Natural product antibiotics: synthesis and next generation analogues

Lysenko, V.

2A

Chapter 2A

Exploring the Chemical Space of *Paenibacillus* NRPs and Discovery of Paenilipoheptin B

Abstract

A combination of genomic and metabolomic analyses paired with molecular networking was applied to a collection of *Paenibacillus* spp. to identify the producers of a little-studied class of lipopeptides known as paenilipoheptins. Mass spectrometry and NMR spectroscopy allowed revision of the structure of previously reported paenilipoheptin A and elucidation of the structure of novel paenilipoheptin B.

Parts of this chapter have been published in:

Machushynets, N. V.*; Lysenko, V.*; Du, C.; Slingerland, C. J.; Elsayed, S. S.; Liles, M. R.; Martin, N. I.; van Wezel, G. P. Exploring the chemical space of *Paenibacillus* NRPs: discovery of paenilipoheptin B. *Org. Lett.* **2025**, *27* (12), 2821–2825.

* - these authors contributed equally.

Introduction

The rise in antimicrobial resistance (AMR) has led to an increased interest in novel antibiotics. An important class of natural products with therapeutic potential produced by bacteria is nonribosomal peptides (NRPs).[1] NRPs currently used in the clinic, such as bacitracin, daptomycin, polymyxin, and vancomycin, constitute effective treatments for infections caused by multidrug-resistant pathogens, though for these compounds, AMR also becomes an issue.[2,3] A wide variety of bacteria produce NRPs,[4] whereby *Paenibacillus* spp. have yielded several potent antimicrobial lipopeptides, such as polymyxins, tridecaptins, paenibacterins, octapeptins, and pelgipeptins, among others.[5,6] We previously bioinformatically analyzed 785 complete genomes from *Paenibacillus* spp. to identify biosynthetic gene clusters (BGCs) that encode the biosynthesis of nonribosomal peptide synthetases (NRPSs).[7] NRPSs are large, multifunctional enzymes that have modular structures, with each NRPS module catalyzing the incorporation of a specific substrate into the growing peptide.[8,9] A typical module consists of three enzymatic domains, namely, adenylation (A), thiolation (T), condensation (C), and epimerization (E) domains and the terminal thioesterase (TE).[10] The collinearity rule of NRPS systems, combined with knowledge of the specificity-conferring code of the A domain, allows for the prediction of the peptide structures synthesized by the corresponding NRPS.

To visualize the diversity, distribution, and NRPS novelty, a sequence similarity network (SSN) was constructed using BiG-SCAPE.[11] Besides known classes of BGCs for among others polymyxins, tridecaptins, fusaricidins, paenibacterins, octapeptins, bacitracins, and cilagicins,[7] the analysis also identified BGCs for unknown or partially characterized classes of NRPs. The paenilipoheptins are notable examples (**Figure 1A**). While paenilipoheptin A was detected in the extracts of *P. polymyxa* E681 using LC-MS, the compound itself had not been isolated, and thus, the specific bioactivity was also unknown.[12]

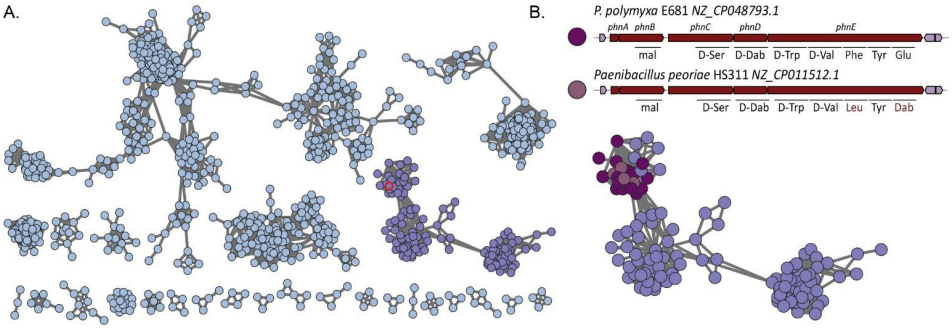

Figure 1. Sequence similarity network of *Paenibacillus* NRPS BGCs and predicted paenilipoheptin BGCs. (A) BiG-SCAPE SSN (c 0.25) containing validated NRPS BGCs of *Paenibacillus* spp. visualized in Cytoscape. Each node represents one NRPS BGC predicted by antiSMASH. Singletons and single links are not shown. The node corresponding to the paenilipoheptin BGC from *P. polymyxa* E681 is emphasized in red. (B) Expanded view of gene cluster families (GCFs) predicted to specify paenilipoheptin-like compounds. The BGCs and predicted amino acid sequences for paenilipoheptin A and the newly identified paenilipoheptin B are shown above the network. Notable differences in the NRPS assembly lines of paenilipoheptin B compared to paenilipoheptin A are highlighted in modules 6 and 8, shown in red.

Results and discussion

The paenilipoheptin BGC of *P. polymyxa* E681 encodes a hybrid NRPS-Trans-AT-PKS that produces paenilipoheptin A.[12] The entire paenilipoheptin assembly line involves three peptide synthetases, PhnC, PhnD, and PhnE (**Figure 1B**). The one-module enzymes PhnC and PhnD were predicted to mediate the incorporation of amino acids Ser[1] and Dab[2], respectively. PhnE consists of five modules with predicted substrate specificity of A domains for the amino acids Trp[3], Val[4], Phe[5], Tyr[6], and Glu[7]. Modules 2–5 contain epimerization domains, indicating that Ser[1], Dab[2], Trp[3], and Val[4] may be converted to the D-configuration. The SSN highlighted numerous BGCs with architectures similar to the paenilipoheptin BGCs found in the genome of *P. polymyxa* E681. To predict the amino acid sequences of paenilipoheptins encoded by these BGCs, an in silico analysis of A-domain substrate specificities was conducted using antiSMASH 7.1.0.[13] Based on the predicted amino acid sequences, we classified the detected paenilipoheptin-like BGCs into two groups.

The first group consisted of paenilipoheptin BGCs with predicted amino acid sequences identical with those of paenilipoheptin A from *P. polymyxa* E681, with the predicted sequence D-Ser–D-Dab–D-Trp–D-Val–L-Phe–L-Tyr–L-Glu (**Figure 1B**). The second group included BGCs with variations at amino acid positions 5 and 7, resulting in the predicted sequence D-Ser–D-Dab–D-Trp–D-Val–L-Leu–L-Tyr–L-Dab (**Figure 1B**). This variation suggests the production of a new analogue, designated paenilipoheptin B, thereby expanding the diversity of known paenilipoheptins. Sequence similarity network indicates that paenilipoheptins were mainly produced by species of *P. polymyxa*, *P. peoriae*, and *P. jamilae*. Therefore, a subset of strains from the Auburn University (USA) Plant-Associated Microbial strain collection, showing high 16S rRNA gene sequence

similarity (greater than 99%) with these species, was selected to identify potential paenilipoheptin producers. Isolates were cultured in tryptic soy agar (TSA) media in 96 deep-well plates for 72 h and extracted with isopropyl alcohol supplemented with 0.1% (v/v) formic acid. The extracts were then subjected to LC-MS/MS analysis to detect the produced secondary metabolites. LC-MS/MS data were processed with MzMine 2[14] and exported for global natural product social (GNPS) feature-based molecular networking (FBMN).[15] FBMN analysis resulted in a molecular network consisting of 514 parent ions (nodes) connected through 738 edges (**Figure 2A**). Characteristics of the natural products, such as annotation, *m/z* value, and species-specific molecule production were visualized using Cytoscape 3.9.1.[16] Mass spectrometry-based molecular networking allows clustering of molecules with similar MS/MS fragmentation patterns, which stem from the similarity in their structures.[17] We focused on a molecular family of compounds with a node of *m/z* 562.3133, which corresponds to the exact mass of the previously reported paenilipoheptin A (**Figure 2A**).[12]

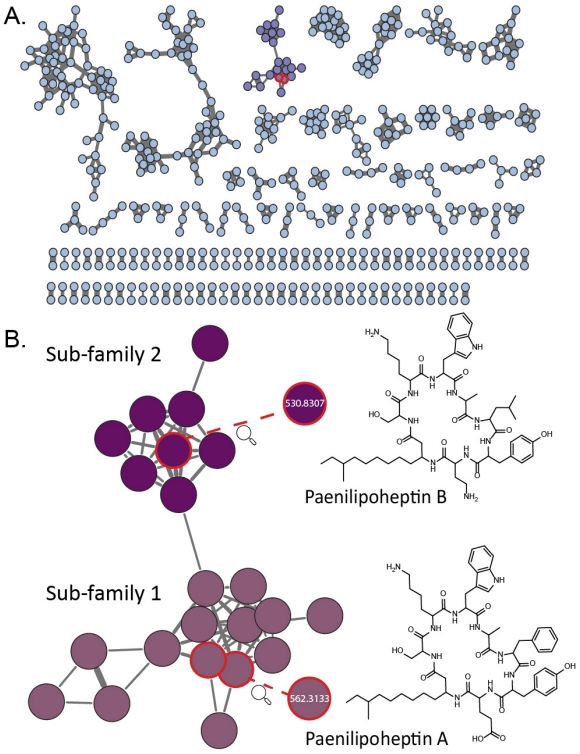

Figure 2. Molecular network of the mass features detected in the bacterial extracts from a subset of 25 plant-associated *Paenibacillus* isolates. The highlighted spectral family represents paenilipoheptin-like compounds (A). Subfamily 1 contains nodes that represent ions with masses corresponding to paenilipoheptin A (nodes highlighted with red circles). Exploring the mass features from subfamilies 1 and 2 led to the discovery of paenilipoheptin A and B, respectively (B).

Subfamily 1 represents doubly charged ions with a mass range of 1080 to 1150 Da and contains 2 nodes of *m/z* 562.3133 (**Figure 2B**). However, the MS/MS fragmentation pattern of both of these nodes indicated differences in the amino acid sequences as compared to that reported for paenilipoheptin A.[12] The differences lie in residues 2 and 4, which were previously annotated as Dab and Val, while our data suggested that they are Lys and Ala, respectively. The latter were confirmed by high-resolution MS/MS spectroscopy, which indeed revealed an ion corresponding to a Lys residue (*m/z* 129.1021), together with the fragment ions Ala–Trp (*m/z* 258.1240) and Ala–Phe (*m/z* 219.1122) (**Figure S1**). Further analysis of the *b* ions of the parent mass with *m/z* 562.3133 corroborated the lipopeptide sequence as FAA(C13H26)–Ser–Lys–Trp–Ala–Phe–Tyr–Glu (**Figure S1**). In the original study, low-resolution MS was used to assign the structure of paenilipoheptin A,[12] which might explain the discrepancy. To further verify that *Paenibacillus* sp. JJ-21 indeed produces paenilipoheptin A, we compared the compound to the one produced by P. *polymyxa* E681. For this, we cultured *P. polymyxa* E681 on TSA agar, employing the original extraction methods.[12] Subsequent LC-MS/MS analysis of the crude extract from *P. polymyxa* E681 identified a paenilipoheptin with an *m/z* of 562.3142, which is identical to that of the paenilipoheptin produced by Paenibacillus sp. JJ-21. Additionally, comparative mirror plot analyses of the MS/MS fragmentation patterns from both strains confirmed that both of them produce paenilipoheptin A, with the sequence FAA(C13H26)–Ser–Lys–Trp–Ala–Phe–Tyr–Glu (**Figure S2**).

The node with *m/z* 555.305 in subfamily 1 of the Lys-containing molecular family was linked to subfamily 2 via the node with *m/z* 523.8231. A mirror plot of the MS/MS spectra revealed that the two mass features shared numerous peaks in the low mass region (**Figure S3**). The common fragment ions correspond to the amino acids Lys, Tyr, and Trp, together with the dipeptide fragments Ser–Lys, Lys–Trp, and Trp–Ala. Interestingly, in addition to Lys, a fragment ion for Dab was detected in the MS/MS spectrum of the mass feature with *m/z* 523.8231. At the same time, no fragment ions for Glu or Phe were detected. These observations suggested that compounds from the second subfamily contained two positively charged amino acids, namely, Lys and Dab, and might be the products of the paenilipoheptin B BGC (**Figure 1B**). Upon further investigation of the *b* ions of the parent mass with *m/z* = 530.8307, we conclude that the lipopeptide sequence is FAA(C13H26)–Ser–Lys–Trp–Ala–Leu–Tyr–Dab (**Figure S4**). *Paenibacillus* sp. JJ-21 and *Paenibacillus* sp. JJ-1722 stood out as they were found to produce substantial amounts of paenilipoheptins, and these strains were therefore chosen for larger scale fermentation in search of novel paenilipoheptin congeners.

Paenibacillus sp. JJ-21 and *Paenibacillus* sp. JJ-1722 were grown in 10 L Mueller-Hinton Broth (MHB) medium, and the biomass was collected by centrifugation. Specialized metabolites were extracted from the cells with isopropyl alcohol supplemented with 0.1% (v/v) formic acid. The most abundant compounds, with *m/z* values of 562.3133 and 530.8307, were isolated from the extracts of *Paenibacillus* sp. JJ-21 and JJ-1722 through

multiple rounds of HPLC and were designated as paenilipoheptin A (**1**) and B (**2**), respectively. Compound **1** showed a molecular ion peak for an $[M + H]^+$ ion having *m/z* 1123.6180 (calcd. for $C_{59}H_{83}N_{10}O_{12}$, 1123.6192) in the ESI-HRMS spectrum. Analysis of the 1D and 2D NMR spectra of compound **1** indicated the presence of seven amino acids: Ser, Lys, Trp, Ala, Phe, Tyr, Glu, and the β-amino fatty acyl chain (**Table S1**, **Figure 3A**).

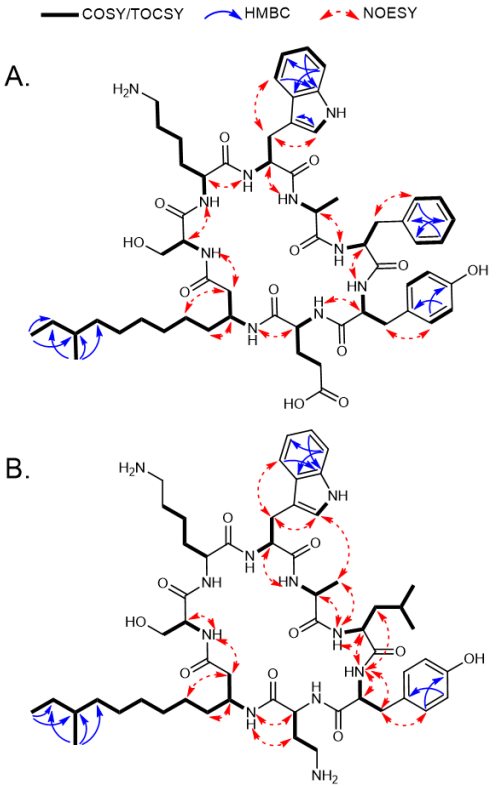

Figure 3. Correlations obtained by COSY, HMBC, and NOESY measurements in NMR of **1** (A) and **2** (B).

The fatty β-amino acid was identified as 3-amino-10-methyldodecanoic acid upon detailed NMR analysis. The NMR analysis thus confirmed that paenilipoheptin A contains Ala instead of Val at position 4 and Lys instead of Dab at position 2. For compound **2**, ESI-HRMS analysis revealed a molecular ion $[M + H]^+$ with an *m/z* of 1060.6555 (calcd. for $C_{55}H_{86}N_{11}O_{10}$, 1060.6559). The 1H NMR spectra of **1** and **2** were similar, indicating similar structures. Due to the low yield of **2**, the NMR signals were quite weak. Nevertheless, we could confirm the presence of Leu and Dab at positions 5 and 7, respectively, which differs from compound **1** that contains Phe and Glu residues, respectively, at these positions (**Table S2**, **Figure 3B**).

The absolute configurations of the amino acids in compounds **1** and **2** were determined by Marfey's analysis.[18] This revealed Ser[1], Phe[5], and Tyr[6] to be L-amino acids, whereas Lys[2], Ala[4], and Glu[7] were found to be D-amino acids (**Tables S3** and **S4**). The

stereochemistry of Trp3 could not be established due to degradation of the Trp residue under the conditions used for generating the Marfey's derivatives.

To characterize the paenilipoheptin BGCs, the genomes of *Paenibacillus* sp. JJ-21 and *Paenibacillus* sp. JJ-1722 were sequenced using the PacBio platform. Assembly of the PacBio reads with Falcon (version 1.8.1)[19] resulted in single contigs of 6.2 and 6.1 Mb for *Paenibacillus* sp. JJ-21 (GenBank accession number: CP132974) and *Paenibacillus* sp. JJ-1722 (GenBank accession number: CP182500), respectively. An in silico analysis of the A domain substrate specificity and stereochemistry predictions for these BGCs was conducted with antiSMASH 7.1.0.[13] The predicted amino acid composition of paenilipoheptin A of *Paenibacillus* sp. JJ-21 was identical with that of *P. polymyxa* E681, which was D-Ser–D-Dab–D-Trp–D-Val–L-Phe–L-Tyr–L-Glu (**Table S5**). The amino acid sequence predicted for paenilipoheptin B differed from paenilipoheptin A at positions 5 and 7, and was D-Ser–D-Dab–D-Trp–D-Val–L-Leu–L-Tyr–L-Dab.

Marfey's analysis, MS/MS, and NMR studies revealed structural discrepancies with the previously predicted primary sequences for paenilipoheptin A (**Tables S3** and **S4**). Specifically, Dab was predicted at position 2, but Lys was identified in the actual product. Similarly, Val was predicted at position 4, while in fact it is an Ala residue. Discrepancies between predicted and actual structures are likely due to database limitations in the bioinformatic tools used in making the previous structure predictions.[13] Notably, the genomic predictions also differed from the actual paenilipoheptin structures in stereochemistry. For instance, in both paenilipoheptins A and B, L-Ser was found at position 1 instead of the previously predicted D-Ser. Amino acid sequences of E domains from modules 2–5 of paenilipoheptin BGCs of *Paenibacillus* sp. JJ-21 and *Paenibacillus* sp. JJ-1722 were also aligned to compare their active site motifs (HHxxxD).[20] This revealed that the conserved active site motif HHxxxD of Ser epimerization domains is replaced with DPxxxD in both strains (**Figure S5**). Therefore, although epimerization (E) domains were detected in module 2 of the NRPSs encoded by these Paenibacillus genomes, we hypothesize that they are nonfunctional. Furthermore, D-Glu was detected at position 7 of paenilipoheptin A, rather than the predicted L-Glu, suggesting alternative mechanisms such as noncanonical epimerization or the involvement of an external enzyme. Additional studies are required to confirm these hypotheses.

Both paenilipoheptin A and B inhibited the growth of *Bacillus subtilis* 168 with moderate bioactivity (minimal inhibitory concentration of 16–32 μg/mL), while no bioactivity against *Escherichia coli* ATCC 25922 was observed. A detailed investigation of the structure and antimicrobial activity of paenilipoheptin A is presented in **Chapter 2B** and a corresponding publication.[21]

Conclusions

In conclusion, genomic analysis and mass spectral networking revealed the high biosynthetic potential of *Paenibacillus* spp. as producers of NRPs with diverse chemistry.

Bioinformatic analysis using BIG-SCAPE revealed that the paenilipoheptin GCFs contained BGCs associated not only with the previously described paenilipoheptin A but also with a new analogue here assigned as paenilipoheptin B. GNPS networking allowed us to identify the producers of these compounds for further purification and structure elucidation. This led to the structural revision of previously described paenilipoheptin A as well as the structure elucidation of the novel antibiotic paenilipoheptin B. These results further highlight the potential of *Paenibacillus* spp. for discovering NRPs and advancing the development of novel antibiotics.

Acknowledgement

We are grateful to Choong-Min Ryu for providing *P. polymyxa* E681. The work was supported by the NACTAR program of the Netherlands Organization for Scientific Research (NWO), Grant 16440 to G.P.vW. and N.I.M.

Materials and methods

General information

Extended supporting information, which includes NMR and HRMS figures, is available free of charge at https://doi.org/10.1021/acs.orglett.5c00231.

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). HPLC purification was performed on a Waters preparative HPLC system composed of a 1525 pump, a 2707 autosampler, a 2998 photo diode array (PDA) detector, and a Waters fraction collector III or BESTA-Technik preparative HPLC system equipped with an ECOM Flash UV detector. All solvents and chemicals were of HPLC or LC-MS grade, depending on the experiment. Structural assignments were made with additional information from NOESY, gCOSY, and gHSQC experiments.

Global genome mining of NRPSs

Genomes of all *Paenibacillus* spp. available from RefSeq (release 213)[22] were downloaded from the NCBI FTP site. All genomes with >400 contigs were considered as low-quality assemblies and were removed from the collection. All genomes were analyzed using AntiSMASH (version 6.0.1) to obtain BGC predictions. These predictions were then used as input into BiG-SCAPE (version 1.1.4),[11,23] for the creation of a sequence similarity network, with a distance matrix cutoff set to 0.25. The resulting full network was visualized by Cytoscape (3.9.1).[16]

Isolation and identification of bacterial strains

Paenibacillus strains were obtained from the Auburn University Plant-Associated Microbial strain collection. The similarity to the *P. polymyxa*, *P. peoriae* and *P. jamilae spp.* was calculated using 16S rRNA gene sequences of *Paenibacillus* spp. from Auburn University Plant-Associated Microbial strain collection obtained through PCR amplification with universal bacterial primers 27F and 1492R.[24] Each PCR amplicon was purified and used for Sanger sequencing. Subsequently, 16S rRNA gene sequences were assembled into consensus sequences and compared with reference sequences.

Growth of *Paenibacillus* spp. and natural product extraction

Cultivation and extraction of specialized metabolites was done following previously described method, with modifications.[25] Briefly, liquid pre-cultures of 25 *Paenibacillus* spp. from the Auburn University strain collection, namely, *Paenibacillus spp.* JJ-16, JJ-21, JJ-195, JJ-226, JJ-227, JJ-228, JJ-1580, JJ-1582, JJ-1603, JJ-1614, JJ-1638, JJ-1640, JJ-1652, JJ-1715, JJ-1720, JJ-1722, JJ-1724, JJ-333, JJ-845, JJ-1604, JJ-1650, JJ-1683, JJ-1729, JJ-1743, and JJ-1747 were used to inoculate as an inoculum for the 2.0 mL 96-deep well plate (Thermo Scientific, Nunc 2.0 mL DeepWell Plate) containing 600 µL TSA (TSA, Bacto Soybean-Casein Digest Medium, 30 g/L). Plates were sealed with 96 Well-Cap Mats (Thermo Scientific, Nunc 96 Well-Cap Mats), and incubated at 30 °C for 72 h before extraction. The cultures were extracted twice with 300 µL 100% isopropanol acidified with 0.1% formic acid. The crude extracts were transferred into a pre-washed 96-well plate (Agilent Technologies, 96-well plates, 0.5 mL, polypropylene) and lyophilized to dryness. Dried samples were redissolved in 160 µL of methanol/water (1:1 v/v). *Paenibacillus polymyxa* E681 was cultivated on TSA, and metabolites were extracted following the protocol described in our previous study.[12]

Genome sequencing, assembly and annotation

Paenibacillus sp. JJ-21 and *Paenibacillus* sp. JJ-1722 were grown in tryptic soy broth (TSB) at 30 °C and 220 rpm for 24 h. DNA was extracted as described.[26] 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).[19] BGCs in these genomes were annotated using AntiSMASH (version 7.1.0).[13]

Data-dependent LC-ESI-HRMS/MS

LC-MS/MS acquisition was performed using Shimadzu Nexera X2 UHPLC system, with an attached PDA detector, coupled to Shimadzu 9030 QTOF mass spectrometer, equipped with a standard 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.[27] 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. 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 with fixed collision energy (20 eV) and excluded for 1 s before being re-selected for fragmentation.

MZmine 2 parameters

Prior to statistical analysis, mzXML files were imported into Mzmine 2.53 and processed as previously described.[14,27] 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. The detected peaks were deisotoped (monotonic shape; maximum charge: 3; representative isotope: most intense). Peak lists from different samples were aligned (weight for t_R = weight for *m/z* = 50; compare the isotopic pattern with a minimum score of 50%). Only the features with MS/MS data were exported to a GNPS-FBMN. In addition, all features originating from the culture medium, as well as those with an *m/z* of less than 300, were removed.

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).[17] 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 and displayed using an unweighted force-directed layout.[16] The data are publicly accessible in the MassIVE repository (MSV000094386).

Up-scale fermentation, extraction, and isolation

Paenibacillus sp. JJ-21 and JJ-1722 were grown at 30 °C on 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-21 and JJ-1722, cells were collected by centrifugation (8000 rpm, 30 min, 4 °C) and washed with H_2O. Then, the cells were sonicated for 30 min and extracted with 100% isopropyl alcohol supplemented with 0.1% (v/v) formic acid for 6 h. The crude extracts were collected, concentrated under reduced pressure and reconstituted in 50% MeCN. The crude extract of *Paenibacillus* sp. JJ-21 was subjected to the preparative HPLC BESTA-Technik system

(Dr. Maisch Reprosil Gold 120 C18 column (25 × 250 mm, 10 μm), buffer A (H_2O/MeCN/TFA = 95:5:0.1)/buffer B (H_2O/MeCN/TFA = 5:95:0.1) = 20-100, flow rate = 12.0 mL/min, l = 214 nm), t_R = 33.8 minutes to yield paenilipoheptin A (**1**, 1 mg, 0.9%). The crude extract of *Paenibacillus* sp. JJ-1722 was subjected to the Waters preparative HPLC system (SunFire C18 column (19 × 150 mm, 10 μm), buffer A (H_2O)/buffer B (MeCN) = 10-60, flow rate = 15.0 mL/min, l = 214 nm), resulting in three fractions. The third fraction was further purified (SunFire C18 column (10 × 250 mm, 5 μm), buffer A (H_2O/MeCN = 75:25)/buffer B (H_2O/MeCN = 65:35) = 25-35, flow rate = 3.0 mL/min, l = 214 nm) t_R = 13.48 minutes to yield paenilipoheptin B (**2**, 0.3 mg, 0.3%).

Paenilipoheptin A (**1**): white powder; ^{1}H (DMSO-d_6, 850 MHz) and ^{13}C NMR (DMSO-d_6, 212.5 MHz) data, see **Table S1**; HRMS (ESI): *m/z* $[M+H]^+$ 1123.6180 (calcd. for $C_{59}H_{83}N_{10}O_{12}$, 1123.6192).

Paenilipoheptin B (**2**): white powder; ^{1}H (DMSO-d_6, 850 MHz) and ^{13}C NMR (DMSO-d_6, 212.5 MHz) data, see **Table S2**; HRMS (ESI): *m/z* $[M+H]^+$ 1060.6555 (calcd. for $C_{55}H_{86}N_{11}O_{10}$, 1060.6559).

Marfey's analysis

The stereochemistry of chiral centers present at α carbons was 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.[18] The general method for Marfey's analysis was conducted as described).[28] Briefly, a sample of peptide (30 μg) in 6 M HCl (100 μL) was heated to 100 °C in a sealed vial for 8−12 h using a 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/MeCN (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.

Supplementary information

Table S1. NMR data of paenilipoheptin A (**1**), measured in DMSO-d_6 at 298 K*.

	Residue	NH	H_α (C_α, type)	H_β (C_β, type)	Other
1	**Fatty amino acid**	7.20	2.21 (41.7, CH_2)	4.00 (47.2, CH)	1 C: (170.4) 4 CH_2: 1.39, 1.33 (35.11) 5-8 CH_2: 1.05-1.25 (25.5) 9 CH_2: 1.20, 1.00 (36.30) 10 CH: 1.25 (33.8) 11 CH_2: 1.26, 1.07 (29.0) 12 CH_3: 0.81 (11.18) 13 CH_3: 0.79 (19.07)
2	**1-Ser**	7.74	4.48 (54.4, CH)	3.44, 3.41 (61.63, CH_2)	1C: ND OH: 4.80
3	**2-Lys**	8.42	3.90 (54.7, CH)	1.49 (30.0, CH_2)	1C: ND 4 CH_2: 1.16, 1.05 (22.3) 5 CH_2: 1.39 (26.66) 6 CH_2: 2.62 (38.8) NH_2: ND
4	**3-Trp**	8.02	4.48 (54.4, CH)	3.13, 3.06 (27.17, CH_2)	1 C: ND 4 C: (110.5) 5 CH: 7.14 (123.5) 6 C: (136.1) 7 CH: 7.33 (111.3) 8 CH: 7.06 (120.9) 9 CH: 6.98 (118.27) 10 CH: 7.54 (118.3) 11 C: (127.2) NH: 10.81
5	**4-Ala**	7.70	4.03 (49.14, CH)	1.09 (17.85, CH_3)	1 C: (172.1)
6	**5-Phe**	7.92	4.33 (54.7, CH)	3.05, 2.90 (37.2, CH_2)	1 C: ND 4 C: 138.1 5 & 9 CH: 7.29 (129.5) 6 & 8 CH: 7.22 (128.0) 7 CH: 7.14 (126.1)
7	**6-Tyr**	8.24	4.25 (55.9, CH)	2.89, 2.85 (36.0, CH_2)	1 C: (170.75) 4 C: (127.05) 5 & 9 CH: 7.05 (130.1) 6 & 8 CH: 6.67 (115.0) 7 C: (156.3) OH: 9.22
8	**7-Glu**	7.95	4.11 (51.8, CH)	1.88, 1.57 (26.6, CH_2)	1 C: ND 4 CH_2: 2.00 (29.9) 5 C: (174.0)

* ^1H 850 MHz and ^{13}C 212.5 MHz

ND – not determined under these experimental conditions.

Table S2. NMR data of paenilipoheptin B (**2**), measured in DMSO-d_6 at 298 K*.

	Residue	NH	H_α(C_α, type)	H_β(C_β, type)	Other
1	**Fatty amino acid**	8.20	2.23, 2.17 (43.5, CH_2)	3.87 (47.1, CH)	4 CH_2: 1.49 (34.9) 5-8 CH_2: 1.23 (28.7) 9 CH_2: 1.24 (35.7) 10 CH: 1.28 (33.5) 11 CH_2: 1.26, 1.07 (29.0) 12 CH_3: 0.83 (10.9) 13 CH_3: 0.82 (18.8)
2	**1-Ser**	7.58	4.53 (55.3, CH)	3.79, 3.65 (62.0, CH_2)	
3	**2-Lys**	ND	3.83 (56.3, CH)	1.59, 1.53 (29.5, CH_2)	4 CH_2: 1.21, 1.12 (22.2) 5 CH_2: 1.34 (29.6) 6 CH_2: ND NH_2: ND
4	**3-Trp**	8.60	4.23 (54.8, CH)	3.16, 3.09 (25.7, CH_2)	4 C: ND 5 CH: 7.26, (123.5) 6 NH: 10.8 7 C: 135.8 8 CH: 7.32 (111.1) 9 CH: 7.04 (120.6) 10 CH: 6.96 (118.1) 11 CH: 7.50 (117.9) 12 C: 126.9
5	**4-Ala**	7.16	4.15 (47.7, CH)	1.12 (16.9, CH_3)	
6	**5-Leu**	7.23	3.89 (52.1, CH)	1.35, 1.05 (40.0, CH_2)	4 CH: 1.45 (23.8) 5 CH_3: 0.68 (20.8) 6 CH_3: 0.78 (22.4)
7	**6-Tyr**	7.65	4.53 (51.7, CH)	2.85, 2.65 (37.3, CH_2)	4 C: 127.7 5 & 9 CH: 6.92 (129.8) 6 & 8 CH: 6.56 (114.3) 7 C: 155.5
8	**7-Dab**	ND	3.91 (52.8, CH)	1.65 (33.2, CH_2)	4 CH_2: 2.56 (37.4) NH_2: ND

* ^{1}H 850 MHz and ^{13}C chemical shifts inferred from HSQC and HMBC spectra

ND – not determined under these experimental conditions (and also all carbonyl carbons).

Table S3. Retention times (t_R, min) of the FDAA derivatives for natural paenilipoheptin A and standard amino acids.

Amino acid	$[M+H]^+$	t_R, min			
		L-AA (standard)	D-AA (standard)	Paenilipoeptin A	Stereochemical assignment
Ser	358.0994	4.66	4.72	4.66	**L**
Lys	651.2121	6.50	6.67	6.68	**D**
Trp	457.1466	6.37	6.63	-	**ND**
Ala	342.1044	5.28	5.62	5.62	**D**
Phe	418.1357	6.51	6.90	6.52	**L**
Tyr*	686.1801	7.09	7.48	7.09	**L**
Glu	400.1099	4.97	5.12	5.12	**D**
FAA	482.2609	NA	NA	10.17	**ND**

* Product of the double-addition of Marfey's reagent
ND – not determined

Table S4. Retention times (t_R, min) of the FDAA derivatives for natural paenilipoheptin B and standard amino acids.

Amino acid	$[M+H]^+$	t_R, min			
		L-AA (standard)	D-AA (standard)	Paenilipoeptin B	Stereochemical assignment
Ser	358.0994	4.66	4.72	4.66	**L**
Lys	651.2121	6.50	6.67	6.68	**D**
Trp	457.1466	6.37	6.63	-	**ND**
Ala	342.1044	5.28	5.62	5.62	**D**
Leu	384.1514	6.45/6.55	6.98/7.04	6.45/6.55	**L**
Tyr*	686.1801	7.09	7.48	7.09	**L**
Dab*	623.1804	6.23	6.38	6.23	**L**
FAA	482.2609	NA	NA	10.17	**ND**

* Product of the double-addition of Marfey's reagent
ND – not determined

Domain	Residues in the binding pocket			Amino acid prediction			Amino acid detected		
	PaenA BGC of *P. polymyxa* E681	PaenA BGC of *Paenibacillus* sp. JJ-21	PaenB BGC of *Paenibacillus* sp. JJ-1722	PaenA BGC of *P. polymyxa* E681	PaenA BGC of *Paenibacillus* sp. JJ-21	PaenB BGC of *Paenibacillus* sp. JJ-1722	PaenA of *P. polymyxa* E681	PaenA *Paenibacillus* sp. JJ-21	PaenB *Paenibacillus* sp. JJ-1722
M 2	DVWHFSLVDK	DVWHFSLVDK	DVWHFSLVDK	Ser	Ser	Ser	Ser	Ser	Ser
M 3	DVCETGTIEK	DVCETGTIEK	DVCETGTIEK	Dab	Dab	Dab	Lys	Lys	Lys
M 4	DAWAFAGVAK	DAWAFAGVAK	DAWAFAGVAK	Trp	Trp	Trp	Trp	Trp	Trp
M 5	DVFWLGGTFK	DVFWMGGTFK	DVFWLGGTFK	Val	Val	Val	Ala	Ala	Ala
M 6	DAWTFAAIAK	DAWTFAAIAK	DAWIFGAITK	Phe	Phe	Leu	Phe	Phe	Leu
M 7	DTSTLAAVAK	DTSTLAAVAK	DTSTLAAVAK	Tyr	Tyr	Tyr	Tyr	Tyr	Tyr
M 8	DAKDIGVVDK	DAKDIGVVDK	DVGEISSIDK	Glu	Glu	Dab	Glu	Glu	Dab

Table S5 Analysis of A-domain specificities of paenilipoheptin BGCs.

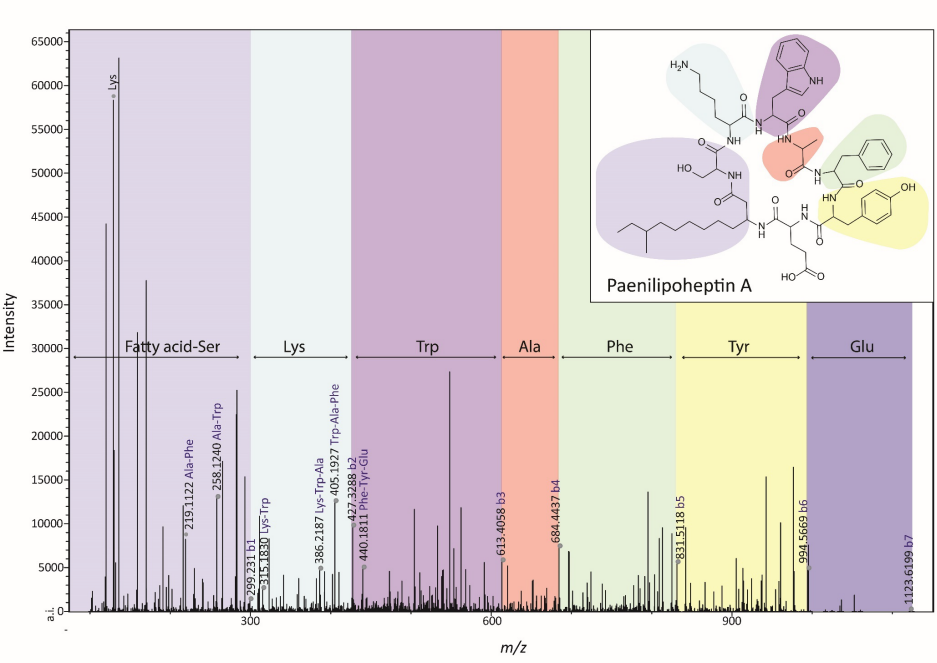

Figure S1. MS/MS spectrum of paenilipoheptin A (precursor ion $[M + 2H]^{2+}$ *m/z* 562.3133). The assignment of the sequence of amino acid residues is based on the mass differences between the consecutive *b* ions.

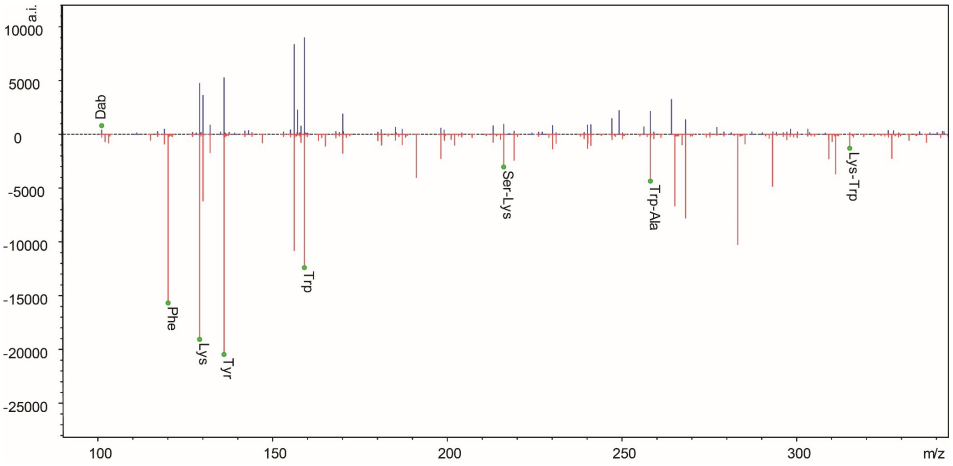

Figure S2. Direct MS/MS spectra comparison of the mass feature having *m/z* of 555.305 (red) with the mass feature having *m/z* of 523.8231 (blue).

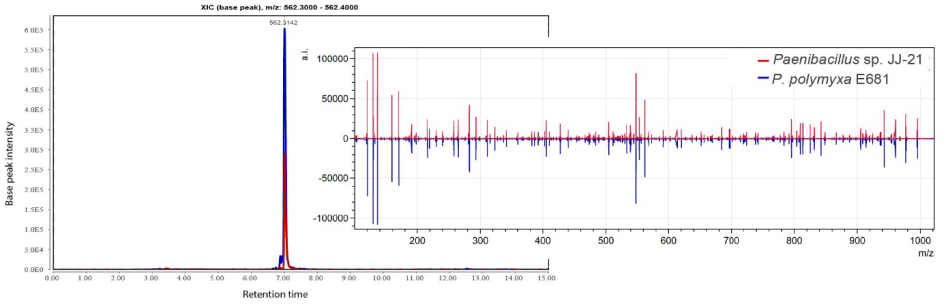

Figure S3. Comparison of the retention time and fragmentation patterns of paenilipoheptin from the extract of *Paenibacillus* sp. JJ-21 and *P. polymyxa* E681.

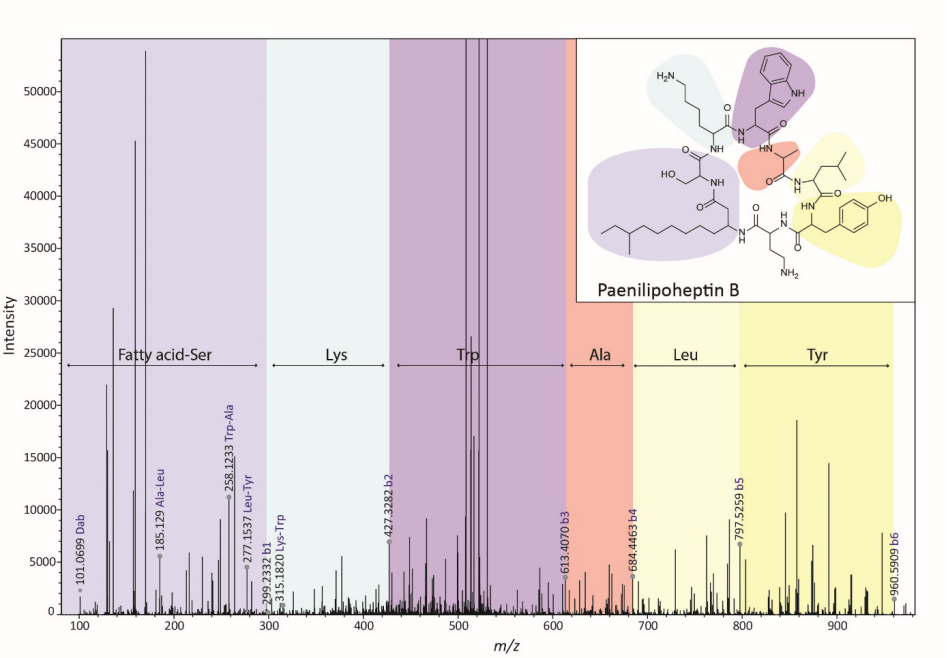

Figure S4. MS/MS spectrum of paenilipoheptin B (precursor ion $[M + 2H]^{2+}$ *m/z* 530.8307). The assignment of the sequence of amino acid residues is based on the mass differences between the consecutive *b* ions.

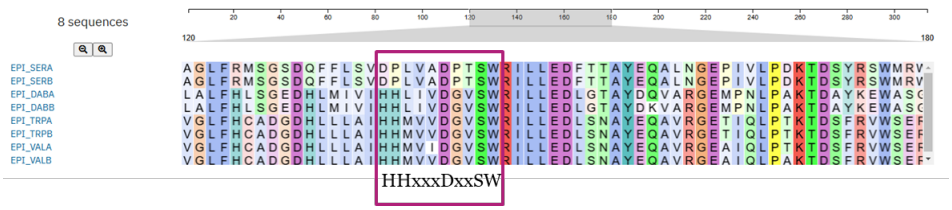

Figure S5. Clustal2 multiple sequence alignment of the epimerization domains of the paenilipoheptin A and B BGCs from *Paenibacillus* sp. JJ-21 and *Paenibacillus* sp. JJ-1722, respectively.

References

(1) Clardy, J.; Fischbach, M. A.; Walsh, C. T. New antibiotics from bacterial natural products. *Nat. Biotechnol.* **2006**, *24* (12), 1541−1550.

(2) Felnagle, E. A.; Jackson, E. E.; Chan, Y. A.; Podevels, A. M.; Berti, A. D.; McMahon, M. D.; Thomas, M. G. Nonribosomal peptide synthetases involved in the production of medically relevant natural products. *Mol. Pharm.* **2008**, *5* (2), 191−211.

(3) Liu, Y.; Ding, S.; Shen, J.; Zhu, K. Nonribosomal antibacterial peptides that target multidrug-resistant bacteria. *Nat. Prod. Rep.* **2019**, *36* (4), 573−592.

(4) Wang, H.; Fewer, D. P.; Holm, L.; Rouhiainen, L.; Sivonen, K. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (25), 9259−9264.

(5) Cochrane, S. A.; Vederas, J. C. Lipopeptides from *Bacillus* and *Paenibacillus* spp.: A gold mine of antibiotic candidates. *Med. Res. Rev.* **2016**, *36* (1), 4−31.

(6) Olishevska, S.; Nickzad, A.; Déziel, E. *Bacillus* and *Paenibacillus* secreted polyketides and peptides involved in controlling human and plant pathogens. *Appl. Microbiol. Biotechnol.* **2019**, *103* (3), 1189− 1215.

(7) Machushynets, N. V.; Al Ayed, K.; Terlouw, B. R.; Du, C.; Buijs, N. P.; Willemse, J.; Elsayed, S. S.; Schill, J.; Trebosc, V.; Pieren, M.; Alexander, F. M.; Cochrane, S. A.; Liles, M. R.; Medema, M. H.; Martin, N. I.; van Wezel, G. P. Discovery and derivatization of tridecaptin antibiotics with altered host specificity and enhanced bioactivity. *ACS Chem. Biol.* **2024**, *19* (5), 1106−1115.

(8) Sieber, S. A.; Marahiel, M. A. Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem. Rev.* **2005**, *105* (2), 715−738.

(9) Fischbach, M. A.; Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* **2006**, *106* (8), 3468−3496.

(10) Marahiel, M. A. A structural model for multimodular NRPS assembly lines. *Nat. Prod. Rep.* **2016**, *33* (2), 136−140.

(11) Navarro-Muñoz, J. C.; Selem-Mojica, N.; Mullowney, M. W.; Kautsar, S. A.; Tryon, J. H.; Parkinson, E. I.; De Los Santos, E. L. C.; Yeong, M.; Cruz-Morales, P.; Abubucker, S.; Roeters, A.; Lokhorst, W.; Fernandez-Guerra, A.; Cappelini, L. T. D.; Goering, A. W.; Thomson, R. J.; Metcalf, W. W.; Kelleher, N. L.; Barona-Gomez, F.; Medema, M. H. A computational framework to explore large-scale biosynthetic diversity. *Nat. Chem. Biol.* **2020**, *16* (1), 60−68.

(12) Vater, J.; Herfort, S.; Doellinger, J.; Weydmann, M.; Borriss, R.; Lasch, P. Genome mining of the lipopeptide biosynthesis of *Paenibacillus polymyxa* E681 in combination with mass spectrometry: discovery of the lipoheptapeptide paenilipoheptin. *ChemBioChem.* **2018**, *19* (7), 744−753.

(13) Blin, K.; Shaw, S.; Augustijn, H. E.; Reitz, Z. L.; Biermann, F.; Alanjary, M.; Fetter, A.; Terlouw, B. R.; Metcalf, W. W.; Helfrich, E. J. N.; van Wezel, G. P.; Medema, M. H.; Weber, T. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res.* **2023**, *51* (W1), W46− W50.

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

(15) Nothias, L.-F.; Petras, D.; Schmid, R.; Dührkop, K.; Rainer, J.; Sarvepalli, A.; Protsyuk, I.; Ernst, M.; Tsugawa, H.; Fleischauer, M.; Aicheler, F.; Aksenov, A. A.; Alka, O.; Allard, P.-M.; Barsch, A.; Cachet, X.; Caraballo-Rodriguez, A. M.; Da Silva, R. R.; Dang, T.; Garg, N.; et al. Feature-based molecular networking in the GNPS analysis environment. *Nat. Microbiol.* **2020**, 17 (9), 905−908.

(16) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.; Wang, J.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, 13 (11), 2498−2504.

(17) Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kapono, C. A.; Luzzatto-Knaan, T.; Porto, C.; Bouslimani, A.; Melnik, A. V.; Meehan, M. J.; Liu, W. T.; Crusemann, M.; Boudreau, P. D.; Esquenazi, E.; SandovalCalderon, M.; Kersten, R. D.; et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34* (8), 828−837.

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

(19) Chin, C. A.-O.; Peluso, P.; Sedlazeck, F. A.-O.; Nattestad, M.; Concepcion, G. T.; Clum, A.; Dunn, C. A.-O.; O'Malley, R.; Figueroa-Balderas, R.; Morales-Cruz, A.; Cramer, G. A.-O.; Delledonne, M.; Luo, C.; Ecker, J. R.;

Cantu, D. A.-O.; Rank, D. R.; Schatz, M. C. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* **2016**, *13*, 1050−1054.

(20) Samel, S. A.; Czodrowski, P.; Essen, L.-O. Structure of the epimerization domain of tyrocidine synthetase A. *Acta Crystallographica Section D* **2014**, *70* (5), 1442−1452.

(21) Lysenko, V.; Machushynets, N. V.; van Dam, J. L.; Sterk, F. A. C.; Speer, A.; Ram, A. F. J.; Slingerland, C. J.; van Wezel, G. P.; Martin, N. I. Total synthesis, structure elucidation, and bioactivity evaluation of the cyclic lipopeptide natural product paenilipoheptin A. *Org. Lett.* **2025**, *27* (12), 2826–2831.

(22) O'Leary, N. A.; Wright, M. W.; Brister, J. R.; Ciufo, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; Astashyn, A.; Badretdin, A.; Bao, Y.; Blinkova, O.; Brover, V.; Chetvernin, V.; Choi, J.; Cox, E.; Ermolaeva, O.; Farrell, C. M.; et al. Reference Sequence (RefSeq) Database at NCBI: Current Status, Taxonomic Expansion, and Functional Annotation. *Nucleic Acids Res.* **2016**, *44* (D1), D733–D745.

(23) Blin, K.; Shaw, S.; Kloosterman, A. M.; Charlop-Powers, Z.; van Wezel, G. P.; Medema, Marnix H.; Weber, T. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* **2021**, *49* (W1), W29-W35.

(24) Lane, D. *16S/23S rRNA sequencing*; Wiley, **1991**.

(25) Nguyen, D. D.; Melnik, A. V.; Koyama, N.; Lu, X.; Schorn, M.; Fang, J.; Aguinaldo, K.; Lincecum, T. L.; Ghequire, M. G. K.; Carrion, V. J.; Cheng, T. L.; Duggan, B. M.; Malone, J. G.; Mauchline, T. H.; Sanchez, L. M.; Kilpatrick, A. M.; Raaijmakers, J. M.; De Mot, R.; Moore, B. S.; Medema, M. H.; Dorrestein, P. C. Indexing the Pseudomonas Specialized Metabolome Enabled the Discovery of Poaeamide B and the Bananamides. *Nat. Microbiol.* **2016**, *2* (1), 16197.

(26) Kieser, T.; Bibb, M. J.; Buttner, M. J.; Chater, K. F.; Hopwood, D. A. *Practical Streptomyces genetics*; John Innes Foundation, **2000**.

(27) van Bergeijk, D. A.; Elsayed, S. S.; Du, C.; Santiago, I. N.; Roseboom, A. M.; Zhang, L.; Carrión, V. J.; Spaink, H. P.; van Wezel, G. P. The ubiquitous catechol moiety elicits siderophore and angucycline production in *Streptomyces*. *Commun. Chem.* **2022**, *5* (14).

(28) Lysenko, V.; Son, S.; Theriault, M. E.; Slingerland, C. J.; Hauk, G.; Cleenewerk, L.; Speer, A.; Berger, J. M.; Lewis, K.; Martin, N. I. Total synthesis and structural reassignment of the antitubercular natural product evybactin. *Chem. Eur. J.* **2024**, *31* (1), e202403767.