



OPEN ACCESS

EDITED BY

Jørgen J. Leisner,
University of Copenhagen, Denmark

REVIEWED BY

Maria Carla Martini,
Worcester Polytechnic Institute, United States
Sajid Iqbal,
Oujiang Laboratory (Zhejiang Laboratory for
Regenerative Medicine, Vision, and Brain
Health), China

*CORRESPONDENCE

Blaise Waongo
✉ wangblaise85@gmail.com

RECEIVED 23 October 2024

ACCEPTED 21 May 2025

PUBLISHED 16 June 2025

CITATION

Waongo B, Ndayishimiye L, Tapsoba F, Zongo W-SA, Li J and Savadogo A (2025) Prospection for potential new non-ribosomal peptide gene clusters in *Bacillus* genus isolated from fermented foods and soil through genome mining. *Front. Microbiol.* 16:1515483. doi: 10.3389/fmicb.2025.1515483

COPYRIGHT

© 2025 Waongo, Ndayishimiye, Tapsoba, Zongo, Li and Savadogo. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Prospection for potential new non-ribosomal peptide gene clusters in *Bacillus* genus isolated from fermented foods and soil through genome mining

Blaise Waongo^{1*}, Libère Ndayishimiye², François Tapsoba¹, Wend-Soo Abel Zongo³, Jinquan Li² and Aly Savadogo¹

¹Laboratory of Applied Biochemistry and Immunology, University Joseph KI-ZERBO, Ouagadougou, Burkina Faso, ²National Key Laboratory of Agricultural Microbiology, Key Laboratory of Environment Correlative Dietology, College of Food Science and Technology, College of Biomedicine and Health, Huazhong Agricultural University, Wuhan, China, ³College of Food Science and Technology, Zhejiang University of Technology, Hangzhou, China

Experimental studies, though often very costly, lead to the discovery of known antimicrobial products. Yet, pathogenic microorganisms are proving increasingly resistant to pre-existing antimicrobial molecules, and this is a cause for worldwide concern. Therefore, it is necessary to search for new molecules that could serve as alternatives in the food, medical and agricultural sectors. Thus, 123 complete genomes of *Bacillus* strains isolated from soil and fermented foods were analyzed and annotated using bioinformatics prediction and characterization tools. The view was to discover new gene clusters for the biosynthesis of non-ribosomal peptides (lipopeptides, siderophores, antibiotics). This study revealed that 83% of the genomes analyzed possess biosynthetic gene clusters for the production of siderophore bacillibactin, 61% for surfactins, 37% for fengycins, 23% for iturins, 15% for kurstakins and 3% for bacitracin. Besides, seven new biosynthetic gene clusters coding Non Ribosomal Peptide Synthetases (NRPS) have been identified in *B. velezensis* ATR2, *B. velezensis* DSYZ, *B. velezensis* CGMCC11640, *B. amyloliquefaciens* HM618, *B. amyloliquefaciens* WF02, *B. cereus* CMCC P0011, *B. cereus* CMCC P0021, *B. subtilis* SJ-10 and *B. anthracis* CMF9. The results of this study revealed a significant potential of the genus *Bacillus* to produce new non-ribosomally synthesized peptides. Now, these predicted new antimicrobial molecules can be easily studied experimentally as many new gene clusters have been identified.

KEYWORDS

Bacillus, nonribosomal peptide, lipopeptide, biosynthetic gene clusters, genome mining

Highlights

- Seven novel biosynthetic gene clusters that encode Non-Ribosomal Peptide Synthetases (NRPS) were discovered.
- A strain of *Bacillus* can potentially co-produce 4 families of lipopeptides.
- *Bacillus* genus has a significant potential to develop new non-ribosomally produced peptides.
- Genome mining strategy makes it possible to discover new metabolites.

1 Introduction

Bacillus are Gram-positive bacteria belonging to the firmicute genus and *Bacillaceae* family. They are spore-forming, aerobic or facultative aero-anaerobic and get energy by respiration or fermentation (Tamang et al., 2016). Their sporulating ability allows them to resist adverse environmental conditions. They are found in water, soil, dust, plants, food and in animals digestive tract. Some species of the genus *Bacillus* have important roles as antibiotics or antifungals producers (Khurana et al., 2020). For example, locillomycins and surfactin from *B. subtilis* 916 show antibacterial activity against *Xanthomonas oryzae* and *Fusarium oxysporum* (Zhao et al., 2018). Nowadays, 788 species from the genus *Bacillus* are sequenced, leading to various studies and explorations (www.ncbi.nlm.nih.gov/genome/?term=Bacillus, consulted 05/07/2024).

Indeed, many *Bacillus* strains have the ability to produce a wide variety of extracellular enzymes and nonribosomal secondary metabolites such as lipopeptides, the bacillibactin siderophore or antimicrobial compounds such as bacitracin which is a semi-cyclic peptide antibiotic commercialized as a mixture of polypeptides for the treatment of Gram-positive bacterial infections (May et al., 2001; Tapi et al., 2010). Other peptides, such as the lipopeptides have antimicrobial properties and can be used in food or soil biopreservation (Zhao et al., 2018).

Bacillus lipopeptides are subdivided into five families (fengycin, iturin, surfactin, kurstakin, and locillomycin) in which various subfamilies and variants can be found (Lam et al., 2021; Waongo et al., 2023). They are produced by modular mega-enzymes or assembly lines referred to as NRPSs (Iqbal et al., 2021). The NRPSs are composed of modules responsible for the incorporation of amino acids in the final nonribosomal peptide (Th   tre et al., 2021). The modules contain the following main catalytic domains: the adenylation domain (A), responsible for the amino acid selection and activation, the carrier protein (CP), responsible for tethering the amino acid to the enzyme, and the condensation domain (C), forming a peptide bond between two amino acids attached to two consecutive modules (Lecl  re et al., 2016). Some NRPSs also include optional domains that can modify the incorporated amino acids during the synthesis, as the Epimerisation domain (E) leading to the D-isomery (Duban et al., 2022). The enzymatic domains work step by step to assemble the monomers into the peptide, so that NRPSs can be compared to assembly lines. A particular trait of NRPSs synthesising Lipopeptides (LP), except those belonging to the iturin and locillomycin families, is the presence of a condensation domain starting the assembly line. The role of this so-called C-starter is to condense the fatty acid into the first amino acid of the peptide chain. The fengycin family includes decapeptides fengycin A, B, C, and plipastatins A, B and S. The iturin family includes bacillomycins (D, DC, F, L and Lc), mycosubtilin, mojavensin A, subtilene A, mixirin and iturins (A, A_L and C), all containing 7 amino acids. Heptapeptides are also found in the surfactin family and variant forms of lichenysin, pumilacidin, esperin and kurstakin. The fifth family contains the nonapeptides locillomycins A, B, and C (Luo et al., 2015a; Waongo et al., 2023).

Many drugs remain effective; however, the emergence of resistant organisms has become a major concern, drawing global attention and promoting the “One Health” concept (Aslam et al., 2021). This concept highlights the close links between human, animal, and environmental health. In this context, nonribosomal

secondary metabolites produced by the genus *Bacillus* can be a source of interesting active compounds. Indeed, some antimicrobial peptides have been shown to be antibiotics with a broad spectrum against pathogens and lipopeptides and siderophores, which are not antibiotics but represent an alternative for food preservation and plant protection against phytopathogenic fungi (Huan et al., 2020).

An urgent challenge today is to discover new natural drugs to tackle emerging human, animal and plant pathogens. The identification of these biomolecules is usually performed by experimental studies including isolation, characterization, purification and activity tests (Fanaei et al., 2021). With the development of bioinformatics tools, this approach can now be accelerated and complemented by sequencing and subsequent genome or metagenome mining to identify natural product biosynthetic pathways (Blin et al., 2021).

Regarding nonribosomal peptide synthesis, specific tools have been developed to identify biosynthetic gene clusters in the genome sequences, to decipher the organization of NRPS into modules and domains, and to predict the nature of the monomers activated by each Adenylation-domain and their isomery in regard to the presence of epimerization domains (Blin et al., 2023). Notably, bacterial genome mining via *in silico* analysis, using the program antiSMASH, offers an attractive opportunity to discover new secondary metabolites such as NRPs (Rahman et al., 2014; Aleti et al., 2015). AntiSMASH 7 allows better visualization of enzyme assembly chain and good structuring of predicted molecules. Predicted peptides are likely to be characterized by comparison with other peptides available in Norine database (Flissi et al., 2020). The questions addressed here are (i) to review the structural and functional information together with annotation of gene clusters for known lipopeptides (LPs), antibiotics and siderophores produced by the genus *Bacillus* and (ii) to elucidate through genome mining the potential products of yet uncharacterized nonribosomal gene clusters. In addition, a bibliographical search was carried out to better understand the lipopeptides produced by *Bacillus*. The objective of this study is to combine literature exploration and genome analysis to identify and characterize the diversity of nonribosomal peptides as antibiotics, siderophores, and lipopeptides, which are potentially produced by *Bacillus* strains isolated from fermented foods and soil.

2 Materials and methods

2.1 An overview of known structures of nonribosomal lipopeptides produced by *Bacillus* strains

An overview of known structures of nonribosomal lipopeptides generated by *Bacillus* strains was performed. In brief, Scopus, PubMed, Web of Science, and ResearchGate databases were queried on October 2nd, 2024 using the keywords “lipopeptides AND *Bacillus*, lipopeptides AND structure, surfactin AND *Bacillus*, surfactin AND structure, fengycin AND *Bacillus*, Fengycin AND structure, kurstakin AND *Bacillus*, Kurstakin AND structure, locillomycin AND *Bacillus*, locillomycin AND structure, iturin AND *Bacillus*, iturin AND structure, New lipopeptides AND *Bacillus*, New lipopeptides *Bacillus* AND structure, non-ribosomal peptide AND *Bacillus*.”

2.2 Database search for genome sequences

Sequences (including chromosome and plasmid) of *Bacillus* isolated from soil and fermented foods were retrieved from the NCBI nucleotide database.¹ A total of 123 complete genomes from *Bacillus* strains isolated from soil and fermented food samples, were selected according to the quality (Table 1). Thus, information on assembly, genome size, number of contigs/scaffolds and N50 were examined.

2.3 Detection of nonribosomal peptide biosynthetic gene clusters (BGCs)

The search for nonribosomal peptide sequences from the genome sequences of 123 *Bacillus* strains was performed following a workflow previously described (Leclère et al., 2016). First, the prediction of biosynthetic gene clusters (BGCs) was performed using antiSMASH version 7 (Blin et al., 2023). The regions encoding nonribosomal synthetases were further analyzed. The monomer composition of the predicted peptide was compared to all known NRPs annotated in the Norine database² (Flissi et al., 2020) in order to identify them or predict them as a new peptide or new variant. When it was incomplete with truncated or fragmented clusters, we further investigated antiSMASH results to reconstruct partial or complete BGCs by assembling cluster fragments scattered in the shotgun genome. The number of genes, the order of modules, and the domains in the NRPSs and the predictions of the A-domain specificity helped identify clusters of known NRP families, even when they were fragmented.

Since poor sequence assembly can affect the quality of predicted metabolites, only high-quality complete genomes were considered in this study.

3 Results

Out of 123 *Bacillus* chromosomes selected, 115 were found to carry BGC NRPSs. Among these NRPSs, those responsible for the synthesis of a siderophore (102), antibiotics (4) or lipopeptides (93) were found.

3.1 Prediction of NRP siderophore

The screened sequences harbor genetic potential to produce the siderophore bacillibactin. The gene cluster is composed of three modules with A-domain specificity for 2,3-dihydroxybenzoate (Dhb or diOH-Bz in Norine), glycine and threonine. The corresponding NRPS is known to follow an iterative mode of biosynthesis, which could lead to the production of active iron-chelating compounds named bacillibactins. This small cluster was predicted in 83% of the studied genomes.

3.2 Overview of lipopeptide BGCs produced by *Bacillus*

The literature search revealed diverse non-ribosomal lipopeptides synthesized by the genus *Bacillus*. The structures of the different molecules are shown in Table 2.

3.3 Prediction of the lipopeptide structures

Within BGCs detected by antiSMASH, special attention was paid to traits specific to LP synthesis. Thus, a condensation domain allowing the incorporation of fatty acid (FA) (so-called C-starter) was sought. C-starter is usually present in LP NRPS except for LP belonging to iturin and locillomycin families (Figure 1). As up to now, no bioinformatics tool enables the prediction of fatty acid (FA) structure, the work was focused on peptide moieties (Table 3).

In silico analysis of the genomic sequences revealed the potential production of four lipopeptide families (Table 2). These are the surfactin, iturin, fengycin and kurstakin families depending on the strains. Over the 123 chromosomes analyzed, 75 (61%) possessed gene clusters responsible for surfactin production, 29 (23%) for iturin, 46 (37%) for fengycin and 19 (15%) for kurstakin (Table 3). No locillomycin synthetic gene cluster was detected in the genomes explored in this study. All chromosomes carrying kurstakin BGCs came from species belonging to the *Bacillus cereus* group. When this cluster was present, the other clusters (surfactin, fengycin, iturin) were absent. A kurstakin variant was detected in *Bacillus mycoide* BGSC4BQ1 (Table 3, Figure 2). A variation was found in monomer 6, which is a threonine instead of Gln or Glu, which are most often found in this position. In addition, 25 chromosomes contained only clusters of surfactin genes. No chromosomes contained only fengycin or iturin gene clusters. A co-existence of gene clusters was also observed. As far surfactin and fengycin gene clusters are concerned, they were found in 21 chromosomes, while surfactin and iturin gene clusters were found in 4 chromosomes. A total of 25 chromosomes contained surfactin, fengycin and iturin gene clusters (Figure 3). In addition, six chromosomes contained four gene clusters (surfactin, fengycin, iturin, and novel gene clusters). All *B. subtilis*, *B. velezensis*, *B. licheniformis*, *B. paralicheniformis*, *B. amyloliquefaciens*, *B. pumilus*, *B. cellulasensis*, *B. altitudinis*, *B. spizizenii*, *B. safensis*, *B. vallismortis*, *B. halotolerans*, *B. inaquosorum*, *B. stratosphericus*, *B. aerophilus* and *B. atrophaeus* have gene clusters to produce a known surfactin or variant. In addition, all *B. licheniformis* and *B. paralicheniformis* have gene clusters which produce lichenysin A (Table 3). All 29 iturin-producing strains belong to the *B. subtilis* group (*B. subtilis*, *B. velezensis*, *B. amyloliquefaciens*, *B. spizizenii*, *B. atrophaeus*). Gene clusters for the biosynthesis of bacillomycin D and bacillomycin L, variants of iturin were detected in the chromosomes of *B. subtilis*, *B. amyloliquefaciens* and *B. velezensis* species. In species such as *B. atrophaeus* and *B. spizizenii*, chromosomes possess gene clusters producing iturin A and mycosubtilin, respectively. For the fengycin family, A or B production gene clusters were detected in 1/3 of the chromosomes of producing strains. In fact, the 6th monomer, which may be Val in the case of fengycin B or Ala in the case of fengycin A, was not predicted during the study. A fengycin B variant was detected in the genomes of 10 strains belonging to the *B. subtilis* species (Table 3). However, all the genomes of *B. paralicheniformis* strains contained clusters of fengycin B biosynthesis genes (Table 3). The D-allothreonine

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

² <https://bioinfo.cristal.univ-lille.fr/norine>

TABLE 1 Genome sequences including chromosomes and plasmids for the different *Bacillus* strains.

Number	Strains	Molecules	Accession number	Isolation source	Country	Molecule size	Contig N50	Completeness (%)	Contamination (%)
1	<i>B. aerophilus</i> KJ82	Chromosome	CP091093.1	Soil	China	3,754,440 bp	3.8 Mb	98.92	0.59
2	<i>B. albus</i> B-9	Chromosome	NZ_JAAAWQ000000000.1	Soil	China	5,371,297 bp	200.4 kb	99.33	0.1
3	<i>B. albus</i> YK87	Chromosome	NZ_CP142670.1	Soil	China	5,347,990 bp	5.3 Mb	98.82	0.10
4	<i>B. altitudinis</i> B4133	Chromosome	NZ_JXC�000000000.1	Fermented food	Netherlands	3,718,745 bp	95.4 kb	99.12	0.29
5	<i>B. altitudinis</i> GR-8	Chromosome	CP009108.1	Soil	China	3,674,849 bp	3.7 Mb	99.12	0.00
6	<i>B. altitudinis</i> GR-8	Plasmid	NZ_CP009109.1	Soil	China	6,935 bp	3.7 Mb	99.12	0.00
7	<i>B. altitudinis</i> NJ-V2	Chromosome	NZ_CP012482.1	Soil	China	3,787,818 bp	3.8 Mb	99.12	0.29
8	<i>B. altitudinis</i> NJ-M2	Chromosome	NZ_CP012329.1	Soil	China	3,793,419 bp	3.8 Mb	99.12	0.29
9	<i>B. altitudinis</i> NJ-V	Chromosome	NZ_CP012330.1	Soil	China	3,792,999 bp	3.8 Mb	99.12	0.29
10	<i>B. altitudinis</i> G6S2	Chromosome	NZ_CP126096.1	Soil	Unknown	3,797,173 bp	3.8 Mb	99.41	0.29
11	<i>B. amyloliquefaciens</i> S499	Chromosome	CP014700.1	Soil	Congo	3,927,922 bp	3.9 Mb	97.18	0.98
12	<i>B. amyloliquefaciens</i> S499	Plasmid	NZ_CP014701.1	Soil	Congo	8,008 bp	3.9 Mb	97.18	0.98
13	<i>B. amyloliquefaciens</i> 205	Chromosome	CP054415.1	Soil	China	4,006,790 bp	4 Mb	95.66	1.32
14	<i>B. amyloliquefaciens</i> WF02	Chromosome	CP053376.1	Soil	Taiwan	4,026,648 bp	4 Mb	98.24	1.46
15	<i>B. amyloliquefaciens</i> bm1	Chromosome	CP088005.1	Soil	China	3,929,792 bp	3.9 Mb	98.03	1.25
16	<i>B. amyloliquefaciens</i> LS1-002-014 s	Chromosome	NZ_CP089530.1	Fermented food	China	4,214,996 bp	4.2 Mb	98.26	0.94
17	<i>B. amyloliquefaciens</i> HM618	Chromosome	CP029466.1	Soil	China	4,021,851 bp	4 Mb	97.71	0.78
18	<i>B. amyloliquefaciens</i> B3	Chromosome	NZ_JAJNPN000000000.1	Fermented food	China	4,001,777 bp	171.8 kb	97.25	2.09
19	<i>B. anthracis</i> HYU01	Chromosome	CP008846.1	Soil	South Korea	5,213,498 bp	5.2 Mb	98.09	0.48
20	<i>B. anthracis</i> HYU01	Plasmid	CP008847.1	Soil	South Korea	181,894 bp	5.2 Mb	98.09	0.48
21	<i>B. anthracis</i> HYU01	Plasmid	CP008848.1	Soil	South Korea	94,732 bp	5.2 Mb	98.09	0.48
22	<i>B. anthracis</i> 9080-G	Chromosome	CM002398.1	Soil	Georgia	5,232,192 bp	195.9 kb	98.43	0.71
23	<i>B. anthracis</i> 9,080-G	Plasmid	CM002399.1	Soil	Georgia	181,656 bp	195.9 kb	98.43	0.71
24	<i>B. anthracis</i> 9,080-G	Plasmid	NZ_CM002400.1	Soil	Georgia	94,815 bp	195.9 kb	98.43	0.71
25	<i>B. anthracis</i> CMF9	Chromosome	CP085402.1	Soil	China	5,324,354 bp	5.3 Mb	96.09	3.04
26	<i>B. arachidis</i> YX15	Chromosome	CP127376.1	Soil	China	4,913,320 bp	4.9 Mb	98.53	0.00
27	<i>B. atrophaeus</i> PENSVC20	Chromosome	CP050705.1	Soil	Canada	4,148,820 bp	4.1 Mb	99.12	0.00
28	<i>B. badius</i> NBPM-293	Chromosome	CP082363.1	Soil	Unknown	3,868,812 bp	3.9 Mb	98.88	1.16

(Continued)

TABLE 1 (Continued)

Number	Strains	Molecules	Accession number	Isolation source	Country	Molecule size	Contig N50	Completeness (%)	Contamination (%)
29	<i>B. bombysepticus</i> F12	Chromosome	CP085406.1	Soil	China	5,244,749 bp	5.2 Mb	99.12	0.05
30	<i>B. bombysepticus</i> Cuernavaca_S2	Chromosome	CP126590.1	Soil	Mexico	5,288,525 bp	5.3 Mb	98.82	0.00
31	<i>B. cellulasensis</i> NJ-V2	Chromosome	CP012482.1	Soil	China	3,787,818 bp	3.8 Mb	99.12	0.29
32	<i>B. cellulasensis</i> NJ-M2	Chromosome	CP012329.1	Soil	China	3,793,419 bp	3.8 Mb	99.12	0.29
33	<i>B. cereus</i> CMCC P0011	Chromosome	CP011153.1	Soil	China	5,506,876 bp	5.5 Mb	98.97	0.06
34	<i>B. cereus</i> CMCC P0011	Plasmid	NZ_CP011154.1	Soil	China	591,112 bp	5.5 Mb	98.97	0.06
35	<i>B. cereus</i> M3	Chromosome	CP016316.1	Fermented food	South Korea	5,218,997 bp	5.2 Mb	98.43	0.22
36	<i>B. cereus</i> CMCC P0021	Chromosome	CP011151.1	Soil	China	5,521,782 bp	5.5 Mb	99.13	0.06
37	<i>B. cereus</i> CMCC P0021	Plasmid	NZ_CP011152.1	Soil	China	591,110 bp	5.5 Mb	99.13	0.06
38	<i>B. cereus</i> NJ-W	Chromosome	CP012483.1	Soil	China	5,370,032 bp	5.4 Mb	99.22	0.06
39	<i>B. cereus</i> NJ-W	Plasmid	NZ_CP012485.1	Soil	China	11,744 bp	5.4 Mb	99.22	0.06
40	<i>B. cereus</i> NJ-W	Plasmid	NZ_CP012484.1	Soil	China	9,032 bp	5.4 Mb	99.22	0.06
41	<i>B. cereus</i> NJ-W	Plasmid	NZ_CP012486.1	Soil	China	7,703 bp	5.4 Mb	99.22	0.06
42	<i>B. cereus</i> C-1	Chromosome	CP089601.1	Soil	China	5,268,934 bp	5.3 Mb	97.79	0.06
43	<i>B. gobiensis</i> FJAT-4402	Chromosome	CP012600.1	Soil	China	4,597,707 bp	2.8 Mb	98.77	1.69
44	<i>B. haikouensis</i> MNJ12	Chromosome	CP076017.1	Soil	China	4,454,189 bp	4.5 Mb	99.41	1.78
45	<i>B. halotolerans</i> ZB201702	Chromosome	CP029364.1	Soil	China	4,154,245 bp	4.2 Mb	99.41	0.59
46	<i>B. inaquosorum</i> DE111	Chromosome	CP013984.1	Soil	USA	4,143,890 bp	4.1 Mb	98.82	0.00
47	<i>B. inaquosorum</i> LBA001	Chromosome	NZ_CP127095.1	Soil	Mexico	4,200,707 bp	4.2 Mb	98.82	0.00
48	<i>B. infantis</i> 63-11	Chromosome	NZ_JAJBAP020000001.1	Fermented food	Thailand	4,841,671 bp	4.8 Mb	99.41	0.21
49	<i>B. licheniformis</i> SCDB 14	Chromosome	CP014842.1	Fermented food	South Korea	4,136,986 bp	4.1 Mb	97.23	1.98
50	<i>B. licheniformis</i> P8_B2	Chromosome	CP045814.1	Soil	Denmark	4,343,379 bp	4.3 Mb	97.89	2.24
51	<i>B. licheniformis</i> SCK B11	Chromosome	CP014795.1	Soil	South Korea	4,300,706 bp	4.3 Mb	98.33	1.24
52	<i>B. licheniformis</i> 14ADL4	Chromosome	CP026673.1	Fermented food	South Korea	4,332,232 bp	4.3 Mb	97.73	3.06
53	<i>B. methanolicus</i> MGA3	Chromosome	CP007739.1	Soil	USA	3,337,035 bp	3.3 Mb	97.88	2.66
54	<i>B. mojavensis</i> B-41341	Chromosome	NZ_JARLZB000000000.1	Soil	Israel	3,856,892 bp	148 kb	99.41	0.00
55	<i>B. mojavensis</i> B-41812	Chromosome	NZ_JARSHG000000000.1	Soil	Argentina	3,924,370 bp	428.5 kb	98.82	0.00
56	<i>B. mycoides</i> KBAB4	Chromosome	NC_010184.1	Soil	Unknown	5,262,775 bp	5.3 Mb	98.82	0.32
57	<i>B. mycoides</i> KBAB4	Plasmid	NC_010183.1	Soil	Unknown	52,830 bp	5.3 Mb	98.82	0.32

(Continued)

TABLE 1 (Continued)

Number	Strains	Molecules	Accession number	Isolation source	Country	Molecule size	Contig N50	Completeness (%)	Contamination (%)
58	<i>B. mycoides</i> KBAB4	Plasmid	NC_010182.1	Soil	Unknown	64,977 bp	5.3 Mb	98.82	0.32
59	<i>B. mycoides</i> KBAB4	Plasmid	NC_010181.1	Soil	Unknown	75,107 bp	5.3 Mb	98.82	0.32
60	<i>B. mycoides</i> BGSC 4BQ1	Chromosome	NZ_NFDI00000000.1	Soil	Spain	5,673,703 bp	639 kb	99.41	0.15
61	<i>B. mycoides</i> Gnyt1	Chromosome	CP020743.1	Soil	China	5,597,907 bp	5.6 Mb	99.41	0.00
62	<i>B. mycoides</i> ATCC 6462	Chromosome	CP009692.1	Soil	Unknown	5,255,868 bp	5.3 Mb	99.41	0.00
63	<i>B. pacificus</i> AT31	Chromosome	NZ_CP142003.1	Soil	China	4,903,194 bp	4.9 Mb	99.41	0.00
64	<i>B. pacificus</i> MP6	Chromosome	CP093424.1	Fermented food	China	5,038,982 bp	5 Mb	99.41	0.00
65	<i>B. pacificus</i> MP6	Plasmid	CP093425.1	Fermented food	China	291,500 bp	5 Mb	99.41	0.00
66	<i>B. paralicheniformis</i> UBBLI-30	Chromosome	NZ_SULF00000000.1	Fermented food	India	4,404,722 bp	332.9 kb	99.41	0.06
67	<i>B. paralicheniformis</i> 14DA11	Chromosome	NZ_CP023168.1	Fermented food	South Korea	4,535,069 bp	4.5 Mb	95.35	0.05
68	<i>B. paralicheniformis</i> CP47	Chromosome	NZ_CP133705.1	Fermented food	South Korea	4,537,254 bp	4.5 Mb	98.19	0.05
69	<i>B. paranthracis</i> SL1	Chromosome	CP093423.1	Fermented food	China	5,243,627 bp	5.2 Mb	99.41	0.29
70	<i>B. paranthracis</i> SL1	Plasmid	CP093422.1	Fermented food	China	325,335 bp	5.2 Mb	99.41	0.29
71	<i>B. paranthracis</i> Gxun-30	Chromosome	NZ_CP065149.1	Soil	China	5,149,464 bp	5.1 Mb	99.12	0.00
72	<i>B. paranthracis</i> KF11	Chromosome	CP085413.1	Soil	China	5,276,246 bp	5.3 Mb	99.41	0.29
73	<i>B. paranthracis</i> Bt C4	Chromosome	NZ_CP101135.1	Soil	China	5,245,328 bp	5.2 Mb	98.82	0.29
74	<i>B. pumilus</i> NJ-V	Chromosome	CP012330.1	Soil	China	3,792,999 bp	3.8 Mb	99.12	0.29
75	<i>B. pumilus</i> DSM 1794	Chromosome	CP187664.1	Soil	Unknown	3,786,030 bp	3.8 Mb	95.31	2.17
76	<i>B. pumilus</i> MS32	Chromosome	NZ_CP092829.1	Soil	Germany	3,824,664 bp	3.8 Mb	95.53	2.6
77	<i>B. pumilus</i> B4127	Chromosome	NZ_JXCL00000000.1	Fermented food	Netherlands	3,886,280 bp	142 kb	95.37	3.01
78	<i>B. pumilus</i> B4127	Plasmid	MH581228.1	Fermented food	Netherlands	6,580 bp	142 kb	95.37	3.01
79	<i>B. safensis</i> G6S3	Chromosome	CP128114.1	Soil	Germany	3,670,853 bp	3.7 Mb	99.41	0.29
80	<i>B. spizizenii</i> T30	Chromosome	NZ_CP011051.1	Soil	Russia	4,031,727 bp	4 Mb	95.76	0.00
81	<i>B. spizizenii</i> AS2	Chromosome	NZ_MUXL00000000.1	Soil	Oman	4,041,230 bp	573.3 kb	98.82	0.29
82	<i>B. spizizenii</i> HUK15	Chromosome	NZ_LSMU00000000.1	Soil	France	4,254,363 bp	322.1 kb	98.82	0.29
83	<i>B. stratosphericus</i> MRPD-01	Chromosome	PPGB00000000.1	Soil	India	3,687,760 bp	322 kb	99.22	0.29
84	<i>B. subtilis</i> SJ-10	Chromosome	NZ_CP025258.1	Fermented food	South Korea	4,041,647 bp	4 Mb	99.36	0.59
85	<i>B. subtilis</i> MEC_B298	Chromosome	NZ_CP100436.1	Soil	China	4,030,831 bp	4 Mb	97.41	0.67
86	<i>B. subtilis</i> s-16	Chromosome	NZ_CP063150.1	Soil	China	4,209,504 bp	4.2 Mb	97.67	1.35
87	<i>B. subtilis</i> BSP1	Chromosome	NZ_CP160396.1	Soil	Lebanon	4,043,759 bp	4 Mb	96.28	1.21

(Continued)

TABLE 1 (Continued)

Number	Strains	Molecules	Accession number	Isolation source	Country	Molecule size	Contig N50	Completeness (%)	Contamination (%)
88	<i>B. subtilis</i> KH2	Chromosome	CP018184.1	Soil	China	4,138,265 bp	4.1 Mb	96.73	1.28
89	<i>B. subtilis</i> FUA2231	Chromosome	NZ_CP154918.1	Fermented food	Zimbabwe	4,475,230 bp	4.5 Mb	97.31	1.62
90	<i>B. subtilis</i> UD1022	Chromosome	CP011534.1	Soil	USA	4,025,326 bp	4 Mb	97.53	0.91
91	<i>B. subtilis</i> SFA-H43	Chromosome	KZ836066.1	Soil	China	4,018,162 bp	2.1 Mb	96.73	1.48
92	<i>B. subtilis</i> CGMCC 2108	Chromosome	CP014471.1	Soil	China	4,122,154 bp	4.1 Mb	96.73	1.08
93	<i>B. subtilis</i> CGMCC 2108	Plasmid	NZ_CP014473.1	Soil	China	65,774 bp	4.1 Mb	96.73	1.08
94	<i>B. subtilis</i> CGMCC 2108	Plasmid	NZ_CP014472.1	Soil	China	5,820 bp	4.1 Mb	96.73	1.08
95	<i>B. subtilis</i> T30	Chromosome	CP011051.1	Soil	Russia	4,031,727 bp	4 Mb	95.76	0.00
96	<i>B. subtilis</i> NCIB 3610	Chromosome	NZ_CP094361.1	Soil	China	4,210,909 bp	4.2 Mb	98.51	1.33
97	<i>B. subtilis</i> ZD01	Chromosome	NZ_CP046448.1	Soil	China	4,015,360 bp	4 Mb	97.5	0.67
98	<i>B. subtilis</i> 73	Chromosome	CP045826.1	Soil	Germany	4,166,516 bp	4.2 Mb	97.33	1.97
99	<i>B. subtilis</i> G8	Chromosome	NZ_AP025224.1	Fermented food	Indonesia	4,017,503 bp	4 Mb	96.73	1.08
100	<i>B. subtilis</i> FUA2232	Chromosome	NZ_CP154920.1	Fermented food	Zimbabwe	4,475,228 bp	4.5 Mb	97.31	1.61
101	<i>B. subtilis</i> SRCM103517	Chromosome	NZ_CP035226.1	Fermented food	South Korea	4,192,706 bp	4.2 Mb	96.85	1.87
102	<i>B. subtilis</i> S1	Chromosome	NZ_JAGFPW000000000.1	Fermented food	India	4,488,261 bp	153.3 kb	95.77	1.85
103	<i>B. thuringiensis</i> HD12	Chromosome	CP014847.1	Soil	China	5,776,895 bp	5.8 Mb	98.39	0.96
104	<i>B. thuringiensis</i> HD12	Plasmid	NZ_CP014853.1	Soil	China	345,196 bp	5.8 Mb	98.39	0.96
105	<i>B. thuringiensis</i> HD12	Plasmid	NZ_CP014852.1	Soil	China	161,353 bp	5.8 Mb	98.39	0.96
106	<i>B. thuringiensis</i> HD12	Plasmid	NZ_CP014851.1	Soil	China	112,429 bp	5.8 Mb	98.39	0.96
107	<i>B. thuringiensis</i> HD12	Plasmid	NZ_CP014850.1	Soil	China	39,023 bp	5.8 Mb	98.39	0.96
108	<i>B. thuringiensis</i> HD12	Plasmid	NZ_CP014848.1	Soil	China	17,228 bp	5.8 Mb	98.39	0.96
109	<i>B. thuringiensis</i> Bt185	Chromosome	CP014282.1	Soil	China	5,243,635 bp	5.2 Mb	98.26	1.05
110	<i>B. thuringiensis</i> Bt185	Plasmid	NZ_CP014284.1	Soil	China	293,705 bp	5.2 Mb	98.26	1.05
111	<i>B. thuringiensis</i> Bt185	Plasmid	NZ_CP014285.1	Soil	China	55,372 bp	5.2 Mb	98.26	1.05
112	<i>B. thuringiensis</i> Bt185	Plasmid	NZ_CP014286.1	Soil	China	54,205 bp	5.2 Mb	98.26	1.05
113	<i>B. thuringiensis</i> Bt185	Plasmid	NZ_CP014287.1	Soil	China	41,937 bp	5.2 Mb	98.26	1.05
114	<i>B. thuringiensis</i> Bt185	Plasmid	NZ_CP014288.1	Soil	China	12,487 bp	5.2 Mb	98.26	1.05
115	<i>B. thuringiensis</i> Bt185	Plasmid	NZ_CP014290.1	Soil	China	7,486 bp	5.2 Mb	98.26	1.05
116	<i>B. thuringiensis</i> c25	Chromosome	NZ_CP022345.1	Soil	South Korea	5,334,660 bp	5.3 Mb	98.91	1.07
117	<i>B. thuringiensis</i> 97–27	Chromosome	NC_005957.1	Soil	Unknown	5,237,682 bp	5.2 Mb	99.44	0.32
118	<i>B. thuringiensis</i> YWC2-8	Chromosome	CP013055.1	Soil	China	5,674,369 bp	5.7 Mb	98.38	1.01

(Continued)

TABLE 1 (Continued)

Number	Strains	Molecules	Accession number	Isolation source	Country	Molecule size	Contig N50	Completeness (%)	Contamination (%)
119	<i>B. thuringiensis</i> YWC2-8	Plasmid	NZ_CP013056.1	Soil	China	250,706 bp	5.7 Mb	98.38	1.01
120	<i>B. thuringiensis</i> YWC2-8	Plasmid	NZ_CP013057.1	Soil	China	84,491 bp	5.7 Mb	98.38	1.01
121	<i>B. thuringiensis</i> YWC2-8	Plasmid	NZ_CP013058.1	Soil	China	82,531 bp	5.7 Mb	98.38	1.01
122	<i>B. thuringiensis</i> YWC2-8	Plasmid	NZ_CP013059.1	Soil	China	80,699 bp	5.7 Mb	98.38	1.01
123	<i>B. thuringiensis</i> YWC2-8	Plasmid	NZ_CP013060.1	Soil	China	46,634 bp	5.7 Mb	98.38	1.01
124	<i>B. thuringiensis</i> YWC2-8	Plasmid	NZ_CP013061.1	Soil	China	8,512 bp	5.7 Mb	98.38	1.01
125	<i>B. thuringiensis</i> BGSC 4AA1	Chromosome	CP010577.1	Soil	China	5,652,292 bp	5.7 Mb	97.24	3.40
126	<i>B. thuringiensis</i> BGSC 4AA1	Plasmid	NZ_CP010578.1	Soil	China	232,994 bp	5.7 Mb	97.24	3.40
127	<i>B. thuringiensis</i> BGSC 4AA1	Plasmid	NZ_CP010579.1	Soil	China	92,619 bp	5.7 Mb	97.24	3.40
128	<i>B. thuringiensis</i> BGSC 4AA1	Plasmid	NZ_CP010580.1	Soil	China	76,979 bp	5.7 Mb	97.24	3.40
129	<i>B. thuringiensis</i> BGSC 4AA1	Plasmid	NZ_CP010581.1	Soil	China	68,444 bp	5.7 Mb	97.24	3.40
130	<i>B. thuringiensis</i> BGSC 4AA1	Plasmid	NZ_CP010582.1	Soil	China	51,723 bp	5.7 Mb	97.24	3.40
131	<i>B. thuringiensis</i> BGSC 4AA1	Plasmid	NZ_CP010583.1	Soil	China	4,845 bp	5.7 Mb	97.24	3.40
132	<i>B. thuringiensis</i> CTC	Chromosome	CP013274.1	Soil	China	5,327,397 bp	5.3 Mb	97.63	1.10
133	<i>B. thuringiensis</i> CTC	Plasmid	NZ_CP013273.1	Soil	China	25,529 bp	5.3 Mb	97.63	1.10
134	<i>B. thuringiensis</i> HS18-1	Chromosome	CP012099.1	Soil	China	5,292,526 bp	5.3 Mb	98.68	0.55
135	<i>B. thuringiensis</i> HS18-1	Plasmid	NZ_CP012101.1	Soil	China	337,579 bp	5.3 Mb	98.68	0.55
136	<i>B. thuringiensis</i> HS18-1	Plasmid	NZ_CP012102.1	Soil	China	92,085 bp	5.3 Mb	98.68	0.55
137	<i>B. thuringiensis</i> HS18-1	Plasmid	NZ_CP012103.1	Soil	China	94,695 bp	5.3 Mb	98.68	0.55
138	<i>B. thuringiensis</i> HS18-1	Plasmid	NZ_CP012104.1	Soil	China	42,726 bp	5.3 Mb	98.68	0.55
139	<i>B. thuringiensis</i> HS18-1	Plasmid	NZ_CP012105.1	Soil	China	14,336 bp	5.3 Mb	98.68	0.55
140	<i>B. thuringiensis</i> HS18-1	Plasmid	NZ_CP012106.1	Soil	China	4,669 bp	5.3 Mb	98.68	0.55
141	<i>B. thuringiensis</i> HS18-1	Plasmid	NZ_CP012107.1	Soil	China	8,287 bp	5.3 Mb	98.68	0.55
142	<i>B. thuringiensis</i> HS18-1	Plasmid	NZ_CP012108.1	Soil	China	7,386 bp	5.3 Mb	98.68	0.55
143	<i>B. thuringiensis</i> JW-1	Chromosome	NZ_CP045030.1	Soil	China	5,500,376 bp	5.5 Mb	99.30	0.74
144	<i>B. thuringiensis</i> HD-1	Chromosome	CP004870.1	Soil	USA	5,631,672 bp	5.6 Mb	97.67	1.75
145	<i>B. thuringiensis</i> YBT-1518	Chromosome	CP005935.1	Soil	China	6,002,284 bp	6 Mb	98.60	3.18
146	<i>B. toyonensis</i> UTDF19-29B	Chromosome	CP081872.1	Soil	USA	5,240,743 bp	5.2 Mb	99.41	0.10
147	<i>B. toyonensis</i> Monterrey_S3	Chromosome	CP126524.1	Soil	Mexico	5,309,620 bp	5.3 Mb	99.41	0.10
148	<i>B. toyonensis</i> Cuernavaca_S4	Chromosome	CP126520.1	Soil	Mexico	5,270,039 bp	5.3 Mb	98.82	0.10

(Continued)

TABLE 1 (Continued)

Number	Strains	Molecules	Accession number	Isolation source	Country	Molecule size	Contig N50	Completeness (%)	Contamination (%)
149	<i>B. tropicus</i> CK18	Chromosome	CP085399.1	Soil	China	5,237,233 bp	5.2 Mb	99.41	0.00
150	<i>B. tropicus</i> T36S-23	Chromosome	NZ_CP119875.1	Soil	South Korea	5,262,398 bp	5.3 Mb	98.82	0.62
151	<i>B. vallismortis</i> NBIF-001	Chromosome	CP020893.1	Soil	China	3,929,787 bp	3.9 Mb	98.82	0.00
152	<i>B. vallismortis</i> DSM 11031	Chromosome	CP026362.1	Soil	USA	4,286,362 bp	4.3 Mb	99.22	0.00
153	<i>B. velezensis</i> ATR2	Chromosome	CP018133.1	Soil	China	4,006,746 bp	4 Mb	99.41	0.00
154	<i>B. velezensis</i> SYP-B637	Chromosome	CP043546.1	Soil	China	3,915,550 bp	3.9 Mb	99.41	0.00
155	<i>B. velezensis</i> CGMCC 11640	Chromosome	CP026610.1	Soil	China	4,322,979 bp	4.3 Mb	99.41	0.00
156	<i>B. velezensis</i> Lzh-a42	Chromosome	CP025308.1	Soil	China	4,246,605 bp	4.2 Mb	99.41	0.00
157	<i>B. velezensis</i> YJ0-1	Chromosome	NZ_CP128184.1	Soil	China	4,043,341 bp	4 Mb	99.41	1.24
158	<i>B. velezensis</i> L-S60	Chromosome	CP011278.1	Soil	China	3,903,017 bp	3.9 Mb	99.41	0.00
159	<i>B. velezensis</i> CBMB205	Chromosome	CP014838.1	Soil	South Korea	3,929,745 bp	3.9 Mb	99.41	0.00
160	<i>B. velezensis</i> SB1216	Chromosome	CP015417.1	Soil	USA	3,814,720 bp	3.8 Mb	97.56	0.00
161	<i>B. velezensis</i> DSYZ	Chromosome	CP030150.1	Soil	China	4,258,978 bp	4.3 Mb	99.41	0.00
162	<i>B. velezensis</i> DR-08	Chromosome	CP028437.1	Soil	South Korea	3,929,794 bp	3.9 Mb	99.36	0.00
163	<i>B. velezensis</i> CMF18	Chromosome	CP085388.1	Soil	China	3,963,155 bp	4 Mb	99.41	0.00
164	<i>B. velezensis</i> CK17	Chromosome	CP085706.1	Soil	China	3,921,806 bp	3.9 Mb	99.34	0.00
165	<i>B. velezensis</i> DMB06	Chromosome	NZ_CP083763.1	Fermented food	South Korea	4,157,945 bp	4.2 Mb	99.4	0.00
166	<i>B. velezensis</i> UD6-2	Chromosome	NZ_JAGFMB000000000.1	Fermented food	Thailand	3,951,373 bp	1 Mb	99.41	0.00
167	<i>B. velezensis</i> DMB05	Chromosome	NZ_CP083715.1	Fermented food	South Korea	3,262,563 bp	3.3 Mb	99.12	0.00
168	<i>B. velezensis</i> DMB05	Plasmid	NZ_CP083716.1	Fermented food	South Korea	806,695 bp	3.3 Mb	99.12	0.00
169	<i>B. velezensis</i> DMB05	Plasmid	NZ_CP083717.1	Fermented food	South Korea	72,020 bp	3.3 Mb	99.12	0.00
170	<i>B. wiedmannii</i> LN15	Chromosome	CP126099.1	Soil	China	5,391,143 bp	5.4 Mb	98.82	0.10
171	<i>B. wiedmannii</i> JAS08/1	Chromosome	NZ_CP036073.1	Soil	Poland	5,513,922 bp	5.5 Mb	99.41	0.10
172	<i>B. wiedmannii</i> JAS07/5	Chromosome	CP036070.1	Soil	Poland	5,239,489 bp	5.2 Mb	99.41	1.76
173	<i>B. wiedmannii</i> JAS07/5	Plasmid	NZ_CP036072.1	Soil	Poland	536,465 bp	5.2 Mb	99.41	1.76
174	<i>B. wiedmannii</i> JAS07/5	Plasmid	NZ_CP036071.1	Soil	Poland	79,335 bp	5.2 Mb	99.41	1.76
175	<i>B. wiedmannii</i> PL1	Chromosome	AP022643.1	Soil	Japan	5,309,441 bp	5.3 Mb	97.22	0.00
176	<i>B. safensis</i> ZK-1	Chromosome	NZ_CP095759.1	Soil	China	3,735,476 bp	3.7 Mb	99.06	0.29
177	<i>B. safensis</i> SRCM125915	Chromosome	NZ_CP116774.1	Soil	South Korea	3,769,976 bp	3.8 Mb	98.92	0.59
178	<i>B. safensis</i> PLA 1006	Chromosome	NZ_CP114177.1	Soil	China	3,891,304 bp	3.9 Mb	98.92	0.59

(Continued)

TABLE 2 Structure of different lipopeptides classified by families.

Name of lipopeptides	Monomer composition of the peptide moiety	Main fatty acid chains	References
Fengycin family	Decapeptide with a lactone ring between carboxy-terminal group	β -OH fatty acids	
Fengycin A1	L-Glu,D-Orn,D-Tyr,D-aThr,L-Glu,D-Ala,L-Pro,L-Gln,L-Tyr,L-Ile	C ₁₄ -C ₁₉	Ait Kaki et al. (2020)
Fengycin A2	L-Glu,D-Orn,D-Tyr,D-aThr,L-Glu,D-Ala,L-Pro,L-Gln,L-Tyr,L-Val	C ₁₄ , C ₁₅ , C ₁₇	Ongena and Jacques (2007)
Fengycin B1	L-Glu-D-Orn-D-Tyr-D-aThr,L-Glu,D-Val,L-Pro,L-Gln,L-Tyr,L-Ile	C ₁₄ -C ₁₇	Fanaei et al. (2021)
Fengycin B2	L-Glu,D-Orn,D-Tyr,D-aThr,L-Glu,D-Val,L-Pro,L-Gln,L-Tyr,L-Val	C ₁₄ -C ₂₁	Ongena and Jacques (2007)
Fengycin C1	L-Glu,D-Orn,D-Tyr,D-aThr,L-Glu,D-Leu/D-Ile,L-Pro,L-Gln,L-Tyr,L-Val	C ₁₅ -C ₁₉	Vater et al. (2002)
Fengycin C2	L-Glu,D-Orn,D-Tyr,D-aThr,L-Glu,D-Val,L-Pro,L-Gln,L-aThr,L-Ile	C ₁₄ -C ₁₈	Ongena and Jacques (2007)
Fengycin S	L-Glu,D-Orn,D-Tyr,D-Ser,L-Glu,D-Val,L-Pro,L-Gln,L-Tyr,L-Ile	C ₁₇	Sang-cheol et al. (2010)
Plipastatin A1, A2	L-Glu,D-Orn,D-Tyr,D-aThr,L-Glu,D-Ala,L-Pro,L-Gln,D-Tyr,L-Ile	nC ₁₆ , aiC ₁₇	Nishikiori et al. (1986)
Plipastatin B1, B2	L-Glu,D-Orn,D-Tyr,D-aThr,L-Glu,D-Val,L-Pro,L-Gln,L-Tyr,L-Ile	nC ₁₆ , aiC ₁₇	Nishikiori et al. (1986)
Iturin family	Heptapeptide cyclised by a lactam ring with the β -NH ₂ group of the acid chain	β -NH ₂ fatty acids	
Iturin A, A _L	L-Asn,D-Tyr,D-Asn,L-Gln,L-Pro,D-Asn,L-Ser	nC ₁₄ , iC ₁₅ , aiC ₁₅ , nC ₁₆ , iC ₁₆	Pathak and Keharia (2014)
Iturin C	L-Asp,D-Tyr,D-Asn,L-Gln,L-Pro,D-Asn,L-Ser	nC ₁₄ , iC ₁₅ , aiC ₁₅	Pathak and Keharia (2014)
Bacillomycin D	L-Asn,D-Tyr,D-Asn,L-Pro,L-Gln,D-Ser,L-Thr	nC ₁₄ , iC ₁₅ , aiC ₁₅	Jin et al. (2017)
Bacillomycin DC	L-Asn,D-Tyr,D-Asn,L-Pro,L-Glu,D-Ser,L-Thr	C ₁₃	Jin et al. (2017)
Bacillomycin F	L-Asn,D-Tyr,D-Asn,L-Gln,L-Pro,D-Asn,L-Thr	C ₁₆ , iC ₁₇ , aiC ₁₇	Peypoux et al. (1985)
Bacillomycin L	L-Asp,D-Tyr,D-Asn,L-Ser,L-Glu,D-Ser,L-Thr	nC ₁₄ , iC ₁₅ , aiC ₁₅	Volpon et al. (2007)
Bacillomycin Lc ^a	L-Asn,D-Tyr,D-Asn,L-Ser,L-Glu,D-Ser,L-Thr	nC ₁₄ , iC ₁₅ , aiC ₁₅ , iC ₁₆	Volpon et al. (2007)
Mycosubtilin	L-Asn,D-Tyr,D-Asn,L-Gln,L-Pro,D-Asn/D-Ser,L-Ser/L-Asn	nC ₁₆ , iC ₁₆ , aiC ₁₇	Peypoux et al. (1986)
Mojavensin A	L-Asp,D-Tyr,D-Asn,L-Gln,L-Pro,D-Asn,L-Asn	aiC ₁₂	Ma et al. (2012)
Subtulene A	L-Asn,D-Tyr,D-Asn,L-Gln,L-Pro,D-Asn,L-Ser	(CH ₂) ₄ H ₂ C = CH(CH ₂) ₃ CH(CH ₃) ₂	Thasana et al. (2010)
Mixinin	L-Asn,D-Tyr,D-Asn,L-Gln,L-Ser,D-Asn,L-Pro	(A) C ₁₁ , (B) C ₈ , (C) aiC ₁₀	Zhang et al. (2004)
Kurstakin family	Heptapeptide with a lactone ring between carboxy-terminal group of Gln7 and OH group of Ser4		
Kurstakin ^b	D-Thr,L-Gly,D-Ala,L-Ser,L-His,D-Gln,L-Gln	iC ₁₁ , nC ₁₂ , iC ₁₂ , iC ₁₃	Hathout et al. (2000)
Surfactin family	Heptapeptide closed by a lactone ring with the b-OH group of the fatty acid chain	β -OH fatty acids	
Bamylocin A ^c	Glu,Leu,Met,Leu,Pro,Leu,Leu,Leu	C ₁₃	Lee et al. (2007)
Surfactin	L-Glu,L-Leu,D-Leu,L-Val,L-Asp,D-Leu,L-Leu	iC ₁₄ , nC ₁₄ , iC ₁₅ , aC ₁₅	Bonmatin et al. (2003)
[Orn2] surfactin	L-Glu,L-Orn,D-Leu,L-Val,L-Asp,D-Leu,L-Ile	C ₁₃ -C ₁₅	Bonmatin et al. (2003)
[Orn7] surfactin	L-Glu,L-Leu,D-Leu,L-Val,L-Asp,D-Leu,L-Orn	–	Bonmatin et al. (2003)
[Phe7] surfactin	L-Glu,L-Leu,D-Leu,L-Val,L-Asp,D-Leu,L-Phe	–	Bonmatin et al. (2003)
[Cys7] surfactin	L-Glu,L-Leu,D-Leu,L-Val,L-Asp,D-Leu,L-Cys	–	Bonmatin et al. (2003)
[Ala4] surfactin	L-Glu,L-Leu,D-Leu,L-Ala,L-Asp,D-Leu,L-Leu	iC ₁₄ , nC ₁₄ , iC ₁₅ , aiC ₁₅	Bonmatin et al. (2003)
[Ile4] surfactin	L-Glu,L-Leu,D-Leu,L-Ile,L-Asp,D-Leu,L-Leu	aiC ₁₅	Bonmatin et al. (2003)

(Continued)

TABLE 2 (Continued)

Name of lipopeptides	Monomer composition of the peptide moiety	Main fatty acid chains	References
[Ile2, 7] surfactin	L-Glu,L-Ile,D-Leu,L-Val,L-Asp,D-Leu,L-Ile	C ₁₃ -C ₁₅	Bonmatin et al. (2003)
[Ile4, 7] surfactin	L-Glu,L-Leu,D-Leu,L-Ile,L-Asp,D-Leu,L-Ile	C ₁₃ -C ₁₅	Bonmatin et al. (2003)
[Leu4] surfactin	L-Glu,L-Leu,D-Leu,L-Leu,L-Asp,D-Leu,L-Leu	iC ₁₅	Bonmatin et al. (2003)
[Val7] surfactin	L-Glu,L-Leu,D-Leu,L-Val,L-Asp,D-Leu,L-Val	iC ₁₄ , nC ₁₄ , iC ₁₅ , aiC ₁₅	Bonmatin et al. (2003)
[Ile7] surfactin	L-Glu,L-Leu,D-Leu,L-Val,L-Asp,D-Leu,L-Ile	iC ₁₄ , nC ₁₄ , iC ₁₅ , aiC ₁₅	Bonmatin et al. (2003)
[Ile2, 4] surfactin	L-Glu,L-Ile,D-Leu,L-Ile,L-Asp,D-Leu,L-Leu	aiC ₁₅	Bonmatin et al. (2003)
[Val2, 7] surfactin	L-Glu,L-Val,D-Leu,L-Val,L-Asp,D-Leu,L-Val	C ₁₃ -C ₁₅	Bonmatin et al. (2003)
[Val2, Ile7] surfactin	L-Glu,L-Val,D-Leu,L-Val,L-Asp,D-Leu,L-Ile	C ₁₃ -C ₁₅	Bonmatin et al. (2003)
[Ile2, Val7] surfactin	L-Glu,L-Ile,D-Leu,L-Val,L-Asp,D-Leu,L-Val	C ₁₃ -C ₁₅	Bonmatin et al. (2003)
[Ile2, 4, 7] surfactin	L-Glu,L-Ile,D-Leu,L-Ile,L-Asp,D-Leu,L-Ile	aiC ₁₅	Bonmatin et al. (2003)
Lichenysin ^c	L-Gln,L-Leu,D-Leu,L-Val,L-Asp,D-Leu,L-Ile	iC ₁₃ , aC ₁₃ , nC ₁₄ , iC ₁₅ , aC ₁₅	Bonmatin et al. (2003)
[Ile4] lichenysin	L-Gln,L-Leu,D-Leu,L-Ile,L-Asp,D-Leu,L-Ile	aiC ₁₅	Bonmatin et al. (2003)
[Val7] lichenysin	L-Gln,L-Leu,D-Leu,L-Val,L-Asp,D-Leu,L-Val	iC ₁₃ , aiC ₁₃ , nC ₁₄ , iC ₁₅ , aiC ₁₅	Bonmatin et al. (2003)
[Ile2, 4] lichenysin	L-Gln,L-Ile,D-Leu,L-Ile,L-Asp,D-Leu,L-Ile	aiC ₁₅	Bonmatin et al. (2003)
Esperin ^d	L-Glu,L-Leu,D-Leu,L-Val,L-Asp,D-Leu,L-XE7-COOH	C ₁₃ , C ₁₄ , C ₁₅	Thomas and Ito (1969)
Pumilacidin	L-Glu,L-Leu,D-Leu,L-Leu,L-Asp,D-Leu,L-XP7	aC ₁₅ , iC ₁₅ , nC ₁₆ , iC ₁₆ , aC ₁₇ , iC ₁₇	Naruse et al. (1990)
Locillomycin family	Partially cyclized nonapeptides		
Locillomycin A, B, C	L-Thr,D-Gln,L-Asp,L-Gly,L-Asn,L-Asp,L-Gly,L-Tyr,L-Val	C ₁₃ , C ₁₄ , C ₁₅	Luo et al. (2015a, 2015b)
Cerexin family	Linear decapeptides		
Cerexin A	D-Asp,D-Val,D-val,L-Asn,D-Asn,L-γHyl,D-Thr,L-Ser,D-Trp,D-Ile	iC ₁₁ h ³	Shoji et al. (1976b)
Cerexin B	D-Asn,D-val,D-Phe,L-Asn,D-Asn,L-γHyl,D-Thr,L-Gly,D-Trp,D-Ile	iC ₁₀ h ³ , nC ₁₀ h ³ , iC ₁₁ h ³ , aiC ₁₁ h ³	Shoji and Kato (1976)
Cerexin C	D-Asp,D-Val,D-val,L-Asn,D-Asn,L-Lys,D-Thr,L-Ser,D-Trp,D-Ile	iC ₁₁ h ³	Shoji et al. (1976a)
Cerexin D	D-Asp,D-Val,D-Phe,L-Asn,D-Asn,L-Lys,D-Thr,L-Gly,D-Trp,D-Ile	iC ₁₀ h ³ , nC ₁₀ h ³ , iC ₁₁ h ³ , aiC ₁₁ h ³	Shoji et al. (1976a)
Marihysin A family	Cyclic heptapeptide		
Marihysin A	Pro,Gln,Asn,Ser,Asn,Tyr,Asn	C ₁₄	Liu et al. (2010)
Octapeptin family			
Octapeptin A1, A2, A3, A4	D-Dab,L-Dab,L-Dab,L-Dab,L-Dab,D-Leu,L-Leu,L-Leu	aiC ₁₁ h ³ , iC ₁₀ h ³ , nC ₁₀ h ³ , iC ₁₁ h ³	Meyers et al. (1976)
Octapeptin B1, B2, B3, B4	D-Dab,L-Dab,L-Dab,L-Dab,L-Dab,D-Leu,L-Leu,L-Phe	aiC ₁₁ h ³ , iC ₁₀ h ³ , nC ₁₀ h ³ , iC ₉ h ³	Meyers et al. (1976)
Octapeptin C1, C2, C3, C4	D-Dab,L-Dab,L-Dab,L-Dab,L-Dab,L-Leu,L-Leu,D-Phe	aiC ₉ h ³ , aiC ₉ h ³ , iC ₁₀ h ³ , nC ₁₀ h ³	Meyers et al. (1976)
Octapeptin D1, D2, D3, D4	D-Ser-L-Dab-L-Dab-D-Leu-L-Leu-L-Dab-L-Dab-L-Leu	aiC ₁₁ h ³ , iC ₁₀ h ³ , nC ₁₀ h ³ , iC ₁₁ h ³	Kato and Shoji (1980)

^aBacillomycin Lca; or bacillopeptin.

^bD forms of amino acid residue are deduced from NRPS modular structure.

^cL and D forms are not specified.

^dThe b-carboxyl of Asp5 is engaged in the lactone.

^eOr halobacillin With XE7 = Leu or Val; XP7 = Val or Ile.

ai, anteiso; i, iso; n, normal; h³, 3-hydroxy; L-Hyl, L-threo—hydroxylysine; Dab, 2,4-Diaminobutyric acid.

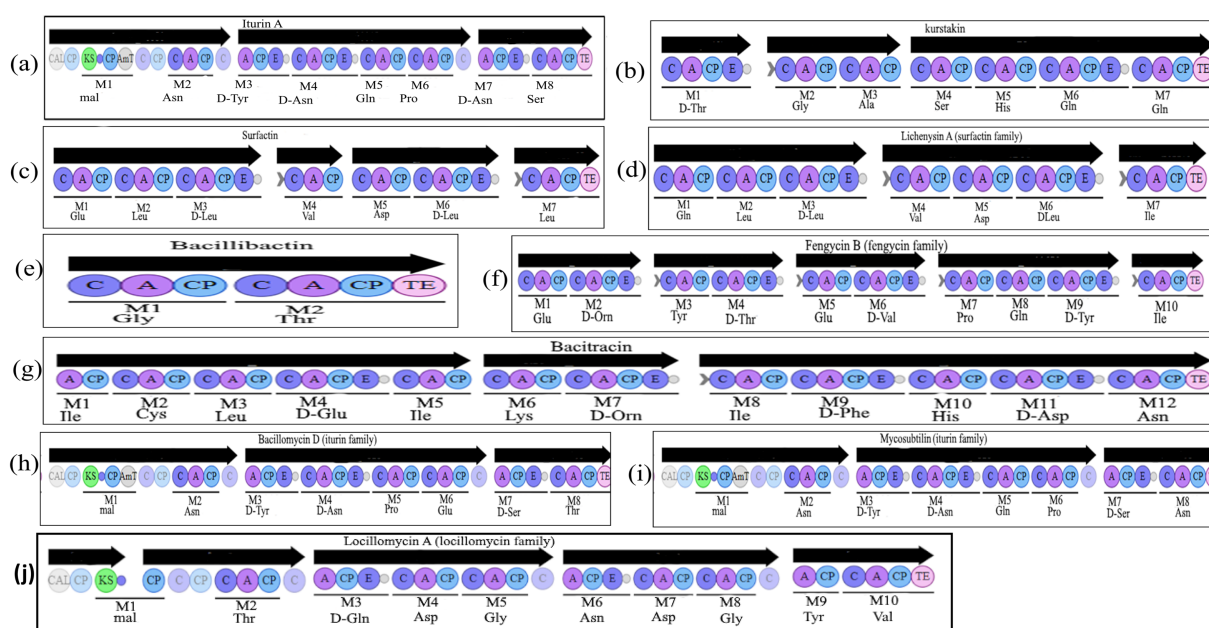


FIGURE 1

Organization of the known non-ribosomal peptide synthetases (NRPS) encoding lipopeptides, siderophore bacillibactin and antibiotic bacitracin in *Bacillus*. Iterative domains: A, adenylation domain; C, condensation domain; CP, Carrier Protein; E, epimerization domain; TE, thioesterase domain; mal, malonyl-CoA; CAL, Co-enzyme A ligase; KS, ketosynthetase domain; AmT, aminotransferase. Organization of the known NRPSs: (a) iturin A in *B. amyloliquefaciens* S499, (b) known kurstakin, (c) surfactin in *B. velezensis* UD6-2, (d) lichenysin A in *B. paralicheniformis* CP47, (e) bacillibactin in *B. aerophilus* KJ82, (f) fengycin B in *B. paralicheniformis* CP47, (g) bacitracin in *B. paralicheniformis* CP47, (h) bacillomycin D in *B. velezensis* DMB06, (i) mycosubtilin in *B. subtilis* T30, (j) locillomycin in *B. subtilis* 916.

observed in fengycin B is predicted to be a D-threonine by antiSMASH because these two threonine isomers can be selected by the same A-domain. A small difference is noticed in the beta carbon. Beta carbon of D-allo-threonine (C α D, C β D) belongs to the D series whereas the beta carbon of D-threonine (C α D, C β L) belongs to the L series.

3.4 Prevalence of BGC NRPSs in chromosomes of the different species

The 123 strains represent 33 species with distributions as follows: 19 species were found in the soil samples only, 12 species were found in both soil and fermented food isolation media and 2 species were found only in the fermented food. *Bacillus* species that were only found in soil samples included *B. albus*, *B. arachidis*, *B. atrophaeus*, *B. badius*, *B. bombysepticus*, *B. cellulasensis*, *B. gobiensis*, *B. haikouensis*, *B. halotolerans*, *B. methanolicus*, *B. mycoides*, *B. inaquosorum*, *B. stratosphericus*, *B. thuringiensis*, *B. toyoninsis*, *B. tropicus*, *B. anthracis*, *B. vallismortis* and *B. aerophilus* (Figure 4A). Species found in soil and fermented foods were: *B. licheniformis*, *B. paralicheniformis*, *B. paranthracis*, *B. pacificus*, *B. pumilus*, *B. safensis*, *B. spizizenii*, *B. subtilis*, *B. velezensis*, *B. wiedmannii*, *B. amyloliquefaciens* and *B. cereus*. Two species, *B. altitudinis* and *B. infantis*, were only found in fermented foods (Figure 4B). Among genomes of soil *Bacillus* species, only *B. atrophaeus* had a gene cluster for iturin biosynthesis. Furthermore, only the genome of *B. thuringiensis* species carried a gene clusters for the biosynthesis of the antibiotic bacitracin and a new gene cluster. Analysis of genomes of the 12 species found in isolation media (soil and fermented food) revealed the presence of new gene clusters in species such as *B. velezensis*, *B. amyloliquefaciens*, *B. subtilis*

and *B. cereus*. Prevalence of new gene clusters was higher in *B. velezensis* and *B. amyloliquefaciens*, respectively (Figure 4B) and the genomes of these two strains carried 3 new BGC NRPSs. The genome of *B. infantis*, a species that was only present in fermented foods had no NRPSs gene cluster (Figure 4B). Moreover, bacitracin biosynthesis genes were carried only by the genomes of *B. paralicheniformis* and *B. thuringiensis*. Genomes of *B. pacificus*, *B. cereus*, *B. wiedmannii* and *B. paranthracis* lacked surfactin biosynthesis genes (Figure 4).

3.5 Bacitracin production

Among the 123 genomes screened, only 4 (*B. paralicheniformis* 14DA11, *B. paralicheniformis* CP47, *B. paralicheniformis* UBBLI-30, and *B. thuringiensis* Bt185) bear BGC corresponding to the synthetic pathway for antibiotic bacitracin. This antibiotic is a semi-cyclic peptide constituted of 12 amino acids. In *B. paralicheniformis* CP47 chromosome sequence, a BGC encoding an NRPS containing 12 modules was found. These 12 modules correspond exactly to the 12 modules of bacitracin synthetase. In the genomes of *B. thuringiensis* Bt185, *B. paralicheniformis* 14DA11, and *B. paralicheniformis* UBBLI-30, modular structuring and amino acid composition were similar to bacitracin A1 (Figure 1), similarities ranged from 85 to 100%.

3.6 BGC potentially producing new metabolites

A total of 7 new molecules carried by 7 new gene clusters were detected in this study, based on the exploration of 123 complete

TABLE 3 Structures of predicted and identified lipopeptides among different genomic sequences of *Bacillus* strains.

Predicted LP	Structure analysis	Comments	Strains
[Ile7] surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Ile)	Complete	<i>B. aerophilus</i> KJ82
[Ile7] surfactin	(Glu,Leu,D-Leu)(Val,Asp,Leu)(Ile)	Complete	<i>B. altitudinis</i> B4133
[Ile7] surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Ile)	Complete	<i>B. altitudinis</i> GR-8
[Ile7] surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Ile)	Complete	<i>B. altitudinis</i> NJ-V2
[Ile7] surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Ile)	Complete	<i>B. altitudinis</i> NJ-M2
[Ile7] surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Ile)	Complete	<i>B. altitudinis</i> NJ-V
[Ile7] surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Ile)	Complete	<i>B. altitudinis</i> G6S2
Iturin A	(Asn)(D-Tyr,D-Asn,Gln,Pro)(D-Asn,Ser)	Complete	<i>B. amyloliquefaciens</i> S499
Fengycin A/B	(Glu,D-Orn)(Tyr, D-Thr)(Glu,D-X)(Pro,Gln,D-Tyr)(Ile)	Incomplete (X could be Val or Ala)	<i>B. amyloliquefaciens</i> S499
Surfactin	(Glu,Leu,D-Leu)(...)(Leu)	Partial and fragmented	<i>B. amyloliquefaciens</i> S499
[Ile7] surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Ile)	Complete	<i>B. amyloliquefaciens</i> 205
Fengycin	(.....)(...)(...)(...Glu,D-Tyr)(Ile)	Partial and fragmented (Glu instead of Gln)	<i>B. amyloliquefaciens</i> 205
Iturin A	(Asn)(D-Tyr,D-Asn,Gln,Pro)(D-Asn,Ser)	Complete	<i>B. amyloliquefaciens</i> 205
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. amyloliquefaciens</i> WF02
Fengycin A/B	(Glu,D-Orn)(Tyr, D-Thr)(Glu,D-X)(Pro,Gln,D-Tyr)(Ile)	Incomplete (X could be Val or Ala)	<i>B. amyloliquefaciens</i> WF02
Iturin A	(Asn)(D-Tyr,D-Asn,Gln,Pro)(D-Asn,Ser)	Complete	<i>B. amyloliquefaciens</i> WF02
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,Leu)(Leu)	Complete	<i>B. amyloliquefaciens</i> bm1
Fengycin A/B	(Glu,D-Orn,Tyr, D-Thr)(Glu,D-X)(Pro,Gln,D-Tyr)(Ile)	Incomplete (X could be Val or Ala)	<i>B. amyloliquefaciens</i> bm1
Iturin A	(Asn)(D-Tyr,D-Asn,Gln,Pro)(D-Asn,Ser)	Complete	<i>B. amyloliquefaciens</i> bm1
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. amyloliquefaciens</i> LS1-002-014 s
Fengycin A/B	(Glu,D-Orn)(Tyr, D-Thr)(Glu,D-X)(Pro,Gln,D-Tyr)(Ile)	Incomplete (X could be Val or Ala)	<i>B. amyloliquefaciens</i> LS1-002-014 s
Iturin A	(Asn)(D-Tyr,D-Asn,Gln,Pro)(D-Asn,Ser)	Complete	<i>B. amyloliquefaciens</i> LS1-002-014 s
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. amyloliquefaciens</i> HM618
Fengycin	(.....)(Tyr, D-Thr)(Glu,D-X)(Pro,Gln,D-Tyr)(Ile)	Partial (X could be Val or Ala)	<i>B. amyloliquefaciens</i> HM618
Iturin A	(Asn)(D-Tyr,D-Asn,Gln,Pro)(D-Asn,Ser)	Complete	<i>B. amyloliquefaciens</i> HM618
Surfactin	(Glu,Leu,...)(Val,Asp,...)(Leu)	Partial and fragmented	<i>B. amyloliquefaciens</i> B3
Fengycin	(.....)(...)(...,D-X)(Pro,Gln,D-Tyr)(Ile)	Partial	<i>B. amyloliquefaciens</i> B3
[Ile7] surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Ile)	Complete	<i>B. atrophaeus</i> PENS20
Fengycin A/B	(Glu,D-Orn)(Tyr, D-Thr)(Glu,D-X)(Pro,Gln,D-Tyr)(Ile)	Incomplete (X could be Val or Ala)	<i>B. atrophaeus</i> PENS20
Iturin A	(Asn)(D-Tyr,D-Asn,Gln,Pro)(...,Ser)	Partial	<i>B. atrophaeus</i> PENS20
Kurstakin	(D-Thr)(X,Ala)(Ser,Leu,D-X,Glu)	Incomplete (Leu instead of His; Glu instead of Gln; X, X could be Gly and Gln respectively)	<i>B. bombysepticus</i> F12

(Continued)

TABLE 3 (Continued)

Predicted LP	Structure analysis	Comments	Strains
Kurstakin	(D-Thr)(X,Ala)(Ser,Leu,D-X,Glu)	Incomplete (Leu instead of His; Glu instead of Gln; X, X could be Gly and Gln respectively)	<i>B. bombysepticus</i> Cuernavaca_S2
[Ile7] surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Ile)	Complete	<i>B. cellulasensis</i> NJ-V2
[Ile7] surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Ile)	complete	<i>B. cellulasensis</i> NJ-M2
Kurstakin	(D-Thr)(X,Ala)(Ser,Leu,D-X,Glu)	Incomplete (Leu instead of His; Glu instead of Gln; X, X could be Gly and Gln respectively)	<i>B. cereus</i> CMCC P0011
Kurstakin	(D-Thr)(X,Ala)(Ser,Leu,D-X,Glu)	Incomplete (Leu instead of His; Glu instead of Gln; X, X could be Gly and Gln respectively)	<i>B. cereus</i> CMCC P0021
Kurstakin	(D-Thr)(X,Ala)(Ser,Leu,D-X,Glu)	Incomplete (Leu instead of His; Glu instead of Gln; X, X could be Gly and Gln respectively)	<i>B. cereus</i> NJ-W
Kurstakin	(D-Thr)(X,Ala)(Ser,Leu,D-X,Glu)	Incomplete (Leu instead of His; Glu instead of Gln; X, X could be Gly and Gln respectively)	<i>B. cereus</i> C-1
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. halotolerans</i> ZB201702
Fengycin	(Glu,D-Orn)(Tyr,D-Thr)(Glu,D-X)(Pro,Glu,...)(...)	Partial (Glu instead of Gln and X could be Val or Ala)	<i>B. halotolerans</i> ZB201702
Fengycin	(...,D-Orn)(Tyr,D-Thr)(...,...)(Pro,...,...)(Ile)	Partial and fragmented	<i>B. inaquosorum</i> DE111
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,...)(Leu)	Partial and fragmented	<i>B. inaquosorum</i> DE111
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. inaquosorum</i> LBA001
Fengycin	(...,D-Orn)(Tyr,D-Thr)(Glu,...)(Pro,Glu,...)(Ile)	Partial and fragmented (Glu instead of Gln)	<i>B. inaquosorum</i> LBA001
Bacillomycin F	(Asn)(D-Tyr,D-Asn,Pro,Gln)(D-Asn,Thr)	Complete	<i>B. inaquosorum</i> LBA001
Lichenysin A	(Gln,Leu,D-Leu)(Val,Asp,D-Leu)(Ile)	Complete	<i>B. licheniformis</i> SCDB 14
Lichenysin A	(Gln,Leu,D-Leu)(Val,Asp,D-Leu)(Ile)	Complete	<i>B. licheniformis</i> P8_B2
Lichenysin A	(Gln,Leu,D-Leu)(Val,Asp,D-Leu)(Ile)	Complete	<i>B. licheniformis</i> SCK B11
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. licheniformis</i> 14ADL4
Fengycin	(...)(...)(...,Glu,D-Tyr)(Ile)	Partial and fragmented	<i>B. licheniformis</i> 14ADL4
Fengycin	(...,D-Orn)(Tyr,D-Thr)(...,D-X)(Pro,Glu,D-Tyr)(Ile)	Partial and fragmented (Glu instead of Gln and X could be Val)	<i>B. mojavensis</i> B-41341
Surfactin	(Glu,Leu,...)(Val,Asp,...)(Leu)	Partial and fragmented	<i>B. mojavensis</i> B-41341
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. mojavensis</i> B-41812
Fengycin B variant	(Glu,D-Orn)(Tyr,D-Thr)(Glu,D-X)(Pro,Glu,D-Tyr)(Ile)	Incomplete (Glu instead of Gln and X could be Val)	<i>B. mojavensis</i> B-41812
Kurstakin variant	(D-Thr)(X,Ala)(Ser,Leu,D-Thr,Glu)	Incomplete (Leu instead of His; Thr instead of Gln; Glu instead of Gln; X could be Gly)	<i>B. mycoides</i> BGSC 4BQ1
Kurstakin	(D-Thr)(X,Ala)(Ser,Leu,D-X,Glu)	Incomplete (Leu instead of His; Glu instead of Gln; X, X could be Gly and Gln respectively)	<i>B. mycoides</i> Gnyt1
Lichenysin A	(Gln,Leu,D-Leu)(Val,Asp,...)(Ile)	Partial and fragmented	<i>B. paralicheniformis</i> UBBLI-30
Fengycin B	(...,D-Orn)(Tyr,D-Thr)(...,D-Val)(Pro,Gln,D-Tyr)(Ile)	Partial and fragmented	<i>B. paralicheniformis</i> UBBLI-30
Lichenysin A	(Gln,Leu,D-Leu)(Val,Asp,D-Leu)(Ile)	Complete	<i>B. paralicheniformis</i> 14DA11
Fengycin B	(Glu,...)(Tyr,D-Thr)(Glu,Val)(Pro,Gln,D-Tyr)(Ile)	Partial and fragmented	<i>B. paralicheniformis</i> 14DA11
Lichenysin A	(Gln,Leu,D-Leu)(Val,Asp,D-Leu)(Ile)	Complete	<i>B. paralicheniformis</i> CP47
Fengycin B	(Glu,D-Orn)(Tyr,D-Thr)(Glu,D-Val)(Pro,Gln,D-Tyr)(Ile)	Complete	<i>B. paralicheniformis</i> CP47

(Continued)

TABLE 3 (Continued)

Predicted LP	Structure analysis	Comments	Strains
[Ile7] surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Ile)	Complete	<i>B. pumilus</i> NJ-V
[Ile7] surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Ile)	Complete	<i>B. pumilus</i> DSM 1794
[Ile7] surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Ile)	Complete	<i>B. pumilus</i> MS32
[Ile7] surfactin	(Glu,Leu,D-Leu)(Val,Asp,...)(Ile)	Partial	<i>B. pumilus</i> B4127
[Ile4,7] surfactin	(Glu,Leu,D-Leu)(Ile,Asp,D-Leu)(Ile)	Complete	<i>B. safensis</i> G6S3
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,Leu)(Leu)	Complete	<i>B. spizizenii</i> T30
Mycosubtilin	(Asn)(D-Tyr,D-Asn,Gln,Pro)(D-Ser,Asn)	Complete	<i>B. spizizenii</i> T30
Mycosubtilin	(Asn)(D-Tyr,D-Asn,Gln,Pro)(D-Ser,Asn)	Complete	<i>B. spizizenii</i> AS2
Surfactin	(Glu,Leu,...)(Val,Asp,...)(Leu)	Partial and fragmented	<i>B. spizizenii</i> AS2
Mycosubtilin	(Asn)(D-Tyr,D-Asn,Gln,Pro)(D-Ser,Asn)	Complete	<i>B. spizizenii</i> HUK15
Surfactin	(Glu,Leu,...)(Val,Asp,...)(Leu)	Partial and fragmented	<i>B. spizizenii</i> HUK15
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. subtilis</i> SJ-10
Fengycin B	(Glu,D-Orn)(Tyr, D-Thr)(Glu,D-Val)(Pro,Glu,D-Tyr)(Ile)	Complete	<i>B. subtilis</i> SJ-10
Bacillomycin D	(Asn)(D-Tyr,D-Asn,Pro,Glu)(D-Ser,Thr)	Complete	<i>B. subtilis</i> SJ-10
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. subtilis</i> MEC_B298
Fengycin B variant	(.....)(Tyr, D-Thr)(Glu,D-Val)(Pro,Glu,D-Tyr)(Ile)	Partial (Glu instead of Gln)	<i>B. subtilis</i> MEC_B298
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. subtilis</i> s-16
Fengycin B variant	(Glu,...)(Tyr, D-Thr)(Glu,D-Val)(Pro,Glu,D-Tyr)(Ile)	Partial and fragmented (Glu instead of Gln)	<i>B. subtilis</i> s-16
Fengycin B variant	(Glu,...)(Tyr, D-Thr)(Glu,D-Val)(Pro,Glu,D-Tyr)(Ile)	Partial and fragmented (Glu instead of Gln)	<i>B. subtilis</i> BSP1
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. subtilis</i> BSP1
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. subtilis</i> KH2
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. subtilis</i> FUA2231
Fengycin B variant	(Glu,D-Orn,Tyr, D-Thr)(Glu,D-Val)(Pro,Glu,D-Tyr)(Ile)	Complete (Glu instead of Gln)	<i>B. subtilis</i> FUA2231
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. subtilis</i> UD1022
Fengycin B variant	(Glu,D-Orn)(Tyr, D-Thr)(Glu,D-Val)(Pro,Glu,D-Tyr)(Ile)	Complete (Glu instead of Gln)	<i>B. subtilis</i> UD1022
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. subtilis</i> SFA-H43
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. subtilis</i> CGMCC 2108
Mycosubtilin	(Asn)(D-Tyr,D-Asn,Gln,Pro)(D-Ser,Asn)	Complete	<i>B. subtilis</i> T30
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. subtilis</i> T30
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. subtilis</i> NCIB 3610
Fengycin	(Glu,D-Orn)(Tyr, D-Thr)(.....)(Pro,Glu,D-Tyr)(Ile)	Partial and fragmented (Glu instead of Gln)	<i>B. subtilis</i> NCIB 3610

(Continued)

TABLE 3 (Continued)

Predicted LP	Structure analysis	Comments	Strains
Fengycin B variant	(Glu,D-Orn)(Tyr, D-Thr)(Glu,D-Val)(Pro,Glu,D-Tyr)(Ile)	Complete (Glu instead of Gln)	<i>B. subtilis</i> ZD01
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. subtilis</i> ZD01
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. subtilis</i> 73
Fengycin	(...)(Glu,D-Tyr)(Ile)	Partial and fragmented	<i>B. subtilis</i> 73
[Ile7] surfactin	(Glu,Leu,...)(Val,Asp,...)(Ile)	Partial and fragmented	<i>B. subtilis</i> G8
Fengycin	(...)(...)(...,Glu,D-Tyr)(...)	Partial and fragmented	<i>B. subtilis</i> G8
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. subtilis</i> FUA2232
Fengycin B variant	(Glu,D-Orn,Tyr, D-Thr)(Glu,D-Val)(Pro,Glu,D-Tyr)(Ile)	Complete (Glu instead of Gln)	<i>B. subtilis</i> FUA2232
Surfactin	(Glu,Leu, D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. subtilis</i> SRCM103517
Fengycin B variant	(Glu,D-Orn,Tyr, D-Thr)(Glu,D-Val)(Pro,Glu,D-Tyr)(Ile)	Complete (Glu instead of Gln)	<i>B. subtilis</i> SRCM103517
Surfactin	(Glu,Leu,...)(Val,Asp,...)(Leu)	Partial and fragmented	<i>B. subtilis</i> S1
Kurstakin	(D-Thr)(X,Ala)(Ser,Leu,D-X,Glu)	Incomplete (Leu instead of His; Glu instead of Gln; X, X could be Gly and Gln respectively)	<i>B. thuringiensis</i> HD12
Kurstakin	(D-Thr)(X,Ala)(Ser,Leu,D-X,Glu)	Incomplete (Leu instead of His; Glu instead of Gln; X, X could be Gly and Gln respectively)	<i>B. thuringiensis</i> Bt185
Kurstakin	(D-Thr)(X,Ala)(Ser,Leu,D-X,Glu)	Incomplete (Leu instead of His; Glu instead of Gln; X, X could be Gly and Gln respectively)	<i>B. thuringiensis</i> YWC2-8
Kurstakin	(D-Thr)(X,Ala)(Ser,Leu,D-X,Glu)	Incomplete (Leu instead of His; Glu instead of Gln; X, X could be Gly and Gln respectively)	<i>B. thuringiensis</i> BGSC 4AA1
Kurstakin	(D-Thr)(X,Ala)(Ser,Leu,D-X,Glu)	Incomplete (Leu instead of His; Glu instead of Gln; X, X could be Gly and Gln respectively)	<i>B. thuringiensis</i> HS18-1
Kurstakin	(D-Thr)(X-Ala)(Ser,Leu,D-X,Glu)	Incomplete (Leu instead of His; Glu instead of Gln; X, X could be Gly and Gln respectively)	<i>B. thuringiensis</i> c25
Kurstakin	(D-Thr)(X-Ala)(Ser,Leu,D-X,Glu)	Incomplete (Leu instead of His; Glu instead of Gln; X, X could be Gly and Gln respectively)	<i>B. thuringiensis</i> JW-1
Kurstakin	(D-Thr)(X,Ala)(Ser,Leu,D-X,Glu)	Incomplete (Leu instead of His; Glu instead of Gln; X, X could be Gly and Gln respectively)	<i>B. thuringiensis</i> HD-1
Kurstakin	(D-Thr)(X,Ala)(Ser,Leu,D-X,Glu)	Incomplete (Leu instead of His; Glu instead of Gln; X, X could be Gly and Gln respectively)	<i>B. thuringiensis</i> YBT-1518
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. vallismortis</i> NBIF-001
Fengycin A/B	(Glu,D-Orn)(Tyr, D-Thr)(Glu,D-X)(Pro,Gln,D-Tyr)(Ile)	Incomplete (X Could be Val or Ala)	<i>B. vallismortis</i> NBIF-001
Iturin A	(Asn)(D-Tyr,D-Asn,Gln,Pro)(D-Asn,Ser)	Complete	<i>B. vallismortis</i> NBIF-001
Fengycin B variant	(Glu,D-Orn)(Tyr, D-Thr)(Glu,D-Val)(Pro,Glu,D-Tyr)(Ile)	Complete (Glu instead of Gln)	<i>B. vallismortis</i> DSM 11031
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. vallismortis</i> DSM 11031
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. velezensis</i> ATR2
Fengycin	(Glu,D-Orn)(Tyr, D-Thr)(.....)(Pro,Gln...)(Ile)	Partial and fragmented	<i>B. velezensis</i> ATR2
Bacillomycin D	(Asn)(D-Tyr,D-Asn,Pro,Glu)(D-Ser,Thr)	Complete	<i>B. velezensis</i> ATR2
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. velezensis</i> SYP-B637
Fengycin A/B	(Glu,D-Orn)(Tyr, D-Thr)(Glu,D-X)(Pro,Gln,D-Tyr)(Ile)	Incomplete (X could be Val or Ala)	<i>B. velezensis</i> SYP-B637
Bacillomycin D	(Asn)(D-Tyr,D-Asn,Pro,Glu)(D-Ser,Thr)	Complete	<i>B. velezensis</i> SYP-B637

(Continued)

TABLE 3 (Continued)

Predicted LP	Structure analysis	Comments	Strains
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. velezensis</i> CGMCC 11640
Fengycin A/B	(Glu,D-Orn)(Tyr, D-Thr)(Glu,D-X)(Pro,Gln,D-Tyr)(Ile)	Incomplete (X could be Val or Ala)	<i>B. velezensis</i> CGMCC 11640
Iturin A	(Asn)(D-Tyr,D-Asn,Gln,Pro)(D-Asn,Ser)	Complete	<i>B. velezensis</i> CGMCC 11640
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,...)(Leu)	Partial and fragmented	<i>B. velezensis</i> Lzh-a42
Fengycin A/B	(Glu,D-Orn)(Tyr, D-Thr)(Glu,D-X)(Pro,Gln,D-Tyr)(Ile)	Incomplete (X could be Val or Ala)	<i>B. velezensis</i> Lzh-a42
Bacillomycin D	(Asn)(D-Tyr,D-Asn,Pro,Glu)(D-Ser,Thr)	Complete	<i>B. velezensis</i> Lzh-a42
Fengycin A/B	(Glu,D-Orn)(Tyr, D-Thr)(Glu,D-X)(Pro,Gln,D-Tyr)(Ile)	Incomplete (X could be Val or Ala)	<i>B. velezensis</i> YJ0-1
Bacillomycin D	(Asn)(D-Tyr,D-Asn,Pro,Glu)(D-Ser,Thr)	Complete	<i>B. velezensis</i> YJ0-1
Surfactin	(Glu,Leu,D-Leu)(Val)(Asp,D-Leu)(Leu)	Complete	<i>B. velezensis</i> YJ0-1
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. velezensis</i> L-S60
Fengycin	(Glu,D-Orn)(Tyr, D-Thr)(.....)(Pro,Gln,D-Tyr)(Ile)	Partial and fragmented	<i>B. velezensis</i> L-S60
Iturin A	(Asn)(D-Tyr,D-Asn,Gln,Pro)(D-Asn,Ser)	Complete	<i>B. velezensis</i> L-S60
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. velezensis</i> CBMB205
Fengycin A/B	(Glu,D-Orn)(Tyr, D-Thr)(Glu,D-X)(Pro,Gln,D-Tyr)(Ile)	Incomplete (X could be Val or Ala)	<i>B. velezensis</i> CBMB205
Iturin A	(Asn)(D-Tyr,D-Asn,Gln,Pro)(D-Asn,Ser)	Complete	<i>B. velezensis</i> CBMB205
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. velezensis</i> SB1216
Bacillomycin L	(Asn)(D-Tyr,D-Asn,Ser,Glu)(D-Ser,Thr)	Complete	<i>B. velezensis</i> SB1216
Fengycin	(.....)(Tyr, D-Thr)(.....)(Pro,.....)(Ile)	Partial and fragmented	<i>B. velezensis</i> SB1216
Surfactin	(Glu,Leu,...)(Val,Asp,D-Leu)(Leu)	Partial and fragmented	<i>B. velezensis</i> DSYZ
Fengycin A/B	(Glu,D-Orn)(Tyr, D-Thr)(Glu,D-X)(Pro,Gln,D-Tyr)(Ile)	Incomplete (X could be Val or Ala)	<i>B. velezensis</i> DSYZ
Iturin A	(Asn)(D-Tyr,D-Asn,Gln,Pro)(D-Asn,Ser)	Complete	<i>B. velezensis</i> DSYZ
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. velezensis</i> DR-08
Fengycin A/B	(Glu,D-Orn)(Tyr, D-Thr)(Glu,D-X)(Pro,Gln,D-Tyr)(Ile)	Incomplete (X could be Val or Ala)	<i>B. velezensis</i> DR-08
Iturin A	(Asn)(D-Tyr,D-Asn,Gln,Pro)(D-Asn,Ser)	Complete	<i>B. velezensis</i> DR-08
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. velezensis</i> CMF18
Fengycin A/B	(Glu,D-Orn)(Tyr, D-Thr)(Glu,D-X)(Pro,Gln,D-Tyr)(Ile)	Incomplete (X could be Val or Ala)	<i>B. velezensis</i> CMF18
Iturin A	(Asn)(D-Tyr,D-Asn,Gln,Pro)(D-Asn,Ser)	Complete	<i>B. velezensis</i> CMF18
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. velezensis</i> CK17
Fengycin A/B	(Glu,D-Orn)(Tyr, D-Thr)(Glu,D-X)(Pro,Gln,D-Tyr)(Ile)	Incomplete (X could be Val or Ala)	<i>B. velezensis</i> CK17
Iturin A	(Asn)(D-Tyr,D-Asn,Gln,Pro)(D-Asn,Ser)	Complete	<i>B. velezensis</i> CK17
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. velezensis</i> DMB06

(Continued)

TABLE 3 (Continued)

Predicted LP	Structure analysis	Comments	Strains
Bacillomycin D	(Asn)(D-Tyr,D-Asn,Pro,Glu)(D-Ser,Thr)	Complete	<i>B. velezensis</i> DMB06
Fengycin	(.....)(...)(D-X)(Pro,...)(Ile)	Partial and fragmented	<i>B. velezensis</i> DMB06
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. velezensis</i> UD6-2
Fengycin	(.....)(...)(D-X)(Pro,...)(Ile)	Partial and fragmented	<i>B. velezensis</i> UD6-2
Bacillomycin D	(Asn)(D-Tyr,D-Asn,Pro,Glu)(D-Ser,Thr)	Complete	<i>B. velezensis</i> UD6-2
Surfactin	(Glu,Leu,D-Leu)(Val,Asp,D-Leu)(Leu)	Complete	<i>B. velezensis</i> DMB05
Fengycin	(.....)(.....)(...X)(Pro,...)(Ile)	Partial and fragmented	<i>B. velezensis</i> DMB05
Bacillomycin D	(Asn)(D-Tyr,D-Asn,Pro,Glu)(D-Ser,Thr)	Complete	<i>B. velezensis</i> DMB05
Kurstakin	(D-Thr)(X,Ala)(Ser,Leu,D-X,Glu)	Incomplete (Leu instead of His; Glu instead of Gln; X, X could be Gly and Gln respectively)	<i>B. wiedmannii</i> JAS07/5
Kurstakin	(D-Thr)(X,Ala)(Ser,Leu,D-X,Glu)	Incomplete (Leu instead of His; Glu instead of Gln; X, X could be Gly and Gln respectively)	<i>B. wiedmannii</i> PL1
[Ile4,7] surfactin	(Glu,Leu,D-Leu)(Ile,Asp,D-Leu)(Ile)	Complete	<i>B. safensis</i> ZK-1
[Ile4,7] surfactin	(Glu,Leu,D-Leu)(Ile,Asp,D-Leu)(Ile)	Complete	<i>B. safensis</i> SRCM125915
[Ile4,7] surfactin	(Glu,Leu,D-Leu)(Ile,Asp,D-Leu)(Ile)	Complete	<i>B. safensis</i> PLA 1006

The monomers in brackets are those synthesized by a single gene. Monomers in bold differ from the known variants described in the same position. When the gene cluster was incomplete because scattered on several contigs or located on a too small contig, it was specify in the comment column and symbolized the truncated genes by "...".

genomes of strains belonging to the *Bacillus* genus and isolated either from fermented foods or from soil. These newly identified gene clusters would be capable of producing other new lipopeptide families and new antibiotics. All newly predicted molecules were first compared with those identified subsequently, and then with the non-ribosomal peptides available in the Norine database. Low similarities ranged from 28.6 to 50%. The first new gene cluster detected consisted of 3 genes and 4 modules incorporating Thr, X, Asn and Asn monomers, respectively (Figure 2). Amino acid incorporated by module 2 was not detected, hence the letter X at this position (Figure 2). The predicted tetrapeptide showed a low similarity of 39.4% to Cis-7-tetradecenoyl-D-Asparagine. This new gene cluster was carried by plasmids from *B. cereus* CMCC P0011 and *B. cereus* CMCC P0021 strains. Two new gene clusters consisting of 6 modules each were detected by exploring the genomes of *B. velezensis* ATR2, *B. velezensis* DSYZ, *B. velezensis* CGMCC 11640 and *B. amyloliquefaciens* WF02. The hexapeptide predicted in the *B. velezensis* ATR2 genome was structured as follows: D-Cys-Ser-Cys-Ala-X-D-Asn, and contained a CAL (Co-enzyme A ligase) domain. The hexapeptide predicted in the genomes of *B. velezensis* DSYZ, *B. velezensis* CGMCC 11640 and *B. amyloliquefaciens* WF02 was structured: Val-D-Phe-Asp-D-Asn-Gly-D-Glu. Formation is ensured by the combination of an NRPS system and a PKS (polyketide synthase) system (Figure 2). CAL and C-starter domains were not detected in this hexapeptide. Therefore, it is probably an antibiotic. Literature searches and structural analyses in Norine did not reveal any similar molecules.

In silico analyses also revealed the presence of 3 new NRPS clusters capable of producing 3 different heptapeptides depending on the modular organization and amino acid composition of each. The first predicted heptapeptide detected was carried by *B. velezensis* DSYZ and *B. subtilis* SJ-10 genomes and is structured as follows: D-Cys-Ser-Cys-Ala-X-Asn-D-Asn (Figure 2). Monomer 5 has not been predicted. A CAL domain was detected at the start of the peptide chain (Figure 2). Structural analysis of this heptapeptide in Norine showed little similarity to iturin A1 (32.3%). Furthermore, literature searches revealed no molecules with a similar structure. For the last two predicted heptapeptides, each gene cluster consists of 7 modules incorporating 7 amino acids. One had the following structure: D-Phe-D-Leu-Phe-D-Thr-Val-Ala-Thr and was carried by the *B. amyloliquefaciens* HM618 genome while the other had the following structural architecture: D-Phe-Tyr-Ile-X-D-Phe-Leu-Leu and was carried by the *B. anthracis* CMF9 genome (Figure 2). Their structures were similar to kahalalide A (39.4%) and axinastatin 5 (40%) respectively. The heptapeptide identified in the *B. amyloliquefaciens* HM618 genome contained a thioesterase (TE) domain marking the end of the peptide chain. The heptapeptide identified in the *B. anthracis* genome contained a CAL domain at the start of the peptide chain and a C condensation domain at the end of module 7 (Figure 2).

Exploration of the plasmid sequence of *B. cereus* CMCCP0011 and *B. cereus* CMCCP0021 revealed a BGC probably responsible for the synthesis of a new octapeptide nonribosomal peptide (Figure 2). The chemical structure of unpredictable monomers are represented by the letter X in the peptide chain. The structure search on Norine showed little similarity (50%) with cyanostatin B. No molecules with a similar structure were found in the literature.

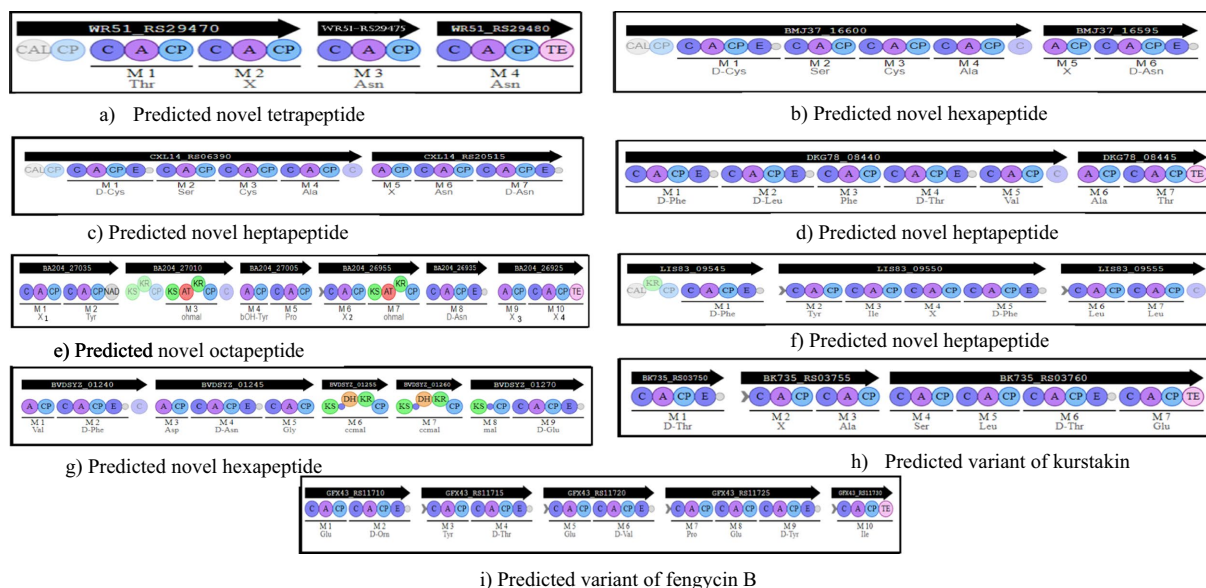


FIGURE 2

Organization of the predicted novel NRPSs and NRPS/PKS. Iterative domains: A, adenylation domain; C, condensation domain; CP, Carrier Protein; E, epimerization domain; TE, thioesterase domain; AT, acetyltransferase domain; KS, ketosynthetase domain; KR, ketoreductase domain; CAL, Coenzyme A ligase; C surrounded in light blue, condensation domain allowing condensation between a PKS monomer and a NRPS monomer. Predicted amino acid specificity is shown under each A domain. (a) a tetrapeptide in *B. cereus* CMCCP0011 (plasmid) and *B. cereus* CMCCP021 (plasmid); (b) a hexapeptide in *B. velezensis* ATR2 (chromosome); (c) a heptapeptide in *B. velezensis* DSYZ (chromosome), *B. subtilis* SJ-10 (chromosome) and *B. velezensis* CGMCC 11640 (chromosome); (d) a heptapeptide in *B. amyloliquefaciens* HM618 (chromosome); (e) an octapeptide in *B. cereus* CMCC P0011 (plasmid) and *B. cereus* CMCC P0021 (plasmid); (f) a heptapeptide in *B. anthracis* CMF9 (chromosome); (g) a hexapeptide in *B. amyloliquefaciens* WF02 (chromosome), *B. velezensis* CGMCC 11640 (chromosome) and *B. velezensis* DSYZ (chromosome); (h) predicted variant of kurstakin in *B. mycoides* BGSC 4BQ1, the difference with the other kurstakin predicted concerning the composition amino acids is located at the level of monomer 6; (i) predicted variant of fengycin B in *B. vallismortis* DSM 11031, *B. subtilis* UD1022, *B. subtilis* FUA2231, *B. subtilis* FUA2232, *B. subtilis* SRCM103517 and *B. subtilis* ZD01 genomes. Modular architecture is similar to the known fengycin B, the difference with fengycin B regarding the composition in amino acids is located at the level monomer 8.

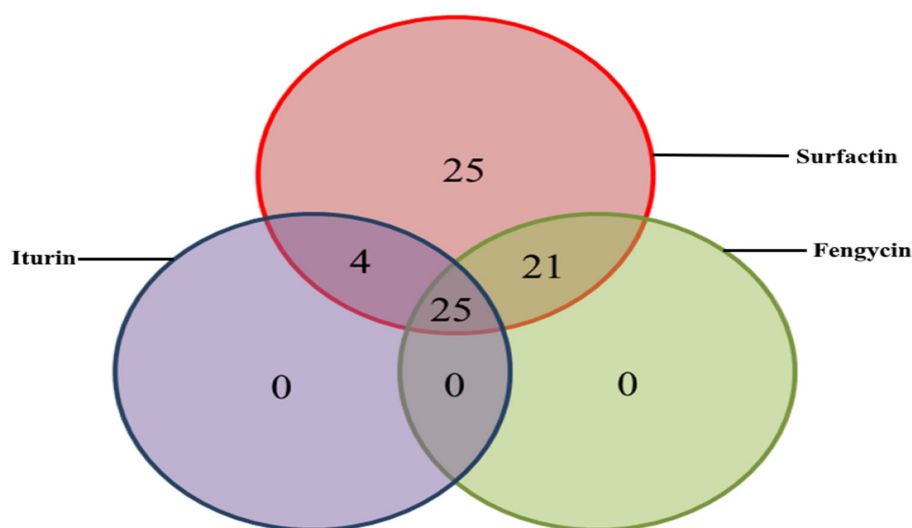


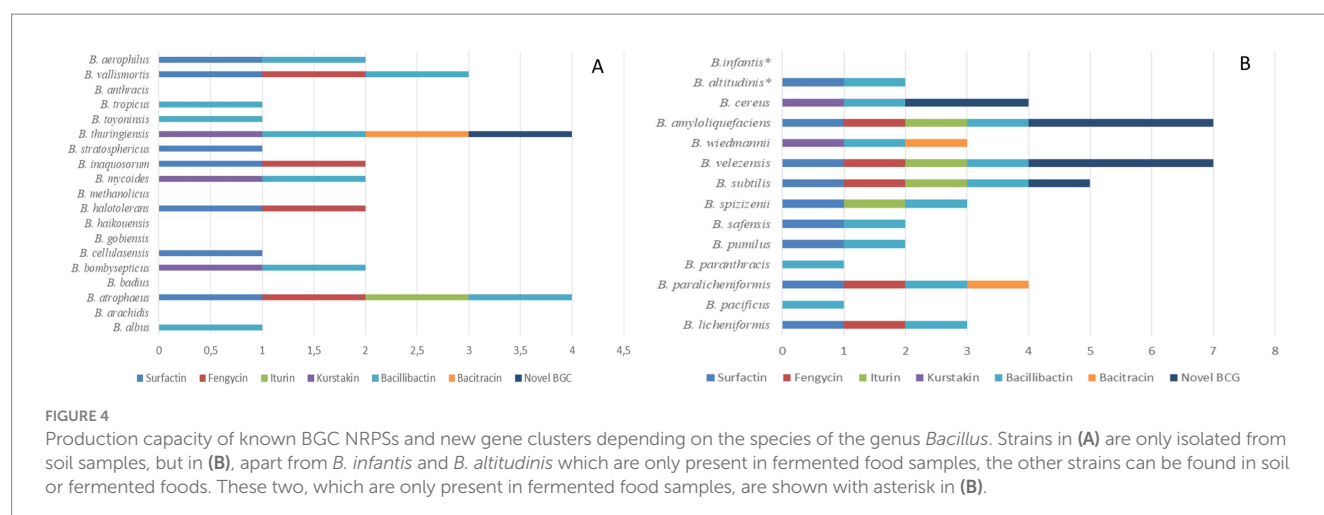
FIGURE 3

Prevalence and coexistence of lipopeptide biosynthesis gene clusters in 123 chromosomes analysis from *Bacillus* (Venn diagram, $n = 75$).

4 Discussion

Bacillus species are known to produce a large variety of secondary metabolites. This production is influenced by environmental conditions. Thus, this study explored 123 complete genomes of

Bacillus isolated from soil and fermented foods. Thus, *in silico* analysis of chromosomal and plasmid sequences using bio-informatics tools specific to non-ribosomal peptides revealed a potential gene cluster responsible for the biosynthesis of lipopeptides (sufactin, fengycin, iturin, kurstakin), the antibiotic bacitracin and the siderophore



bacillibactin. Only *B. subtilis* 916 (Luo et al., 2015a) produced locillomycin, which was not detected in this study. However, surfactins are lipopeptides with variants such as esperin, lichenysin, pumilacidin and surfactin (Waongo et al., 2023). Two distinct amino acids were predicted to be incorporated in the first module: either a glutamate (Glu) or a glutamic acid (Gln) (Hu et al., 2019). As for iturins, they are lipopeptides whose main variants are: iturins A and C, bacillomycins D, F and L, mycosubtilin and mojavensin (Waongo et al., 2023). Bacillibactin is a siderophore whose sequence is Dhb-Gly-Thr (May et al., 2001). As a result, the modular organization and monomer composition of predicted peptides from the surfactin and iturin groups as well as bacillibactin, were similar to literature data. Compared with fengycin, the lipodecapeptide predicted in genomes of *B. subtilis* UD1022, *B. vallismortis* DSM11031, *B. subtilis* ZD01, *B. halotolerans* ZB201702, *B. mojavensis* B-41812, *B. mojavensis* B-41341, *B. subtilis* MEC_B298, *B. subtilis* s-16, *B. subtilis* FUA2231, *B. subtilis* FUA2232, *B. subtilis* SRCM103517, and *B. subtilis* BSP1, contains at position 8 a glutamic acid (Glu). In contrast, various authors have reported that module 8 incorporates glutamine (Gln) during peptide chain formation (Ait Kaki et al., 2020; Hussein, 2019). This observed variability shows that certain *Bacillus* could produce a fengycin B variant. Also, structural analysis of predicted kurstakins shows the presence of a threonine (Thr) at position 6 in the *B. mycoide* BGSC 4BQ1 genome. Generally, the amino acid occupying this position in the case of kurstakin is glutamine (Gln) (Béchet et al., 2012). Although monomers 2 and 6 have not been predicted in the genomes of other *B. cereus* group species (*B. thuringiensis*, *B. cereus*, *B. wiedmannii*, *B. bombysepticus*), it is quite possible that *B. mycoide* BGSC 4BQ1 would be capable of producing a kurstakin variant. Nevertheless, the two putative variants identified through this *in silico* screening approach will need to be further investigated through *in vitro* experiments.

This study revealed the coexistence of lipopeptide biosynthesis genes. Indeed, Luo et al. (2015b) reported the coexistence of surfactin (sfr), bacillomycin (bmy), fengycin (fen) and locillomycin (Loc) gene clusters in *B. subtilis* 916 genome. In this study, *B. amyloliquefaciens* WF02, *B. amyloliquefaciens* HM618, *B. velezensis* ATR2, *B. velezensis* CGMCC 11640, *B. velezensis* Lzh-a42, *B. subtilis* SJ-10 and *B. velezensis* DSYZ genomes also contained three known lipopeptide biosynthesis gene clusters and novel gene cluster that could produce a new lipopeptide family. To our knowledge, no study has demonstrated the

coexistence of four or five gene clusters of lipopeptide in genomes of species such as *B. amyloliquefaciens* and *B. velezensis*. According to Luo et al. (2015b), a co-production of surfactin, fengycin, iturin and locillomycin by *B. subtilis* 916 is at the origin of its inhibitory capacity against multi-resistant *Staphylococcus aureus*. Similarly, the co-production of lipopeptides would reduce hemolytic activity of the producing strains (Luo et al., 2019; Waongo et al., 2023).

The emergence of multi-resistant strains to commonly used antibiotics is a major challenge (Eduardo-Correia et al., 2020). The exploration of new molecules that could serve as alternatives remains a necessity. Thus, the study identified seven new gene clusters (NRPS and NRPS/PKS), two of which were carried by plasmid sequences. These new gene clusters were responsible for the biosynthesis of a tetrapeptide, two hexapeptides, three heptapeptides and octapeptide. The predicted tetrapeptide and octapeptide were synthesized by gene clusters carried by plasmids. The structural architecture and monomer composition of all predicted molecules differed from the peptides available in the database (Flissi et al., 2023). Furthermore, literature searches did not reveal any molecules with similar structure and monomer composition. Except for the hexapeptide predicted in the chromosomes of *B. velezensis* DSYZ and *B. amyloliquefaciens* WF02, all the others carried a CAL or C-starter domain in the first module. In fact, lipopeptides are made up of chains of amino acids and fatty acids. Their structure is characterized by the presence of C-starter and CAL domains, which play a role in fatty acid chain activation (Ongena and Jacques, 2007). Thus, their presence in the first module of the seven predicted new peptides means that they could belong to the lipopeptide family. However, certain antibiotics of purely peptidic clinical interest, such as vancomycin and bacitracin, are characterized by the presence of an adenylation domain (A) in the first module (Konz et al., 1997; Wageningen et al., 1998). This A domain is responsible for selecting and activating the first amino acid to be integrated into the growing peptide chain. Consequently, the new hexapeptide predicted in the chromosomes of *B. velezensis* DSYZ and *B. amyloliquefaciens* WF02 would be a peptide antibiotic. Esmael et al. (2016) reported the synthesis of a new lipopeptide by analyzing the genomes of genus *Burkholderia*. On the other hand, since lipopeptides are synthesized non-ribosomally, the presence of gene clusters responsible for their synthesis in plasmids is rare. According to the literature, no study has revealed the presence of NRPS lipopeptide synthase genes on a *Bacillus* plasmid, only on the *Burkholderia* plasmid (Esmael et al., 2016). As a result, the putative new BGCs would

constitute new families of lipopeptides as well as new antibiotics, which could have unique and interesting biological properties (e.g., antifungal, antibacterial, antiviral). Biomolecules newly predicted in this study could be at the center of current research in the interest of their future use in several fields such as agri-food, cosmetics and medicine. The results of this study reveal that the genomes of *Bacillus* strains available in databases contain many unknown molecules that could play important roles in antimicrobial control. However, it would be interesting to elucidate the biological functions of all these new molecules. Although the bioinformatic discovery of novel molecules often captures attention, rigorous experimental validation remains the crucial step in translating promising ideas into tangible scientific findings and ensuring genuine progress.

5 Conclusion

Exploration of microorganisms genomes allows for the rapid identification of gene clusters located on chromosomes coding for new beneficial molecules. Thus, in addition to chromosomes, it would be interesting to analyze plasmids that could have BGC clusters of NRPs or lipopeptides. Then, this study allowed us to identify seven new gene clusters synthesizing new non-ribosomal peptides. Our results suggest that several NRPs capable of being produced by *Bacillus* strains isolated from fermented foods and soil are still uncharacterized and their properties still unknown. Therefore, targeted research should be conducted experimentally to validate these different predicted molecules.

Data availability statement

Genome mining work was done base on the whole genome sequences of 123 *Bacillus* strains is listed in Table 1, which can be obtained from NCBI nucleotide database (<https://www.ncbi.nlm.nih.gov/nucleotide>). All the other data supporting the findings is contained within the manuscript.

Ethics statement

The manuscript presents research on animals that do not require ethical approval for their study.

References

- Ait Kaki, A., Smargiasso, N., Ongena, M., Kara Ali, M., Moula, N., de Pauw, E., et al. (2020). Characterization of new fengycin cyclic lipopeptide variants produced by *Bacillus amyloliquefaciens* (ET) originating from a salt lake of eastern Algeria. *Curr. Microbiol.* 77, 443–451. doi: 10.1007/s00284-019-01855-w
- Aleti, G., Sessitsch, A., and Brader, G. (2015). Genome mining: prediction of lipopeptides and polyketides from *Bacillus* and related Firmicutes. *Comput. Struct. Biotechnol. J.* 13, 192–203. doi: 10.1016/j.csbj.2015.03.003
- Aslam, B., Khurshid, M., Arshad, M. I., Muzammil, S., Rasool, M., Yasmeen, N., et al. (2021). Antibiotic resistance: one health one world outlook. *Front. Cell. Infect. Microbiol.* 11, 1–20. doi: 10.3389/fcimb.2021.771510
- Béchet, M., Caradec, T., Hussein, W., Abderrahmani, A., Chollet, M., Leclère, V., et al. (2012). Structure, biosynthesis, and properties of kurstakins, nonribosomal lipopeptides from *Bacillus* spp. *Appl. Microbiol. Biotechnol.* 95, 593–600. doi: 10.1007/s00253-012-4181-2

Author contributions

BW: Conceptualization, Investigation, Methodology, Writing – original draft, Writing – review & editing. LN: Investigation, Methodology, Writing – review & editing. FT: Methodology, Writing – review & editing. W-SAZ: Methodology, Writing – review & editing. JL: Supervision, Writing – review & editing. AS: Supervision, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Acknowledgments

We would like to thank University Joseph KI-ZERBO of Burkina Faso and Huazhong Agricultural University of China for their financial support.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Blin, K., Shaw, S., Augustijn, H. E., Reitz, Z. L., Biermann, F., Alanjary, M., et al. (2023). antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res.* 51, 46–50. doi: 10.1093/nar/gkad344

- Blin, K., Shaw, S., Kloosterman, A. M., Charlop-Powers, Z., Van Wezel, G. P., Medema, M. H., et al. (2021). AntiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* 49, W29–W35. doi: 10.1093/nar/gkab335

- Bonmatin, J.-M., Laprevote, O., and Peypoux, F. (2003). Diversity among microbial cyclic Lipopeptides: Iturins and Surfactins. Activity-structure relationships to design new bioactive agents. *Comb. Chem. High Throughput Screen.* 6, 541–556. doi: 10.2174/138620703106298716

- Duban, M., Cociancich, S., and Leclère, V. (2022). Nonribosomal peptide synthesis definitely working out of the rules. *Microorganisms*. 10, 1–19. doi: 10.3390/microorganisms10030577

- Eduardo-Correia, B., Morales-Fillo, H., and Abad, J. P. (2020). Bacteria from the multi-contaminated Tinto river estuary (SW, Spain) show high multi-resistance to

antibiotics and point to *Paenibacillus* spp. as antibiotic-resistance-dissemination players. *Front. Microbiol.* 10, 1–18. doi: 10.3389/fmicb.2019.03071

Esmael, Q., Pupin, M., Kieu, N. P., Chataigné, G., Béchet, M., Deravel, J., et al. (2016). *Burkholderia* genome mining for nonribosomal peptide synthetases reveals a great potential for novel siderophores and lipopeptides synthesis. *MicrobiologyOpen*. 5, 512–526. doi: 10.1002/mbo3.347

Fanaei, M., Jurcic, K., and Emtiazi, G. (2021). Detection of simultaneous production of kurstakin, fengycin and surfactin lipopeptides in *Bacillus mojavensis* using a novel gel-based method and MALDI-TOF spectrometry. *World J. Microbiol. Biotechnol.* 37, 1–11. doi: 10.1007/s11274-021-03064-9

Flissi, A., Duban, M., Jacques, P., Leclère, V., and Pupin, M. (2023). Norine: bioinformatics methods and tools for the characterization of newly discovered nonribosomal peptides A. Flissi, M. Duban, P. Jacques, V. Leclère and M. Pupin *Methods Mol. Biol.* 2670, 303–318. doi: 10.1007/978-1-0716-3214-7_16

Flissi, A., Ricart, E., Campart, C., Chevalier, M., Dufresne, Y., Michalik, J., et al. (2020). Norine: update of the nonribosomal peptide resource. *Nucleic Acids Res.* 48, 465–469. doi: 10.1093/nar/gkz1000

Hathout, Y., Ho, Y., Ryzhov, V., Demirev, P., and Fenselau, C. (2000). Kurstakins: a new class of lipopeptides isolated from *Bacillus thuringiensis*. *J. Nat. Prod.* 63, 1492–1496. doi: 10.1021/np000169q

Hu, F., Liu, Y., and Li, S. (2019). Rational strain improvement for surfactin production: enhancing the yield and generating novel structures. *Microb. Cell Factories*. 18, 1–13. doi: 10.1186/s12934-019-1089-x

Huan, Y., Kong, Q., Mou, H., and Yi, H. (2020). Antimicrobial peptides: classification, design, application and research progress in multiple fields. *Front. Microbiol.* 11, 1–21. doi: 10.3389/fmicb.2020.582779

Hussein, W. (2019). Fengycin or plipastatin? A confusing question in bacilli. *Biotechnologia*. 100, 47–55. doi: 10.5114/bta.2019.83211

Iqbal, S., Ullah, N., and Janjua, H. A. (2021). In vitro evaluation and genome mining of *Bacillus subtilis* strain RS10 reveals its biocontrol and plant growth-promoting potential. *Agriculture*. 11:1273. doi: 10.3390/agriculture11121273

Jin, P., Wang, H., Liu, W., Fan, Y., and Miao, W. (2017). A new cyclic lipopeptide isolated from *Bacillus amyloliquefaciens* HAB-2 and safety evaluation. *Pestic. Biochem. Physiol.* 147, 40–45. doi: 10.1016/j.pestbp.2017.08.015

Kato, T., and Shoji, J. (1980). The structure of octapeptin D (studies on antibiotics from the genus *Bacillus*. XXVIII). *J. Antibiot.* 33, 186–191. doi: 10.7164/antibiotics.33.186

Khurana, H., Sharma, M., Verma, H., Lopes, B. S., Lal, R., and Negi, R. K. (2020). Genomic insights into the phylogeny of *Bacillus* strains and elucidation of their secondary metabolic potential. *Genomics*. 112, 3191–3200. doi: 10.1016/j.ygeno.2020.06.005

Konz, D., Klens, A., Schörgendorfer, K., and Marahiel, M. A. (1997). The bacitracin biosynthesis operon of *Bacillus licheniformis* ATCC 10716: molecular characterization of three multi-modular peptide synthetases. *Chem. Biol.* 4, 927–937. doi: 10.1016/S1074-5521(97)90301-X

Lam, V. B., Meyer, T., Arias, A. A., Ongena, M., Oni, F. E., and Höfte, M. (2021). *Bacillus* cyclic lipopeptides iturin and fengycin control rice blast caused by *Pyricularia oryzae* in potted and acid sulfate soils by direct antagonism and induced systemic resistance. *Microorganisms* 9, 1–25. doi: 10.3390/microorganisms9071441

Leclère, V., Weber, T., Jacques, P., and Pupin, M. (2016). Bioinformatics tools for the discovery of new nonribosomal peptides. *Methods Mol. Biol.* 1401, 209–232. doi: 10.1007/978-1-4939-3375-4_14

Lee, S. C., Kim, S. H., Park, I. H., Chung, S. Y., and Choi, Y. L. (2007). Isolation and structural analysis of bamylocin A, novel lipopeptide from *Bacillus amyloliquefaciens* LP03 having antagonistic and crude oil-emulsifying activity. *Arch. Microbiol.* 188, 307–312. doi: 10.1007/s00203-007-0250-9

Liu, R., Zhang, D., Li, Y., Tao, L., and Tian, L. (2010). A new antifungal cyclic lipopeptide from *Bacillus marinus* B-9987. *Helv. Chim. Acta*. 93, 2419–2425. doi: 10.1002/hlca.201000094

Luo, C., Chen, Y., Liu, X., Wang, X., Wang, X., Li, X., et al. (2019). Engineered biosynthesis of cyclic lipopeptide locillomycins in surrogate host *Bacillus velezensis* FZB42 and derivative strains enhance antibacterial activity. *Appl. Microbiol. Biotechnol.* 103, 4467–4481. doi: 10.1007/s00253-019-09784-1

Luo, C., Liu, X., Zhou, X., Guo, J., Truong, J., Wang, X., et al. (2015a). Unusual biosynthesis and structure of locillomycins from *Bacillus subtilis* 916. *Appl. Environ. Microbiol.* 81, 6601–6609. doi: 10.1128/AEM.01639-15

Luo, C., Liu, X., Zhou, H., Wang, X., and Chen, Z. (2015b). Nonribosomal peptide synthase gene clusters for lipopeptide biosynthesis in *Bacillus subtilis* 916 and their phenotypic functions. *Appl. Environ. Microbiol.* 81, 422–431. doi: 10.1128/AEM.02921-14

Ma, Z., Wang, N., Hu, J., and Wang, S. (2012). Isolation and characterization of a new iturinic lipopeptide, mojavensin A produced by a marine-derived bacterium *Bacillus mojavensis* B0621A. *J. Antibiot.* 65, 317–322. doi: 10.1038/ja.2012.19

May, J. J., Wendrich, T. M., and Marahiel, M. A. (2001). The dbb operon of *Bacillus subtilis* encodes the biosynthetic template for the catecholic siderophore 2,

3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin *. *J. Biol. Chem.* 276, 7209–7217. doi: 10.1074/jbc.M009140200

Meyers, E., Parker, W. L., and Brown, W. E. (1976). A nomenclature proposal for the octapeptin antibiotics. *J. Antibiot.* 29, 1241–1242. doi: 10.7164/antibiotics.29.1241

Naruse, N., Tenmyo, O., Kobaru, S., Kamei, H., Miyaki, T., Konishi, M., et al. (1990). Pumilacidin, a complex of new antiviral antibiotics: production, isolation, chemical properties, structure and biological activity. *J. Antibiot.* 43, 267–280. doi: 10.7164/antibiotics.43.267

Nishikiori, T., Naganawa, H., Muraoka, Y., Aoyagi, T., and Umezawa, H. (1986). Plipastatins: new inhibitors of phospholipase A2, produced by *Bacillus cereus* BMG302-ff67. *J. Antibiot.* 39, 755–761. doi: 10.7164/antibiotics.39.755

Ongena, M., and Jacques, F. (2007). *Bacillus* lipopeptides: versatile weapons for plant disease biocontrol. *Trends Microbiol.* 16, 115–125. doi: 10.1016/j.tim.2007.12.009

Pathak, K. V., and Keharia, H. (2014). Identification of surfactins and iturins produced by potent fungal antagonist, *Bacillus subtilis* K1 isolated from aerial roots of banyan (*Ficus benghalensis*) tree using mass spectrometry. 3 *Biotech* 4, 283–295. doi: 10.1007/s13205-013-0151-3

Peypoux, F., Marion, D., Maget-Dana, R., Ptak, M., Das, B. C., and Michel, G. (1985). Structure of bacillomycin F, a new peptidolipid antibiotic of the iturin group. *Eur. J. Biochem.* 153, 335–340. doi: 10.1111/j.1432-1033.1985.tb09307.x

Peypoux, F., Pommier, M. T., Marion, D., Ptak, M., Das, B. C., and Michel, G. (1986). Revised structure of mycosubtilin, a peptidolipid antibiotic from *Bacillus subtilis*. *J. Antibiot.* 39, 636–641. doi: 10.7164/antibiotics.39.636

Rahman, M. A., Noore, M. S., Hasan, M. A., Ullah, M. R., Rahman, M. H., Hossain, M. A., et al. (2014). Identification of potential drug targets by subtractive genome analysis of *Bacillus anthracis* A0248: an in silico approach. *Comput. Biol. Chem.* 52, 66–72. doi: 10.1016/j.compbiolchem.2014.09.005

Sang-cheol, L., Kim, S., Park, I., Chung, S., Chandra, M. S., and Choi, Y. (2010). Isolation, purification, and characterization of novel fengycin S from *Bacillus amyloliquefaciens* LSC04 degrading-crude oil. *Biotechnol. Bioprocess Eng.* 15, 246–253. doi: 10.1007/s12257-009-0037-8

Shoji, J., and Kato, T. (1976). The structure of cerexin B (studies of antibiotics from the genus *Bacillus*. XVII). *J. Antibiot.* 29, 1275–1280.

Shoji, J., Kato, T., Matsumoto, K., Takahashi, Y., and Mayama, M. (1976a). Production and isolation of cerexins C and D (studies on antibiotics from the genus *Bacillus*. XVIII). *J. Antibiot.* 29, 1281–1285. doi: 10.7164/antibiotics.29.1268

Shoji, J., Kato, T., and Sakazaki, R. (1976b). The total structure of cerexin A (studies on antibiotics from the genus *Bacillus*. XVI). *J. Antibiot.* 29, 1268–1274. doi: 10.7164/antibiotics.29.1275

Tamang, J. P., Shin, D.-H., Jung, S. J., and Chae, S.-W. (2016). Functional properties of microorganisms in fermented foods. *Front. Microbiol.* 7, 1–13. doi: 10.3389/fmicb.2016.00578

Tapi, A., Chollet, M., Scherens, B., and Jacques, P. (2010). New approach for the detection of non-ribosomal peptide synthetase genes in *Bacillus* strains by polymerase chain reaction. *Appl. Microbiol. Biotechnol.* 85, 1521–1531. doi: 10.1007/s00253-009-2176-4

Thasana, N., Prapagdee, B., Rangkadilok, N., Sallabhan, R., Aye, S. L., Ruchirawat, S., et al. (2010). *Bacillus subtilis* SSE4 produces subtilene A, a new lipopeptide antibiotic possessing an unusual C15 unsaturated β -amino acid. *FEBS Lett.* 584, 3209–3214. doi: 10.1016/j.febslet.2010.06.005

Théâtre, A., Cano-Prieto, C., Bartolini, M., Laurin, Y., Deleu, M., Niehren, J., et al. (2021). The surfactin-like lipopeptides from *Bacillus* spp.: natural biodiversity and synthetic biology for a broader application range. *Front. Bioeng. Biotechnol.* 9:623701. doi: 10.3389/fbioe.2021.623701

Thomas, D. W., and Ito, T. (1969). The revised structure of the peptide antibiotic esperin, established by mass spectrometry. *Tetrahedron*. 25, 1985–1990. doi: 10.1016/S0040-4020(01)82819-2

Vater, J., Kablitz, B., Wilde, C., Franke, P., Mehta, N., and Cameotra, S. S. (2002). Matrix-assisted laser desorption/ionization-time of flight mass spectrometry of lipopeptide biosurfactants in whole cells and culture filtrates of *Bacillus subtilis* C-1 isolated from petroleum sludge. *Appl. Environ. Microbiol.* 68, 6210–6219. doi: 10.1128/AEM.68.12.6210-6219.2002

Volpon, L., Tsan, P., Majer, Z., Vass, E., Hollósi, M., Noguera, V., et al. (2007). NMR structure determination of a synthetic analogue of bacillomycin Lc reveals the strategic role of I-Asn1 in the natural iturinic antibiotics. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 67, 1374–1381. doi: 10.1016/j.saa.2006.10.027

Wageningen, A. A., Kirkpatrick, P. N., Williams, D. H., Harris, B. R., Kershaw, J. K., Lennan, N. J., et al. (1998). Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic. *Chem. Biol.* 5, 155–162. doi: 10.1016/S1074-5521(98)90060-6

Waongo, B., Pupin, M., Duban, M., Chataigne, G., Zongo, O., Cisse, H., et al. (2023). Kawal: a fermented food as a source of *Bacillus* strain producing antimicrobial peptides. *Sci. Afr.* 20:e01714. doi: 10.1016/j.sciaf.2023.e01714

Zhang, H. L., Hua, H. M., Pei, Y. H., and Yao, X. S. (2004). Three new cytotoxic cyclic acylpeptides from marine *Bacillus* sp. *Chem. Pharm. Bull.* 52, 1029–1030. doi: 10.1248/cpb.52.1029

Zhao, P., Xue, Y., Gao, W., Li, J., Zu, X., Fu, D., et al. (2018). Bacillaceae-derived peptide antibiotics since 2000. *Peptides*. 101, 10–16. doi: 10.1016/j.peptides.2017.12.018