Paenitracins, a novel family of bacitracin-type nonribosomal peptide antibiotics produced by plant-associated *Paenibacillus* species

Nataliia V. Machushynets [a, *], Somayah S. Elsayed [a], Chao Du [a], Vladyslav Lysenko [b], Mercedes de la Cruz [c], Pilar Sanchez [c], Olga Genilloud [c], Nathaniel I. Martin [b], Mark R. Liles [d], and Gilles P. van Wezel [a, e, *]

[a] Molecular Biotechnology Group, Institute of Biology, Leiden University, Sylviusweg 72, 2333 BE, Leiden, The Netherlands

[b] Biological Chemistry Group, Institute of Biology, Leiden University, Sylviusweg 72, 2333 BE, Leiden, The Netherlands

[c] Fundación MEDINA, Health Sciences Technology Park, Avda Conocimiento 34, 18016, Granada, Spain.

[d] Department of Biological Sciences, Auburn University, 120 W Samford Av, AL 36849, Auburn, AL, USA

[e] Department of Microbial Ecology, Netherlands Institute of Ecology, Droevendaalsesteeg 10, 6708 PB, Wageningen, The Netherlands

*Corresponding authors:

Nataliia V. Machushynets, email: n.v.machushynets@biology.leidenuniv.nl; Gilles P. van Wezel, email: g.wezel@biology.leidenuniv.nl.

ABSTRACT

The growing threat of antimicrobial resistance necessitates the discovery of novel antibiotics with activity against drug-resistant pathogens. Members of the genus *Paenibacillus* are a rich source of nonribosomal peptides (NRPs), including well-known antibiotics such as polymyxins, paenibacterin and tridecaptins. Here we use a targeted Mass-QL-based mass spectrometry approach to identify the NRPs produced by a collection of 227 taxonomically diverse plant-associated *Paenibacillus* strains, providing detailed insights into their NRP-producing potential. Using MassQL to zoom in specifically on NRPs containing basic amino acids, we discovered a novel family of bacitracins, which we designated paenitracins. The paenitracins are the first bacitracin-type peptides reported in *Paenibacillus*, and are distinguished from canonical bacitracins by three previously unseen amino acid substitutions. The paenitracins exhibit potent activity against Gram-positive pathogens, including vancomycin-resistant *Enterococcus faecium* E155. Our work provides a novel metabolomics- and genomics-guided workflow for the discovery of bioactive NRPs as a strategy to prioritize natural product chemical space and accelerate antibiotic discovery.

IMPORTANCE

Members of the genus *Paenibacillus* play an important role in soil ecology, producing a range of important nonribosomal peptides (NRPs) that protect their eukaryotic host. A collection of plant-associated *Paenibacillus* spp. was analyzed for their phylogenetic and metabolic diversity. We developed a novel discovery pipeline that combines feature-based molecular networking with MassQL queries to systematically prioritize bioactive

NRPs containing basic amino acids. Thus we provide a comprehensive genus-wide inventory of NRPs produced by *Paenibacillus* spp. We thereby identified the paenitracins, a new subfamily of bacitracins active against multidrug-resistant Gram-positive pathogens. Our pipeline enables the discovery of novel peptidic natural products to accelerate the prioritization of chemical space for antibiotics.

Keywords:

Paenibacillus; antimicrobial resistance; natural product biosynthesis; nonribosomal peptides; biosynthetic gene cluster; metabolic networking.

INTRODUCTION

Antimicrobial resistance (AMR) continues to challenge global health, and the discovery of new antibiotics is therefore a prerequisite to be able to keep fighting infectious diseases caused by multidrug-resistant pathogens (1). Bacterial natural products (NPs) represent a rich reservoir of antibacterial compounds, many of which have been successfully translated into clinical applications (2). A major challenge is to prioritize the huge extant genomic space for natural products (3).

Members of the bacterial phylum Bacillota (formerly Firmicutes) are well known for their potential to produce a wide range of antimicrobials, in particular nonribosomally synthesized peptides (NRPs) (4-6). Numerous bioactive NRPs produced by *Bacillus* and *Paenibacillus* have been identified, including octapeptins, paenibacterins, polypeptins, tridecaptins, etc (5-7). NRPs such as polymyxins and bacitracins that are derived from these microorganisms, have found application in clinical settings (8, 9). A large fraction of the NRPs contain positively charged amino acids and possess potent antimicrobial activities against Gram-positive and Gram-negative pathogens (10, 11). The mode of action polymyxins, involves binding to the lipid A component of lipopolysaccharides (LPSs) in the outer membrane of Gram-negative bacteria. Other NRPs, like bacitracin produced by *Bacillus subtilis* and *Bacillus licheniformis*, act as inhibitors of cell wall biosynthesis. (9) Specifically, in the presence of a divalent metal ion, most commonly Zn^{2+}, bacitracin binds to and sequesters the membrane-associated phospholipid undecaprenyl pyrophosphate ($C_{55}PP$) (12, 13). Bacitracin A is the major bioactive congener with the highest clinical relevance, and it is mainly used topically to treat infections caused by Gram-positive bacteria (14). The bacitracin family is composed of structurally similar macrocyclic dodecapeptides containing a unique N-terminal

aminothiazoline moiety (15). Previously reported bacitracin A-F congeners vary at positions 1, 5, and 8 of the peptide sequence, with substitutions among Ile, Leu, and Val (16).

One of the challenges in NP discovery is the rapid identification of previously reported compounds, termed structural dereplication (17). Liquid chromatography coupled with high-resolution mass spectrometry (LC-MS) is widely used at the early stages of dereplication studies. Molecular formulas of the detected molecules are determined through the exact mass using the Seven Golden Rules, Sirius 2, and MSFINDER software; and subsequently queried in various NP databases like DNP, UNPD, ChemSpider, and REAXYS; to annotate the molecular structures (18).

In order to further identify compounds in LC-MS-based metabolite profiling workflows, the introduction of the Global Natural Products Social Molecular Networking (GNPS MN) has appeared as a powerful tool to process tandem mass spectrometry (MS) data and perform molecular networking (MN) (19). The MN concept is based on organizing and visualizing tandem MS data through a spectral similarity map, revealing the presence of homologous MS^2 fragmentations. As structurally related compounds share similar fragmentation spectra, their nodes tend to gather together and create clusters of analogs (20). The known compounds are annotated by matching their MS^2 spectra against an available spectral library. Moreover, combination of MN with annotation methods, such as MS2LDA (21), NAP (22), or DEREPLICATOR (23) accelerates the identification of unannotated ions (24). Mass Spec Query Language (MassQL) has been developed based on advancements in molecular networking to explore underutilized MS/MS data further (25). MassQL captures the unique characteristics of MS data, including isotopic patterns, diagnostic fragmentation, and

neutral loss; and establishes a comprehensive MS terminology for searching MS patterns across datasets. Additionally, NP prioritization strategies have been applied based on the combination of molecular networking with different sources of information (e.g., genomic, taxonomic, geographic, bioactivity, and/or spectral data) (26-29). Despite the ability of current tools to dereplicate NPs upstream of the isolation process, effective strategies for the early detection of potentially bioactive NPs within crude extracts are still lacking.

In this study, we provide a full phylogenomic analysis of 227 *Paenibacillus* spp. and their potential to produce NRPs. We introduce an early-stage prioritization method that combines feature-based MN (FBMN) and MassQL search to target basic amino acid-containing NRPs with potential antimicrobial properties that are produced by a collection of plant-associated *Paenibacillus* spp. Basic amino acid signatures were detected and mapped onto a massive multi-informative molecular network. This led to the discovery of new bacitracin congeners, which we designated paenitracins, with potent bioactivity against Gram-positive pathogens, including vancomycin-resistant *E. faecium* E155.

RESULTS AND DISCUSSION

Taxonomic and chemical diversity of a collection of *Paenibacillus* strains

Considering the great pharmacological and biotechnological potential of the metabolites produced by *Paenibacillus*, we here analyzed 227 *Paenibacillus* isolates from the Auburn University (USA) plant-associated microbial strain collection. These *Paenibacillus* species had been isolated from the root surface of field-grown corn (*Z. mays*) and soybean (*G. max*) grown in Dunbar (Nebraska), Carrol (Iowa), Whitewater (Wisconsin), Sparta (Illinois), Troy (Ohio) and Tallassee (Alabama) (Table S1). Taxonomic analysis of the strains was conducted by comparing the 16S rRNA genes from the isolates, and a maximum-likelihood tree was constructed (Figure 1).

In search of novel NRPs, we investigated the chemical diversity of the specialized metabolites produced by *Paenibacillus* isolates. For this purpose, all 227 isolates of the collection were cultured individually on TSA media in 96 deep-well plates for 72 h and extracted with isopropyl alcohol (IPA) supplemented with 0.1% (v/v) formic acid (FA). Subsequently, crude extracts were tested against Gram-positive (*Staphylococcus aureus* ATCC 29213) and Gram-negative (*Escherichia coli* ATCC 25922) indicator strains to assess the antibacterial activity of the strains. Nine crude extracts exhibited significant growth inhibition of *S. aureus*, while 22 crude extracts showed bioactivity against *E. coli* (Figure 1, Table S1). These extracts were then subjected to LC-MS/MS analysis to identify the natural products accountable for the observed bioactivities. LC-MS/MS data were processed with MzMine 2 (30) and exported for GNPS FBMN. Prior to uploading the data to GNPS, media-derived mass features were eliminated. A mass exclusion threshold of 300 *m/z* was implemented to minimize the occurrence of mass features originating from small metabolites that are not relevant to our study. FBMN

analysis resulted in generating molecular network consisting of 3096 parent ions (nodes) connected through 3692 edges (Figure 2). Characteristics of the natural products, such as annotation, *m/z* value and species-specific molecule production were visualized using Cytoscape 3.9.1 (31).

Mass spectrometry-based molecular networking allows clustering of molecules with similar MS/MS fragmentation patterns, which stem from the similarity in their structures (19). The clusters allow visual exploration of the metabolites produced by the strains, enabling rapid dereplication of known compounds and visualization of unknown compounds (20). Known metabolites were dereplicated by matching the experimentally derived MS/MS fragmentation patterns with the annotated and curated MS/MS spectra within the GNPS database. Since the spectral library search did not work well for NRPs from *Paenibacillus*, an in-house database was created of the MS data for known lipopeptides from *Paenibacillus* and used to dereplicate known metabolites. This identified lipopeptides belonging to the families of fusaricidins, polymyxins and tridecaptins (Table S3, Figure 2, Figures S1-S3). These dereplicated mass features were used to propagate the annotation to other connected mass features. Despite the identification of a substantial number of putative congeners of known lipopeptides in the network, the majority of the produced metabolites remained unidentified.

Targeting cationic lipopeptides using MassQL-assisted fragment search

Several of the identified lipopeptides are well-known antibiotics with potent activity, typically attributed to their net positive charge (Table S3), which enhances electrostatic interactions with the negatively charged components on the surface of bacterial cells, such as the cell membrane or lipopolysaccharides (LPSs) (32). Therefore, we sought to

detect specifically those lipopeptides with potential bioactive properties that contained one or multiple positively charged amino acid residues. To accomplish this, a feature-based spectral similarity network was paired with a MassQL (33) search query to target the MS/MS spectra that contained the most common positively charged amino acids. The fragment search of *b* ions of 2,4-diaminobuturic acid (Dab) (*m/z* 101.0709), ornithine (Orn) (*m/z* 115.0871), lysine (Lys) (*m/z* 129.1027), arginine (Arg) (*m/z* 157.1089) and the immonium ion of histidine (His) (*m/z* 110.0718) was conducted with the accuracy of *m/z* 0.005 and at minimum intensity of 5% (Figure 3).

MassQL search of Dab-containing parent ions highlighted previously dereplicated polymyxin and tridecaptin spectral families (Figure 4). The identification of different polymyxin (34, 35) and tridecaptin (36) congeners within numerous *Paenibacillus* extracts underscored the chemical diversity of our strain collection. Despite the structural heterogeneity of polymyxins and tridecaptins, a shared characteristic is the occurrence of multiple Dab amino acid residues. Furthermore, MassQL revealed a spectral family of compounds that contained two distinct amino acid residues, namely ornithine and lysine (Figure 5A). This cluster represented triply-charged ions with a mass range of 1500–1650 Da. Interestingly, the compounds from this cluster were produced exclusively by two isolates from the entire strain collection, namely *Paenibacillus* sp. JJ-602 and *Paenibacillus* sp. JJ-1115. These compounds could not be dereplicated by spectral library and in-house database search. Extensive literature search and manual MS/MS dereplication resulted in the annotation of these compounds as paenibacterin A congeners (Figure 5B) (10).

Multiple positively charged lysine and ornithine residues contribute to the overall charge of paenibacterins and enhance interactions with negatively charged LPS from the outer membrane of Gram-negative bacteria (37).

Discovery of a novel family of bacitracins

In addition to the annotated lipopeptide classes, we identified multiple unknown molecular families that contained compounds with basic amino acids such as lysine, ornithine, and histidine (Figure 3). One molecular family of Lys-containing compounds stood out, as the production was specific to only a limited number of isolates (Figure S4). The Lys-containing molecular family could be distinguished into three sub-families (Figure 6A). Sub-family 1 and 2 represent doubly charged ions with a mass range of 1080 to 1150 Da. These compounds were similar to recently discovered lipopeptide paenilipoheptins A and B produced by *P. polymyxa* E681 (38, 39). The third sub-family comprised doubly charged molecular ions with significantly larger masses compared to paenilipoheptins (Figure 6A). The edge connecting a node in sub-family 3 (*m/z* 712.8836) with the node due to paenilipoheptin D (*m/z* 530.8307) has a cosine score of 0.6, indicating low similarity. A mirror plot comparing the MS/MS spectra of the two mass features showed no matching dipeptide fragment ions, and the only shared ions were due to a Lys residue and fragments of Trp (Figure S6).

MS/MS spectra of all the mass features within this sub-family exhibited an ion with an *m/z* 199.0897. However, these mass features could not be matched to known microbial natural products. Analysis of the MS/MS spectra of the parent compound with *m/z* 712.8836 revealed that a fragment ion with an *m/z* 199.0897 has a molecular formula of $C_9H_{15}N_2OS$, which corresponds to a thiazoline-containing substructure

commonly found in the peptide antibiotic bacitracin, so far exclusively identified from the genus *Bacillus*. (40) Based on the fact that the b_2, b_3, b_4, and b_5 ions were shared between the parent mass with *m/z* 712.8836 and bacitracin A (41), we concluded that their pentapeptide side chains are identical. The differences observed in b_6–b_{11} ions indicated that the structural distinctions between the two metabolites lie in the heptapeptide ring. We did not observe fragment ions corresponding to Orn, Phe or His, which are found in the structure of bacitracin A, while we could ascertain the presence of an ion corresponding to a Trp residue (*m/z* 187.0864). Moreover, the dipeptide fragment ions Trp-Leu (*m/z* 300.171), Trp-Thr (*m/z* 288.1352), and Leu-Leu (*m/z* 227.1764) were detected. By tracing the series of both *b* and *y* ions, we elucidated the amino acid sequence of the compound with *m/z* 712.8836 as (Ile-Cys)–Leu/Ile–Glu–Leu/Ile–Lys–Leu/Ile–Leu/Ile–Trp–Thr–Asp–Asn (Figure S7). The MS/MS studies used did not allow us to discriminate between Leu and Ile in positions 3, 5, 7 and 8 of the peptide amino acid sequence. Notably, this sequence diverges from any known bacitracin analogues, indicating the discovery of a novel branch of the important bacitracin-family of NRPs.

To purify the compound for more detailed studies, *Paenibacillus* sp. JJ-778 was cultured in a 10 L MHB media and subjected to extraction using Diaion® HP-20 beads, followed by elution with 100% IPA supplemented with 0.1% (v/v) formic acid. Multiple rounds of chromatographic purification resulted in the isolation of compounds (**1**) and (**2**) (Figures S8, S9). Both compounds had the same molecular formula of $C_{67}H_{106}N_{15}O_{17}S$, based on their exact mass of *m/z* 1424.7584 for $[M+H]^+$, highly similar to the calculated mass of 1424.7611. 1D and 2D NMR analysis of **1** confirmed the amino acid sequence (Figure 7), and together with the molecular formula dictated by the accurate mass, the structure of **1** was determined as a new congener of bacitracin (Table S4, Figures S10-

S16). NMR analysis identified Leu residues in positions 3 and 7, and Ile in positions 5 and 8. The ^{1}H NMR spectra of **1** and **2** were almost identical, suggesting that they are diastereomers (Figure S17). Accordingly, the new compounds were designated as paenitracin A (**1**) and B (**2**).

To verify the identity and stereochemistry of specific residues, compounds 1 and 2 were analyzed using Marfey's method (Table S5, S6) (42). This analysis confirmed the presence of D-Glu, L-Lys, L-Thr, and L-Ile as single isomers. Additionally, it was found that both L and D isomers of Leu were present in the molecules, and it was also observed that the Asn and Asp residues exhibit different stereochemistry. However, the stereochemistry of the thiazoline moiety and Trp could not be determined due to their instability under the conditions used for Marfey's analysis. The 1H NMR spectra of molecules **1** and **2** were nearly identical, with the exception of the region corresponding to the thiazoline moiety, indicating that these molecules are diastereomers (Figure S17) with stereochemical ambiguity arising from the thiazoline portion. Previous studies on bacitracin have demonstrated that racemization at the Cys position within the thiazoline moiety completely abolishes activity, whereas racemization at the Ile position only slightly impacts activity, showing a preference for L-stereochemistry (43). Consequently, we propose that the variation between molecules **1** and **2** is due to the racemization of the terminal Ile residue, with compound **1** exhibiting D-stereochemistry and compound **2** showing L-stereochemistry, as the latter was found to be marginally more potent.

Bacitracin, originally isolated from *Bacillus licheniformis* and *B. subtilis*, targets primarily Gram-positive bacteria such as *Staphylococcus*, *Streptococcus*, *Enterococcus* and *Clostridium difficile* (44). Bacitracin and its congeners are synthesized by a nonribosomal peptide synthetase (NRPS) complex encoded by the bacitracin synthetase

operon, which consists of *bacA*, *bacB*, and *bacC* genes. To explore the biosynthesis of the novel paenitracin, the genome of *Paenibacillus* sp. JJ-778 was sequenced and assembled using Falcon (version 1.8.1) (45). Analysis using antiSMASH 7.1.0 (46) readily identified a 12-modular gene cluster with 77% similarity to the bacitracin BGC of *B. licheniformis* ATCC 10716. In silico analysis of the A domain substrate specificity of BacA, BacB, and BacC predicted variations in amino acid composition at positions 7, 9, and 10 compared to the well-known bacitracin A (Figure 6B).

All known bacitracin congeners isolated and identified so far exhibit similar amino acid compositions, with the Ile residues occasionally replaced by Val (Figure 6B). In contrast, paenitracin contains Leu, Trp and Thr instead of Orn, Phe and His at the amino acid positions 7, 9 and 10, respectively. This makes paenitracin a structurally distant relative of the bacitracins.

Mode of action and bioactivity of paenitracins

Bacitracins specifically target the unique bacterial phospholipid undecaprenyl pyrophosphate (C_{55}-PP). C_{55}PP is the membrane-associated phospholipid that anchors lipid II and shuttles it between the cell's cytoplasm where it is synthesized, and the membrane exterior, where it is incorporated into the growing cell wall (47). To see if paenitracin indeed has the same mode of action as bacitracin, the C_{55}PP dependence of paenitracin activity was tested using antibacterial activity antagonization assays in the presence or absence of zinc (48). For this, paenitracin was mixed with C_{10}-PP, a more soluble analogue of C_{55}-PP, (15) and incubated with cell suspension of *S. aureus* ATCC 29213. Antagonization assays showed that the activity of paenitracin B was antagonized by C_{10}PP exclusively in the presence of Zn^{2+} (Table S7). This indicates that paenitracin

binds to C_{55}-PP of Gram-positive bacteria in a zinc-dependent manner and thus has the same mode of actions as bacitracins.

To obtain insights into the antibacterial spectrum for paenitracin A and B, we performed minimum inhibitory concentration (MIC) assays against a panel of representative pathogenic strains, namely, *E. coli* ATCC 25922, *B. subtilis* 168, *S. aureus* ATCC 29213, *S. aureus* USA300, *E. faecium* E155 and *E. faecium* E980. The media used in the MIC assay was supplemented with 0.3 mM $ZnSO_4$ to provide the divalent metal ion required for bacitracin's activity (43). The activity of the paenitracin diastereomers was compared to that of commercially sourced bacitracin A, which served as a positive control. Paenitracin B exhibited activity against Gram-positive bacteria, with MICs ranging from 1 to 8 µg/mL, comparable to commercial bacitracin (Figure 6C). *B. subtilis* 168, previously reported to be resistant to bacitracin, also exhibited reduced susceptibility to paenitracins A and B, with MIC values of 32 µg/mL. Importantly, paenitracin B exhibited activity against vancomycin-resistant *E. faecium* E155 with MICs of 2 µg/mL, which is lower compared to the MIC of the commercially available bacitracin A of natural origin. Considering the divergent structure of paenitracin as compared to the previously reported congeners, the mode of action should be investigated further.

In conclusion, MS-based indexing of NRPs from 227 plant-associated *Paenibacillus* isolates revealed extensive chemical diversity. Molecular networking combined with MassQL identified several classes of lipopeptides, whereby we specifically focused on NRPs containing basic residues. This revealed the presence of distant relatives of bacitracin, which we designated paenitracins. Paenitracin A and B showed potent bioactivity against Gram-positive ESKAPE pathogens, including vancomycin-resistant *E. faecium* E155. Unlike previously described bacitracin variants,

paenitracins show substantial structural divergence and represent the first example of bacitracin-family NRPs in *Paenibacillus*. These findings highlight the untapped chemical space within this genus and demonstrate the power of combining metabolomic indexing with targeted mining strategies to accelerate the discovery of structurally unique antibiotics.

MATERIALS AND METHODS

General experimental procedures

NMR spectra were recorded on a Bruker Ascend 850 MHz NMR spectrometer (Bruker BioSpin GmbH). Data were analyzed using MestReNova 14 software (Mestrelab Research, Santiago de Compostela, Spain). High-performance liquid chromatography (HPLC) purification was performed on a Waters preparative HPLC system composed of a 1525 pump, a 2707 autosampler, a 2998 PDA detector, and a Water fraction collector III. All solvents and chemicals were of HPLC or LC-MS grade, depending on the experiment.

Isolation and identification of bacterial strains

Paenibacillus strains were obtained from the Auburn University Plant-Associated Microbial strain collection and had previously been isolated from the root surface of field-grown corn (*Zea mays*) and soybean (*Glycine max*) plants grown in Dunbar (Nebraska), Carrol (Iowa), Whitewater (Wisconsin), Sparta (Illinois), Troy (Ohio) and Tallassee (Alabama) (Table S1).

The phylogenetic tree was constructed using 16S rRNA gene sequences obtained through PCR amplification with universal bacterial primers 27F and 1492R (49). Each PCR amplicon was purified and used for Sanger sequencing. Subsequently, 16S rRNA gene sequences were assembled into consensus sequences and aligned with reference sequences. These include 220 sequences from our study, 41 sequences from type strains of *Paenibacillus* spp. downloaded from GenBank (Table S2) and sequence of *Bacillus subtilis* IAM 12118 (GenBank accession NR_112116.2). The alignment is done using ClustalOmega (v 1.2.2) (50). Aligned sequences are trimmed to

the length of 1,452 bp to remove extending regions from a few sequences and also to preserve as many variable regions as possible. The trimmed alignment is re-aligned and a neighbor-joining consensus tree is generated using Geneious (v 2024.0.3) tree function, with Tamura-Nei genetic distance model and *B. subtilis* as outgroup. The generated tree is then uploaded to ITOL for visualization (51).

Growth of *Paenibacillus* spp. and natural product extraction

Our method for cultivation and extraction of specialized metabolites was adapted from previously described procedures (52). Briefly, frozen stocks of 227 *Paenibacillus* spp. from the Auburn University strain collection were inoculated into 1600 µL of Tryptic Soy Agar (TSA, Bacto Soybean-Casein Digest Medium, 30 g/L) in 2.0 mL 96 deep well plates (Thermo Scientific, Nunc 2.0 mL DeepWell Plate). After 72 h, cultures were used to inoculate 200 µL Tryptic Soy Broth (TSB, Bacto Soybean-Casein Digest Medium, 30 g/L) with a pin replicator in 96 well plates. These cultures were incubated overnight at 30 °C and used as an inoculum for the second 2.0 mL 96-deep well plate containing 600 µL TSA. Plates were sealed with 96 Well-Cap Mats (Thermo Scientific, Nunc 96 Well-Cap Mats), and incubated at 30 °C for 72 h. The cultures were extracted with 300 µL 100% isopropanol acidified with 0.1% formic acid. The plates were resealed with the same 96 Well-Cap Mats, sonicated for 10 min, and extracted for an additional 50 min. 200 µL of these crude extracts were transferred into a pre-washed 96 well plate (Agilent Technologies, 96 well plates, 0.5 mL, polypropylene) and lyophilized to dryness. The extraction protocol was repeated once more for a total extract volume of 400 µL. Dried samples were redissolved in 160 µL of methanol:water (1:1 v/v).

Genome sequencing, assembly and annotation

Paenibacillus sp. JJ-778 was grown in TSB at 30 °C and 220 rpm for 24 h. DNA was extracted as described (53). DNA quality was verified by agarose gel electrophoresis. PacBio sequencing and assembly was performed by Novogene (UK). Generally, libraries were prepared using SMRTbell template prep kit (PacBio, USA) according to manufacturer instructions. Sequencing was then performed using PacBio Sequel platform in continuous long reads mode. Assembly was done using Falcon (version 1.8.1) (45). The complete genome sequence of *Paenibacillus* sp. JJ-778 has been deposited in GenBank under the accession number CP194855. BGCs in these genomes were annotated using AntiSMASH (version 7.1.0) (46).

Data-dependent LC-ESI-HRMS/MS

LC-MS/MS acquisition was performed using Shimadzu Nexera X2 ultra-high-performance liquid chromatography (UPLC) system, with an attached photodiode array detector (PDA), coupled to Shimadzu 9030 QTOF mass spectrometer, equipped with a standard electrospray ionization (ESI) source unit. A total of 2 µL was injected into a Waters Acquity HSS C_{18} column (1.8 µm, 100 Å, 2.1 × 100 mm) and data acquisition was performed as previously described (54). Briefly, the gradient used was 5% B for 1 min, 5–85% B for 9 min, 85–100% B for 1 min, and 100% B for 4 min. The column was re-equilibrated to 5% B for 3 min before the next run was started. All the samples were analyzed in positive polarity, using data-dependent acquisition mode. In this regard, full scan MS spectra (*m/z* 100–1700, scan rate 10 Hz, ID enabled) were followed by two data-dependent MS/MS spectra (*m/z* 100–1700, scan rate 10 Hz, ID disabled) for the

two most intense ions per scan. The ions were fragmented using collision-induced dissociation (CID) with fixed collision energy (CE 20 eV) and excluded for 1 s before being re-selected for fragmentation. The parameters used for the ESI source were: interface voltage 4 kV, interface temperature 300 °C, nebulizing gas flow 3 L/min, and drying gas flow 10 L/min.

MZmine 2 parameters

Prior to statistical analysis, mzXML files were imported into Mzmine 2.53 (30) and processed as previously described (54). Briefly, mass ion peaks were detected for MS^1 and MS^2 at a noise level of 2.0E2 and 0.0E0, respectively (positive polarity, mass detector: centroid), and their chromatograms were built using ADAP chromatogram builder. The detected peaks were smoothed, and the chromatograms were deconvoluted (algorithm: local minimum search; chromatographic threshold: 90; search minimum in RT range: 0.05; minimum relative height: 1%; minimum ratio of peak top/edge: 2; peak duration: 0.03–3.00 min). The detected peaks were deisotoped (monotonic shape; maximum charge: 3; representative isotope: most intense). Peak lists from different samples were aligned (weight for RT = weight for *m/z* = 50; compare the isotopic pattern with a minimum score of 50%). Missing peaks detected in at least one of the samples were filled with the gap-filling algorithm (RT tolerance: 0.1 min, *m/z* tolerance was set to 0.002 *m/z* or 15 ppm). Duplicate peaks were filtered. Only the features with MS/MS data were exported to a GNPS-FBMN.

In addition, all features that originate from the culture medium were removed by retaining only features with an average peak intensity of at least 100 times greater in the bacterial extracts than in the culture medium extracts. Features that are present in the

blanks with an intensity higher than 5000 were also removed. In the end, features with *m/z* less than 300 were filtered out.

Molecular network analysis and MassQL search

The resulting feature quantification table (CSV file) and MS/MS spectrum files (in mgf format) were uploaded to the GNPS webserver (http://gnps.ucsd.edu) (19). The LC-MS/MS data for the 227 *Paenibacillus* isolates were analyzed using the Feature-Based Molecular Networking tool (FBMN) (20). Briefly, the precursor ion mass tolerance was set to 0.005 Da and the MS/MS fragment ion tolerance to 0.05 Da. A molecular network was then created where edges were filtered to have a cosine score above 0.5 and more than 3 matched peaks. The molecular networks were visualized using Cytoscape software version 3.9.1 (31) and displayed using an unweighted force-directed layout. The data are publicly accessible in the MassIVE repository (MSV000094386) and the networking results and parameters can be found at https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=2f1e0ae1728142249ac4e841d5a72ef4 . The MassQL tool version 31.4 (33) was employed to search for specific MS/MS fragments that correspond to the *b* or immonium ions of positively charged amino acids. The fragment search of *b* ions of 2,4-diaminobuturic acid (*m/z* 101.0709), ornithine (*m/z* 115.0871), lysine (*m/z* 129.1027), and arginine (*m/z* 157.1089); together with the immonium ion of histidine (*m/z* 110.0718), was conducted with an accuracy of *m/z* 0.005 and minimum intensity of 5%.

Up-scale fermentation, extraction, and isolation

Paenibacillus sp. JJ-778 was grown at 30 °C on tryptic soy agar (TSA) for 72 h and three colonies were inoculated into TSB and incubated at 30 °C overnight. This inoculum (1%) was used to inoculate thirteen 2 L Erlenmeyer flasks containing 0.75 L of sterile MHB and fermented at 30 °C while shaking at 200 rpm for 72 h. To extract the specialized metabolites produced by *Paenibacillus* sp. JJ-778 supernatant was collected by centrifugation and subjected to extraction using Diaion® HP-20 (Resindion, Mitsubishi), followed by elution with 100% IPA supplemented with 0.1% (v/v) formic acid. The crude extract was adsorbed onto silica gel (pore size 60 Å, 70–230 mesh, Sigma Aldrich), and loaded on a silica column, which was eluted using varying gradients of EtOAc/MeOH. The fractions eluted with EtOAc-MeOH (1:3) supplemented with 0.1% (v/v) formic acid were combined, reconstituted in 50% ACN and injected into the preparative SunFire column (19 × 150 mm), which was eluted with H_2O:ACN gradient of 15–60% in 30 min, at a flow rate of 15 mL/min, to yield three fractions. Fraction 3 was further separated on the semi-preparative SunFire column (10 × 250 mm) at 3 mL/min using a H_2O–ACN gradient of 30–50% in 30 min to yield paenitracin A (**1**, 0.3 mg) and paenitracin B (**2**, 0.2 mg).

Marfey`s analysis

The stereochemistry of chiral centers present at α carbons were assigned by applying derivatization methods coupled with chromatographic analysis. The advanced Marfey's method using L-FDAA (1-fluoro-2-4-dinitrophenyl-5-L-alanine amide) established the absolute configurations of amino acids (55). The general method for Marfey`s analysis

was conducted as described) (42). Briefly, a sample of peptide (30 µg) in 6M HCl (100 µL) was heated to 100 °C in a sealed vial for 8−12 h using heating block, after which the hydrolysate was concentrated to dryness at 40 °C under a stream of dry N_2. The hydrolysate was then treated with 1 M $NaHCO_3$ (20 µL) and L-FDAA (1% solution in acetone, 40 µL) at 40 °C for 1 h, after which the reaction was neutralized with 1 M HCl (20 µL). An aliquot of the analyte was diluted 50 times with H_2O/ACN (1:1) and injected (2 µL) into an HRMS instrument following the standard protocol of the analysis. The analyte amino acid content was assessed by comparison to authentic standards. The authentic standards were prepared via a similar procedure.

Antibacterial activity assays and MIC determination

For *E. coli* ATCC 25922 and *S. aureus* ATCC 29213, single colonies were inoculated into 20 mL of MHII broth medium in 250 mL Erlenmeyer flask and incubated overnight at 37 °C with shaking at 220 rpm and then diluted to obtain an assay inoculum of 5 x 10^5 CFU/mL. Aztreonam and vancomycin were used as reference positive controls for each pathogen. Absorbance at OD_{612} nm was measured at T_0 (zero time) using an EnVision™ Microplate Reader (PerkinElmer), and assay plates were incubated at 37 °C. After that, absorbance at OD_{612} nm was measured at T_f (final time) using the same microplate reader equipment. The percentage of growth inhibition was calculated using the following normalization:

$$\%\text{Inhibition} = 100 \times \left\{1 - \frac{([T_{f\text{Sample}} - T_{0\text{Sample}}] - [T_{f\text{Blank}} - T_{0\text{Blank}}])}{([T_{fGrowth} - T_{0\text{Growth}}] - [T_{f\text{Blank}} - T_{0\text{Blank}}])}\right\}$$

$T_{0\text{Sample}}$, the absorbance of the strain growth in the presence of sample measured at zero time;

$T_{f\text{Sample}}$, the absorbance of the strain growth in the presence of sample measured at final time;

$T_{0\text{Growth}}$, the absorbance of the strain growth in the absence of sample measured at zero time;

$T_{f\text{Growth}}$, the absorbance of the strain growth in the absence of sample measured at the final time;

$T_{0\text{Blank}}$, the absorbance of the broth medium (blank) measured at zero time;

$T_{f\text{Blank}}$, the absorbance of the broth medium (blank) measured at the final time.

The Genedata Screener software (Genedata, Inc., Basel, Switzerland) was used to process and analyze the data. The activity of the samples was expressed as a percentage of growth inhibition, where (-100%) represented the total growth inhibition of the target microorganism and 0% represented the total growth of the target microorganism in the assay. The Z' factor predicts the robustness of an assay by considering the mean and standard deviation of both positive and negative controls (56). The robust Z' factor (RZ' factor) is based on the Z' factor, but standard deviations and means are replaced by the robust standard deviations and medians, respectively. In all experiments performed in this work, the RZ' factor obtained was between 0.9173 and 0.9684.

For MIC tests, blood agar plates were inoculated with glycerol stocks of *E. coli* ATCC 25922, *B. subtilis* 168, *S. aureus* ATCC 29213, *S. aureus* USA300 (clinical isolate from Texas Children's Hospital, USA), *E. faecium* E155 and *E. faecium* E980 and incubated overnight at 37 °C. Following incubation, 3 mL of tryptic soy broth (TSB) was inoculated with an individual colony and incubated at 37 °C with shaking at 220 rpm until

the cultures reached the exponential phase. The bacterial suspensions were then diluted 100-fold in cation-adjusted Mueller-Hinton broth (MHB) or Luria-Bertani (LB) broth supplemented with 0.3 M $ZnSO_4$ to reach a bacterial cell density of 10^6 CFU/mL. In parallel, test compounds were 2-fold serially diluted with cation-adjusted MHB or LB supplemented with 0.3 M $ZnSO_4$ in polypropylene 96-well plates to achieve a final volume of 50 µL per well. An equal volume of bacterial suspension (10^6 CFU/mL) was added to the wells. The well plates were sealed with an adhesive membrane and incubated for 20 hours at 37 °C. The reported MIC values are based on three technical triplicates and are defined as the lowest concentration of the compound that prevented the visible growth of bacteria.

In vitro antagonization of paenitracin antibacterial activity

Antibacterial activity antagonization assay was performed as previously described (48). Briefly, geranyl pyrophosphate ammonium salt ($C_{10}PP$) (1 mg/mL in methanol [aqueous 10 mM NH_4OH (7:3), Sigma Aldrich, Taufkirchen, Germany] was used as an antagonist of paenitracin A in either the presence or absence of 0.3 mM $ZnSO_4$. Paenitracin A was mixed with $C_{10}PP$ (five-fold molar excess relative to paenitracin) or added to empty wells (absence of antagonist) giving in all cases 50 µL of total volume in each well with paenitracin A at 4× MIC in the presence and absence of antagonist as well as in the presence and absence of 0.3 mM $ZnSO_4$, in triplicate. Subsequently, 50 µL of *S. aureus* ATCC 29213 with OD_{600} of 0.005 was added to the test compounds in the microtiter plate to achieve a final antibiotic concentration of 2× MIC. The samples were incubated at 37 °C for 24 h with constant shaking and inspected for visible bacterial growth.

525

Acknowledgments

We thank D. van der Horst for technical assistance with LC-MS analysis, to Jesús Martín for assistance in compound annotation and to Marnix Medema for stimulating discussions. The work was supported by the NACTAR program of The Netherlands Organization for Scientific Research (NWO), Grant 16440.

Competing Interests Declaration

The authors declare no conflicts of interests.

REFERENCES

1. Jesudason T. 2024. WHO publishes updated list of bacterial priority pathogens. *The Lancet Microbe* 5.

2. Newman DJ, Cragg GM. 2020. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J Nat Prod* 83:770-803.

3. Medema MH, van Wezel GP. 2025. New solutions for antibiotic discovery: Prioritizing microbial biosynthetic space using ecology and machine learning. *PLOS Biology* 23:e3003058.

4. Aleti G, Sessitsch A, Brader G. 2015. Genome mining: prediction of lipopeptides and polyketides from *Bacillus* and related Firmicutes. *Comput Struct Biotechnol J* 13:192-203.

5. Cochrane SA, Vederas JC. 2016. Lipopeptides from *Bacillus* and *Paenibacillus* spp.: A gold mine of antibiotic candidates. *Med Res Rev* 36:4-31.

6. Olishevska S, Nickzad A, Déziel E. 2019. *Bacillus* and *Paenibacillus* secreted polyketides and peptides involved in controlling human and plant pathogens. *Appl Microbiol Biotechnol* 103:1189-1215.

7. Pilz M, Cavelius P, Qoura F, Awad D, Brück T. 2023. Lipopeptides development in cosmetics and pharmaceutical applications: A comprehensive review. *Biotechnol Adv* 67:108210.

8. Falagas ME, Kasiakou SK. 2005. Colistin: the revival of polymyxins for the management of multidrug-resistant Gram-negative bacterial infections. *Clin Infect Dis* 40:1333-41.

9. Johnson BA, Anker H, Meleney FL. 1945. Bacitracin: a new antibiotic produced by a member of the *B. subtilis gr*oup. *Science* 102:376-377.

10. Li YX, Zhong Z, Zhang WP, Qian PY. 2018. Discovery of cationic nonribosomal peptides as Gram-negative antibiotics through global genome mining. *Nat Commun* 9:3273.

11. Cardoso P, Glossop H, Meikle TG, Aburto Medina A, Conn CE, Sarojini V, Valery C. 2021. Molecular engineering of antimicrobial peptides: microbial targets, peptide motifs and translation opportunities. *Biophys Rev* 13:35-69.

12. Stone KF, Strominger JL. 1971. Mechanism of action of bacitracin: complexation with metal ion and C_{55}-isoprenyl pyrophosphate. *Proc Natl Acad Sci U S A* 68.

13. Siewert G, Strominger JL. 1967. Bacitracin: an inhibitor of the dephosphorylation of lipid pyrophosphate, an intermediate in the biosynthesis of the peptidoglycan of bacterial cell walls. *Proc Natl Acad Sci U S A* 57.

14. Katz BE, Fisher AA. 1987. Bacitracin: a unique topical antibiotic sensitizer. *Journal of the American Academy of Dermatology* 17:1016-1024.

15. Economou NJ, Cocklin S, Loll PJ. 2013. High-resolution crystal structure reveals molecular details of target recognition by bacitracin. *Proc Natl Acad Sci USA* 110.

16. Ikai Y, Oka H, Hayakawa J, Matsumoto M, Saito M, Harada K, Mayumi Y, Suzuki M. 1995. Total structures and antimicrobial activity of bacitracin minor components. *J Antibiot (Tokyo)* 48:233-42.

17. Hubert J, Nuzillard JM, Renault JH. 2017. Dereplication strategies in natural product research: How many tools and methodologies behind the same concept? *Phytochem Rev* 16:55-95.

18. Gaudêncio SP, Bayram E, Lukić Bilela L, Cueto M, Díaz-Marrero AR, Haznedaroglu BZ, Jimenez C, Mandalakis M, Pereira F, Reyes F, Tasdemir D. 2023. Advanced methods for natural products discovery: bioactivity screening, dereplication, metabolomics profiling, genomic sequencing, databases and informatic tools, and structure elucidation. *Mar Drugs* 21:308.

19. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu WT, Crusemann M, Boudreau PD, Esquenazi E, Sandoval-Calderon M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu CC, Floros DJ, Gavilan RG, Kleigrewe K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw CC, Yang YL, Humpf HU, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, et al. 2016. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* 34:828-837.

20. Nothias LF, Petras D, Schmid R, Dührkop K, Rainer J, Sarvepalli A, Protsyuk I, Ernst M, Tsugawa H, Fleischauer M, Aicheler F, Aksenov AA, Alka O, Allard PM, Barsch A, Cachet X, Caraballo-Rodriguez AM, Da Silva RR, Dang T, Garg N, Gauglitz JM, Gurevich A, Isaac G, Jarmusch AK, Kameník Z, Kang KB, Kessler N, Koester I, Korf A, Le Gouellec A, Ludwig M, Martin H C, McCall L-I, McSayles J, Meyer SW, Mohimani H, Morsy M, Moyne O, Neumann S, Neuweger H, Nguyen NH, Nothias-Esposito M, Paolini J, Phelan VV, Pluskal T, Quinn RA, Rogers S, Shrestha B, Tripathi A, van der Hooft JJJ, et al. 2020. Feature-based molecular networking in the GNPS analysis environment. *Nat Microbiol* 17:905-908.

21. van der Hooft JJJ, Wandy J, Young F, Padmanabhan S, Gerasimidis K, Burgess KEV, Barrett MP, Rogers S. 2017. Unsupervised discovery and comparison of structural families across multiple samples in untargeted metabolomics. *Anal Chem* 89:7569-7577.

22. da Silva RR, Wang M, Nothias LF, van der Hooft JJJ, Caraballo-Rodríguez AM, Fox E, Balunas MJ, Klassen JL, Lopes NP, Dorrestein PC. 2018. Propagating annotations of molecular networks using in silico fragmentation. *PLOS Comput Biol* 14:e1006089.

23. Mohimani H, Gurevich A, Mikheenko A, Garg N, Nothias LF, Ninomiya A, Takada K, Dorrestein PC, Pevzner PA. 2017. Dereplication of peptidic natural products through database search of mass spectra. *Nat Chem Biol* 13:30-37.

24. de Jonge NF, Mildau K, Meijer D, Louwen JJR, Bueschl C, Huber F, van der Hooft JJJ. 2022. Good practices and recommendations for using and benchmarking computational metabolomics metabolite annotation tools. *Metabolomics* 18:103.

25. Damiani T, Jarmusch AK, Aron AT, Petras D, Phelan VV, Zhao HN, Bittremieux W, Acharya DD, Ahmed MMA, Bauermeister A, Bertin MJ, Boudreau PD, Borges RM, Bowen BP, Brown CJ, Chagas FO, Clevenger KD, Correia MSP, Crandall WJ, Crüsemann M, Fahy E, Fiehn O, Garg N, Gerwick WH, Gilbert JR, Globisch D, Gomes PWP, Heuckeroth S, James CA, Jarmusch SA, Kakhkhorov SA, Kang KB, Kessler N, Kersten RD, Kim H, Kirk RD, Kohlbacher O, Kontou EE, Liu K, Lizama-Chamu I, Luu GT, Luzzatto Knaan T, Mannochio-Russo H, Marty MT, Matsuzawa Y, McAvoy AC, McCall L-I, Mohamed OG, Nahor O, Neuweger H, et al. 2025. A universal language for finding mass spectrometry data patterns. *Nature Methods* 22:1247-1254.

26. Olivon F, Allard PM, Koval A, Righi D, Genta-Jouve G, Neyts J, Apel C, Pannecouque C, Nothias LF, Cachet X, Marcourt L, Roussi F, Katanaev VL, Touboul D, Wolfender JL, Litaudon M. 2017. Bioactive natural products prioritization using massive multi-informational molecular networks. *ACS Chem Biol* 12:2644-2651.

27. Nothias LF, Nothias-Esposito M, da Silva R, Wang M, Protsyuk I, Zhang Z, Sarvepalli A, Leyssen P, Touboul D, Costa J, Paolini J, Alexandrov T, Litaudon M, Dorrestein PC. 2018. Bioactivity-based molecular networking for the discovery

of drug leads in natural product bioassay-guided fractionation. *J Nat Prod* 81:758-767.

28. Olivon F, Retailleau P, Desrat S, Touboul D, Roussi F, Apel C, Litaudon M. 2020. Isolation of picrotoxanes from austrobuxus carunculatus using taxonomy-based molecular networking. *J Nat Prod* 83:3069-3079.

29. Crüsemann M, O'Neill EC, Larson CB, Melnik AV, Floros DJ, da Silva RR, Jensen PR, Dorrestein PC, Moore BS. 2017. Prioritizing natural product diversity in a collection of 146 bacterial strains based on growth and extraction protocols. *J Nat Prod* 80:588-597.

30. Pluskal T, Castillo S, Villar-Briones A, Oresic M. 2010. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* 11:395.

31. Shannon P, Markiel A, Ozier O, Baliga N, Wang J, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498-2504.

32. Straus SK, Hancock REW. 2006. Mode of action of the new antibiotic for Gram-positive pathogens daptomycin: Comparison with cationic antimicrobial peptides and lipopeptides. *Biochim Biophys Acta* 1758:1215-1223.

33. Jarmusch A, Aron A, Petras D, Phelan V, Bittremieux W, Acharya D, Ahmed M, Bauermeister A, Bertin M, Boudreau P, Borges R, Bowen B, Brown C, Chagas F, Clevenger K, Correia M, Crandall W, Crüsemann M, Damiani T, Fiehn O, Garg N,

Gerwick W, Gilbert J, Globisch D, Gomes PW, Heuckeroth S, James A, Jarmusch S, Kakhkhorov S, Kang KB, Kersten R, Kim H, Kirk R, Kohlbacher O, Kontou E, Liu K, Lizama-Chamu I, Luu G, Knaan TL, Marty M, McAvoy A, McCall L-I, Mohamed O, Nahor O, Niedermeyer THJ, Northen T, Overdahl K, Pluskal T, Rainer J, Reher R, et al. 2022. A universal language for finding mass spectrometry data patterns doi:10.1101/2022.08.06.503000. bioRxiv.

34. Vogler K, Studer RO. 1966. The chemistry of the polymyxin antibiotics. *Experientia* 22:345-354.

35. Jian Li, Roger L Nation, John D Turnidge, Robert W Milne, Kingsley Coulthard, Craig R Rayner, Paterson DL. 2006. Colistin: the re-emerging antibiotic for multidrug-resistant Gram-negative bacterial infections. *Lancet Infect Dis* 6:589-601.

36. Bann SJ, Ballantine RD, Cochrane SA. 2021. The tridecaptins: non-ribosomal peptides that selectively target Gram-negative bacteria. *RSC Med Chem* 12:538-551.

37. Knolhoff AM, Zheng J, McFarland MA, Luo Y, Callahan JH, Brown EW, Croley TR. 2015. Identification and structural characterization of naturally-occurring broad-spectrum cyclic antibiotics isolated from *Paenibacillus*. *J Am Soc Mass Spectrom* 26:1768-1779.

38. Machushynets NV, Lysenko V, Du C, Slingerland CJ, Elsayed SS, Liles MR, Martin NI, van Wezel GP. 2025. Exploring the chemical space of *Paenibacillus* NRPs and discovery of paenilipoheptin B. *Organic Letters* 27:2821-2825.

39. Lysenko V, Machushynets NV, van Dam JL, Sterk FAC, Speer A, Ram AFJ, Slingerland CJ, Van Wezel GP, Martin NI. 2025. Total synthesis, structure elucidation, and bioactivity evaluation of the cyclic lipopeptide natural product paenilipoheptin A. *Organic Letters* 27:2826-2831.

40. Qin F, Zhang H, Liu H, Zhang X. 2020. High performance liquid chromatography-quadrupole/Time of Flight–tandem mass spectrometry for the characterization of components in bacitracin. *Chromatographia* 83.

41. Suleiman SA, Song F, Su M, Hang T, Song M. 2017. Analysis of bacitracin and its related substances by liquid chromatography tandem mass spectrometry. *J Pharm Anal* 7:48-55.

42. Lysenko V, Son S, Theriault ME, Slingerland CJ, Hauk G, Cleenewerk L, Speer A, Berger JM, Lewis K, Martin NI. 2024. Total synthesis and structural reassignment of the antitubercular natural product evybactin. *Chem Eur J* doi:10.1002/chem.202403767:e202403767.

43. Buijs N, Vlaming HC, van Haren MJ, Martin NI. 2022. Synthetic studies with bacitracin A and preparation of analogues containing alternative zinc binding groups. *ChemBioChem* 23:e202200547.

44. André B. 2005. Antimicrobial agents: antibacterials and antifungals. ASM Press, Washington, D.C.

45. Chin CAO, Peluso P, Sedlazeck FA-O, Nattestad M, Concepcion GT, Clum A, Dunn CA-O, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, Cramer GA-O, Delledonne M, Luo C, Ecker JR, Cantu DA-O, Rank DR, Schatz MC. 2016. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods* 13:1050–1054.

46. Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M, Fetter A, Terlouw BR, Metcalf WW, Helfrich EJN, van Wezel GP, Medema MH, Weber T. 2023. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res* 51:W46-W50.

47. Storm DR. 1974. Mechanism of bacitracin action: a specific lipid-peptide interaction. *Annals of the New York Academy of Sciences* 235:387-398.

48. Buijs NP, Vlaming HC, Kotsogianni I, Arts M, Willemse J, Duan Y, Alexander FM, Cochrane SA, Schneider T, Martin NI. 2024. A classic antibiotic reimagined: Rationally designed bacitracin variants exhibit potent activity against vancomycin-resistant pathogens. *Proceedings of the National Academy of Sciences* 121:e2315310121.

49. Lane D. 1991. 16S/23S rRNA sequencing. Wiley.

50. Sievers F, Higgins DG. 2018. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci* 27:135-145.

51. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 49:W293-W296.

52. Nguyen DD, Melnik AV, Koyama N, Lu X, Schorn M, Fang J, Aguinaldo K, Lincecum TL, Ghequire MGK, Carrion VJ, Cheng TL, Duggan BM, Malone JG, Mauchline TH, Sanchez LM, Kilpatrick AM, Raaijmakers JM, De Mot R, Moore BS, Medema MH, Dorrestein PC. 2016. Indexing the *Pseudomonas* specialized metabolome enabled the discovery of poaeamide B and the bananamides. *Nat Microbiol* 2:16197.

53. Kieser T, Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. 2000. Practical *Streptomyces* genetics. John Innes Foundation.

54. van Bergeijk DA, Elsayed SS, Du C, Santiago IN, Roseboom AM, Zhang L, Carrión VJ, Spaink HP, van Wezel GP. 2022. The ubiquitous catechol moiety elicits siderophore and angucycline production in *Streptomyces*. *Commun Chem* 5.

55. Marfey P. 1984. Determination of D-amino acids. II. Use of a bifunctional reagent, 1, 5-difluoro-2, 4-dinitrobenzene. *Carlsberg Res Commun* 49:591-596.

56. Zhang J-H, Chung TDY, Oldenburg KR. 1999. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J Biomol Screen* 4:67-73.

FIGURES

Figure 1. Phylogenetic tree based on 16S rRNA sequences of plant-associated *Paenibacillus* isolates and closely related type strains. The tree was obtained by neighbor-joining analysis. For an overview of the strains see Tables S1 and S2. Type strains are indicated with green squares. The tree was rooted using *Bacillus subtilis* IAM 12118 as an outgroup. Antibacterial activity of the strains against *E. coli* ATCC 25922 and *S. aureus* ATCC 29213 is indicated with purple and red squares, respectively. The tree was generated using Geneious v2024.0.3.

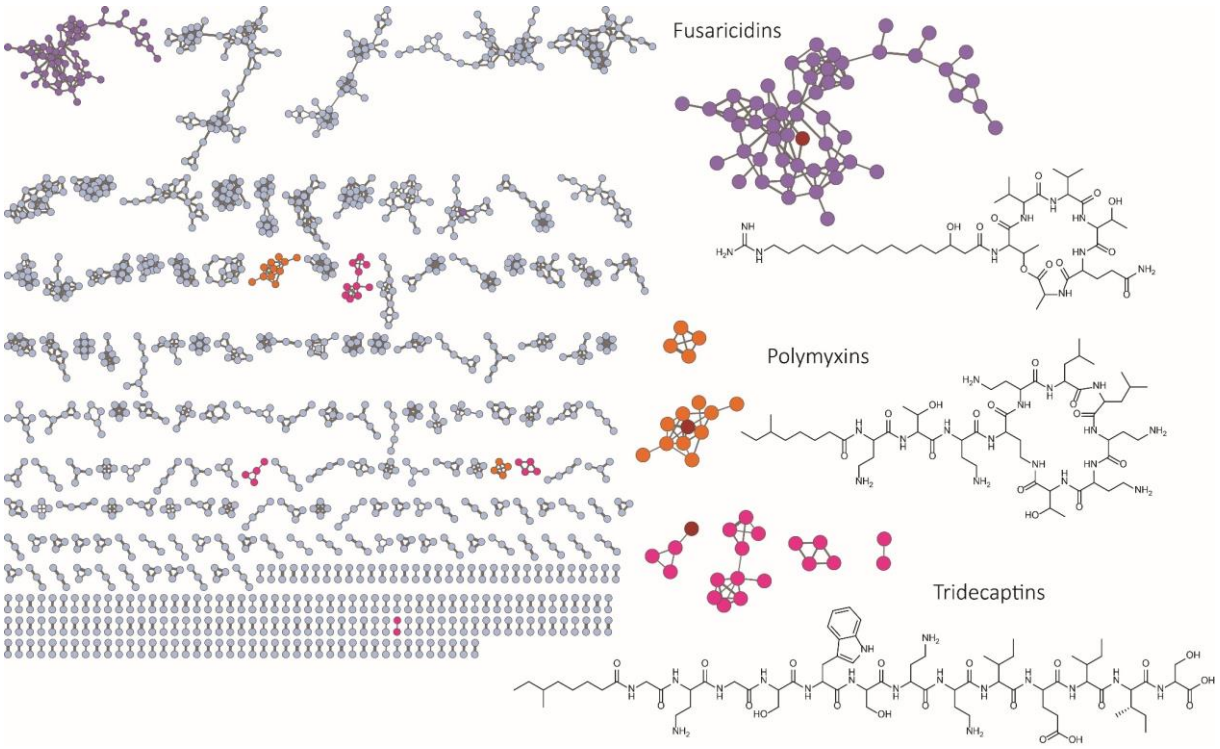

Figure 2. Molecular network of mass features detected in the bacterial extracts from the *Paenibacillus* isolates. Each node represents a parent ion detected in one or more strains. Different node colors represent distinct molecular families. The red nodes highlight the specific annotated compound in each family, with the corresponding chemical structure displayed next to the network.

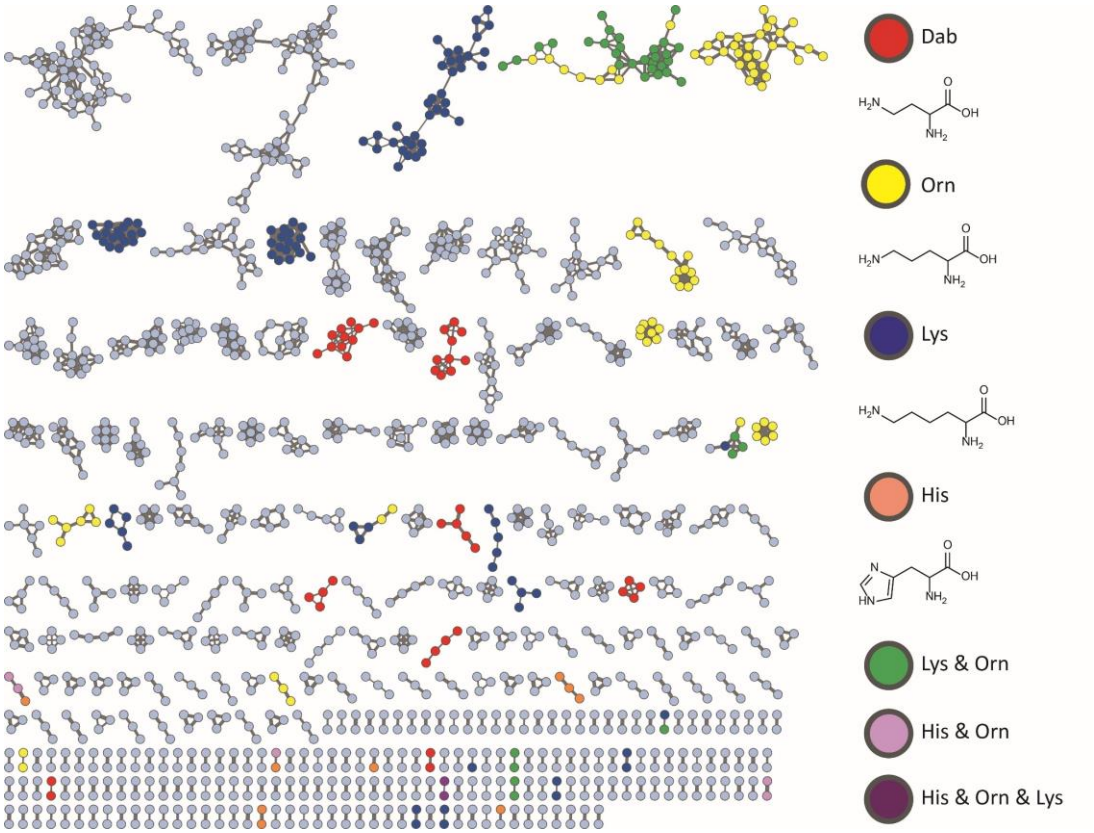

Figure 3. MassQL-annotated molecular network of the ions detected in the bacterial extracts from *Paenibacillus* isolates. Colors of the nodes, as indicated in the figure, highlight parent ions that contain specific basic amino acid residues (Dab, Orn, His, Lys or combinations thereof) in the structures.

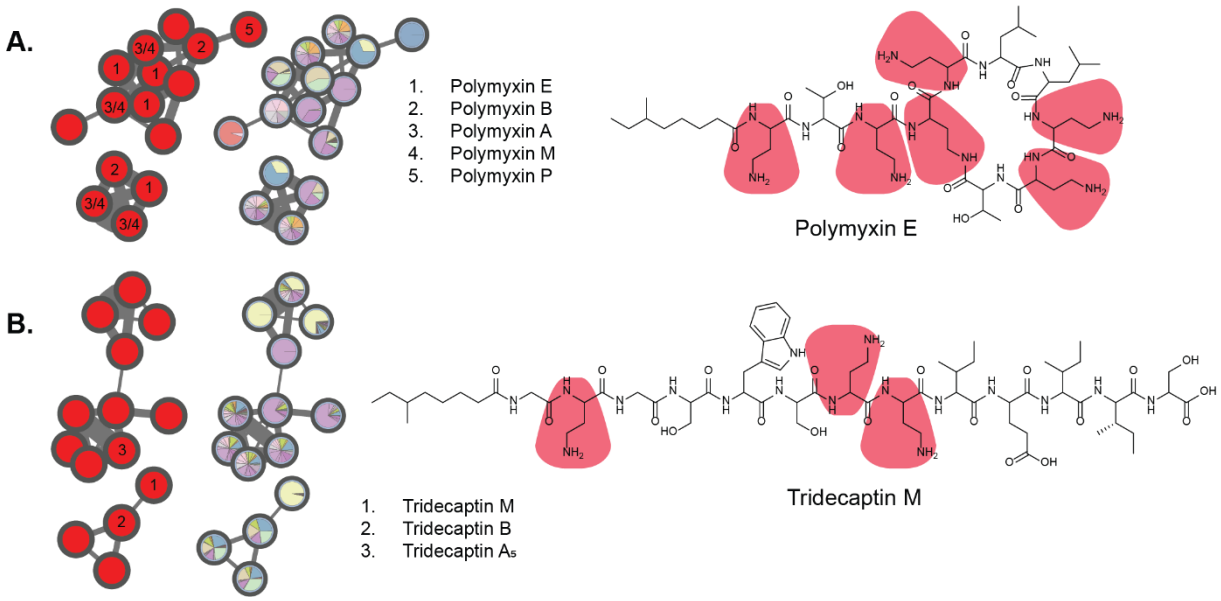

Figure 4. Dab-containing lipopeptide families highlighted by MassQL search. Molecular families of polymyxins (A) and tridecaptins (B) consist of multiple structural analogues, reflecting high chemical diversity. Pie charts were mapped to the nodes to represent the relative precursor ion intensities within each extract. Note that both polymyxins and tridecaptins contain multiple Dab residues.

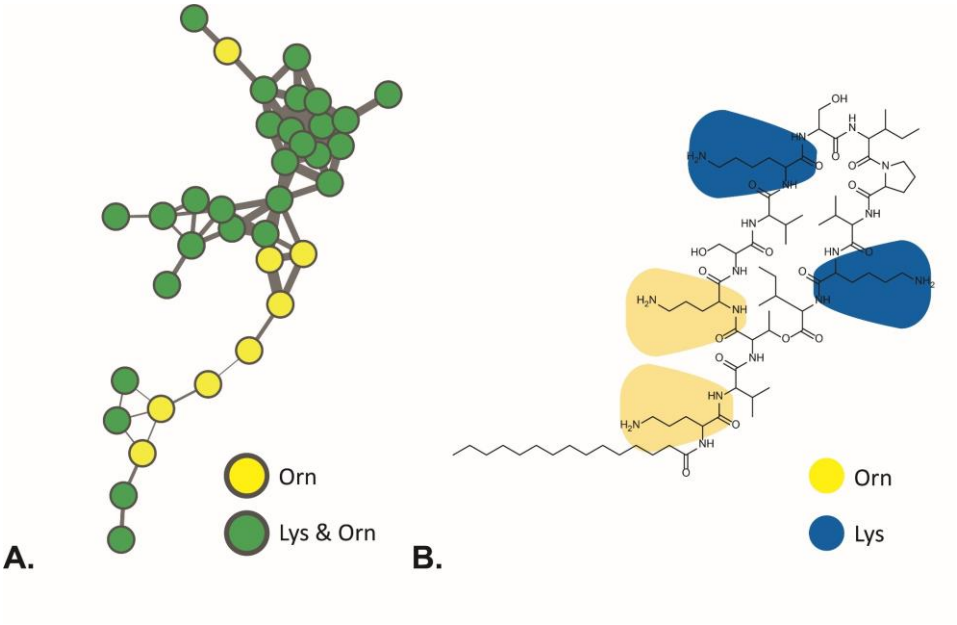

Figure 5. MassQL-based detection of paenibacterins. A. MassQL-annotated molecular family of paenibacterins. **B.** Chemical structure of paenibacterin A.

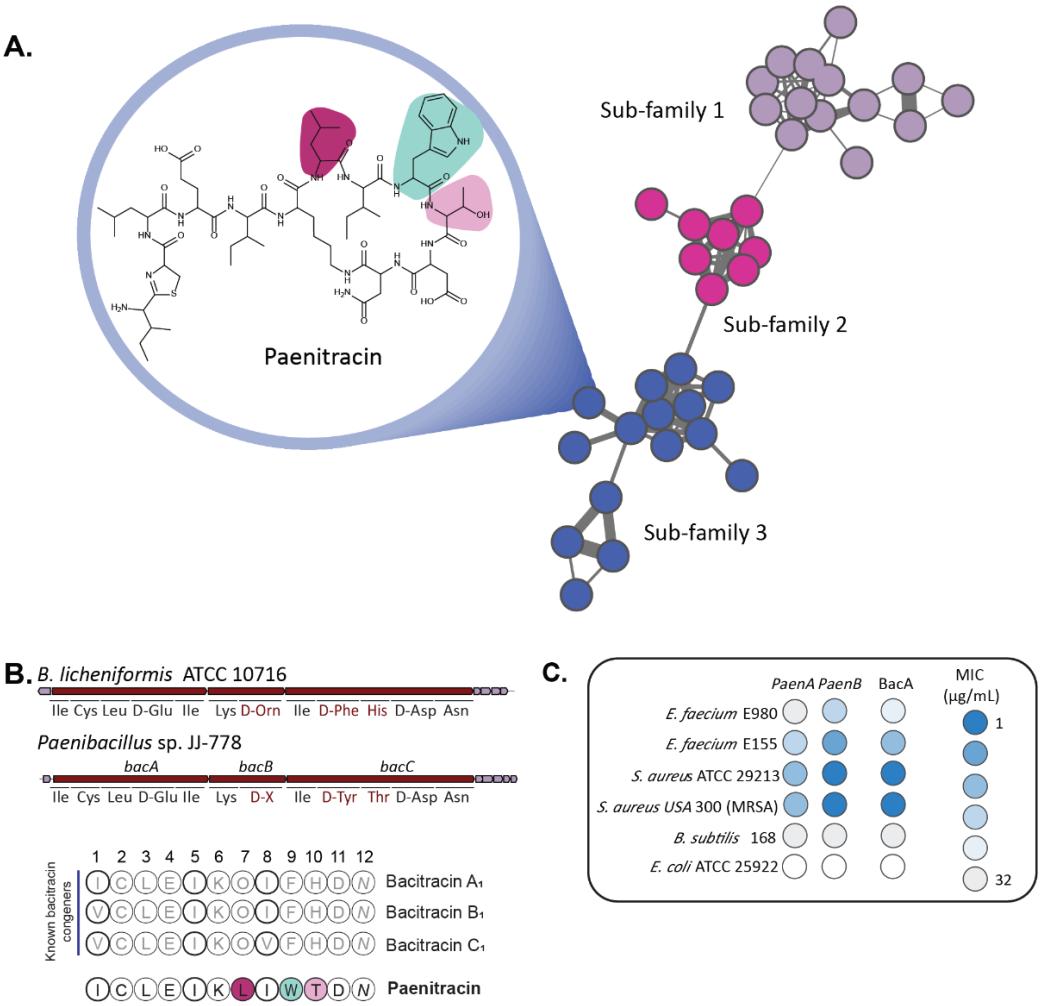

Figure 6. Molecular family of Lys-containing NRPs highlighted by MassQL search. **A.** Exploring the mass features from sub-family 3 led to the discovery of new bacitracin analog, designated as paenitracin. **B.** The BGC and its predicted amino acid substrates of new paenitracin from the genome of *Paenibacillus* sp. JJ-778 was compared to the reference BGC from *B. licheniformis* ATCC 10716. The amino acid sequence of paenitracin was compared to those of known bacitracins from *B. licheniformis*. Colored amino acids are unique to paenitracin. **C.** MIC data of paenitracin A and B and bacitracin against a panel of human pathogenic bacteria.

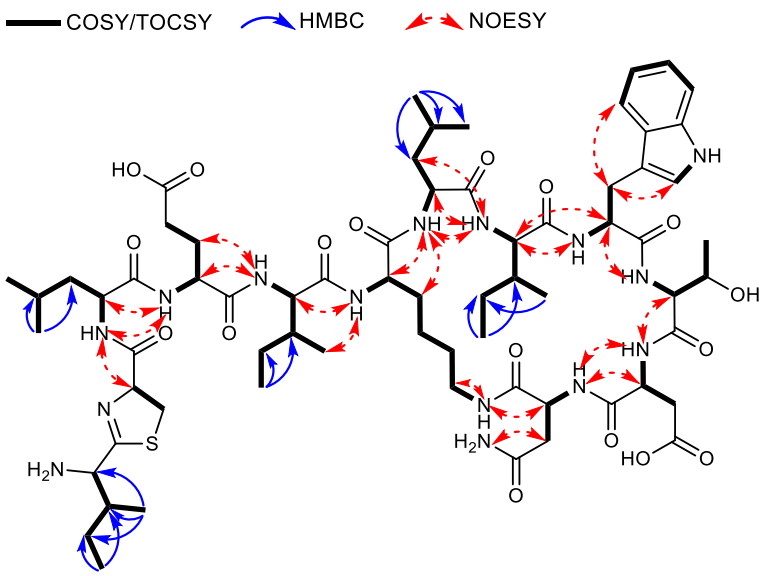

Figure 7. Correlations obtained by COSY, HMBC, and NOESY measurements in NMR of paenitracin A (1).