 2007.
- ALETI, Gajender et al. Identification of the bacterial biosynthetic gene clusters of the oral microbiome illuminates the unexplored social language of bacteria during health and disease. **MBio**, v. 10, n. 2, p. e00321-19, 2019.
- ALETI, Gajender; SESSITSCH, Angela; BRADER, Günter. Genome mining: prediction of lipopeptides and polyketides from *Bacillus* and related Firmicutes. **Computational and structural biotechnology journal**, v. 13, p. 192-203, 2015.
- ASH, Carol et al. Phylogenetic heterogeneity of the genus *Bacillus* revealed by comparative analysis of small-subunit-ribosomal RNA sequences. **Letters in Applied Microbiology**, v. 13, n. 4, p. 202-206, 1991.
- ASH, Carol; PRIEST, Fergus, G.; COLLINS, David, M. Molecular identification of rRNA group 3 *Bacilli* (Ash, Farrow, Wallbanks and Collins) using a PCR probe test: proposal for the creation of a new genus *Paenibacillus*. **Antonie van leeuwenhoek**, v. 64, p. 253-260, 1993.
- BATISTA, Irineu et al. Produção de esqualeno e ácidos gordos polinsaturados por microrganismos do grupo dos *Thraustochytrids*. **biotecnologia**, p. 44, 2014.
- 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, 2013.
- BILLS, Gerald, F.; GLOER, James, B. Biologically active secondary metabolites from the fungi. **The fungal kingdom**, p. 1087-1119, 2017.
- BLIN, Kai et al. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. **Nucleic acids research**, v. 47, n. 1, p. 81-87, 2019.
- BLIN, Kai et al. antiSMASH 6.0: improving cluster detection and comparison capabilities. **Nucleic acids research**, v. 49, n. 1, p. 29-35, 2021.
- BOGDANOVE, Adam, J. et al. Two new complete genome sequences offer insight into host and tissue specificity of plant pathogenic *Xanthomonas* spp. **Journal of Bacteriology**, v. 193, n. 19, p. 5450-5464, 2011.
- BUTCHER, Rebecca, A. et al. The identification of bacillaene, the product of the PksX megacomplex in *Bacillus subtilis*. **Proceedings of the national academy of sciences**, v. 104, n. 5, p. 1506-1509, 2007.

- CABRAL, Célia; PITA, J. R. Alcaloides: relevância na farmácia e no medicamento. **Ciclo de exposições**, 2015.
- CAVALCANTE, Danilo, A. et al. Ultrastructural analysis of spores from diverse *Bacillales* species isolated from Brazilian soil. **Environ Microbiol Rep.** v. 11, n.2, p. 155 – 164, 2019.
- CHAKRABORTY, Kajal et al. Polyketide-derived macrobrevins from marine macroalga-associated *Bacillus amyloliquefaciens* as promising antibacterial agents against pathogens causing nosocomial infections. **Phytochemistry**, v. 193, p. 112983, 2022.
- CHENEBY, Dominique et al. 16S rDNA analysis for characterization of denitrifying bacteria isolated from three agricultural soils. **FEMS Microbiology Ecology**, v. 34, n. 2, p. 121-128, 2000.
- CHRISTIE, Graham. Initiation of germination in *Bacillus* and *Clostridium* spores. **Bacterial spores: current research and applications**, p. 89-106, 2012.
- COCHRANE, Stephen, A.; VEDERAS, John, C. Lipopeptides from *Bacillus* and *Paenibacillus* spp.: a gold mine of antibiotic candidates. **Medicinal research reviews**, v. 36, n. 1, p. 4-31, 2016.
- CUADRAT, Rafael, R.C. **Exploração da diversidade de policetídeo sintases (PKSs) ambientais**. 2010. Tese de Doutorado. Instituto Oswaldo Cruz-RJ.
- DARKO, Rodney et al. Comparative genome analysis of *Bacillus sporothermodurans* with its closest phylogenetic neighbor, *Bacillus oleronius*, and *Bacillus cereus* and *Bacillus subtilis* groups. **Microorganisms**, v. 8, n. 8, p. 1185, 2020.
- DEANE, Caitlin. Come-hither compounds. **Nature Chemical Biology**, v. 15, n. 7, p. 651-651, 2019.
- DEINHARD, G. et al. *Bacillus cycloheptanicus* sp. nov., a new thermoacidophile containing ω -cycloheptane fatty acids. **Systematic and Applied Microbiology**, v. 10, n. 1, p. 68-73, 1987.
- DE VOS Paul et al. *Bergey's Manual of Systematic Bacteriology*. 2ª Ed, Nova Iorque, Springer 2009. 1450 p.
- DRIKS, Adam. The *Bacillus* spore coat. **Phytopathology**, v. 94, n. 11, p. 1249-1251, 2004.
- DUNLAP, Christopher, A. et al. *Bacillus velezensis* is not a later heterotypic synonym of *Bacillus amyloliquefaciens*; *Bacillus methylotrophicus*, *Bacillus amyloliquefaciens* subsp. *plantarum* and '*Bacillus oryzicola*' are later heterotypic synonyms of *Bacillus velezensis* based on phylogenomics. **International journal of systematic and evolutionary microbiology**, v. 66, n. 3, p. 1212-1217, 2016.
- EHLING-SCHULZ, Monika; MESSELHÄUSSER, Ute. *Bacillus* “next generation” diagnostics: moving from detection toward subtyping and risk-related strain profiling. **Frontiers in Microbiology**, v. 4, p. 32, 2013.

- EHLING-SCHULZ, Monika; LERECLUS, Didier; KOEHLER, Theresa, M. The *Bacillus cereus* group: *Bacillus* species with pathogenic potential. **Microbiology spectrum**, v. 7, n. 3, p. 7.3. 6, 2019.
- ELTOKHY, Mohamed, A. et al. Exploring the Nature of the Antimicrobial Metabolites Produced by *Paenibacillus ehimensis* Soil Isolate MZ921932 Using a Metagenomic Nanopore Sequencing Coupled with LC-Mass Analysis. **Antibiotics**, v. 11, n. 1, p. 12, 2021.
- FALQUETO, Silvia, A. et al. *Bacillus* spp. metabolites are effective in eradicating *Aedes aegypti* (Diptera: *Culicidae*) larvae with low toxicity to non-target species. **Journal of Invertebrate Pathology**, v. 179, p. 107525, 2021.
- FRITZE, Dagmar. Taxonomy of the genus *Bacillus* and related genera: the aerobic endosporeforming bacteria. **Phytopathology**, v. 94, n. 11, p. 1245-1248, 2004.
- GALPERIN, Michael, Y. Genome Diversity of Spore-Forming Firmicutes. **Microbiol Spectr**, v. 1, n. 2, p. 1-15, 2013.
- GARCIA-ASUA, Guillermo et al. Carotenoid diversity: a modular role for the phytoene desaturase step. **Trends in plant science**, v. 3, n. 11, p. 445-449, 1998.
- GARRITY, George, M. et al. Systematic bacteriology. **The Proteobacteria, Part C: The Alpha-, Beta-, Delta-, and Epsilonproteobacteria, Bergey's Manual Trust, Department of Microbiology and Molecular Genetics**, v. 2, 2005.
- GIBBONS, N. E.; MURRAY, R. G. E. Proposals concerning the higher taxa of bacteria. **International Journal of Systematic and Evolutionary Microbiology**, v. 28, n. 1, p. 1-6, 1978.
- GIORNO, Rebecca et al. Morphogenesis of the *Bacillus anthracis* spore. **Journal of bacteriology**, v. 189, n. 3, p. 691-705, 2007.
- GLARE, Travis, R. et al. Phylogenetic determinants of toxin gene distribution in genomes of *Brevibacillus laterosporus*. **Genomics**, v. 112, n. 1, p. 1042-1053, 2020.
- GRADY, Elliot, N. et al. Current knowledge and perspectives of *Paenibacillus*: a review. **Microbial cell factories**, v. 15, p. 1-18, 2016.
- 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, 2020.
- GUSTAFSON, Kirk et al. The macrolactins, a novel class of antiviral and cytotoxic macrolides from a deep-sea marine bacterium. **Journal of the american chemical society**, v. 111, n. 19, p. 7519-7524, 1989.

- GUZMÁN-CHÁVEZ, Fernando et al. Engineering of the filamentous fungus *Penicillium chrysogenum* as cell factory for natural products. **Frontiers in microbiology**, v. 9, p. 2768, 2018.
- GUTBROD, Katharina; ROMER, Jill; DÖRMANN, Peter. Phytol metabolism in plants. **Progress in lipid research**, v. 74, p. 1-17, 2019.
- HAMDACHE, Ahlem et al. Comparative genome analysis of *Bacillus* spp. and its relationship with bioactive nonribosomal peptide production. **Phytochemistry reviews**, v. 12, p. 685716, 2013.
- HEILBRONNER, Simon; KRISMER, Bernhard; OESTERHELT, Heike, B.; PESCHEL, Andreas. The microbiome-shaping roles of bacteriocins. **Nature Reviews Microbiology**, v. 19, n.11, p. 726-739, 2021.
- HELFRICH, Eric, J.N.; PIEL, Jörn. Biosynthesis of polyketides by trans-AT polyketide synthases. **Natural product reports**, v. 33, n. 2, p. 231-316, 2016.
- HELFRICH, Eric, J.N. et al. Bipartite interactions, antibiotic production and biosynthetic potential of the *Arabidopsis* leaf microbiome. **Nature Microbiology**, v. 3, n. 8, p. 909-919, 2018.
- HEYNDRIKX, Marc et al. Emended descriptions of *Bacillus sporothermodurans* and *Bacillus oleronius* with the inclusion of dairy farm isolates of both species. **International journal of systematic and evolutionary microbiology**, v. 62, n. 2, p. 307-314, 2012.
- HILBERT, David, W.; PIGGOT, Patrick J. Compartmentalization of gene expression during *Bacillus subtilis* spore formation. **Microbiology and molecular biology reviews**, v. 68, n. 2, p. 234-262, 2004.
- HUANG, En; YOUSEF, Ahmed, E. The lipopeptide antibiotic paenibacterin binds to the bacterial outer membrane and exerts bactericidal activity through cytoplasmic membrane damage. **Applied and environmental microbiology**, v. 80, n. 9, p. 2700-2704, 2014.
- HUR, Gene, H.; VICKERY, Christopher, R.; BURKART, Michael, D. Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. **Natural product reports**, v. 29, n. 10, p. 1074-1098, 2012.
- HUU-DAT, Huu et al. Diversity and antimicrobial activity of Vietnamese sponge-associated bacteria. **Marine drugs**, v. 19, n. 7, p. 353, 2021.
- ISLAM, Tarequl et al. Biosynthesis, Molecular Regulation, and Application of Bacilysin Produced by *Bacillus* Species. **Metabolites**, v. 12, n. 5, p. 397, 2022.
- JAHNE, Jennifer et al. Novel Plant-Associated *Brevibacillus* and *Lysinibacillus* Genomespecie Harbor a Rich Biosynthetic Potential of Antimicrobial Compounds. **Microorganisms**, v. 11, n. 1, p. 168, 2023.
- JARMUDA, Stanisław et al. Correlation between serum reactivity to Demodex-associated *Bacillus oleronius* proteins, and altered sebum levels and Demodex populations in

- erythematotelangiectatic rosacea patients. **Journal of medical microbiology**, v. 63, n. 2, p. 258-262, 2014.
- KANG, Beom, R.; PARK, Joon, S.; JUNG, Woo-Jin. Antifungal evaluation of fengycin isoforms isolated from *Bacillus amyloliquefaciens* PPL against *Fusarium oxysporum* f. sp. lycopersici. **Microbial Pathogenesis**, v. 149, p. 104509, 2020.
- KARP, Peter, D. et al. Pathway Tools version 23.0 update: software for pathway/genome informatics and systems biology. **Briefings in bioinformatics**, v. 22, n. 1, p. 109-126, 2021.
- KELLER, Nancy, P.; TURNER, Geoffrey; BENNETT, Joan, W. Fungal secondary metabolism - from biochemistry to genomics. **Nature Reviews Microbiology**, v. 3, n. 12, p. 937-947, 2005.
- KIESEWALTER, Heiko T. et al. Genomic and chemical diversity of *Bacillus subtilis* secondary metabolites against plant pathogenic fungi. **Msystems**, v. 6, n. 1, p. 00770-20, 2021.
- KUHNIGK, Thomas et al. *Bacillus oleronius* sp. nov., a member of the hindgut flora of the termite *Reticulitermes santonensis* (Feytaud). **Canadian journal of microbiology**, v. 41, n. 8, p. 699-706, 1995.
- KUMAR, Sudhir et al. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. **Briefings in bioinformatics**, v. 9, n. 4, p. 299-306, 2008.
- KUMAR, Sudhir et al. MEGA X: molecular evolutionary genetics analysis across computing platforms. **Molecular biology and evolution**, v. 35, n. 6, p. 1547, 2018.
- LARSON, Joakim, D.G. Antibiotics in the environment. **Upsala Journal of Medical Sciences**. v. 119, p.108–112, 2014.
- LEE, Jung, Y. et al. Regulation of petrobactin and bacillibactin biosynthesis in *Bacillus anthracis* under iron and oxygen variation. **PLoS One**, v. 6, n. 6, p. e20777, 2011.
- LEE, Namil et al. Mini review: genome mining approaches for the identification of secondary metabolite biosynthetic gene clusters in *Streptomyces*. **Computational and Structural Biotechnology Journal**, v. 18, p. 1548-1556, 2020.
- LI, Lei et al. Biotechnological production of lycopene by microorganisms. **Applied Microbiology and Biotechnology**, v. 104, p. 10307-10324, 2020.
- LIANG, Zijian; ZHI, Hang; FANG, Zhongxiang; ZHANG, Pangzhen. Genetic engineering of yeast, filamentous fungi and bacteria for terpene production and applications in food industry. **Food Research International**, v. 147, p. 110487, 2021.
- LIU, Zhenhua et al. Formation and diversification of a paradigm biosynthetic gene cluster in plants. **Nature Communications**, v. 11, n. 1, p. 5354, 2020.
- LOGAN, Niall, A. et al. Proposed minimal standards for describing new taxa of aerobic, endospore-forming bacteria. *Int J Syst Evol Microbiol*, v. 59, n.8, p. 2114-2121, 2009.

- LOGAN, Niall A.; HALKET, Gillian. Developments in the taxonomy of aerobic, endospore-forming bacteria. In: **Endospore-forming Soil Bacteria**. Springer, Berlin, Heidelberg, 2011. p. 1-29.
- LUDWIG, Wolfgang et al. Taxonomic outline of the phylum Firmicutes. **Bergey's Manual of Systematic bacteriology**, v. 2, p. 15-17, 2011.
- LUO, Hao; BAO, Yihong; ZHU, Ping. Development of a novel functional yogurt rich in lycopene by *Bacillus subtilis*. **Food Chemistry**, v. 407, p. 135142, 2023.
- MADIGAN, Michael T. et al. **Microbiologia de Brock-14ª Edição**. Artmed Editora, 2016.
- MAGALHÃES, Maria, I.A. **Avaliação do perfil toxigênico da Coleção de Bactérias aeróbias formadoras de endósporos isoladas de solo**. 2019. 83 p. Dissertação (Mestrado em Patologia Molecular), Universidade de Brasília, Brasília, 2019.
- MAHER, Amie; STAUNTON, Kara; KAVANAGH, Kevin. Analysis of the effect of temperature on protein abundance in Demodex-associated *Bacillus oleronius*. **Pathogens and disease**, v. 76, n. 4, p. fty032, 2018.
- MANDIC-MULEC, Ines; PROSSER, James I. Diversity of endospore-forming bacteria in soil: characterization and driving mechanisms. In: **Endospore-forming soil bacteria**. Springer, Berlin, Heidelberg, 2011. p. 31-59.
- MARESCA, Julia, A. et al. Identification of a fourth family of lycopene cyclases in photosynthetic bacteria. **Proceedings of the National Academy of Sciences**, v. 104, n. 28, p. 11784-11789, 2007.
- MARTINS, Paulo, H.R. et al. Protein profiling as a tool for identifying environmental aerobic endospore-forming bacteria. **Open J Bac**, v. 4, n.1, p. 01-07, 2020.
- MARTINS, Paulo, H. R. et al. Biochemical, physiological, and molecular characterisation of a large collection of aerobic endospore-forming bacteria isolated from Brazilian soils. **ARPHA Preprints**, v. 3, p. 53-72, 2023.
- MEDEMA, Marnix, H. et al. Minimum information about a biosynthetic gene cluster. **Nature chemical biology**, v. 11, n. 9, p. 625-631, 2015.
- MIETHKE, Marcus et al. Inhibition of aryl acid adenylation domains involved in bacterial siderophore synthesis. **The FEBS journal**, v. 273, n. 2, p. 409-419, 2006.
- MOIR, Anne. How do spores germinate? **Journal of applied microbiology**, v. 101, n. 3, p. 526-530, 2006.
- MONDOL, Muhammad A.M.; SHIN, Hee J.; ISLAM, Mohammad T. Diversity of secondary metabolites from marine *Bacillus* species: chemistry and biological activity. **Marine drugs**, v. 11, n. 8, p. 2846-2872, 2013.
- MOULOUD, Ghadbane; DAOUD, Harzallah; BASSEM, Jaouadi; ATEF, Ibn, L.; HANI,

- Belhadj. New Bacteriocin from *Bacillus clausii* S trainGM17: Purification, Characterization, and Biological Activity. **Applied biochemistry and biotechnology**, v. 171, p. 2186-2200, 2013.
- NAGAO, Tomokazu et al. Novel macrolactins as antibiotic lactones from a marine bacterium. **The Journal of antibiotics**, v. 54, n. 4, p. 333-339, 2001.
- NAKASHIMA, Takayuki et al. Cloning, expression, and characterization of cDNAs encoding *Arabidopsis thaliana* squalene synthase. **Proceedings of the National Academy of Sciences**, v. 92, n. 6, p. 2328-2332, 1995.
- NANNAN, Catherine et al. Bacilysin within the *Bacillus subtilis* group: Gene prevalence versus antagonistic activity against Gram-negative foodborne pathogens. **Journal of Biotechnology**, v. 327, p. 28-35, 2021.
- NÚÑEZ, Mario, A.; LÓPEZ, Víctor, E. Nonribosomal peptides synthetases and their applications in industry. **Sustainable Chemical Processes**, v. 4, p. 1-8, 2016.
- OREM, Juliana C. et al. Phylogenetic diversity of aerobic spore-forming *Bacillales* isolated from Brazilian soils. **International Microbiology**, v.22, n.4, p. 511 – 520, 2019.
- OREN, Aharon; GARRITY, George, M. Valid publication of the names of forty-two phyla of prokaryotes. **International Journal of Systematic and Evolutionary Microbiology**, v. 71, n. 10, p. 005056, 2021.
- ORTIZ, Aurelio; SANSINENEA, Estibaliz. Macrolactin antibiotics: amazing natural products. **Mini Reviews in Medicinal Chemistry**, v. 20, n. 7, p. 584-600, 2020.
- ÖZCENGİZ, Gülay; ÖĞÜLÜR, İsmail. Biochemistry, genetics and regulation of bacilysin biosynthesis and its significance more than an antibiotic. **New biotechnology**, v. 32, n. 6, p. 612-619, 2015.
- OSMAN, Kamelia, M. et al. Poultry and beef meat as potential seedbeds for antimicrobial resistant enterotoxigenic *Bacillus* species: a materializing epidemiological and potential severe health hazard. **Scientific reports**, v. 8, n. 1, p. 1-15, 2018.
- OTHOUM, Ghofran et al. In silico exploration of Red Sea *Bacillus* genomes for natural product biosynthetic gene clusters. **BMC genomics**, v. 19, p. 1-11, 2018.
- PALEY, Suzanne, M.; KARP, Peter, D. The pathway tools cellular overview diagram and omics viewer. **Nucleic acids research**, v. 34, n. 13, p. 3771-3778, 2006.
- PARK, Kyung, M. et al. Prevalence, enterotoxin genes, and antibiotic resistance of *Bacillus cereus* isolated from raw vegetables in Korea. **Journal of food protection**, v. 81, n. 10, p. 1590-1597, 2018.
- PATEL, Pramathesh, S. et al. Bacillaene, a novel inhibitor of prokaryotic protein synthesis produced by *Bacillus subtilis*: production, taxonomy, isolation, physico-chemical characterization and biological activity. **The Journal of antibiotics**, v. 48, n. 9, p. 997-1003, 1995.

- PAULO, Bruno, S.; SIGRIST, Renata; OLIVEIRA, Luciana G. Avanços Recentes Em Biossíntese Combinatória De Policetídeos: Perspectivas E Desafios. **Química Nova**, v. 42, p. 71-83, 2019.
- PRANAV, Parameswaran, S. et al. Whole-Genome Sequence Analysis of *Paenibacillus alvei* JR949 Revealed Biosynthetic Gene Clusters Coding for Novel Antimicrobials. **Current Microbiology**, v. 78, n. 4, p. 1168-1176, 2021.
- QIANG, Shan et al. Elevated β -carotene synthesis by the engineered *Rhodobacter sphaeroides* with enhanced CrtY expression. **Journal of agricultural and food chemistry**, v. 67, n. 34, p. 9560-9568, 2019.
- QUIN, Maureen, B.; FLYNN, Christopher M.; SCHMIDT-DANNERT, Claudia. Traversing the fungal terpenome. **Natural product reports**, v. 31, n. 10, p. 1449-1473, 2014.
- RABBEE, Muhammad, F. et al. *Bacillus velezensis*: a valuable member of bioactive molecules within plant microbiomes. **Molecules**, v. 24, n. 6, p. 1046, 2019.
- REDDICK, Jason, J.; ANTOLAK, Stephanie, A.; RANER, Gregory, M. PksS from *Bacillus subtilis* is a cytochrome P450 involved in bacillaene metabolism. **Biochemical and biophysical research communications**, v. 358, n. 1, p. 363-367, 2007.
- SABIO, E. et al. Lycopene and β -carotene extraction from tomato processing waste using supercritical CO₂. **Industrial & engineering chemistry research**, v. 42, n. 25, p. 6641-6646, 2003.
- SADAF, K. et al. *Paenibacillus arachidis* sp. nov., isolated from groundnut seeds. **International Journal of Systematic and Evolutionary Microbiology**, v. 66, n. 8, p. 2923-2928, 2016.
- SÁNCHEZ, Lara, M.; SINGH, Kumar, S.; AVALOS, Mariana; WEZEL, Gilles, P.; DICKSCHAT, Jeroen, S.; GARBEVA, Paolina. Phylogenomic analyses and distribution of terpene synthases among *Streptomyces*. **Beilstein journal of organic chemistry**, v. 15, n. 1, p. 1181-1193, 2019.
- SATO, Tsutomu; YOSHIDA, Satoru; HOSHINO, Hiroko; TANNO, Mizuk; NAKAJIMA, Mami; HOSHINO, Tsoyomu. Sesquiterpenes (C₁₅ terpenes) biosynthesized via the cyclization of a linear C₁₅ isoprenoid by a tetraprenyl- β -curcumene synthase and a tetraprenyl- β -curcumene cyclase: identification of a new terpene cyclase. **Journal of the American Chemical Society**, v. 133, n. 25, p. 9734-9737, 2011.
- SATO, Tsutomu. Unique biosynthesis of sesquiterpenes (C₁₅ terpenes). **Bioscience, biotechnology, and biochemistry**, v. 77, n. 6, p. 1155-1159, 2013.
- SCHWARZER, Dirk; FINKING, Robert; MARAHIEL, Mohamed, A. Nonribosomal peptides: from genes to products. **Natural product reports**, v. 20, n. 3, p. 275-287, 2003.
- SHARRAR, Allison, M. et al. Bacterial secondary metabolite biosynthetic potential in soil varies with phylum, depth, and vegetation type. **Mbio**, v. 11, n. 3, p. 1-17, 2020.

- SHI, Jing et al. Discovery and biosynthesis of guanipiperazine from a NRPS-like pathway. **Chemical Science**, v. 12, n. 8, p. 2925-2930, 2021.
- SHIVAJI, S. et al. *Janibacter hoylei* sp. nov., *Bacillus isronensis* sp. nov. and *Bacillus aryabhatai* sp. nov., isolated from cryotubes used for collecting air from the upper atmosphere. **International journal of systematic and evolutionary microbiology**, v. 59, n. 12, p. 2977-2986, 2009.
- SHIDA, Osamu et al. Transfer of *Bacillus alginolyticus*, *Bacillus chondroitinus*, *Bacillus curdlanolyticus*, *Bacillus glucanolyticus*, *Bacillus kobensis*, and *Bacillus thiaminolyticus* to the genus *Paenibacillus* and emended description of the genus *Paenibacillus*. **International Journal of Systematic Bacteriology**, v. 47, n. 2, p. 289-298, 1997.
- SIEDENBURG, Gabriele; JENDROSSEK, Dieter. Squalene-hopene cyclases. **Applied and environmental microbiology**, v. 77, n. 12, p. 3905-3915, 2011.
- SILVA, Waldeyr, M.C. **Método para reconstrução in silico de redes metabólicas de fungos: um estudo de caso para o *Paracoccidioides lutzii***. 2014. 70 p. Dissertação (Mestrado em Informática), Universidade de Brasília, Brasília, 2014.
- SINGH, Bhim, P.; RATEF, Mostafa, E.; COUTO, Susana, R.; POLIZELI, Maria, L.; LI, Wen J. Microbial secondary metabolites: Recent developments and technological challenges. **Frontiers in microbiology**, v. 10, p. 914, 2019.
- SONG, Chunxu et al. Exploring the genomic traits of fungus-feeding bacterial genus *Collimonas*. **Bmc Genomics**, v. 16, p. 1-17, 2015.
- SONG, Yafeng et al. Production of squalene in *Bacillus subtilis* by squalene synthase screening and metabolic engineering. **Journal of agricultural and food chemistry**, v. 68, n. 15, p. 4447-4455, 2020
- SUMI, Chandra D.; YANG, Byung, W.; YEO, In-Cheol; HAMN, Young, T. Antimicrobial peptides of the genus *Bacillus*: a new era for antibiotics. **Canadian journal of microbiology**, v. 61, n. 2, p. 93-103, 2015.
- TANSEY, Terese, R.; SHECHTER, Ishaiahu. Squalene synthase: **structure and regulation**. 2000.
- TILBURG, Amanda Y. et al. Membrane composition and organization of *Bacillus subtilis* 168 and its genome-reduced derivative mini *Bacillus* PG10. **Microbial Biotechnology**, v. 15, n. 5, p. 1633-1651, 2022.
- TRAVIN, Dmitrii, Y. et al. Structure of ribosome-bound azole-modified peptide phazolicin rationalizes its species-specific mode of bacterial translation inhibition. **Nature communications**, v. 10, n. 1, p. 4563, 2019.
- TWAIJ, Baan, M.; HASAN, Nazmul, M. Bioactive secondary metabolites from plant sources: Types, synthesis, and their therapeutic uses. **International Journal of Plant Biology**, v. 13, n. 1, p. 4-14, 2022.

- TYC, Olaf; SONG, Chunxu; DICKSCHAT, Jeroen, S.; VOS, Michiel; GARBEVA, Paolina. The ecological role of volatile and soluble secondary metabolites produced by soil bacteria. **Trends in microbiology**, v. 25, n. 4, p. 280-292, 2017.
- URBAN-CHMIEL, Renata; MAREK, Agnieszka; PYSNIAK, Dagmara, S.; WIECZOREK, Kinga.; DEC, Marta; NOWACZEK, Nowa; OSEK, Jacek. Antibiotic resistance in bacteria—A review. **Antibiotics**, v. 11, n. 8, p. 1079, 2022.
- VAISHNAV, Preeti; DEMAIN, Arnold, L. Unexpected applications of secondary metabolites. **Biotechnology Advances**, v. 29, n. 2, p. 223-229, 2010.
- VAN DER AUWERA Geráldine, A.; TIMMERY, Sophie; HOTON, Florence; MAHILLON, Jacques. Plasmid exchanges among members of the *Bacillus cereus* group in foodstuffs. **Int J Food Microbiol**, v. 113, n.2, p. 164–172, 2007.
- VANITTANAKOM, Nongnuch et al. Fengycin-a novel antifungal lipopeptide antibiotic produced by *Bacillus subtilis* F-29-3. **The Journal of antibiotics**, v. 39, n. 7, p. 888-901, 1986.
- VATTEKKATTE, Abith et al. Enhanced structural diversity in terpenoid biosynthesis: enzymes, substrates and cofactors. **Organic & Biomolecular Chemistry**, v. 16, n. 3, p. 348-362, 2018.
- VILLEGAS-ESCOBAR, Valeska et al. Fengycin C produced by *Bacillus subtilis* EACB0015. **Journal of natural products**, v. 76, n. 4, p. 503-509, 2013.
- WALSH, Christopher, T.; FISCHBACH, Michael A. Natural products version 2.0: connecting genes to molecules. **Journal of the American Chemical Society**, v. 132, n. 8, p. 2469-2493, 2010.
- WILSON, Melissa, K. et al. Siderophores of *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*. **Biochemical and biophysical research communications**, v. 348, n. 1, p. 320-325, 2006.
- WINK, Michael. Potential of DNA intercalating alkaloids and other plant secondary metabolites against SARS-CoV-2 causing COVID-19. **Diversity**, v. 12, n. 5, p. 175, 2020.
- YAMADA, Yuuki; KUZUYAMA, Tomohisa; KOMATSU, Mamoru; IKEDA, Haruo. Terpene synthases are widely distributed in bacteria. **Proceedings of the National Academy of Sciences**, v. 112, n. 3, p. 857-862, 2015.
- YAN, Yuechen et al. Complete genome sequence of *Bacillus velezensis* YYC, a bacterium isolated from the tomato rhizosphere. **Archives of Microbiology**, v. 204, p. 1-5, 2022.
- YEH, Hsu-Hua et al. Molecular genetic analysis reveals that a nonribosomal peptide synthetase-like (NRPS-like) gene in *Aspergillus nidulans* is responsible for microperforanone biosynthesis. **Applied microbiology and biotechnology**, v. 96, p. 739748, 2012.
- ZEIGLER, Daniel, R.; PERKINS, John, B. The genus *Bacillus*. In: **Practical handbook of microbiology**. CRC Press, 2021. p. 249-278.

- ZEVALLOS, Delia, M. P. et al. Induced defenses of *Veronica spicata*: Variability in herbivore-induced volatile organic compounds. **Phytochemistry Letters**, v. 6, n. 4, p. 653-656, 2013.
- ZHANG, Pangzhen et al. Terpene evolution during the development of *Vitis vinifera* L. cv. Shiraz grapes. **Food Chemistry**, v. 204, p. 463-474, 2016.
- ZHAO, Xin; KUIPERS, Oscar, P. Identification and classification of known and putative antimicrobial compounds produced by a wide variety of *Bacillales* species. **BMC genomics**, v. 17, n. 1, p. 1-18, 2016.
- ZOU, Dian et al. Production of a novel lycopene-rich soybean food by fermentation with *Bacillus amyloliquefaciens*. **LWT**, v. 153, p. 112551, 2022.

10. APÊNDICE

KAB0443654.1 polyprenyl synthetase family protein [Lysinibacillus fusiformis]
 Sequence ID: Query_72593 Length: 293 Number of Matches: 1

Range 1: 1 to 293 [Graphics](#) [▼ Next Match](#) [▲ Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
591 bits(1523)	0.0	Compositional matrix adjust.	292/293(99%)	293/293(100%)	0/293(0%)
Query 1		MNEPLKYFIESNIPQIEETMFALVSKIDAPVDLKDSMLYSLKAGGKRIRPLFVLAVLELY			60
Sbjct 1		MNEPLKYFIESNIPQIEETMFALVSKIDAPVDLKDSMLYSLKAGGKRIRPLFVLAVLELY			60
Query 61		QKNLQDGLIVGSVIEIIHTYSLIHDDLPSMDNDDFRRGKPTNHKVYGEALATLAGDALNT			120
Sbjct 61		QKNLQDGLIVGSVIEIIHTYSLIHDDLPSMDNDDFRRGKPTNHKVYGEALATLAGDALNT			120
Query 121		LAFGILARMEVTAKKRIELVQLLSVAAGAEGMVGQVLDMEGEQRQLNLTELEQVHVNKT			180
Sbjct 121		LAFGILARMEVTAEKRIELVQLLSVAAGAEGMVGQVLDMEGEQRQLNLTELEQVHVNKT			180
Query 181		GALLRFSIEAGAVLADASKQDRATLKEYAHHIGLAFQIQDDILDIEGTTEELGKTAGKDV			240
Sbjct 181		GALLRFSIEAGAVLADASKQDRATLKEYAHHIGLAFQIQDDILDIEGTTEELGKTAGKDV			240
Query 241		ASDKSTYPALLTLNGAKEKLAEHYQHAINALDQLPIDASLLRDFAAAYIVHRKN			293
Sbjct 241		ASDKSTYPALLTLNGAKEKLAEHYQHAINALDQLPIDASLLRDFAAAYIVHRKN			293

Figura 25. Alinhamento de seqüências de aminoácidos para a detecção de enzima da família poliprenil sintase em *L. fusiformis* SDF0005

WP_268443628.1 polyprenyl diphosphate synthase, partial [Bacillus pumilus]
 Sequence ID: Query_63168 Length: 168 Number of Matches: 1

Range 1: 1 to 168 [Graphics](#) [▼ Next Match](#) [▲ Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
247 bits(630)	2e-84	Compositional matrix adjust.	114/168(68%)	139/168(82%)	0/168(0%)
Query 91		KMEVDYLMKLPQEFLSAFLPELIEENVKVEMMGHKDKLPVHTMKAVQKAMDDTCNNSGLI			150
Sbjct 1		K+EVD+LMKLP+EFL+ +LPELIEENV+V + G D LP HT +AV+ A+ T NN GL+			60
Query 151		LNFAALNYGGRDEIVNAVQNILQDAKNGQISTEDLTEETISHYLMKKNLQDPDLLIRTSGE			210
Sbjct 61		LNFAALNYGGR EIV+A + I + K G++ ED+TEE S YLM+++LQDPDLLIRTSGE			120
Query 211		IRLSNFMLWQLAYTEFVFTDVLWPDFSEEDLLEAIEEYQRRSRRYGGI			258
Sbjct 121		IRLSNFMLWQ+AY+EFVFTDVLWPDFS+E L+ AI EYQRR RR+GGI			168

Figura 26. Alinhamento de seqüências de aminoácidos para a detecção de enzima da família poliprenil sintase em *B. oleronius* SDF0015

WP_241589686.1 MULTISPECIES: polyprenyl synthetase family protein [unclassified Peribacillus]
 Sequence ID: **Query_492632** Length: 297 Number of Matches: 1

Range 1: 1 to 297 [Graphics](#) ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
593 bits(1530)	0.0	Compositional matrix adjust.	293/297(99%)	297/297(100%)	0/297(0%)
Query 1		MGTASFEMFSKEYKTVIEREIVEYVNLKAPAVVKEAMIYSLEAGGKRIRPLLFAVLEA			60
Sbjct 1		MGTASFEMFSKEYKTVIEREIVEYVNLKAPAVVKEAMIYSLEAGGKRIRPLLFAVLEA			60
Query 61		FGKSLRIGIPAAAAIEMIHTYSLIHDDLPMDDDLRRGKPTNHKVFGEAVAILAGDALL			120
Sbjct 61		FGKSLRIGIPAA+AIEMIHTYSLIHDDLPMDDDLRRGKPTNHKVFGEAVAILAGDALL			120
Query 121		TYSFQLVTDMDPEVTAEMKLNLVSEIAKSAGAEGMVGQVADMEGESKQLTLQELEYIH			180
Sbjct 121		TYSFQLVTDMDPEVTAEMKLNLVSEIAKSAGAEGMVGQVADMEGE+KQLTLQELEYIH			180
Query 181		EHKTGKLLTASILSGAILAGANEEQQLHLRDFAYHLGLAFQIRDDILDIEGSVELIGKPV			240
Sbjct 181		EHKTGKLLTASILSGAILAGANEEQQLHLRDFAYHLGLAFQIRDDILDIEGSVELIGKPV			240
Query 241		GSDVGNHKSTYPSLLTLQGAKEKLEHHIELAHAALGKTNLQTGLNELTDLIANRNH			297
Sbjct 241		GSDVGNHKSTYPSLLTLQGAKEKLEHHIELAHAALGKTNLQTGLNELTDLIANRNH			297

Figura 27. Alinhamento de seqüências de aminoácidos para a detecção de enzima da família poliprenil sintase em *B. simplex* SDF0024

ASK59031.1 polyprenyl synthetase family protein [Bacillus velezensis]
 Sequence ID: **Query_146288** Length: 296 Number of Matches: 1

Range 1: 1 to 296 [Graphics](#) ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
595 bits(1534)	0.0	Compositional matrix adjust.	296/296(100%)	296/296(100%)	0/296(0%)
Query 1		MTNKLDAFLASRKQLIEQQLSLYTERLAMPETLKKSMYLSLEAGGKRLRPLVVLAVLNAY			60
Sbjct 1		MTNKLDAFLASRKQLIEQQLSLYTERLAMPETLKKSMYLSLEAGGKRLRPLVVLAVLNAY			60
Query 61		GKDEKDGIPVGC AVE MIHTYSLIHDDLPCMDLRRGKPTNHKVFGEAAVLAGDGLLT			120
Sbjct 61		GKDEKDGIPVGC AVE MIHTYSLIHDDLPCMDLRRGKPTNHKVFGEAAVLAGDGLLT			120
Query 121		ESFKLITSHVSEAVSAEKRLRLVDELISAAGTEGMVGGQAADMEAEQKRITL E ELES IHE			180
Sbjct 121		ESFKLITSHVSEAVSAEKRLRLVDELISAAGTEGMVGGQAADMEAEQKRITL E ELES IHE			180
Query 181		RKTAKLLGFSVTAGAILADAPEDEVEKRLRFLSSHIGIGFQIRDDILDIEGHEERIGKPVG			240
Sbjct 181		RKTAKLLGFSVTAGAILADAPEDEVEKRLRFLSSHIGIGFQIRDDILDIEGHEERIGKPVG			240
Query 241		SDTSNEKSTYPSLLTLEGAKEKLAYHISEAKQIISGLSLQKELLHDLCDLIAARDH			296
Sbjct 241		SDTSNEKSTYPSLLTLEGAKEKLAYHISEAKQIISGLSLQKELLHDLCDLIAARDH			296

Figura 28. Alinhamento de seqüências de aminoácidos para a detecção de enzima da família poliprenil sintase em *B. velezensis* SDF0141

QWC45887.1 polyprenyl synthetase family protein [Bacillus velezensis]
 Sequence ID: Query_68209 Length: 286 Number of Matches: 1

Range 1: 1 to 286 [Graphics](#) [▼ Next Match](#) [▲ Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
577 bits(1487)	0.0	Compositional matrix adjust.	285/286(99%)	286/286(100%)	0/286(0%)
Query 27		VNKEVSKHVLNFADTKNNFPFAELALDHYLAFGGNDFEEILDGAGIELLILAADIFDDI			86
Sbjct 1		+NKEVSKHVLNFADTKNNFPFAELALDHYLAFGGNDFEEILDGAGIELLILAADIFDDI			60
Query 87		QDKDRPNAEWMKMDQAVSLNIATLIYTIALQVISSRDPEGLLLKTALKYTVQAMEGQHID			146
Sbjct 61		QDKDRPNAEWMKMDQAVSLNIATLIYTIALQVISSRDPEGLLLKTALKYTVQAMEGQHID			120
Query 147		LVDNTLTENQCIEMIKKSGALTALASMLGAVLATKRYQSDVEQYAFYIGVAAQLKNDLF			206
Sbjct 121		LVDNTLTENQCIEMIKKSGALTALASMLGAVLATKRYQSDVEQYAFYIGVAAQLKNDLF			180
Query 207		EIMYVNNTEFFKKTNLAIQFLEKKFNTVSTELLNNNFQQSELKRLLLESGVIHYMSVMI			266
Sbjct 181		EIMYVNNTEFFKKTNLAIQFLEKKFNTVSTELLNNNFQQSELKRLLLESGVIHYMSVMI			240
Query 267		NVYNLKVKNGLNLTLENKNYLIKKICENEKVDKNAGNSKLFSS		312	
Sbjct 241		NVYNLKVKNGLNLTLENKNYLIKKICENEKVDKNAGNSKLFSS		286	

Figura 29. Alinhamento de seqüências de aminoácidos para a detecção de enzima da família poliprenil sintase em em *B. velezensis* SDF0150

Quadro 1 – Tipos de metabólitos especializados identificados pelo antiSMASH a partir de *BGC* detectados

Label	Description	Added	Last updated
2dos	2-deoxy-streptamine aminoglycoside	7.0	7.0
acyl_amino_acids	N-acyl amino acid	4.0	4.1
aminocoumarin	Aminocoumarin	<= 3.0	<= 3.0
aminopolycarboxylic-acid	Aminopolycarboxylic acid metallophores (doi:10.1039/C8MT00009C)	7.0	7.0
amglycycl	Aminoglycoside/aminocyclitol	<= 3.0	<= 3.0
arylpolyene	Aryl polyene	<= 3.0	<= 3.0
betalactone	Beta-lactone containing protease inhibitor	5.0	5.0
blactam	β -lactam	<= 3.0	<= 3.0
bottromycin	Bottromycin	<= 3.0	<= 3.0
butyrolactone	Butyrolactone	<= 3.0	<= 3.0
CDPS	tRNA-dependent cyclodipeptide synthases	5.0	5.0
crocagin	Crocagin-like	7.0	7.0
cyanobactin	Cyanobactins like patellamide (AY986476)	<= 3.0	7.0
cyclic-lactone-autoinducer	agrD-like cyclic lactone autoinducer peptides (AF001782)	6.0	6.0
ectoine	Ectoine	<= 3.0	<= 3.0
epipeptide	D-amino-acid containing RiPPs such as yydF (D78193)	6.0	6.0
fatty_acid	Fatty acid (loose strictness, likely from primary metabolism)	<= 3.0	4.2
furan	Furan	<= 3.0	5.0
fungal-RiPP	Fungal RiPP with POP or UstH peptidase types and a modification	5.0	5.0
fungal-RiPP-like	Fungal RiPP-like	7.0	7.0
glycocin	Glycocin	<= 3.0	<= 3.0
guanidinotides	Pheganomycin-style protein ligase-containing cluster	4.0	6.0
halogenated	Halogenase-containing cluster, potentially generating a halogenated product	5.0	5.0

hgE-KS	Heterocyst glycolipid synthase-like PKS	5.0	5.0
hserlactone	Homoserine lactone	<= 3.0	<= 3.0
indole	Indole	<= 3.0	4.0
LAP	Linear azol(in)e-containing peptides	<= 3.0	6.0
ladderane	Ladderane	<= 3.0	<= 3.0
lanthipeptide class I	Class I lanthipeptides like nisin	4.2	6.0
lanthipeptide class II	Class II lanthipeptides like mutacin II (U40620)	4.2	6.0
lanthipeptide class III	Class III lanthipeptides like labyrinthopeptin (FN178622)	4.2	6.0
lanthipeptide class IV	Class IV lanthipeptides like venezuelin (HQ328852)	4.2	6.0
lanthipeptide class V	Glycosylated lanthipeptide/linaridin hybrids like MT210103	5.1	6.0
lassopeptide	Lasso peptide	<= 3.0	5.0
linaridin	Linear arid peptide such as cypemycin (HQ148718) and salinipeptin (MG788286)	<= 3.0	<= 3.0
lipolanthine	Lanthipeptide class containing N-terminal fatty acids such as MG673929	5.0	5.0
melanin	Melanin	<= 3.0	<= 3.0
methanobactin	Copper-chelating/transporting peptides (doi:10.1126/science.aap9437)	7.0	7.0
microviridin	Microviridin	<= 3.0	<= 3.0
mycosporine-like	Molecules containing mycosporine-like amino acid	7.0	7.0
NAGGN	N-acetylglutaminyglutamine amide	5.0	5.0
NAPAA	Non-alpha poly-amino acids like e-Polylysine	6.0	7.0
NI-siderophore	NRPS-independent, lucA/lucC-like siderophores (<i>siderophore</i> prior to 7.0)	7.0	7.0
NRPS	Non-ribosomal peptide synthetase	<= 3.0	6.0
NRPS-like	NRPS-like fragment	5.0	5.0
NRP-metallophore	Non-ribosomal peptide metallophores	7.0	7.0
nucleoside	Nucleoside	<= 3.0	5.0
oligosaccharide	Oligosaccharide	<= 3.0	<= 3.0
opine-like-metallophore	Opine-like zincophores like staphylopine (doi:10.1128/mSystems.00554-20)	7.0	7.0
other	Cluster containing a secondary metabolite-related protein that does not fit into any other category	4.0	5.0
PBDE	Polybrominated diphenyl ether	4.1	4.1
phenazine	Phenazine	<= 3.0	<= 3.0
phosphoglycolipid	Phosphoglycolipid	<= 3.0	<= 3.0
phosphonate	Phosphonate	<= 3.0	7.0
phosphonate-like	Phosphonate-like (prior to 7.0 this was the phosphonate rule)	7.0	7.0
PKS-like	Other types of PKS	5.0	5.0
PpyS-KS	PPY-like pyrone	4.2	4.2
prodigiosin	Serratia-type non-traditional PKS prodigiosin biosynthesis pathway	6.0	6.0
proteusin	Proteusin	<= 3.0	<= 3.0
PUFA	Polyunsaturated fatty acid	<= 3.0	<= 3.0
pyrrolidine	Pyrrolidines like described in BGC0001510	6.0	6.0
ranthipeptide	Cys-rich peptides (aka. SCIFF: six Cys in fourty-five) like in CP001581:3481278-3502939	6.0	6.0
RaS-RiPP	Streptide-like thioether-bond RiPPs	5.0	5.0
RCDPS	Fungal Arginine-containing cyclic dipeptides	7.0	7.0
redox-cofactor	Redox-cofactors such as PQQ (NC_021985:1458906-1494876)	6.0	6.0
resorcinol	Resorcinol	<= 3.0	<= 3.0
RiPP-like	Other unspecified ribosomally synthesised and post-translationally modified peptide product (RiPP)	4.1	7.0
RRE-containing	RRE-element containing cluster	6.0	6.0
saccharide	Saccharide (loose strictness, likely from primary metabolism)	<= 3.0	<= 3.0
sactipeptide	Sactipeptide	<= 3.0	6.0
spliceotide	RiPPs containing plpX type splicases (NZ_KB235920:17899-42115)	6.0	6.0
T1PKS	Type I PKS (Polyketide synthase)	<= 3.0	<= 3.0
T2PKS	Type II PKS	<= 3.0	5.0

T3PKS	Type III PKS	<= 3.0	<= 3.0
terpene	Terpene	<= 3.0	4.1
thioamitides	Thioamitide RiPPs as found in JOBFO1000011	5.1	6.0
thioamide-NRP	Thioamide-containing non-ribosomal peptide	5.0	5.0
thiopeptide	Thiopeptide	4.2	5.0
transAT-PKS	Trans-AT PKS	<= 3.0	5.0
transAT-PKS-like	Trans-AT PKS fragment, with trans-AT domain not found	<= 5.0	5.0
tropodithietic-acid	Tropodithietic acid	5.0	5.0