

Article

Not peer-reviewed version

# Genome Mining Reveals Pathways for Terpene Production in Aerobic Endospore-Forming Bacteria Isolated from Brazilian Soils

[Felipe de Araujo Mesquita](#) , [Waldeyr Mendes Cordeiro da Silva](#) , Taina Raiol , [Marcelo de Macedo Brigido](#) , [Nalvo Franco de Almeida](#) , [Bruna Fuga](#) , Danilo de Andrade Cavalcante , [Marlene Teixeira De-Souza](#) \*

Posted Date: 10 March 2025

doi: 10.20944/preprints202503.0620.v1

Keywords: Bacillus and related genera; secondary or specialised metabolites; the antiSMASH software; Biosynthetic gene cluster (BGC); conserved signature indels for different clades; methylerythritol-phosphate (MEP) pathway; lycopene; squalene; hopanoids



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

## Article

# Genome Mining Reveals Pathways for Terpene Production in Aerobic Endospore-Forming Bacteria Isolated from Brazilian Soils

Felipe Mesquita <sup>1,2,†</sup>, Waldeyr Silva <sup>3,†</sup>, Taina Raiol <sup>4</sup>, Marcelo Brigido <sup>1</sup>, Nalvo Almeida <sup>5</sup>, Bruna Fuga <sup>1</sup>, Danilo Cavalcante <sup>1,2</sup> and Marlene De-Souza <sup>1,\*</sup>

<sup>1</sup> Department of Cellular Biology, Institute of Biological Sciences, University of Brasilia (UnB), Zip code: 70910-900, Brasilia, Brazil.

<sup>2</sup> Microbial Biology Graduation Program, UnB, Brasilia, Brazil.

<sup>3</sup> Instituto Federal de Goiás, Formosa, GO, Zip code: 73813-816, Brazil.

<sup>4</sup> Fiocruz Brasília, Oswaldo Cruz Foundation, Zip code: 70.904-130, Brasilia, Brazil. Brasilia, Brasilia, Brazil.

<sup>5</sup> Faculdade de Computação, Universidade Federal de Mato Grosso do Sul, Zip code: 79070-900, Campo Grande, MS, Brazil.

\* Correspondence: marlts@unb.br

Marlene Teixeira De-Souza, Cellular Biology Department, University of Brasília, Brasilia - DF, 70.910-900, Brazil, telephone: +55 61 9 9994-5727

ORCID iD: <https://orcid.org/0000-0003-1538-2657>

<sup>†</sup> These authors contributed equally to this work.

**Abstract:** Terpenes are the largest category of specialized metabolites. Aerobic endospore-forming bacteria (AEFB), a diverse group of microorganisms, can thrive in various habitats and produce specialized metabolites, including terpenes. This study investigates the potential for terpene biosynthesis in 10 AEFB strain whole-genome sequences by performing bioinformatics analyses to identify genes associated with these isoprene biosynthesis pathways. Specifically, we focused on the sequences coding for enzymes in the methylerythritol-phosphate (MEP) pathway and the polyprenyl synthase family, which play crucial roles in synthesizing terpene precursors together with terpene synthases. Comparative analysis revealed a unique genetic architecture of these biosynthetic gene clusters (BGCs). Our results indicated that some strains possessed the complete genetic machinery required to produce terpenes such as squalene, hopanoids, and carotenoids. We also reconstructed phylogenetic trees based on the amino acid sequences of terpene synthases, which aligned with the phylogenetic relationships inferred from the whole-genome sequences, suggesting the production of terpenes is an ancestor property in AEFB. Our findings highlight the importance of genome mining as a powerful tool for discovering new biological activities. Furthermore, this research lays the groundwork for future investigations to enhance our understanding of terpene biosynthesis in AEFB and the potential applications of these Brazilian environmental strains.

**Keywords:** bacillus and related genera; secondary or specialised metabolites; the antiSMASH software; biosynthetic gene cluster (BGC); conserved signature indels for different clades; methylerythritol-phosphate (MEP) pathway; lycopene; squalene; hopanoids

## 1. Introduction

Specialised or secondary metabolites are not essential for growth. Nevertheless, they play a relevant ecological role by providing nutrients in competitive environments and offering adaptive advantages to the producing organism [1,2]. Terpenes are hydrocarbons of linked five-carbon units of isoprene [3] and are the largest category of specialised metabolites, with over 80,000 known

structures [4]. These molecules present a wide range of applications in pharmaceuticals, food, and cosmetics, among other relevant industries [3,4].

Originally thought to be exclusively produced by plants, nowadays, terpenes are known to be synthesised by a diversity of organisms, including bacteria, fungi, protozoa, and invertebrates [5–8]. In plant tissues, terpenes are pigments and scents that protect against predation and attract predators of herbivorous parasites [7,9]. They also act as signalling structures and plant hormones. In bacteria, terpenes such as the antimicrobial albaflavenone [10], germacrene D and pentalenene [7] have been described. Nonetheless, the physiological and ecological functions of terpenes in this domain of life are still largely unknown [11,12].

The production of terpenes occurs through diverse biochemical pathways, which can be independent or integrated [13]). The synthesis of these isoprenes starts from the precursor isopentenyl pyrophosphate (IPP), produced via either the mevalonate (MVA) or the methylerythritol-phosphate (MEP) pathways [14]. Both plants and fungi utilise the MVA and MEP pathways for terpene synthesis, while most bacteria rely solely on the MEP route [14]. The enzyme isopentenyl pyrophosphate isomerase (IDI) converts IPP into dimethylallyl pyrophosphate (DMAPP), an IPP isomer. Subsequently, geranyl pyrophosphate synthase (GPPS) condenses these two isomers to produce geranyl pyrophosphate (GPP), which serves as monoterpenes' precursor. GPPS belongs to the enzyme family polyprenyl synthase (PPS), which includes farnesyl pyrophosphate synthase (FPPS) and geranyl geranyl pyrophosphate synthase (GGPPS). FPPS is responsible for the sesquiterpenes and triterpenes syntheses, and GGPPS is involved in diterpene and tetraterpene precursors. An overview of the substrates for the MEP pathway, the enzymes of the PPS family, and their corresponding products is provided in Table 1.

**Table 1. Terpene precursors' enzymes.** Profile of MEP pathway and polyprenyl synthase family molecules.

Substrate	Enzyme code*	Enzyme name (abbreviation)	Product (abbreviation)
Pyruvate and G3P	2.2.1.7	1-deoxy-D-xylulose-5-phosphate synthase (DXS)	1-deoxy-D-xylulose-5-phosphate (DXP)
DXP and NADPH	1.1.1.267	DXP reductoisomerase (DXR)	methylerythritol-phosphate (MEP)
MEP	2.7.7.60	MEP cytidyltransferase (MCT)	4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol (CD-ME)
CD-ME and ATP	2.7.1.148	CD-ME kinase (CMK)	4-difosfocitidil-2-C-metil-Deritritol 2-fosfato (CD-MEP)
CD-MEP	4.6.1.12	2C-methyl-D-erythritol-2,4-cyclodiphosphate synthase (MDS)	2C-methyl-D-erythritol-2,4-cyclodiphosphate (MEC)
MEC and NADPH	1.17.7.3	1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase (HDS)	1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate (HMBPP)
HMBPP and NADPH	1.17.7.4	HMBPP reductase (HDR)	isopentenyl pyrophosphate (IPP)
IPP	5.3.3.2	isopentenyl diphosphate isomerase (IDI)	dimethylallyl pyrophosphate (DMAPP)
IPP and DMAPP	2.5.1.1	GPP synthase (GPPS)**	geranyl diphosphate (GPP)
GPP and IPP	2.5.1.10	FPP synthase (FPPS)**	farnesyl diphosphate (FPP)
FPP and IPP	2.5.1.29	GGPP synthase (GGPPS)**	geranylgeranyl diphosphate (GGPP)

\*Kyoto Encyclopedia of Genes and Genomes (<https://www.genome.jp/kegg/>). \*\*GPPS, FPPS, and GGPPS are enzymes belonging to the polyprenyl synthase family (PPS).

Endospore formation is an outstanding differentiation mechanism that evolved to help some bacteria survive adverse conditions [15,16]. The resulting dormant spore remains sensitive to environmental changes and can germinate to return to active metabolism and reproduction [16]. The ability to form spores has been observed only inside the phylum *Firmicutes*, recently renamed *Bacillota* [17]). This phylum allocates low G+C bacteria, most of which have Gram-positive cell wall structures, distributed in eight classes [18]. Endospore formation is not a universal characteristic inside *Bacillota*, but endospore-formers share a minimal homologous gene set involved in this event [18]. Endospore formation is widespread in the two *Bacillota* major classes—*Bacilli*, which are aerobic or facultative and *Clostridia*, anaerobic [17,18].

Species allocated in the genus *Bacillus* and related genera—whether assigned to the same or different families and orders inside *Bacilli*—are referred to as aerobic endospore-forming bacteria (AEFB), and soil is their major reservoir [19–21]. These bacteria can thrive across large temperatures and pH levels, exhibit metabolic diversity, and demonstrate remarkable spore resistance. These traits contribute to their ubiquity. Researchers extensively study these exceptional characteristics applied in various industrial contexts [22]. AEFB are known to synthesise a diverse array of specialised metabolites, which includes terpenes [20,22–25]. The ecological and socioeconomic significance of these metabolites is well recognised, as they can promote plant growth, manage insect pests and disease vectors, and possess immunosuppressive, antimicrobial, and antitumor activities [22,25–27]). Despite their potential importance, the terpene biosynthetic pathways in AEFB remain largely unexplored.

To acquire knowledge on AEFB and gain insights into their potential as a source of novel bioactive compounds as terpenes, we previously isolated 312 strains through heat-shocking soil samples collected at random areas of the Federal District, Midwest region of Brazil. These environmental strains are designated SDF0001–SDF0312 and are deposited at the Coleção de Bactérias aeróbias formadoras de endósporos (CBafes, or AEFB Collection–AEFBC). The CBafes is hosted at the University of Brasília and is currently undergoing taxonomic classification using a polyphasic approach [28–31].

Genome prediction offers a powerful means for identifying and characterising the genetic basis of terpene production in bacteria. Here, we used a genome mining approach to investigate the potential for terpene biosynthesis in the whole genome of 10 SDF strains (Table 2). We focused on identifying the key genes involved in the MEP pathway and the enzyme of the PPS family by examining 16 BGCs, we previously identified in these genomes employing the antiSMASH *in-silico* pipeline [32]. We also sought to identify genes encoding terpene synthases (TSs), the enzymes responsible for the final steps in terpene biosynthesis. Finally, we reconstructed phylogenetic trees based on corresponding amino acid sequences of the TSs found to liken phylogenetic relationships based on whole genomes of the respective SDF strains. Our findings provide new insights into the diversity and evolution of terpene biosynthesis in AEFB and highlight the potential of these environmental strains as a source of novel terpenes with valuable applications.

**Table 2.** General genomic features of the ten SDF strains analysed.

Strain	Size (bp)	Scaffold #	N50 (bp)	GC content (%)	CDS #	Protein coding regions	Pseudo genes (total)	rRNA genes (5S; 16S; 23S)	tRNA genes	GenBank accession #
<i>Lysinibacillus fusiformis</i> SDF0005	4,472,771	24	392,231	37.6	4,369	4,328	41	13; 7; 7	85	VKHW000000000.1
<i>Bacillus pumilus</i> SDF0011	3,686,817	56	143,274	41.2	3,688	3,617	71	7; 3; 2	73	VKHY000000000.1
<i>Heyndrickxia oleronia</i> SDF0015	5,267,437	75	151,790	34.7	5,127	5,018	109	10; 14; 7	129	VKHZ000000000.1
<i>Bacillus safensis</i> SDF0016	3,674,191	25	484,434	41.6	3,688	3,640	48	4; 1; 1	74	SADW000000000.1
<i>Peribacillus simplex</i> SDF0024	5,376,271	45	497,961	40.2	5,204	5,007	197	14; 7; 6	81	VKHX000000000.1
<i>Paenibacillus popilliae</i> SDF0028	6,580,875	39	611,008	46.5	5,684	5,519	165	2; 2; 3	62	SADY000000000.1
<i>Lysinibacillus sphaericus</i> SDF0037	5,122,785	71	215,682	36.5	4,869	4,643	226	5; 7; 2	71	SADV000000000.1
<i>Brevibacillus brevis</i> SDF0063	6,239,737	31	471,412	47.3	5,789	5,602	187	1; 16; 9	89	SADX000000000.1
<i>Bacillus velezensis</i> SDF0141	3,945,527	15	962,078	46.4	3,887	3,780	107	8; 3; 2	78	VKIB000000000.1
<i>Bacillus velezensis</i> SDF0150	3,927,067	21	271,062	46.4	3,870	3,763	107	8; 6; 2	82	VKIC000000000.1



## 2. Materials and Methods

### *Bacterial Strains*

The 10 SDF strains evaluated in this study (Table 2) are deposited at the *Coleção de Bactérias aeróbias formadoras de endósporos* (CBafes, or Aerobic Endospore-Forming Bacteria—AEFB Collection), hosted at the University of Brasilia, Brazil. They were isolated from Brazilian soils, preserved as dry spores in filter paper, and stored at room temperature as described in Orem et al., 2019 [28] and Cavalcante et al., 2019 [29]. Six genomes were sequenced specifically for this study, plus four previously sequenced genomes, from the same culture collection, accessible at the NCBI.

### *Ethics Statement*

Specific permissions required to collect the SDF strains used in this study were endorsed by the Federal Brazilian Authority (CNPq; Authorisation of Access and Sample of Genetic Patrimony n° 010439/2015-3). Sampling did not involve endangered or protected species.

### *Sequencing, Assembly, Annotation, and Data Availability*

The total DNA of the six SDF strains sequenced specifically for this study (Table 2) was extracted and purified using the Wizard genomic kit (Promega) following the manufacturer's instructions and sequenced using an Illumina Miseq PE platform at the Catholic University of Brasilia (Brazil). MiSeq reads were evaluated for quality control using FASTQC 0.12.0 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), followed by trimming and assembly into contigs/scaffolds using the AS-miseq pipeline [33]. This pipeline automatically processes adapter trimming, quality filtering, error correction, contig, scaffold generation, and misassembly detections. The genomes were deposited at the NCBI (Table 2). Gene annotation was performed using the NCBI prokaryotic genome annotation pipeline [34].

### *Whole Genome-Based Features and Phylogeny*

Whole genome-based phylogeny analysis was performed using the OrthologSorter tool [35]; available in <https://git.facom.ufms.br/bioinfo/orthologsorter>. Orthologsorter generates, among other data, protein families shared across all genomes (core genome). Orthologsorter employs BLASTp [36] and OrthoMCL [37] tools with default parameters to determine orthology. For our set of the 10 SDF strain genomes, plus the included outgroup *Staphylococcus pseudintermedius*, 918 core families have been found. These families were aligned and, after removing poorly aligned positions and divergent regions, using GBlocks [38], the resulting whole alignment was used to build the phylogenetic tree with RAxML [39] with PROTCATJTT substitution model, rapid bootstrapping (1,000 replicates), and a subsequent Maximum Likelihood search.

### *BGC Predictions*

To identify BGCs linked to terpene synthesis, the antiSMASH 6.0 bacterial standalone version [40] optimised for prokaryotic sequences (<https://antismash.secondarymetabolites.org/#!/start>) was run over the 10 genomes of the SDF strains (Table 2). The accuracy parameter of the detected clusters was relaxed (Full featured-run) with algorithms provided by the antiSMASH platform (KnownClusterBlast, ActiveSiteFinder, ClusterPfam, ClusterBlast, and Pfam-based GO term annotation). The BGC similarity level (0-100%) reported for a specific metabolite was obtained by crossing over data available in the Minimum Information about a Biosynthetic Gene cluster (MiBiG) platform (<https://mibig.secondarymetabolites.org>). The percentage index indicates the number of the gene sequences within a BGC that have a hit to any gene in a particular BGC at the MiBiG's reference strain related to terpene production.

### MEP Pathway Reconstruction

The Pathway Tools 26.5, a systems-biology software—associated with the BioCyc Pathway/Genome Database Collection (<http://bioinformatics.ai.sri.com/ptools/>) was used to predict the gene sequences coding for the MEP pathway catalysts (Table 1). The algorithm PathoLogic was used to create a Pathway/Genome Database (PGDB) containing the predicted metabolic pathways of the respective strain. The PGDB was built using a cutoff score of 0.15, and the inference tools Transport Inference Parser, Pathway Hole Filler, Operon Predictor, and Protein Complex Predictor in the activated mode. The Omics Dashboard tool was used to orient metabolomic data to create one diagram showing the aggregated system-oriented view of the metabolic routes of the 10 SDF strains.

### Detection of Polyprenyl Synthase Enzymes

Data from NCBI platform (<https://ncbi.nlm.nih.gov>) were used to investigate the presence of sequences coding for PPS enzymes (Table 1) in the 10 SDF strains studied (Table 2). To this end, a database containing 6,146 files of amino acid sequences (.fasta) of polyprenyl synthases—obtained from species allocated in the four orders explored in this study and deposited at NCBI—was built. The extraction of protein sequences from each SDF strain genome (.gbk) was accomplished using the script available in Bogdanove et al., 2011 [41]. The alignment and comparison of the amino acid sequences among the database created and SDF genome sequences were achieved using BLASTp. Amino acid sequences with the highest hits were considered to detect the enzyme presence in the SDF strains.

### Similarity of the Enzyme Set for Terpene Production

The putative SDF producers and the enzymes of MEP pathway and PPS family (Table 1) in addition to TSs detected for each strain were arranged in heatmaps [42] using the software R. The dichotomous values 0 (for the absence of catalyst) and 1 (for the presence of catalyst) were taken as binary variables representing these associations. Pearson's correlation was employed to cluster the SDF strains, taking a similar set of enzyme results [43].

### Phylogenetic Tree Reconstruction Based on Terpene Synthase Contents

The amino acid sequences of the TSs identified were extracted from the corresponding BGC obtained by the antiSMASH 6.0 [40]—and performed using an in-house script in the Biopython programming language (<http://biopython.org/DIST/docs/tutorial/Tutorial.html>). Inside MEGA software version 11, the amino acid sequences (.fasta) obtained were aligned using ClustalW with default parameters (<https://www.megasoftware.net/ClustalW>). The file generated (.mas) was used to reconstruct phylogenetic trees employing the maximum-likelihood statistical method based on 1,000 bootstrap replicates.

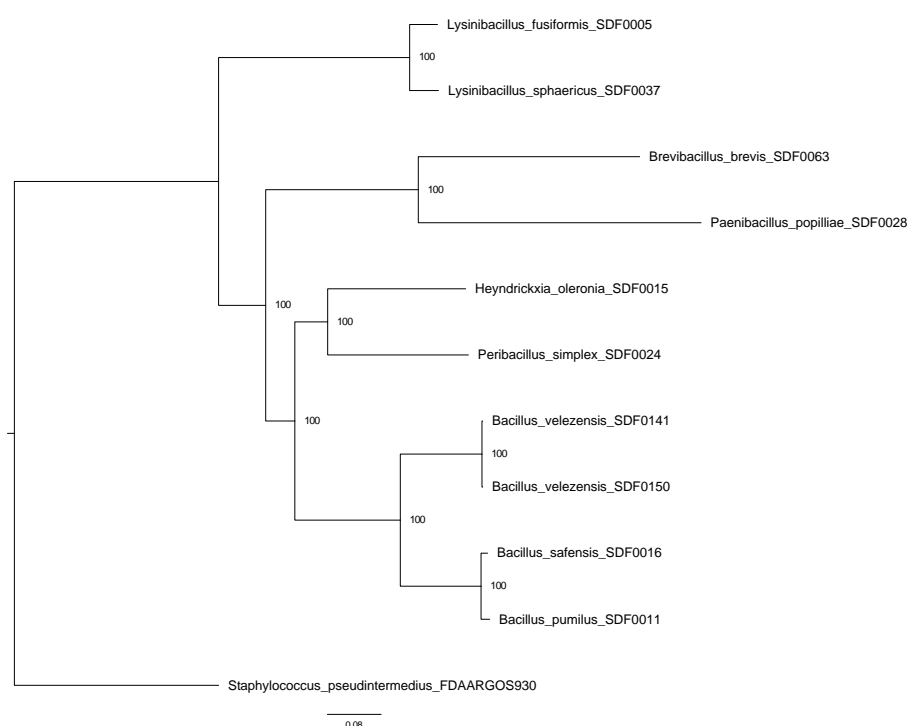
## 3. Results

### 3.1. SDF Strain Genome Features

This study describes genomic resources for 10 cultivable environmental-AEFB samples designated SDF strains (Table 2). We presented high-quality whole-genome sequences from 10 SDF strains based on Illumina. These samples corresponded to four orders, four families, six genera, and nine species allocated in the phylum *Bacillota*, class *Bacilli* (Table 2). Six samples were assigned to order *Bacillales*, family *Bacillaceae*. Among them, four strains belonged to three different *Bacillus* spp.: *Bacillus pumilus* SDF0011, *Bacillus safensis* SDF0016, *Bacillus velezensis* SDF0141, and *Bacillus velezensis* SDF0150. Family *Bacillaceae* was also represented by two other genera and species named strains *Heyndrickxia oleronia* SDF0015 and *Peribacillus simplex* SDF0024, referred to here as *Pe. simplex* SDF0024. Inside the order *Caryophanales*, family *Caryophanaceae*, genus *Lysinibacillus* were represented by two strains *Lysinibacillus fusiformis* SDF0005 and *Lysinibacillus sphaericus* SDF0037. The strains

*Paenibacillus popilliae* SDF0028 belonged to the order *Paenibacillales*, family *Paenibacillaceae*, and genus *Paenibacillus*. Finally, *Brevibacillus brevis* SDF0063 was allocated in the order *Brevibacillales*, family *Brevibacillaceae* and is referred to here as *Br. brevis* SDF0063. Genome analysis of the SDF strains uncovered considerable differences in genome size, scaffold number, N50, GC content, coding sequences (CDS), protein-coding regions, pseudo genes, rRNAs, and tRNAs, as detailed in Table 2. Briefly, genome sizes ranged from 3,674,191 to 6,580,875 bp, with the scaffold numbers varying from 15 up to 75 and GC content (%) spanning from 34.7 to 47.3.

A Maximum Likelihood method was applied to reconstruct a phylogenetic tree based on the results of the OrthologSorter tool [35]; available in <https://git.facom.ufms.br/bioinfo/orthologsorter> in 10 SDF strain whole-genomes (Table 2) and *S. pseudintermedius* as an outgroup (Figure 1), that resulted into two major clades. The *Lysinibacillus* spp., *L. fusiformis* SDF0005 and *L. sphaericus* SDF0037 clustered together in the most distinct branch. The genomes of *H. oleronia* SDF0015 and *Pe. simplex* SDF0024 formed a branch that also included the four *Bacillus* strains: *B. pumilus* SDF0011, *B. safensis* SDF0016, *Bacillus velezensis* SDF0141, and *Bacillus velezensis* SDF0150 on the second major clade. Additionally, the strains *P. popilliae* SDF0028 and *Br. brevis* SDF0063 were positioned as the most distinct SDF strains analysed.



**Figure 1. Whole-genome phylogenetic relationship of the 10 SDF strains.** Whole genome-based unrooted phylogeny built by using RAxML with PROTCATJTT substitution model, rapid bootstrapping, and a subsequent Maximum Likelihood search, having as input 918 core protein families of our 10 SDF strains plus *Staphylococcus pseudintermedius*, as an outgroup. The tree nodes show bootstrap values as percentages of 1,000 replications. SDF strain classifications are indicated in the branches, and the distance scale bar is displayed at the bottom.

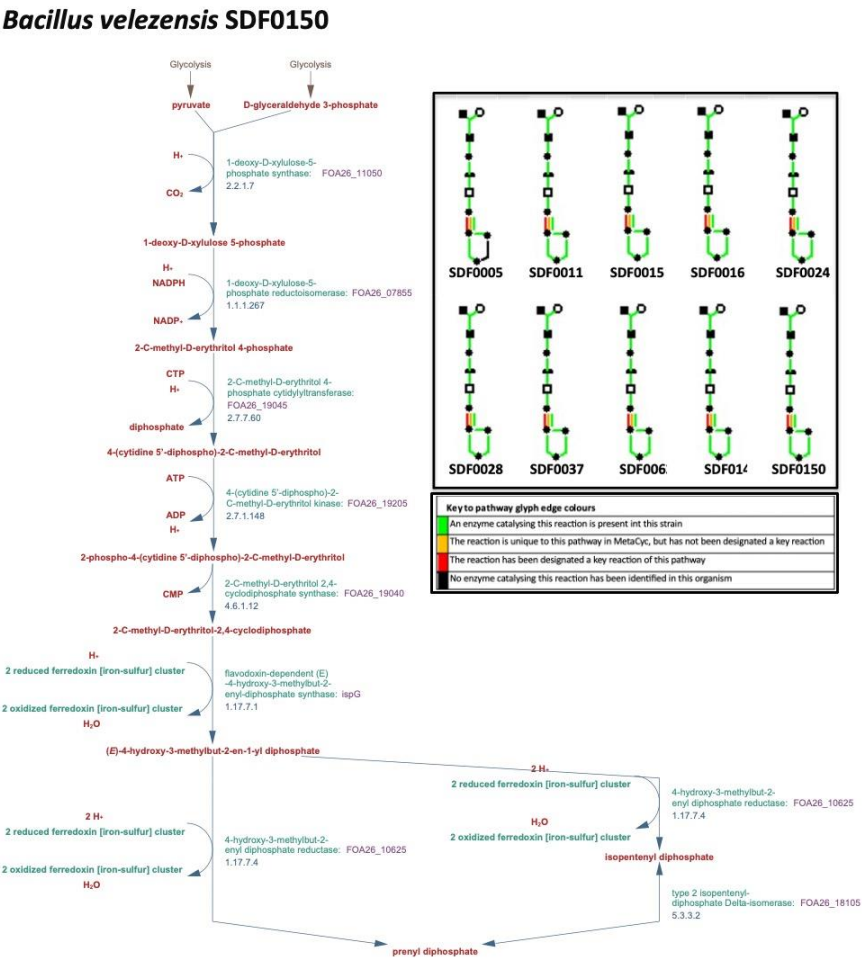
### 3.2. MEP Pathway Reconstruction

BGCs are a locally clustered group of two or more genes in a particular genome. The antiSMASH is an *in-silico* pipeline offering detection and analysis of many BGC types (40). These gene clusters encode biosynthetic pathways for specialised metabolite production with diverse functions, including chemical variants [44]. Previously, using the antiSMASH 6.0 bacterial standalone version [40] we identified 153 putative BGCs codifying for 20 different classes of specialised metabolites synthesis in 10 SDF strains (Table 2) deposited at CBafes [32]. Among these, 16 were related to terpene



synthesis. In this work, the potential of these SDF strains for terpene biosynthesis was further addressed by taking advantage of these 16 high-quality BGC sequences.

The algorithmic PathoLogic (<http://bioinformatics.ai.sri.com/ptools/>) predicted that all the corresponding gene sequences coding for the seven enzymes ( DXS, DXR, MCT, CMK, MDS, HDS, and HDR) that catalyse the MEP pathway reactions (Table 1) were found among the 10 SDF genomes (Table 2). The PGDB obtained is represented in a diagram aggregating a system-oriented view of the metabolic routes of the 10 SDF strains generated by the Omics Dashboard tool (Figure 2). In addition to MEP route enzymes, this tool also detected the enzyme IDI (Table 1) responsible for both IPP isomerization to DMAPP and subsequent IPP and DMAPP condensation that generates the first substrate in the terpenes' production (Table 1). The information coding for the enzyme IDI was present in all SDF strains analysed, except for the *L. fusiformis* SDF0005 strain (Figure 2).



**Figure 2. Methylethylthrotol-phosphate pathway reconstruction.** The Pathway Tools software created a Pathway/Genome Database (PGDB) that includes the predicted metabolic pathways of the respective strain. The Omics Dashboard tool was then applied to align metabolomic data, generating diagrams to provide the aggregated system-oriented view of the predicted metabolic route information found in the genome of *Bacillus velezensis* SDF0150 (larger diagram). Smaller diagrams representing these routes in *Bacillus velezensis* SDF0150 and the remaining nine strains are displayed in the top right corner. The key to the colours of pathway glyph edges are indicated. The SDF strain designation is indicated: *Lysinibacillus fusiformis* SDF0005, *Bacillus pumilus* SDF0011, *Heyndrickxia oleronia* SDF0015, *Bacillus safensis* SDF0016, *Peribacillus simplex* SDF0024, *Paenibacillus popilliae* SDF0028, *Lysinibacillus sphaericus* SDF0037, *Brevibacillus brevis* SDF0063, and *Bacillus velezensis* SDF0141.

### 3.3. Detection of Polyprenyl Synthase Enzymes

Using BLASTp, the alignment and comparison of the amino acid sequences among the database we created (6,146 files; see material and methods section) and SDF sequences revealed the presence of the PPS family —the enzymes responsible for the conversion of IPP to GPP, FPP, and GGPP (Table 1). The strains *L. fusiformis* SDF0005; *H. oleronia* SDF0015; *Pe. simplex* SDF0024; *B. velezensis* SDF0141, and *B. velezensis* SDF0150 presented >98% amino acid similarity, while *H. oleronia* SDF0015 exhibited 67.86% (Table 3). The amino acid sequences for PPS were not detected in the remaining SDF strains.

**Table 3.** Polyprenyl synthase family occurrence in the ten SDF strains analysed.

Strain	Occurrence	identity (%) <sup>*</sup>	Reference species	GeneBank reference sequence#
<i>Lysinibacillus fusiformis</i> SDF0005	+	99.66	<i>Lysinibacillus fusiformis</i>	KAB0443654.1
<i>Bacillus pumilus</i> SDF0011	-	NA	NA	NA
<i>Heyndrickxia oleronia</i> SDF0015	+	67.86	<i>Bacillus pumilus</i>	WP_268443628.1
<i>Bacillus safensis</i> SDF0016	-	NA	NA	NA
<i>Peribacillus simplex</i> SDF0024	+	98.65	<i>Peribacillus</i> sp.	WP_241589686.1
<i>Paenibacillus popilliae</i> SDF0028	-	NA	NA	NA
<i>Lysinibacillus sphaericus</i> SDF0037	-	NA	NA	NA
<i>Brevibacillus brevis</i> SDF0063	-	NA	NA	NA
<i>Bacillus velezensis</i> SDF0141	+	100	<i>Bacillus velezensis</i>	ASK59031.1
<i>Bacillus velezensis</i> SDF0150	+	99.65	<i>Bacillus velezensis</i>	QWC45887.1

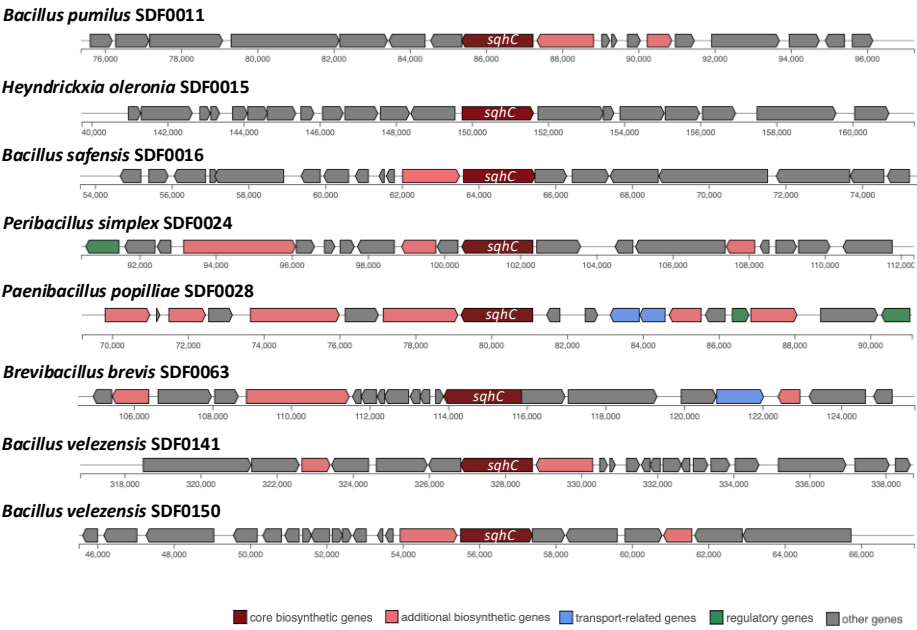
<sup>\*</sup>Highest identity level from the alignments between each SDF amino acid sequence obtained by *in silico* translation and NCBI reference sequence obtained by Blast. The symbols indicate the presence (+) or absence (-) of the corresponding coding gene in the genomes. NA: not applicable.

### 3.4. Prediction of Biosynthetic Gene Clusters Associated with Terpenes Synthesis

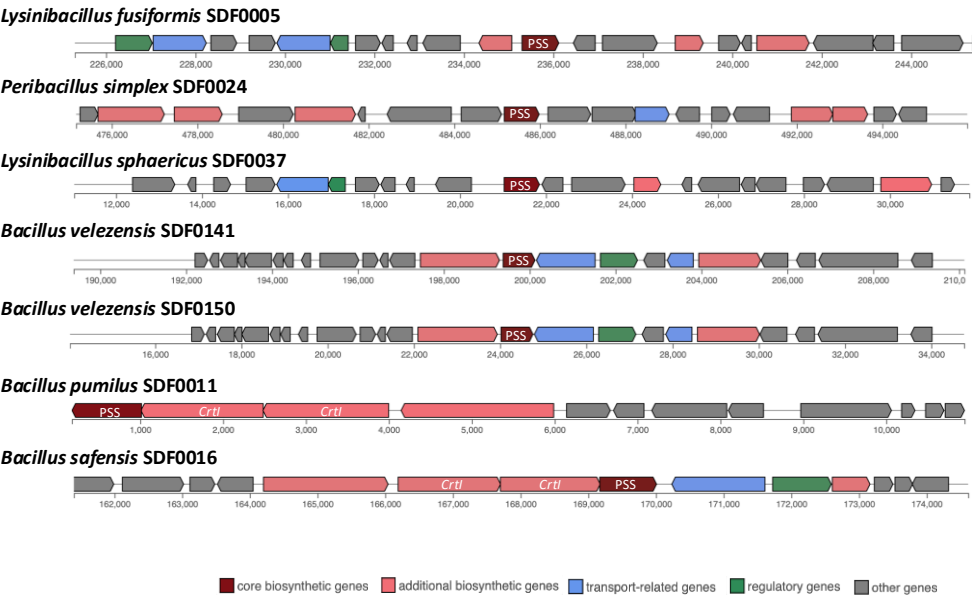
At least three gene sequences inside the 16 BGCs uncovered by the antiSMASH [32] directed the synthesis of three TS: (i) an *sqhC* gene, coding for a squalene-hopene cyclase (SHC); (ii) a gene encoding undetermined activity related to the production of the phytoene and/or squalene synthase (PSS), and (iii) a *crtI* gene, coding for a phytoene desaturase (PDS). Table 3 outlines the TS coding sequences for the strains used in this study.

Out of these 10 SDF strains, the sequences coding for the SHC were detected in eight genomes: *B. pumilus* SDF0011; *H. oleronia* SDF0015; *B. safensis* SDF0016; *Pe. simplex* SDF0024; *P. popilliae* SDF0028; *Br. brevis* SDF0063; *B. velezensis* SDF0141, and *B. velezensis* SDF0150 (Table 4; Figure 3).

The sequences coding for the enzymes of the PSS synthase family were found in seven genomes: *L. fusiformis* SDF0005, *B. pumilus* SDF0011, *B. safensis* SDF0016, *Pe. simplex* SDF0024, *L. sphaericus* SDF0037, *B. velezensis* SDF0141, and *B. velezensis* SDF0150 (Table 4; Figure 4). In the genomes of *B. pumilus* SDF0011 and *B. safensis* SDF0016, these sequences were adjacent to two copies of the gene *crtI* coding for a PDS (Figure 4). The BGC containing these two gene copies in the genomes of these two strains are reported to take part in a carotenoid production [45]. Furthermore, these latter BGC presented 50% similarity compared to the corresponding sequence described for strain *Halobacillus halophylus* DSM2266 in the MIBiG platform (reference number BGC0000645) used by antiSMASH as a reference to predict this metabolite. The similarity percentages (0-100%) indicate the number of genes within the reference that have a hit to any genes in a particular BGC related to terpenes production and were recognised by the antiSMASH.



**Figure 3.** Structure of biosynthetic gene clusters involved in the squalene hopene cyclase expression in the genome of the SDF strains evaluated. The gene *sqhC* that directs the squalene hopene cyclase (SHC) production was identified as a core biosynthetic gene (brown) inside the BGCs detected in eight SDF genomes by the antiSMASH. The colour boxes (bottom) indicate the predicted gene functions in the biosynthesis of terpenes.



**Figure 4.** Structure of biosynthetic gene clusters involved in the phytoene/squalene synthase and phytoene desaturase the genome of the SDF strains evaluated. A gene sequence that directs the phytoene and/or squalene synthase (PSS) family enzyme production was identified as a core biosynthetic gene (brown) inside the BGCs detected in seven SDF genomes by the antiSMASH. In *Bacillus pumilus* SDF0011 and *Bacillus safensis* SDF0016 the gene for PSS were located adjacent to a copy for gene *crtI*, codifying a phytoene desaturase. The colour boxes (bottom) indicate the predicted gene functions in the biosynthesis of terpenes.

**Table 4.** Genes coding for the TS enzymes identified by antiSMASH in the ten SDF strains analysed.

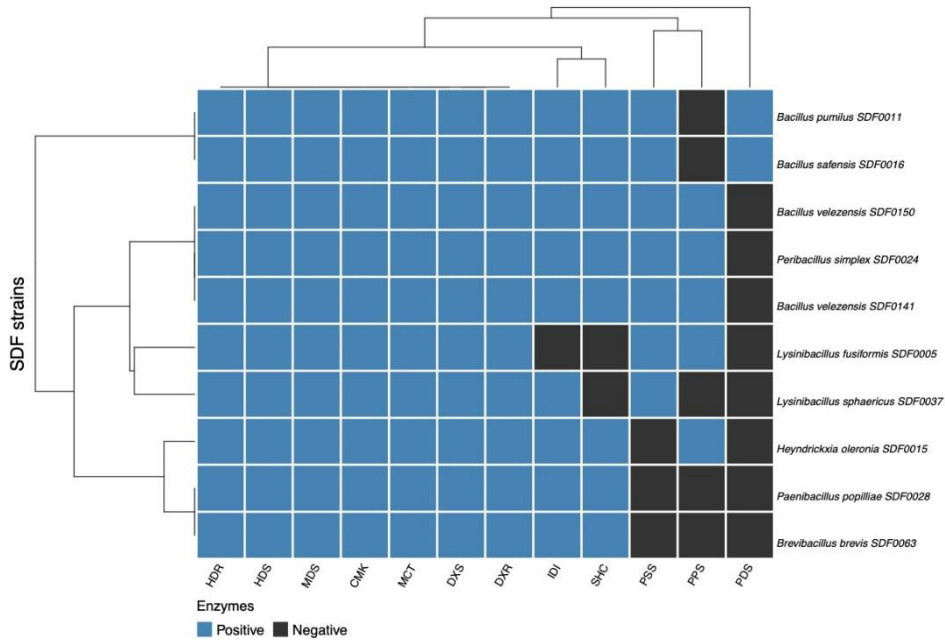
Strain	Gene/TS enzyme
--------	----------------

	sqhC/SHC	Phytoene and/or squalene synthase family gene/PSS	crtI/PDS
<i>Lysinibacillus fusiformis</i> SDF0005	-	+	-
<i>Bacillus pumilus</i> SDF0011	+	+	+
<i>Heyndrickxia oleronia</i> SDF0015	+	-	-
<i>Bacillus safensis</i> SDF0016	+	+	+
<i>Peribacillus simplex</i> SDF0024	+	+	-
<i>Paenibacillus popilliae</i> SDF0028	+	-	-
<i>Lysinibacillus sphaericus</i> SDF0037	-	+	-
<i>Brevibacillus brevis</i> SDF0063	+	-	-
<i>Bacillus velezensis</i> SDF0141	+	+	-
<i>Bacillus velezensis</i> SDF0150	+	+	-

The symbols indicate the presence (+) or absence (-) of the corresponding gene sequence coding for a TS in the genomes.

3.5. Distribution of the Enzyme Set for Terpene Production Among the 10 SDF Strains

The distribution of the predicted enzymes obtained from in silico translation of the corresponding gene sequences coding for the seven MEP pathway enzymes (DXS, DXR, MCT, CMK, MDS, HDS, and HDR), along with the enzyme IDI were also analysed in the SDF strain genomes. PPS family, the three TS enzymes SHC, PSS, and PDS were also included. To this end, Pearson correlation [43] was employed to cluster the 10 SDF strains (Table 2; Figure 5) based on the ensemble of enzymes engaged in the terpene's synthesis detected. We constructed a heatmap [42] representing the presence or absence of an enzyme in a particular strain to enhance the potential visual distribution of the enzyme distributions among the strains. Two major clades were distinguished (Figure 5). The first comprised *B. pumilus* SDF0011 and *B. safensis* SDF0016, which shared an identical profile, or 11 out of the 12 enzymes detected. The second major clade was further split. *Pe. simplex* SDF0024; *B. velezensis* SDF0141, and *B. velezensis* SDF0150 also shared the same profile, bearing the equivalent set of 11 enzymes (Figure 5). The next subclade embraced *Lysinibacillus sphaericus* SDF0037 and *L. fusiformis* SDF0005 despite the dissimilar enzymatic profiles (Figure 5). Although the *H. oleronia* SDF0015 is the only representative of a subclade, the correlation showed that the enzymatic set for terpenes production in this strain was compatible with the subclade encompassed by *P. popilliae* SDF0028 and *Br. brevis* SDF0063, except that one out of nine enzymes was missing.



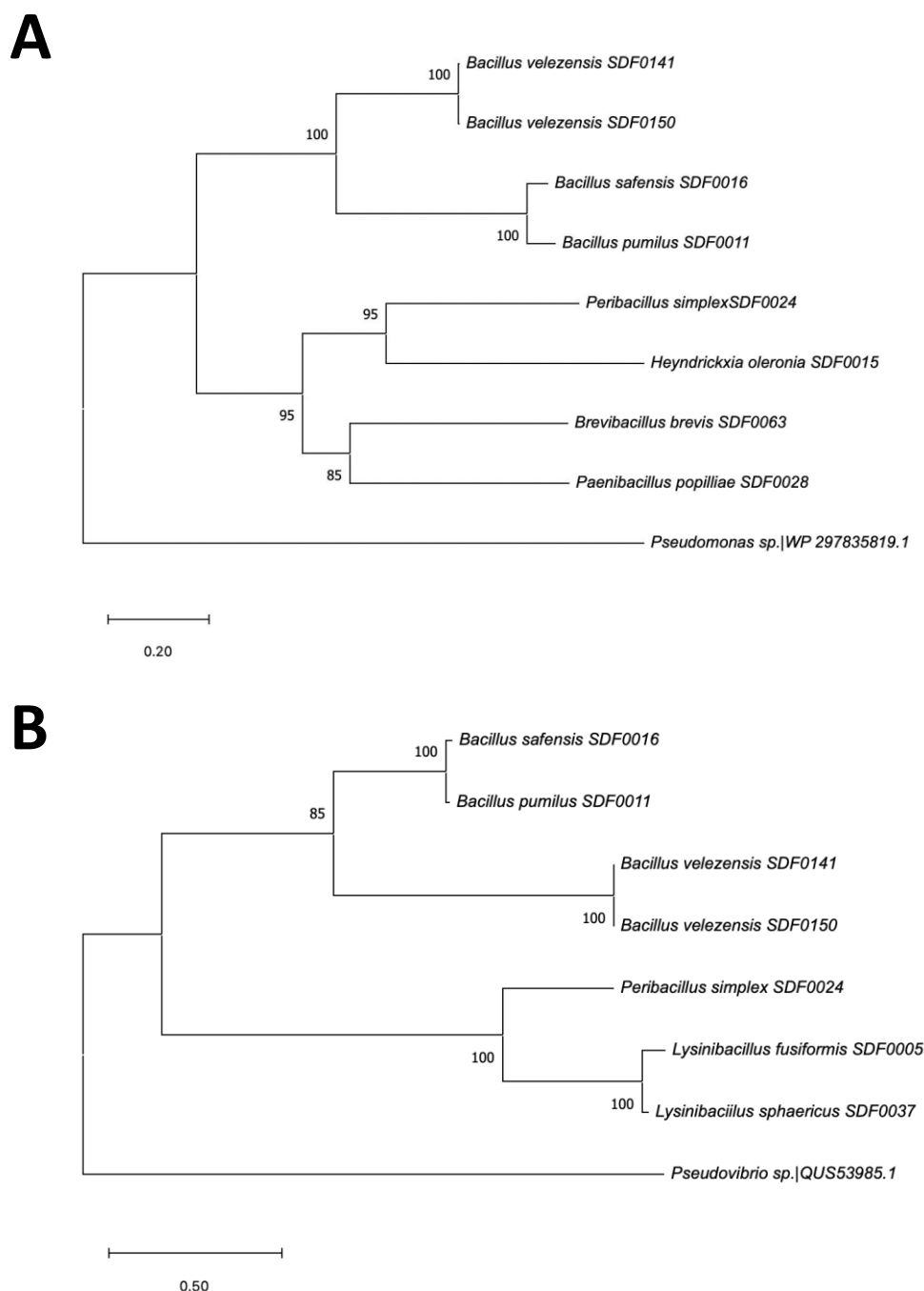
**Figure 5. Distribution of the ensemble of enzymes engaged in the terpenes' synthesis among the ten SDF strains.** A heatmap depicting the association obtained using the Person correlation-based method among the ten SDF strains (right) and 12 genes coding for enzymes involved in terpenes' synthesis (bottom) was constructed. The protein set (Table 1) includes catalysts of the MEP route detected by the Pathways tools, along with polyprenyl synthase family (PPS) detected by BLASTp, and the TS squalene hopene cyclase, phytoene and/or squalene synthase (PSS), and phytoene desaturase (PDS) identified (described on the bottom) by the antiSMASH. The top dendrogram clustered the SDF strains into two sections based on presence (blue squares) or absence (black squares) of the respective gene coding for the enzyme in a particular species (right) genomes.

### 3.6. SDF Strains Evolutionary Relationship Based on Two TS Amino Acid Sequences

As described above, the BGC sequences involved in the terpenes' production detected by the antiSMASH tool encompassed three TS enzyme gene sequences (Table 4). A phylogenetic tree was reconstructed based on multiple alignments of the eight SHC amino acid sequences obtained by the *in-silico* translation from the corresponding gene sequence of these SDF strains (Figure 6A). The respective amino acid sequence of *Pseudomonas* sp. was included as an outgroup. The inferred evolutionary relationship among these SDF strains was clustered in two main clades. In the first, the respective SHC sequences found in the strains *B. velezensis* SDF0141 and *B. velezensis* SDF0150 were equivalent. This enzyme sequence obtained for *B. pumilus* SDF0011 was very close to that obtained for *B. safensis* SDF0016. The SHC sequences from *H. oleronia* SDF0015 and *Pe. simplex* SDF0024 revealed the highest evolutionary relationship in the second main clade generated (Figure 6A). The molecular correlation between the SHC primary chains of these two strains was closer to the corresponding sequences obtained to the strains *P. popilliae* SDF0028 and *Br. brevis* SDF0063, also positioned in the second clade (Figure 6A).

Likewise, the phylogenetic tree generated from seven amino acid sequences corresponding to the PSS enzyme family genomes and *Pseudovibrio brasiliensis* as an outgroup also divided these seven SDF strains into two main clades (Figure 6B). The first was further separated into two subclades. The sequences of this enzyme obtained for the strains *B. velezensis* SDF0141 and *B. velezensis* SDF0150 pointed out a close evolutionary relationship. The sequences of the strain *B. pumilus* SDF0011 and *B. safensis* SDF0016 positioned in the other subclade also displayed a high-level molecular relationship (Figure 6B). The second clade showed that the strains *L. fusiformis* SDF0005 and *L. sphaericus* SDF0037 presented amino acid sequences with the highest molecular relationship. Strain *Pe. simplex* SDF0024 was positioned apart from the other two inside this clade (Figure 6B). Because the gene sequence of PDS was found in two out of 10 SDF strains, it was not considered for further phylogenetic analyses.





**Figure 6. Phylogenetic relationship of the SDF strains based on terpene synthase sequences.** The evolutionary history of SDF strains was determined by aligning deduced amino acid sequences of two terpene synthases using the ClustalW. The sequences were obtained by translating the respective SDF gene sequence in silico. Phylogenetic trees were reconstructed using the maximum-likelihood method in MEGA version 11.0. The tree nodes show bootstrap values as percentages of 1000 replications. SDF strain designations are indicated in the branches. The distance scale bar is displayed at the bottom. The evolutionary relationships among **(A)** eight SDF strains based on the amino acid sequence of the SHC enzyme and the homolog amino acid sequence of *Pseudomonas* sp. as an outgroup and **(B)** seven SDF strains based on the amino acid sequence of the phytoene/squalene synthase family (PSS) and the homolog amino acid sequence of *Pseudovibrio brasiliensis* as an outgroup..

#### 4. Discussion

The phylum Firmicutes was recently renamed *Bacillota* (17). Inside class *Bacilli*, the order *Bacillales*, which allocated AEFB species, displayed an immense diversity, spanning several families, genera, and species [18,22]. Lately, new taxa have been established to reposition strains otherwise considered members of the order *Bacillales* [46]; <https://www.bacterio.net>. This reorganization considers the family *Bacillaceae* the only member of this order, includes *Bacillus* as the type-genus of these taxa, and *Bacillus subtilis* remains the type-species of the genus *Bacillus*.

Historically, the genus *Bacillus* represented a large assemblage of genetically and evolutionarily unrelated microorganisms. Thus, the genus has long been recognized as house members exhibiting extensive polyphyly and apportion very little in common with each other [18,46–48]. To more adequately represent the overall genetic diversity within this genus, it was proposed that the vast majority of *Bacillus* spp. needed to be reclassified into other genera, families, and orders. The revision of the genus *Bacillus* led to reallocating, not limited to, two *Bacillus* spp. to novel genera that better accommodate them. The former *Bacillus oleronius* is now designated *Heyndrickxia oleronia* [49]. Likewise, *Bacillus simplex* was reallocated into a novel genus designated *Peribacillus*, species *Peribacillus simplex* [50].

Other misclassified *Bacillus* spp. were transferred to specific genera and reallocated to the family *Caryophanaceae*, which in turn were transferred to the order *Caryophanales*, including the genus *Lysinibacillus* [47]. In this standpoint, the genera *Paenibacillus* and *Brevibacillus* were moved from the genus *Bacillus*, family *Bacillaceae*, being the genus *Paenibacillus* to the family *Paenibacillaceae*, order *Paenibacillales*, and the genus *Brevibacillus* to the novel family *Brevibacillaceae*, order *Brevibacillales* [48].

#### 4.1. Uncovering Enzymes from the MEP Pathway and the Polyprenyl Synthase Family in the SDF Strains

We previously identified 153 putative BGCs in the genomes of 10 SDF strains (Table 2). Among these, 20 different classes of specialised metabolites were identified [32]. In the current work, we focused on 16 BGCs, which were linked to terpene synthesis, to assess the genomic potential of these environmental AEFB for the biosynthesis of these molecules. Bacteria can produce terpenes through the MEP pathway by synthesising IPP, a precursor for these isoprenes [14]. Therefore, we evaluated the putative genetic information associated with the biosynthesis of essential terpene precursors within the MEP pathway. Through Pathways tools, our study has uncovered that all the 10 SDF strains (Table 2) examined contain a significant number of genetic determinants encoding the enzymes DXS, DXR, MCT, CMK, MDS, HDS, and HDR, which are associated with the MEP pathway (Table 1; Figure 2). These findings suggest that the genetic information coding for these pathway enzymes was conserved among these SDF strains, which can be carriers of the basic apparatus for terpene production. This compelling data underscores the necessity for further investigation to explore the potential implications of these fundamental genetic determinants in terpene biosynthesis within AEFB.

The PPS family, which includes the GPPS, FPPS, and GGPPS enzymes (Table 1) produces template molecules derived from IPP (14). These products are subsequently transformed into terpenes by TS enzymes. Specifically, the GPPS enzyme catalyses the conversion of IPP and DMAPP into GPP, the precursor for monoterpenes. DMAPP, an isomer of IPP, is synthesized by the enzyme IDI (Table 1). Using Pathways tools, we identified the IDI genetic information in all 10 SDF strains, except for the *L. fusiformis* SDF0005 genome (Figure 2). These findings suggested that the latter strain might experience some impairment in terpene biosynthesis, possibly affecting the subsequent steps catalysed by PPS family enzymes to synthesize isoprene.

This result revealed a high similarity (>98% identity) among the amino acid sequences in four SDF strains, supporting the occurrence of genetic information codifying for the PPS family enzyme in the respective genomes. *H. oleronia* SDF0015 exhibited 67.86% identity (Table 3). These findings suggested that this strain may lack the genetic information necessary for synthesizing the PPS family enzyme or possesses a structurally distinct catalyst compared to those found in AEFB, likely due to these unique characteristics associated with *H. oleronia*. Consequently, it is reasonable to infer that the remaining SDF strains lacking at least one enzyme from the PPS family may experience limitations

in terpene production, as they do not possess the necessary enzymatic machinery to synthesize intermediates for isoprene precursors.

#### 4.2. Genomic Potential of Selected SDF Strains for Terpene Production

GPP, FPP, or GGPP are used as precursor molecules in natural terpene syntheses by different TSs. The antiSMASH identified the sequence encoding the enzyme of the PSS in at least seven SDF strains (Figure 4; Table 4). Phytoene and squalene molecules are constructed from two molecules of FPP and two GGPP molecules, respectively [51,52]. The enzymes squalene synthase (SQS) and phytoene synthase (PHS) are closely related [52], and SQS is also reported to the synthesis of phytoene [51]. This connection between these enzymes may explain why antiSMASH did not discriminate the genetic sequences for phytoene and squalene productions. The SDF strains containing genes that encode enzymes from the PPS family and PSS putative presented the necessary genetic apparatus to produce either phytoene or squalene, as illustrated in the genomes of *Pe. simplex* SDF0024, *B. velezensis* SDF0141, and *B. velezensis* SDF0150.

Squalene is known for its antioxidant and antitumor properties, and this terpene also enhances the human immune system, as reported by Sanchez-Quesada et al. 2018 ([53]. This specialised metabolite is commonly used as an additive and supplement in the food and personal care industries [54]. The squalene synthesis has been documented for various organisms, including AEFB, which corroborates the findings in this study. Beyond its benefits, squalene serves as an intermediate compound in the biosynthetic pathway of sterols, such as hopanoids [55].

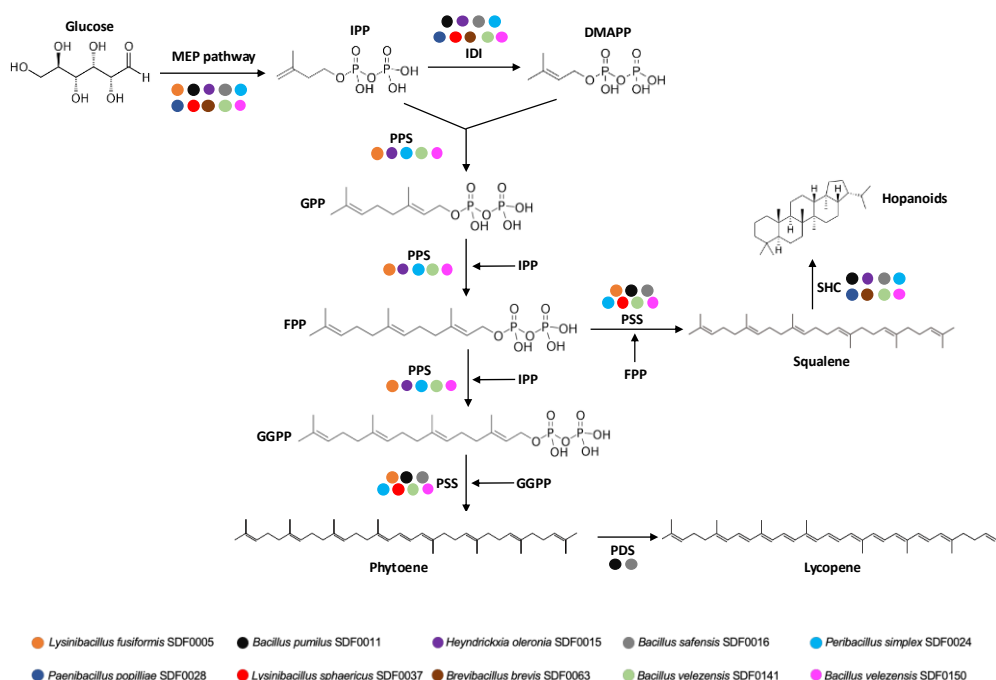
Hopanoids are synthesised from squalene by the enzyme SHC [55]. The *sqhC* gene sequence encodes the SHC enzyme and was identified by antiSMASH in eight SDF strains (Figure 3; Table 4). These data indicate that the SDF strains qualified to produce SHC and the PSS possess the required components for the final enzymatic reactions in hopanoid biosynthesis. This condition was observed in three SDF genomes, specifically *Pe. simplex* SDF0024, *B. velezensis* SDF0141, and *B. velezensis* SDF0150 (Table 4). Hopanoids play a crucial role by integrating into the biological membranes of the producing cells, regulating fluidity and permeability [55,56]. Consequently, these terpenes maintain the bacterial cytoplasmic membrane stable, which is particularly significant given the absence of cholesterol in the membranes of these prokaryotes. While a lack of hopanoids does not hinder bacterial growth, it does affect tolerance to stressful conditions, such as high temperatures and anaerobic or acidic environments [57].

Interestingly, the antiSMASH analysis revealed that the organisation of hopanoid biosynthesis genes in the SDF strains deviated from the typical BGC pattern. While the *sqhC* gene encoding SHC was identified in eight strains, and genes encoding PSS enzymes were found in seven, these genes were not consistently clustered with other hopanoid biosynthesis genes (Figures 3 and 4). This variability suggests that the genetic architecture of hopanoid biosynthesis in SDF strains might be more complex and diverse than previously recognised. Further investigation into the arrangement and regulation of these genes may shed light on this variation in its evolutionary and functional significance. Despite this variability, key hopanoid biosynthesis gene identifications in the genomes of these strains underscore the potential of AEFB as a source of these important membrane components.

The software antiSMASH detected 50% similarity in the BGC sequence involved in the carotenoid's synthesis in the genomes of *B. pumilus* SDF0011 and *B. safensis* SDF0016. The reference strain employed for comparison was *Halobacillus halophilus* DSM2266 (reference number BGC0000645), available through the MIBiG platform. Although the SDF strains did not share the same genetic framework as the reference strain, they possessed the essential genetic information required to express the final enzymatic activities involved in lycopene synthesis. Lycopene, which closely resembles beta-carotene and is widely produced by plants [58], imparts the reddish pigmentation characteristic of tomatoes and watermelons, among other vegetables [59]. This terpene is known for its antioxidant, anti-inflammatory, and antitumor attributes, making it valuable for various applications in the pharmaceutical and food industries.

It is noteworthy that lycopene synthesis in AEFB has been reported in the context of heterologous expression in *B. amyloliquefaciens* and *B. subtilis*, as documented by Zou et al., 2022 [60] and Luo, Bao, and Zhu, 2023 [61]. Additionally, studies have explored natural lycopene production in other AEFB species. For instance, Osawa et al., 2013 [62] investigated the synthesis of an oxidised lycopene in *Cytobacillus firmus* [50], former *Bacillus firmus* [63], while Hwang et al., 2022 [64] also detected genes for lycopene synthesis in *Metabacillus flavus*. The findings of our study are consistent with these previous reports and underscore the potential of AEFB as a source of lycopene. Further investigation is needed to fully elucidate these mechanisms and the evolutionary significance of lycopene production in these bacteria.

Figure 7 summarises the catalytic steps required for terpene production from the MEP pathway to the final reactions by different TSs and the SDF strain carriers for their respective enzymes. Nonetheless, it is essential to note that even though a specific strain of SDF may lack a particular enzyme required for terpene synthesis, this does not automatically imply that the cell is a non-terpene producer. Our study concentrated on the genomic potential of the SDF strain to generate terpenes. Therefore, any inaccuracies in the prior steps, such as purification, extraction, sequencing, and annotation of genome sequences may lead to erroneous results. Additionally, the biosynthetic pathways of terpenes entail multiple enzymatic reactions [65]. Even if the information for a specific enzyme is not detected, the SDF strains examined might still be able to synthesise the predicted terpene because of the promiscuous nature of the enzymes found within the pathway. Otherwise, a complete enzymatic route detection for a given terpene molecule does not assure the synthesis of the corresponding product by this SDF strain due to the complex mechanisms of gene expression. Further research could reveal the synthesis in vitro of the respective molecules detected in this study.



**Figure 7.** Putative biosynthetic pathway of methylerythritol phosphate and terpenes biosynthesis metabolic pathways in the ten SDF strains studied. The reaction steps to synthesize terpenes from isopentenyl diphosphate (IPP), the final product of the methylerythritol phosphate (MEP) route, and the respective catalysts are indicated. The coloured dots indicate the gene sequence coding for an enzyme detected in a particular SDF strain described at the bottom. DMAPP: dimethylallyl pyrophosphate. IDI: isopentenyl diphosphate isomerase. GPP: geranyl pyrophosphate. FPP: farnesyl pyrophosphate. GGPP: geranyl-geranyl pyrophosphate. PSS: phytoene and/or squalene synthase. SHC: squalene hopene cyclase. PDS: phytoene desaturase.

#### 4.3. The Evolutionary Nature of Terpene Production in the SDF Strains

To explore the evolutionary nature of the enzymes involved in terpene production in the evaluated SDF strains, we aligned the amino acid sequences of the SHC (Figure 3) and PSS enzymes (Figure 4) obtained through in silico translation from the gene sequences inside the BGC identified by antiSMASH in this study. This alignment was used to generate phylogenetic trees, as shown in Figures 6A and 6B. However, we did not reconstruct the phylogenetic tree for the PDS enzyme, as it was detected only in two SDF strains analysed (Table 4).

The primary sequences of SHC (Figure 6A) and PSS (Figure 6B) enzymes of *B. velezensis* SDF0140 and *B. velezensis* SDF0150 exhibited a robust molecular relationship, as both strains belong to the same species. In contrast, the strains *B. pumilus* SDF0011 and *B. safensis* SDF0016 showed a significant evolutionary relationship based on their TS amino acid sequences. These two latter species are of biotechnological and pharmaceutical significance and are closely related according to classical phenotypic characteristics and 16S rRNA gene sequences. Consequently, they are challenging to distinguish by these conventional methodologies [28,30,31,66]. In fact, both species comprise a clonally diverse population inside the *B. subtilis* complex [66]. Furthermore, the phylogenetic trees generated using both SHC (Figure 6A) and PSS (Figure 6B) enzymes revealed similar topologies among the *B. pumilus* SDF0011, *B. safensis* SDF0016, *B. velezensis* SDF0141, and *B. velezensis* SDF0150. This result indicates that these four *Bacillus* spp. demonstrate a high level of molecular correlation in their respective TS analyses. Their classification within the same genus likely contributes to this significant conservation of catalytic properties.

In the heatmap (Figure 5), the enzymatic profiles for terpene production in the SDF strains highlighted a strong molecular relationship between *B. pumilus* SDF0011 and *B. safensis* SDF0016, as they share the same enzymatic set. Figure 5 also reveals that *B. velezensis* SDF0141 and *B. velezensis* SDF0150 possess identical catalyst set for terpene production. These results reinforce the molecular relationship among these SDF strains, although the disposition of SDF strains belonged to *Bacillus* spp. did not display the same distribution observed in the phylogenetic tree of TS amino acid sequences (Figure 6).

The strains *H. oleronia* SDF0015 and *Pe. simplex* SDF0024 exhibited a high degree of molecular similarity in their respective amino acid sequences of the SHC enzyme (Figure 6A). The amino acid sequences of SHC from *H. oleronia* SDF0015 and *Pe. simplex* SDF0024 displayed a superior evolutionary distance compared to SHC sequences obtained from the SDF strains within the genus *Bacillus*, therefore positioned in different clades (Figure 6A). *Bacillus*, *Heyndrickxia*, and *Peribacillus* are all classified within the family *Bacillaceae* and species in these genera present significant levels of polyphyly [47]. This observation further corroborated phylogenetic relationships derived from the SHC amino acid sequences (Figure 6A).

*P. popilliae* SDF0028 and *Br. brevis* SDF0063 presented the highest molecular relationship in their SHC amino acid sequences if compared to the SHC sequences of the remaining six SDF strains analysed (Figure 6A). Genera *Bacillus*, *Peribacillus*, and *Heyndrickxia*, which are part of the family *Bacillaceae*, belong to the order *Bacillales*, while the genera *Paenibacillus* and *Brevibacillus* are classified under orders *Paenibacillales* and *Brevibacillales*, respectively [48,49]. Therefore, the minimal molecular relationship between the SHC amino acid sequences of the *P. popilliae* SDF0028 and *Br. brevis* SDF0063 and the other SDF strains aligns with the anticipated evolutionary distance for these species. Additionally, the same enzymatic set for terpene production was observed between these two SDF strains (Figure 5), supporting even more the phylogenetic relationship obtained by the sequence of SHC of *P. popilliae* SDF0028 and *Br. brevis* SDF0063 (Figure 6A). Notably, the amino acid sequences of SHC from *H. oleronia* SDF0015 and *Pe. simplex* SDF0024 were recognised to be evolutionarily closer to those of *P. popilliae* SDF0028 and *Br. brevis* SDF0063 than the SHC sequences from SDF strains within the genus *Bacillus* (Figure 6A).

*L. fusiformis* SDF0005 and *L. sphaericus* SDF0037 shared highly conserved PSS amino acid sequences (Figure 6B). This commonality might be attributed to the two SDF strains from the genus *Lysinibacillus*. The phylogenetic tree generated using the sequences revealed a clear distinction



between the SDF strains of the genus *Bacillus*, which grouped in one clade, and the SDF strains of the genus *Lysinibacillus*, which formed a separate clade (Figure 6B). These results discriminated the members of the families *Bacillaceae* and *Caryophanaceae* based on the amino acid sequence of this catalyst. Furthermore, a relevant molecular relationship for terpene production enzymatic set was already observed between the *Lysinibacillus* spp. as shown in the heatmap (Figure 5). Markedly, the sequences of *L. fusiformis* SDF0005 and *L. sphaericus* SDF0037 demonstrated a closer evolutionary relationship with the sequence of the *Pe. simplex* SDF0024, which is allocated within the family *Bacillaceae*.

The phylogenetic trees derived from the molecular analyses of the TSs detected in the SDF strains examined in this study agreed with the phylogenetic relationships for complete genomes of the SDF strains evaluated (Figure 1). The results indicated that the TS enzymes responsible for terpene production in these environmental strains investigated are evolutionarily conserved. In addition, the production of terpenes—strikingly squalene and hopanoids—appears to be an ancestral characteristic of the AEFB evaluated in this study.

As mentioned above, the phylogenetic trees derived from the amino acid sequences of TS enzymes (Figures 6A and 6B) indicated a closer molecular relationship between *B. velezensis* SDF0141 and *B. velezensis* SDF0150 to *B. pumilus* SDF0011 and *B. safensis* SDF0016. In contrast, the heatmap (Figure 5) groups these two *B. velezensis* strains with *Pe. simplex* SDF0024, along with *L. fusiformis* SDF0005 and *L. sphaericus* SDF0037. Additionally, the strains *P. popilliae* SDF0028 and *Br. brevis* SDF0063 demonstrated similar enzyme content for terpene production and clustered with the strain *H. oleronia* SDF0015 (Figure 5).

The cluster of catalysts implicated in terpene production shown in Figure 5 shares similarities with the phylogenetic trees generated from the TS amino acid sequences (Figures 6A and 6B). Nevertheless, the cluster does not fully align with the expected evolutionary relationships among these species. This analysis suggested that the ability to produce terpene molecules may vary among the SDF strains and might not be influenced by phylogenetic factors. Furthermore, the enzyme set involved in terpene production cannot be used as a molecular marker to establish evolutionary relationships among the SDF strains analysed and, by extension, for AEFB.

## 5. Conclusions

The AEFB are ubiquitous and characterised by producing several specialised metabolites, including terpenes, a significant class. While the synthesis of terpenes has been demonstrated in prokaryotes, research addressing specifically the production of these compounds in AEFB is still scarce. In this study, in silico analysis revealed that strains belonging to these taxa possess the genetic information required to synthesise at least three terpenes: squalene, hopanoids, and lycopene. The metabolic pathway information for synthesising terpene precursors was identified in the genomes of the 10 SDF strains evaluated, and the amino acid sequences of terpene synthases detected revealed a functional equivalence among these catalysts. Although the identified terpene classes represent a small fraction of these isoprene molecules, our findings can support future investigations to broaden the understanding of the physiological and ecological terpene roles in AEFB. Furthermore, our results can help enhance the industrial application of these molecules.

**Authors' contributions:** Conceptualization, Felipe Mesquita, Waldeyr Silva and Marlene DE-SOUZA; Data curation, Felipe Mesquita, Waldeyr Silva, Marcelo Brigido, Bruna Fuga and Marlene DE-SOUZA; Formal analysis, Felipe Mesquita, Waldeyr Silva, Taina Raiol, Nalvo Almeida, Bruna Fuga, Danilo Cavalcante and Marlene DE-SOUZA; Funding acquisition, Marcelo Brigido and Nalvo Almeida; Investigation, Felipe Mesquita, Waldeyr Silva and Marlene DE-SOUZA; Methodology, Felipe Mesquita, Waldeyr Silva, Taina Raiol, Marcelo Brigido, Nalvo Almeida, Bruna Fuga, Danilo Cavalcante and Marlene DE-SOUZA; Project administration, Marlene DE-SOUZA; Resources, Felipe Mesquita, Waldeyr Silva, Danilo Cavalcante and Marlene DE-SOUZA; Software, Felipe Mesquita, Waldeyr Silva, Taina Raiol, Marcelo Brigido, Nalvo Almeida and Bruna Fuga; Supervision, Marlene DE-SOUZA; Validation, Felipe Mesquita, Waldeyr Silva, Nalvo Almeida and Marlene DE-SOUZA; Writing – original draft, Marlene DE-SOUZA; Writing – review & editing, Felipe Mesquita, Waldeyr Silva and Marlene DE-SOUZA.

**Funding:** This work was funded by the Brazilian agencies: NA acknowledges funding from National Council for Scientific and Technological Development (CNPq) grant 304423/2022-04; MB acknowledges funding from Fundação de Amparo a Pesquisa do Distrito Federal (FAP-DF) (FAP-DF) grant 0193-000.560/2009. FM and DC acknowledge CNPq fellowships.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Singh, B.P.; Rateb, M.E.; Rodriguez-Couto, S.; Polizeli, M.D.L.T.D.M.; Li, W.-J. Microbial Secondary Metabolites: Recent Developments and Technological Challenges. *Frontiers in Microbiology* **2019**, *10*, 914. <https://doi.org/10.3389/fmicb.2019.00914>
2. Bills, G.F.; Gloer, J.B. Biologically Active Secondary Metabolites from the Fungi. *Microbiology Spectrum* **2016**, *4*, 10.1128. <https://doi.org/10.1128/microbiolspec.funk-0009-2016>
3. Perveen, S.; Al-Taweel, A. *Terpenes and Terpenoids*; IntechOpen: London, United Kingdom, 2018; pp. 01-152. ISBN 978-1-83881-529-5.
4. Rudolf, J.D.; Aslup, T.; Xu, B.; Li, Z. Bacterial Terpenome. *Natural Product Reports* **2021**, *38*, 905-980. <https://doi.org/10.1039/d0np00066c>
5. Quin, M.B.; Flynn, C.M.; Schmidt-Dannert, C. Traversing the Fungal Terpenome. *Natural Product Reports* **2014**, *31*, 1449-1473. <https://doi.org/10.1039/c4np00075g>
6. Hegazy, M.E.F.; Mohamed, T.A.; Alhammady, M.A.; Shaheen, A.M.; Reda, E.H.; Elshamy, A.I.; Aziz, M.; Paré, P.W. Molecular Architecture and Biomedical Leads of Terpenes from Red Sea Marine Invertebrates. *Marine Drugs* **2015**, *13*, 3154-3181. <https://doi.org/10.3390/md13053154>
7. Yamada, Y.; Kuzuyama, T.; Komatsu, M.; Ikeda, H. Terpene Synthases are Widely Distributed in Bacteria. *PNAS* **2015**, *112*, 857-862. <https://doi.org/10.1073/pnas.1422108112>
8. Morandini, L.; Caulier, S.; Bragard, C.; Mahillon, J. *Bacillus cereus sensu lato* Antimicrobial Arsenal: An Overview. *Microbiological Research* **2024**, *283*, 127697. <http://dx.doi.org/10.1016/j.micres.2024.127697>
9. Pinto-Zevallos, D.M.; Hellén, H.; Hakola, H.; Nouhuys, S.V.; Halopainen, J.K. Induced defenses of *Veronica spicata*: Variability in herbivore-induced volatile organic compounds. *Phytochemistry Letters* **2013**, *6*, 653-656. <https://doi.org/10.1016/j.phytol.2013.08.015>
10. Zheng, D.; Ding, N.; Jiang, Y.; Zhang, J.; Ma, J.; Chen, X.; Liu, J.; Han, L.; Huang, X. Albaflavenoid, a New Tricyclic Sesquiterpenoid from *Streptomyces violascens*. *The Journal of Antibiotics* **2016**, *69*, 773-775.
11. Netzker, T.; Shepherdson, E.M.F.; Zambri, M.P.; Elliot, M.A. Bacterial Volatile Compounds: Functions in Communication, Cooperation, and Competition. *Annual review of microbiology* **2020**, *74*, 409-430. <https://doi.org/10.1146/annurev-micro-011320-015542>
12. Tyc, O.; Song, C.; Dickschat, J.S.; Vos, M.; Garbeva, P. The Ecological Role of Volatile and Soluble Secondary Metabolites Produced by Soil Bacteria. *Trends in microbiology* **2017**, *25*, 280-292. <https://doi.org/10.1016/j.tim.2016.12.002>
13. Twaij, B.M.; Hasan, M.N. Bioactive Secondary Metabolites from Plant Sources: Types, Synthesis, and Their Therapeutic Uses. *International Journal of Plant Biology* **2022**, *13*, 4-14. <https://doi.org/10.3390/ijpb13010003>
14. Liang, Z.; Zhi, H.; Fang, Z.; Zhang, P. Genetic engineering of yeast, filamentous fungi and bacteria for terpene production and applications in food industry. *Food research international* **2021**, *147*, 110487. <https://doi.org/10.1016/j.foodres.2021.110487>
15. Driks, A.; Eichenberger, P. *The Bacterial Spore: From Molecules to Systems*; ASM Press: Washington, DC, United States of America, 2016; pp. 01-397. ISBN 9781555816759.
16. Christie, G.; Setlow, P. *Bacillus* spore germination: Knowns, unknowns and what we need to learn. *Cellular signaling* **2020**, *74*, 109729. <https://doi.org/10.1016/j.cellsig.2020.109729>
17. Oren, A.; Garrity, G.M. Valid publication of the names of forty-two phyla of prokaryotes. *International journal of systematic and evolutionary microbiology* **2021**, *71*, 005056. <https://doi.org/10.1099/ijsem.0.005056>
18. Galperin, M.Y.; Yutin, N.; Wolf, Y.I.; Alvarez, R.V.; Koonin, E.V. Conservation and Evolution of the Sporulation Gene Set in Diverse Members of the Firmicutes. *Journal of Bacteriology* **2022**, *204*, e00079-22. <https://doi.org/10.1128/jb.00079-22>

19. Fritze, D. Taxonomy of the Genus *Bacillus* and Related Genera: The Aerobic Endospore-Forming Bacteria. *Phytopathology* **2004**, *94*, 1245-1248. <https://doi.org/10.1094/phyto.2004.94.11.1245>
20. Mandic-Mulec, I.; Prosser, J.I. Diversity of Endospore-forming Bacteria in Soil: Characterization and Driving Mechanisms. In *Endospore-forming Soil Bacteria*, 1st ed.; Logan, N.A.; Vos, P., Eds.; Springer-Verlag: Berlin, Germany, 2011; Volume 27, pp. 31-59. ISBN 978-3-642-19577-8.
21. Logan, N.A. *Bacillus* and relatives in foodborne illness. *Journal of Applied Microbiology* **2012**, *112*, 417-429. <https://doi.org/10.1111/j.1365-2672.2011.05204.x>
22. Harirchi, S.; Sar, T.; Ramezani, M.; Aliyu, H.; Etemadifar, Z.; Nojoumi, S.A.; Yazdian, F.; Awasthi, M.K.; Taherzadeh, M.J. *Bacillales*: From Taxonomy to Biotechnological and Industrial Perspectives. *Microorganisms* **2022**, *10*, 2355. <https://doi.org/10.3390/microorganisms10122355>
23. Sumi, C.D.; Yang, B.W.; Yeo, I.-C.; Hahm, Y.T. Antimicrobial peptides of the genus *Bacillus*: a new era for antibiotics. *Canadian journal of microbiology* **2015**, *61*, 93-103. <https://doi.org/10.1139/cjm-2014-0613>
24. Heilbronner, S.; Krismer, B.; Brötz-Oesterheld, H.; Peschel, A. The microbiome-shaping roles of bacteriocins. *Natures Reviews Microbiology* **2021**, *19*, 726-739. <https://doi.org/10.1038/s41579-021-00569-w>
25. Salazar, B.; Ortiz, A.; Keswani, C.; Minkina, T.; Mandzhieva, S.; Singh, S.P.; Rekadwad, B.; Borriss, R.; Jain, A.; Singh, H.B.; Sansinenea, E. *Bacillus* spp. as Bio-factories for Antifungal Secondary Metabolites: Innovation Beyond Whole Organism Formulations. *Microbial ecology* **2023**, *86*, 1-24. <https://doi.org/10.1007/s00248-022-02044-2>
26. Mondol, M.A.M.; Shin, H.J.; Islam, M.T. Diversity of Secondary Metabolites from Marine *Bacillus* Species: Chemistry and Biological Activity. *Marine Drugs* **2013**, *11*, 2846-2872. <https://doi.org/10.3390/md11082846>
27. Falqueto, S.A.; Pitaluga, B.F.; Sousa, J.R.; Targanski, S.K.; Campos, M.G.; Mendes, T.A.O.; Silva, G.F.; Silva, D.H.S.; Soares, M.A. *Bacillus* spp. metabolites are effective in eradicating *Aedes aegypti* (Diptera: Culicidae) larvae with low toxicity to non-target species. *Journal of invertebrate pathology* **2021**, *179*, 107525. <https://doi.org/10.1016/j.jip.2020.107525>
28. Orem, J.C.; Silva, W.M.C.; Raiol, T.; Magalhães, M.I.; Martins, P.H.; Cavalcante, D.A.; Kruger, R.H.; Brigido, M.M.; De-Souza, M.T. Phylogenetic diversity of aerobic spore-forming *Bacillales* isolated from Brazilian soils. *International Microbiology* **2019**, *22*, 511-520. <https://doi.org/10.1007/s10123-019-00080-6>
29. Cavalcante, D.A.; De-Souza, M.T.; Orem, J.C.; Magalhães, M.I.A.; Martins, P.H.; Boone, T.J.; Castillo, J.A.; Driks, A. Ultrastructural analysis of spores from diverse *Bacillales* species isolated from Brazilian soil. *Environmental Microbiology Reports* **2019**, *11*, 155-164. <https://doi.org/10.1111/1758-2229.12713>
30. Martins, P.H.R.; Silva, L.P.; Orem, J.C.; Magalhães, M.I.A.; Cavalcante, D.A.; De-Souza, M.T. Protein profiling as a tool for identifying environmental aerobic endospore-forming bacteria. *Open Journal of Bacteriology* **2020**, *4*, 001-007.
31. Martins, P.H.R.; Rabinovitch, L.; Orem, J.C.; Silva, W.M.C.; Mesquita, F.A.; Magalhães, M.I.A.; Cavalcante, D.A.; Vivoni, A.M.; Oliveira, E.J.; Lima, V.C.P.; Brito, J.T.; De-Souza, M.T. Biochemical, physiological, and molecular characterisation of a large collection of aerobic endospore-forming bacteria isolated from Brazilian soils. *Neotropical Biology and Conservation* **2023**, *18*, 53-72. <https://doi.org/10.3897/neotropical.18.e86548>
32. Mesquita, F.A.; Silva, W.M.C.; De-Souza, M.T. In silico Analysis of the Genomic Potential for the Production of Specialized Metabolites of Ten Strains of the *Bacillales* Order Isolated from the Soil of the Federal District, Brazil. In *Advances in Bioinformatics and Computational Biology*, 1st ed.; Scherer, N.M.; Melo-Minardi, R.C., Eds.; Springer-Verlag: Brazil, 2022, Volume 13523, pp. 158-163. ISBN 978-3-031-21175-1.
33. Coil, D.; Jospin, G.; Darling, A.M. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina Miseq data. *Bioinformatics* **2015**, *31*, 587-589. <https://doi.org/10.1093/bioinformatics/btu661>
34. Tatusova, T.; DiCuccio, M.; Badretdin, A.; Chetvernin, V.; Nawrocki, E.P.; Zaslavsky, L.; Lomsadze, A.; Pruitt, K.D.; Borodovsky, M.; Ostell, J. NCBI prokaryotic genome annotation pipeline. *Nucleic Acid Research* **2016**, *44*, 6614-6624. <https://doi.org/10.1093/nar/gkw569>
35. Setubal, J.C.; Almeida, N.F.; Wattam, A.R. Comparative Genomics for Prokaryotes. *Methods in Molecular Biology* **2018**, *1704*, 55-78. [https://doi.org/10.1007/978-1-4939-7463-4\\_3](https://doi.org/10.1007/978-1-4939-7463-4_3)
36. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: architecture and applications. *BMC Bioinformatics* **2009**, *10*, 421. <https://doi.org/10.1186/1471-2105-10-421>

37. Li, L.; Stoeckert Jr, J.; Roos, D.S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research* **2003**, *13*, 2178-2189. <https://doi.org/10.1101/gr.1224503>
38. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* **2000**, *17*, 540-552. <https://doi.org/10.1093/oxfordjournals.molbev.a026334>
39. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312-1313. <https://doi.org/10.1093/bioinformatics/btu033>
40. Blin, K.; Shaw, S.; Kloosterman, A.M.; Charlop-Powers, Z.; vanWezel, G.P.; Medema, M.H.; Weber, T. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Research* **2021**, *49*, W29-W35. <https://doi.org/10.1093/nar/gkab335>
41. Bogdanove, A.J.; Koebnik, R.; Lu, H.; Furutani, A.; Angiuoli, S.V.; Patil, P.B.; Sluys, M.A.V.; Ryan, R.P.; Meyer, D.F.; Han, S.-W.; Aparna, G.; Rajaram, M.; Delcher, A.L.; Phillip, A.M.; Puiu, D.; Schatz, M.C.; Shumway, M.; Sommer, D.D.; Trapnell, C.; Benahmed, F.; Dimitrov, G.; Madupu, R.; Radune, D.; Sullivan, S.; Jha, G.; Ishihara, H.; Lee, S.-W.; Pandey, A.; Sharma, V.; Sriariyanun, M.; Szurek, B.; Vera-Cruz, C.M.; Dorman, K.S.; Ronald, P.C.; Verdier, V.; Dow, M.; Sonti, R.V.; Tsuge, S.; Brendel, V.P.; Rabinowicz, P.D.; Leach, J.E.; White, F.F.; Salzberg, S.L. Two New Complete Genome Sequences Offer Insight into Host and Tissue Specificity of Plant Pathogenic *Xanthomonas* spp. *Journal of Bacteriology* **2011**, *193*, 5450-5464. <https://doi.org/10.1128/JB.05262-11>
42. Gu, Z.; Eils, R.; Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **2016**, *32*, 2847-2849. <https://doi.org/10.1093/bioinformatics/btw313>
43. Hummel, M.; Edelmann, D.; Kopp-Schneider, A. Clustering of samples and variables with mixed-type data. *PLoS ONE* **2017**, *12*, e0188274. <https://doi.org/10.1371/journal.pone.0188274>
44. Medema, M.H.; Kottmann, R.; Yilmaz, P. Minimum Information about a Biosynthetic Gene cluster. *Nature Chemical Biology* **2015**, *11*, 625-631. <https://doi.org/10.1038/nchembio.1890>
45. Köcher, S.; Breitenbach, J.; Müller, V.; Sandmann, G. Structure, function and biosynthesis of carotenoids in the moderately halophilic bacterium *Halobacillus halophilus*. *Archives of Microbiology* **2009**, *191*, 95-104. <https://doi.org/10.1007/s00203-008-0431-1>
46. Xu, X.; Kóvacs, A.T. How to identify and quantify the members of the *Bacillus* genus? *Environmental Microbiology* **2024**, *26*, e16593. <https://doi.org/10.1111/1462-2920.16593>
47. Gupta, R.S.; Patel, S. Robust Demarcation of the Family *Caryophanaceae* (*Planococcaceae*) and Its Different Genera Including Three Novel Genera Based on Phylogenomics and Highly Specific Molecular Signatures. *Frontiers in Microbiology* **2020**, *10*, 2821. <https://doi.org/10.3389/fmicb.2019.02821>
48. Chuvochina, M.; Mussig, A.J.; Chaumeil, P.-A.; Skarkshewski, A.; Rinke, C.; Parks, D.H.; Hugenholtz, P. Proposal of names for 329 higher taxa defined in the Genome Taxonomy Database under two prokaryotic codes. *FEMS Microbiology Letters* **2023**, *370*, 1-33. <https://doi.org/10.1093/femsle/fnad071>
49. Gupta, R.S.; Patel, S.; Saini, N.; Chen, S. Robust demarcation of 17 distinct *Bacillus* species clades, proposed as novel *Bacillaceae* genera, by phylogenomics and comparative genomic analyses: description of *Robertmurraya kyonggiensis* sp. nov. and proposal for an emended genus *Bacillus* limiting it only to the members of the *Subtilis* and *Cereus* clades of species. *International Journal of Systematic and Evolutionary Microbiology* **2020**, *70*, 5753-5798. <https://doi.org/10.1099/ijsem.0.004475>
50. Patel, S.; Gupta, R.S. A phylogenomic and comparative genomic framework for resolving the polyphyly of the genus *Bacillus*: Proposal for six new genera of *Bacillus* species, *Peribacillus* gen. nov., *Cytobacillus* gen. nov., *Mesobacillus* gen. nov., *Neobacillus* gen. nov., *Metabacillus* gen. nov. and *Alkalihalobacillus* gen. nov. *International Journal of Systematic and Evolutionary Microbiology* **2020**, *70*, 406-438. <https://doi.org/10.1099/ijsem.0.003775>
51. Nakashima, T.; Inoue, T.; Oka, A.; Nishino, T.; Osumi, T.; Hata, S. Cloning, expression, and characterization of cDNAs encoding *Arabidopsis thaliana* squalene synthase. *PNAS* **1995**, *98*, 2328-2332. <https://doi.org/10.1073/pnas.92.6.2328>
52. Tansey, T.R.; Shechter, I. Squalene synthase: Structure and regulation. *Progress in Nucleic Acid Research and Molecular Biology* **2000**, *65*, 157-195. [https://doi.org/10.1016/S0079-6603\(00\)65005-5](https://doi.org/10.1016/S0079-6603(00)65005-5)



53. Sánchez-Quesada, C.; López-Biedma, A.; Toledo, E.; Gaforio, J.J. Squalene Stimulates a Key Innate Cell to Foster Wound Healing and Tissue Repair. *Evidence-Based Complementary and Alternative Medicine* **2018**, 2018, 9473094. <https://doi.org/10.1155/2018/9473094>
54. Song, Y.; Guan, Z.; Merkerk, R.V.; Pramastya, H.; Abdallah, I.I.; Setroikromo, R.; Quax, W.J. Production of Squalene in *Bacillus subtilis* by Squalene Synthase Screening and Metabolic Engineering. *Journal of Agricultural and Food Chemistry* **2020**, 68, 4447–4455. <https://dx.doi.org/10.1021/acs.jafc.0c00375?ref=pdf>
55. Siedenburg, G.; Jendrossek, D. Squalene-Hopene Cyclases. *Applied and Environmental Microbiology* **2011**, 77, 3905–3915. <https://doi.org/10.1128/aem.00300-11>
56. Iqbal, S.; Begum, F.; Rabaan, A.A.; Aljeldah, M.; Al Shammari, B.R.; Alawfi, A.; Alshengeti, A.; Sulaiman, T.; Khan, A. Classification and Multifaceted Potential of Secondary Metabolites Produced by *Bacillus subtilis* Group: A Comprehensive Review. *Molecules* **2023**, 28, 927. <https://doi.org/10.3390/molecules28030927>
57. Belin, B.J.; Busset, N.; Giraud, E.; Molinaro, A.; Silipo, A.; Newman, D.K. Hopanoid lipids: from membranes to plant-bacteria interactions. *Nature Reviews Microbiology* **2018**, 16, 304–315. <https://doi.org/10.1038/nrmicro.2017.173>
58. Sabio, E.; Lozano, M.; Espinosa, V.M.; Mendes, R.L.; Pereira, A.P.; Palavra, A.F.; Coelho, J.A. Lycopene and  $\beta$ -Carotene Extraction from Tomato Processing Waste Using Supercritical CO<sub>2</sub>. *Industrial and Engineering Chemistry Research* **2003**, 42, 6641–6646. <http://dx.doi.org/10.1021/ie0301233>
59. Li, L.; Liu, Z.; Jiang, H.; Mao, X. Biotechnological production of lycopene by microorganisms. *Applied Microbiology and Biotechnology* **2020**, 104, 10307–10324. <https://doi.org/10.1007/s00253-020-10967-4>
60. Zou, D.; Ye, C.; Min, Y.; Li, L. Production of a novel lycopene-rich soybean food by fermentation with *Bacillus amyloliquefaciens*. *Food Science and Technology* **2022**, 153, 112551. <https://doi.org/10.1016/j.lwt.2021.112551>
61. Luo, H.; Bao, Y.; Zhu, P. Development of a novel functional yogurt rich in lycopene by *Bacillus subtilis*. *Food Chemistry* **2023**, 407, 135142. <https://doi.org/10.1016/j.foodchem.2022.135142>
62. Osawa, A.; Iki, K.; Sandmann, G.; Shindo, K. Isolation and identification of 4,4'-diapolyycopene-4,4'-diol acid produced by *Bacillus firmus* GB1 and its singlet oxygen quenching activity. *Journal of Oleo Science* **2013**, 62, 955–960. <https://doi.org/10.5650/jos.62.955>
63. Werner, W. Botanische Beschreibung häufiger am Buttersäureabbau beteiligter sporenbildender Bakterienspezies. *Zentralblatt für Bakteriologie, Parasitenkunde, Infektionskrankheiten und Hygiene. Abteilung II* **1933**, 87, 446–475.
64. Hwang, C.Y.; Cho, E.-S.; Yoon, D.J.; Cha, I.-T.; Jung, D.-H.; Nam, Y.-D.; Park, S.-L.; Lim, S.-I.; Seo, M.-J. Genomic and Physiological Characterization of *Metabacillus flavus* sp. nov., a Novel Carotenoid-Producing Bacilli Isolated from Korean Marine Mud. *Microorganisms* **2022**, 10, 979. <https://doi.org/10.3390/microorganisms10050979>
65. Vattekkatte, A.; Garms, S.; Brandt, W.; Boland, W. Enhanced structural diversity in terpenoid biosynthesis: enzymes, substrates and cofactors. *Organic and Biomolecular Chemistry* **2018**, 16, 348–362. <https://doi.org/10.1039/c7ob02040f>
66. Branquinho, R.; Meirinhos-Soares, L.; Carriço, J.A.; Pintado, M.; Peixe, L.V. Phylogenetic and clonality analysis of *Bacillus pumilus* isolates uncovered a highly heterogeneous population of different closely related species and clones. *FEMS Microbiology Ecology* **2014**, 90, 689–698. <https://doi.org/10.1111/1574-6941.12426>

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.