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.

56 **INTRODUCTION**

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

63 Members of the bacterial phylum Bacillota (formerly Firmicutes) are well known for
64 their potential to produce a wide range of antimicrobials, in particular nonribosomally
65 synthesized peptides (NRPs) (4-6). Numerous bioactive NRPs produced by *Bacillus* and
66 *Paenibacillus* have been identified, including octapeptins, paenibacterins, polypeptins,
67 tridecaptins, etc (5-7). NRPs such as polymyxins and bacitracins that are derived from
68 these microorganisms, have found application in clinical settings (8, 9). A large fraction of
69 the NRPs contain positively charged amino acids and possess potent antimicrobial
70 activities against Gram-positive and Gram-negative pathogens (10, 11). The mode of
71 action polymyxins, involves binding to the lipid A component of lipopolysaccharides
72 (LPSs) in the outer membrane of Gram-negative bacteria. Other NRPs, like bacitracin
73 produced by *Bacillus subtilis* and *Bacillus licheniformis*, act as inhibitors of cell wall
74 biosynthesis. (9) Specifically, in the presence of a divalent metal ion, most commonly Zn^{2+},
75 bacitracin binds to and sequesters the membrane-associated phospholipid undecaprenyl
76 pyrophosphate ($C_{55}PP$) (12, 13). Bacitracin A is the major bioactive congener with the
77 highest clinical relevance, and it is mainly used topically to treat infections caused by
78 Gram-positive bacteria (14). The bacitracin family is composed of structurally similar
79 macrocyclic dodecapeptides containing a unique N-terminal aminothiazoline moiety (15).

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

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

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

104 strategies have been applied based on the combination of molecular networking with
105 different sources of information (e.g., genomic, taxonomic, geographic, bioactivity, and/or
106 spectral data) (26-29). Despite the ability of current tools to dereplicate NPs upstream of
107 the isolation process, effective strategies for the early detection of potentially bioactive
108 NPs within crude extracts are still lacking.

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

117 **RESULTS AND DISCUSSION**

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

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

127 In search of novel NRPs, we investigated the chemical diversity of the specialized
128 metabolites produced by *Paenibacillus* isolates. For this purpose, all 227 isolates of the
129 collection were cultured individually on TSA media in 96 deep-well plates for 72 h and
130 extracted with isopropyl alcohol (IPA) supplemented with 0.1% (v/v) formic acid (FA).
131 Subsequently, crude extracts were tested against Gram-positive (*Staphylococcus aureus*
132 ATCC 29213) and Gram-negative (*Escherichia coli* ATCC 25922) indicator strains to
133 assess the antibacterial activity of the strains. Nine crude extracts exhibited significant
134 growth inhibition of *S. aureus*, while 22 crude extracts showed bioactivity against *E. coli*
135 (Figure 1, Table S1). These extracts were then subjected to LC-MS/MS analysis to identify
136 the natural products accountable for the observed bioactivities. LC-MS/MS data were
137 processed with MzMine 2 (30) and exported for GNPS FBMN. Prior to uploading the data
138 to GNPS, media-derived mass features were eliminated. A mass exclusion threshold of
139 300 *m/z* was implemented to minimize the occurrence of mass features originating from
140 small metabolites that are not relevant to our study. FBMN analysis resulted in generating

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

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

158

159 **Targeting cationic lipopeptides using MassQL-assisted fragment search**

160 Several of the identified lipopeptides are well-known antibiotics with potent activity,
161 typically attributed to their net positive charge (Table S3), which enhances electrostatic
162 interactions with the negatively charged components on the surface of bacterial cells, such
163 as the cell membrane or lipopolysaccharides (LPSs) (32). Therefore, we sought to detect
164 specifically those lipopeptides with potential bioactive properties that contained one or

165 multiple positively charged amino acid residues. To accomplish this, a feature-based
166 spectral similarity network was paired with a MassQL (33) search query to target the
167 MS/MS spectra that contained the most common positively charged amino acids. The
168 fragment search of *b* ions of 2,4-diaminobuturic acid (Dab) (*m/z* 101.0709), ornithine (Orn)
169 (*m/z* 115.0871), lysine (Lys) (*m/z* 129.1027), arginine (Arg) (*m/z* 157.1089) and the
170 immonium ion of histidine (His) (*m/z* 110.0718) was conducted with the accuracy of *m/z*
171 0.005 and at minimum intensity of 5% (Figure 3).

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

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

189

190 **Discovery of a novel family of bacitracins**

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

205 MS/MS spectra of all the mass features within this sub-family exhibited an ion with
206 an *m/z* 199.0897. However, these mass features could not be matched to known microbial
207 natural products. Analysis of the MS/MS spectra of the parent compound with *m/z*
208 712.8836 revealed that a fragment ion with an *m/z* 199.0897 has a molecular formula of
209 $C_9H_{15}N_2OS$, which corresponds to a thiazoline-containing substructure commonly found
210 in the peptide antibiotic bacitracin, so far exclusively identified from the genus *Bacillus*.
211 (40) Based on the fact that the b_2, b_3, b_4, and b_5 ions were shared between the parent
212 mass with *m/z* 712.8836 and bacitracin A (41), we concluded that their pentapeptide side

213 chains are identical. The differences observed in b_6–b_{11} ions indicated that the structural
214 distinctions between the two metabolites lie in the heptapeptide ring. We did not observe
215 fragment ions corresponding to Orn, Phe or His, which are found in the structure of
216 bacitracin A, while we could ascertain the presence of an ion corresponding to a Trp
217 residue (*m/z* 187.0864). Moreover, the dipeptide fragment ions Trp-Leu (*m/z* 300.171),
218 Trp-Thr (*m/z* 288.1352), and Leu-Leu (*m/z* 227.1764) were detected. By tracing the series
219 of both *b* and *y* ions, we elucidated the amino acid sequence of the compound with *m/z*
220 712.8836 as (Ile-Cys)–Leu/Ile–Glu–Leu/Ile–Lys–Leu/Ile–Leu/Ile–Trp–Thr–Asp–Asn
221 (Figure S7). The MS/MS studies used did not allow us to discriminate between Leu and
222 Ile in positions 3, 5, 7 and 8 of the peptide amino acid sequence. Notably, this sequence
223 diverges from any known bacitracin analogues, indicating the discovery of a novel branch
224 of the important bacitracin-family of NRPs.

225 To purify the compound for more detailed studies, *Paenibacillus* sp. JJ-778 was
226 cultured in a 10 L MHB media and subjected to extraction using Diaion® HP-20 beads,
227 followed by elution with 100% IPA supplemented with 0.1% (v/v) formic acid. Multiple
228 rounds of chromatographic purification resulted in the isolation of compounds (**1**) and (**2**)
229 (Figures S8, S9). Both compounds had the same molecular formula of $C_{67}H_{106}N_{15}O_{17}S$,
230 based on their exact mass of *m/z* 1424.7584 for $[M+H]^+$, highly similar to the calculated
231 mass of 1424.7611. 1D and 2D NMR analysis of **1** confirmed the amino acid sequence
232 (Figure 7), and together with the molecular formula dictated by the accurate mass, the
233 structure of **1** was determined as a new congener of bacitracin (Table S4, Figures S10-
234 S16). NMR analysis identified Leu residues in positions 3 and 7, and Ile in positions 5 and
235 8. The ^{1}H NMR spectra of **1** and **2** were almost identical, suggesting that they are

236 diastereomers (Figure S17). Accordingly, the new compounds were designated as
237 paenitracin A (**1**) and B (**2**).

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

254 Bacitracin, originally isolated from *Bacillus licheniformis* and *B. subtilis*, targets
255 primarily Gram-positive bacteria such as *Staphylococcus*, *Streptococcus*, *Enterococcus*
256 and *Clostridium difficile* (44). Bacitracin and its congeners are synthesized by a
257 nonribosomal peptide synthetase (NRPS) complex encoded by the bacitracin synthetase
258 operon, which consists of *bacA*, *bacB*, and *bacC* genes. To explore the biosynthesis of
259 the novel paenitracin, the genome of *Paenibacillus* sp. JJ-778 was sequenced and

260 assembled using Falcon (version 1.8.1) (45). Analysis using antiSMASH 7.1.0 (46) readily
261 identified a 12-modular gene cluster with 77% similarity to the bacitracin BGC of *B.*
262 *licheniformis* ATCC 10716. In silico analysis of the A domain substrate specificity of BacA,
263 BacB, and BacC predicted variations in amino acid composition at positions 7, 9, and 10
264 compared to the well-known bacitracin A (Figure 6B).

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

270

271 **Mode of action and bioactivity of paenitracins**

272 Bacitracins specifically target the unique bacterial phospholipid undecaprenyl
273 pyrophosphate (C_{55}-PP). C_{55}PP is the membrane-associated phospholipid that anchors
274 lipid II and shuttles it between the cell's cytoplasm where it is synthesized, and the
275 membrane exterior, where it is incorporated into the growing cell wall (47). To see if
276 paenitracin indeed has the same mode of action as bacitracin, the C_{55}PP dependence of
277 paenitracin activity was tested using antibacterial activity antagonization assays in the
278 presence or absence of zinc (48). For this, paenitracin was mixed with C_{10}-PP, a more
279 soluble analogue of C_{55}-PP, (15) and incubated with cell suspension of *S. aureus* ATCC
280 29213. Antagonization assays showed that the activity of paenitracin B was antagonized
281 by C_{10}PP exclusively in the presence of Zn^{2+} (Table S7). This indicates that paenitracin
282 binds to C_{55}-PP of Gram-positive bacteria in a zinc-dependent manner and thus has the
283 same mode of actions as bacitracins.

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

299 In conclusion, MS-based indexing of NRPs from 227 plant-associated
300 *Paenibacillus* isolates revealed extensive chemical diversity. Molecular networking
301 combined with MassQL identified several classes of lipopeptides, whereby we specifically
302 focused on NRPs containing basic residues. This revealed the presence of distant
303 relatives of bacitracin, which we designated paenitracins. Paenitracin A and B showed
304 potent bioactivity against Gram-positive ESKAPE pathogens, including vancomycin-
305 resistant *E. faecium* E155. Unlike previously described bacitracin variants, paenitracins
306 show substantial structural divergence and represent the first example of bacitracin-family
307 NRPs in *Paenibacillus*. These findings highlight the untapped chemical space within this

308 genus and demonstrate the power of combining metabolomic indexing with targeted
309 mining strategies to accelerate the discovery of structurally unique antibiotics.
310

311 **MATERIALS AND METHODS**

312 **General experimental procedures**

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

320

321 **Isolation and identification of bacterial strains**

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

327 The phylogenetic tree was constructed using 16S rRNA gene sequences obtained
328 through PCR amplification with universal bacterial primers 27F and 1492R (49). Each
329 PCR amplicon was purified and used for Sanger sequencing. Subsequently, 16S rRNA
330 gene sequences were assembled into consensus sequences and aligned with reference
331 sequences. These include 220 sequences from our study, 41 sequences from type strains
332 of *Paenibacillus* spp. downloaded from GenBank (Table S2) and sequence of *Bacillus*
333 *subtilis* IAM 12118 (GenBank accession NR_112116.2). The alignment is done using
334 ClustalOmega (v 1.2.2) (50). Aligned sequences are trimmed to the length of 1,452 bp to

335 remove extending regions from a few sequences and also to preserve as many variable
336 regions as possible. The trimmed alignment is re-aligned and a neighbor-joining
337 consensus tree is generated using Geneious (v 2024.0.3) tree function, with Tamura-Nei
338 genetic distance model and *B. subtilis* as outgroup. The generated tree is then uploaded
339 to ITOL for visualization (51).
340

341 **Growth of *Paenibacillus* spp. and natural product extraction**

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

358

359 **Genome sequencing, assembly and annotation**

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

369

370 **Data-dependent LC-ESI-HRMS/MS**

371 LC-MS/MS acquisition was performed using Shimadzu Nexera X2 ultra-high-performance
372 liquid chromatography (UPLC) system, with an attached photodiode array detector (PDA),
373 coupled to Shimadzu 9030 QTOF mass spectrometer, equipped with a standard
374 electrospray ionization (ESI) source unit. A total of 2 µL was injected into a Waters Acquity
375 HSS C_{18} column (1.8 µm, 100 Å, 2.1 × 100 mm) and data acquisition was performed as
376 previously described (54). Briefly, the gradient used was 5% B for 1 min, 5–85% B for
377 9 min, 85–100% B for 1 min, and 100% B for 4 min. The column was re-equilibrated to 5%
378 B for 3 min before the next run was started. All the samples were analyzed in positive
379 polarity, using data-dependent acquisition mode. In this regard, full scan MS spectra (*m/z*
380 100–1700, scan rate 10 Hz, ID enabled) were followed by two data-dependent MS/MS
381 spectra (*m/z* 100–1700, scan rate 10 Hz, ID disabled) for the two most intense ions per

382 scan. The ions were fragmented using collision-induced dissociation (CID) with fixed
383 collision energy (CE 20 eV) and excluded for 1 s before being re-selected for
384 fragmentation. The parameters used for the ESI source were: interface voltage 4 kV,
385 interface temperature 300 °C, nebulizing gas flow 3 L/min, and drying gas flow 10 L/min.
386

387 **MZmine 2 parameters**

388 Prior to statistical analysis, mzXML files were imported into Mzmine 2.53 (30) and
389 processed as previously described (54). Briefly, mass ion peaks were detected for
390 MS^1 and MS^2 at a noise level of 2.0E2 and 0.0E0, respectively (positive polarity, mass
391 detector: centroid), and their chromatograms were built using ADAP chromatogram
392 builder. The detected peaks were smoothed, and the chromatograms were deconvoluted
393 (algorithm: local minimum search; chromatographic threshold: 90; search minimum in RT
394 range: 0.05; minimum relative height: 1%; minimum ratio of peak top/edge: 2; peak
395 duration: 0.03–3.00 min). The detected peaks were deisotoped (monotonic shape;
396 maximum charge: 3; representative isotope: most intense). Peak lists from different
397 samples were aligned (weight for RT = weight for *m/z* = 50; compare the isotopic pattern
398 with a minimum score of 50%). Missing peaks detected in at least one of the samples
399 were filled with the gap-filling algorithm (RT tolerance: 0.1 min, *m/z* tolerance was set to
400 0.002 *m/z* or 15 ppm). Duplicate peaks were filtered. Only the features with MS/MS data
401 were exported to a GNPS-FBMN.

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

405 blanks with an intensity higher than 5000 were also removed. In the end, features with
406 *m/z* less than 300 were filtered out.
407
408 **Molecular network analysis and MassQL search**
409 The resulting feature quantification table (CSV file) and MS/MS spectrum files (in mgf
410 format) were uploaded to the GNPS webserver (http://gnps.ucsd.edu) (19). The LC-
411 MS/MS data for the 227 *Paenibacillus* isolates were analyzed using the Feature-Based
412 Molecular Networking tool (FBMN) (20). Briefly, the precursor ion mass tolerance was set
413 to 0.005 Da and the MS/MS fragment ion tolerance to 0.05 Da. A molecular network was
414 then created where edges were filtered to have a cosine score above 0.5 and more than
415 3 matched peaks. The molecular networks were visualized using Cytoscape software
416 version 3.9.1 (31) and displayed using an unweighted force-directed layout. The data are
417 publicly accessible in the MassIVE repository (MSV000094386) and the networking
418 results and parameters can be found at
419 https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=2f1e0ae1728142249ac4e841d5a72e
420 f4 . The MassQL tool version 31.4 (33) was employed to search for specific MS/MS
421 fragments that correspond to the *b* or immonium ions of positively charged amino acids.
422 The fragment search of *b* ions of 2,4-diaminobuturic acid (*m/z* 101.0709), ornithine (*m/z*
423 115.0871), lysine (*m/z* 129.1027), and arginine (*m/z* 157.1089); together with the
424 immonium ion of histidine (*m/z* 110.0718), was conducted with an accuracy of *m/z* 0.005
425 and minimum intensity of 5%.

426

427 **Up-scale fermentation, extraction, and isolation**

428 *Paenibacillus* sp. JJ-778 was grown at 30 °C on tryptic soy agar (TSA) for 72 h and three
429 colonies were inoculated into TSB and incubated at 30 °C overnight. This inoculum (1%)
430 was used to inoculate thirteen 2 L Erlenmeyer flasks containing 0.75 L of sterile MHB and
431 fermented at 30 °C while shaking at 200 rpm for 72 h. To extract the specialized
432 metabolites produced by *Paenibacillus* sp. JJ-778 supernatant was collected by
433 centrifugation and subjected to extraction using Diaion® HP-20 (Resindion, Mitsubishi),
434 followed by elution with 100% IPA supplemented with 0.1% (v/v) formic acid. The crude
435 extract was adsorbed onto silica gel (pore size 60 Å, 70–230 mesh, Sigma Aldrich), and
436 loaded on a silica column, which was eluted using varying gradients of EtOAc/MeOH. The
437 fractions eluted with EtOAc-MeOH (1:3) supplemented with 0.1% (v/v) formic acid were
438 combined, reconstituted in 50% ACN and injected into the preparative SunFire column
439 (19 × 150 mm), which was eluted with H_2O:ACN gradient of 15–60% in 30 min, at a flow
440 rate of 15 mL/min, to yield three fractions. Fraction 3 was further separated on the semi-
441 preparative SunFire column (10 × 250 mm) at 3 mL/min using a H_2O–ACN gradient of 30–
442 50% in 30 min to yield paenitracin A (**1**, 0.3 mg) and paenitracin B (**2**, 0.2 mg).

443

444 **Marfey`s analysis**

445 The stereochemistry of chiral centers present at α carbons were assigned by applying
446 derivatization methods coupled with chromatographic analysis. The advanced Marfey's
447 method using L-FDAA (1-fluoro-2-4-dinitrophenyl-5-L-alanine amide) established the
448 absolute configurations of amino acids (55). The general method for Marfey`s analysis
449 was conducted as described) (42). Briefly, a sample of peptide (30 µg) in 6M HCl (100 µL)

450 was heated to 100 °C in a sealed vial for 8−12 h using heating block, after which the
451 hydrolysate was concentrated to dryness at 40 °C under a stream of dry N_2. The
452 hydrolysate was then treated with 1 M $NaHCO_3$ (20 µL) and L-FDAA (1% solution in
453 acetone, 40 µL) at 40 °C for 1 h, after which the reaction was neutralized with 1 M HCl (20
454 µL). An aliquot of the analyte was diluted 50 times with H_2O/ACN (1:1) and injected (2 µL)
455 into an HRMS instrument following the standard protocol of the analysis. The analyte
456 amino acid content was assessed by comparison to authentic standards. The authentic
457 standards were prepared via a similar procedure.
458
459 **Antibacterial activity assays and MIC determination**
460 For *E. coli* ATCC 25922 and *S. aureus* ATCC 29213, single colonies were inoculated into
461 20 mL of MHII broth medium in 250 mL Erlenmeyer flask and incubated overnight at 37
462 °C with shaking at 220 rpm and then diluted to obtain an assay inoculum of 5 x 10^5
463 CFU/mL. Aztreonam and vancomycin were used as reference positive controls for each
464 pathogen. Absorbance at OD_{612} nm was measured at T_0 (zero time) using an EnVision™
465 Microplate Reader (PerkinElmer), and assay plates were incubated at 37 °C. After that,
466 absorbance at OD_{612} nm was measured at T_f (final time) using the same microplate reader
467 equipment. The percentage of growth inhibition was calculated using the following
468 normalization:

469 $$\%\text{Inhibition} = 100 \times \left\{1 - \frac{([T_{f\text{Sample}} - T_{0\text{Sample}}] - [T_{f\text{Blank}} - T_{0\text{Blank}}])}{([T_{f\text{Growth}} - T_{0\text{Growth}}] - [T_{f\text{Blank}} - T_{0\text{Blank}}])}\right\}$$

470 $T_{0\text{Sample}}$, the absorbance of the strain growth in the presence of sample measured at zero
471 time;

472 $T_{f\text{Sample}}$, the absorbance of the strain growth in the presence of sample measured at final
473 time;
474 $T_{0\text{Growth}}$, the absorbance of the strain growth in the absence of sample measured at zero
475 time;
476 $T_{f\text{Growth}}$, the absorbance of the strain growth in the absence of sample measured at the
477 final time;
478 $T_{0\text{Blank}}$, the absorbance of the broth medium (blank) measured at zero time;
479 $T_{f\text{Blank}}$, the absorbance of the broth medium (blank) measured at the final time.

480 The Genedata Screener software (Genedata, Inc., Basel, Switzerland) was used to
481 process and analyze the data. The activity of the samples was expressed as a percentage
482 of growth inhibition, where (-100%) represented the total growth inhibition of the target
483 microorganism and 0% represented the total growth of the target microorganism in the
484 assay. The Z' factor predicts the robustness of an assay by considering the mean and
485 standard deviation of both positive and negative controls (56). The robust Z' factor (RZ'
486 factor) is based on the Z' factor, but standard deviations and means are replaced by the
487 robust standard deviations and medians, respectively. In all experiments performed in this
488 work, the RZ' factor obtained was between 0.9173 and 0.9684.
489 For MIC tests, blood agar plates were inoculated with glycerol stocks of *E. coli*
490 ATCC 25922, *B. subtilis* 168, *S. aureus* ATCC 29213, *S. aureus* USA300 (clinical isolate
491 from Texas Children's Hospital, USA), *E. faecium* E155 and *E. faecium* E980 and
492 incubated overnight at 37 °C. Following incubation, 3 mL of tryptic soy broth (TSB) was
493 inoculated with an individual colony and incubated at 37 °C with shaking at 220 rpm until
494 the cultures reached the exponential phase. The bacterial suspensions were then diluted

495 100-fold in cation-adjusted Mueller-Hinton broth (MHB) or Luria-Bertani (LB) broth
496 supplemented with 0.3 M $ZnSO_4$ to reach a bacterial cell density of 10^6 CFU/mL. In
497 parallel, test compounds were 2-fold serially diluted with cation-adjusted MHB or LB
498 supplemented with 0.3 M $ZnSO_4$ in polypropylene 96-well plates to achieve a final volume
499 of 50 µL per well. An equal volume of bacterial suspension (10^6 CFU/mL) was added to
500 the wells. The well plates were sealed with an adhesive membrane and incubated for 20
501 hours at 37 °C. The reported MIC values are based on three technical triplicates and are
502 defined as the lowest concentration of the compound that prevented the visible growth of
503 bacteria.
504
505 **In vitro antagonization of paenitracin antibacterial activity**
506 Antibacterial activity antagonization assay was performed as previously described (48).
507 Briefly, geranyl pyrophosphate ammonium salt ($C_{10}PP$) (1 mg/mL in methanol [aqueous
508 10 mM NH_4OH (7:3), Sigma Aldrich, Taufkirchen, Germany] was used as an antagonist
509 of paenitracin A in either the presence or absence of 0.3 mM $ZnSO_4$. Paenitracin A was
510 mixed with $C_{10}PP$ (five-fold molar excess relative to paenitracin) or added to empty wells
511 (absence of antagonist) giving in all cases 50 µL of total volume in each well with
512 paenitracin A at 4× MIC in the presence and absence of antagonist as well as in the
513 presence and absence of 0.3 mM $ZnSO_4$, in triplicate. Subsequently, 50 µL of *S. aureus*
514 ATCC 29213 with OD_{600} of 0.005 was added to the test compounds in the microtiter plate
515 to achieve a final antibiotic concentration of 2× MIC. The samples were incubated at 37
516 °C for 24 h with constant shaking and inspected for visible bacterial growth.
517
518

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.

528 **REFERENCES**

529

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

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

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

537 4. Aleti G, Sessitsch A, Brader G. 2015. Genome mining: prediction of lipopeptides
538 and polyketides from *Bacillus* and related Firmicutes. *Comput Struct Biotechnol J*
539 13:192-203.

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

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

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

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

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

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

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

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

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

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

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

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

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

574 18. Gaudêncio SP, Bayram E, Lukić Bilela L, Cueto M, Díaz-Marrero AR,
575 Haznedaroglu BZ, Jimenez C, Mandalakis M, Pereira F, Reyes F, Tasdemir D.
576 2023. Advanced methods for natural products discovery: bioactivity screening,
577 dereplication, metabolomics profiling, genomic sequencing, databases and
578 informatic tools, and structure elucidation. *Mar Drugs* 21:308.

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

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

598 21. van der Hooft JJJ, Wandy J, Young F, Padmanabhan S, Gerasimidis K, Burgess
599 KEV, Barrett MP, Rogers S. 2017. Unsupervised discovery and comparison of
600 structural families across multiple samples in untargeted metabolomics. *Anal Chem*
601 89:7569-7577.

602 22. da Silva RR, Wang M, Nothias LF, van der Hooft JJJ, Caraballo-Rodríguez AM,
603 Fox E, Balunas MJ, Klassen JL, Lopes NP, Dorrestein PC. 2018. Propagating
604 annotations of molecular networks using in silico fragmentation. *PLOS Comput Biol*
605 14:e1006089.

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

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

612 25. Damiani T, Jarmusch AK, Aron AT, Petras D, Phelan VV, Zhao HN, Bittremieux W,
613 Acharya DD, Ahmed MMA, Bauermeister A, Bertin MJ, Boudreau PD, Borges RM,
614 Bowen BP, Brown CJ, Chagas FO, Clevenger KD, Correia MSP, Crandall WJ,
615 Crüsemann M, Fahy E, Fiehn O, Garg N, Gerwick WH, Gilbert JR, Globisch D,
616 Gomes PWP, Heuckeroth S, James CA, Jarmusch SA, Kakhkhorov SA, Kang KB,
617 Kessler N, Kersten RD, Kim H, Kirk RD, Kohlbacher O, Kontou EE, Liu K, Lizama-
618 Chamu I, Luu GT, Luzzatto Knaan T, Mannochio-Russo H, Marty MT, Matsuzawa
619 Y, McAvoy AC, McCall L-I, Mohamed OG, Nahor O, Neuweger H, et al. 2025. A
620 universal language for finding mass spectrometry data patterns. *Nature Methods*
621 22:1247-1254.

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

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

630 drug leads in natural product bioassay-guided fractionation. *J Nat Prod* 81:758-
631 767.

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

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

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

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

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

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

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

657 34. Vogler K, Studer RO. 1966. The chemistry of the polymyxin antibiotics. *Experientia*
658 22:345-354.

659 35. Jian Li, Roger L Nation, John D Turnidge, Robert W Milne, Kingsley Coulthard,
660 Craig R Rayner, Paterson DL. 2006. Colistin: the re-emerging antibiotic for
661 multidrug-resistant Gram-negative bacterial infections. *Lancet Infect Dis* 6:589-
662 601.

663 36. Bann SJ, Ballantine RD, Cochrane SA. 2021. The tridecaptins: non-ribosomal
664 peptides that selectively target Gram-negative bacteria. *RSC Med Chem* 12:538-
665 551.

666 37. Knolhoff AM, Zheng J, McFarland MA, Luo Y, Callahan JH, Brown EW, Croley TR.
667 2015. Identification and structural characterization of naturally-occurring broad-
668 spectrum cyclic antibiotics isolated from *Paenibacillus*. *J Am Soc Mass Spectrom*
669 26:1768-1779.

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

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

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

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

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

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

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

692 45. Chin CAO, Peluso P, Sedlazeck FA-O, Nattestad M, Concepcion GT, Clum A,
693 Dunn CA-O, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, Cramer GA-O,
694 Delledonne M, Luo C, Ecker JR, Cantu DA-O, Rank DR, Schatz MC. 2016. Phased
695 diploid genome assembly with single-molecule real-time sequencing. *Nat Methods*
696 13:1050–1054.

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

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

703 48. Buijs NP, Vlaming HC, Kotsogianni I, Arts M, Willemse J, Duan Y, Alexander FM,
704 Cochrane SA, Schneider T, Martin NI. 2024. A classic antibiotic reimagined:
705 Rationally designed bacitracin variants exhibit potent activity against vancomycin-
706 resistant pathogens. *Proceedings of the National Academy of Sciences*
707 121:e2315310121.

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

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

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

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

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

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

724 55. Marfey P. 1984. Determination of D-amino acids. II. Use of a bifunctional reagent,
725 1, 5-difluoro-2, 4-dinitrobenzene. *Carlsberg Res Commun* 49:591-596.

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

729

730

731 **FIGURES**

732

733

734 **Figure 1. Phylogenetic tree based on 16S rRNA sequences of plant-associated**
735 ***Paenibacillus* isolates and closely related type strains.** The tree was obtained by
736 neighbor-joining analysis. For an overview of the strains see Tables S1 and S2. Type
737 strains are indicated with green squares. The tree was rooted using *Bacillus subtilis* IAM
738 12118 as an outgroup. Antibacterial activity of the strains against *E. coli* ATCC 25922 and
739 *S. aureus* ATCC 29213 is indicated with purple and red squares, respectively. The tree
740 was generated using Geneious v2024.0.3.

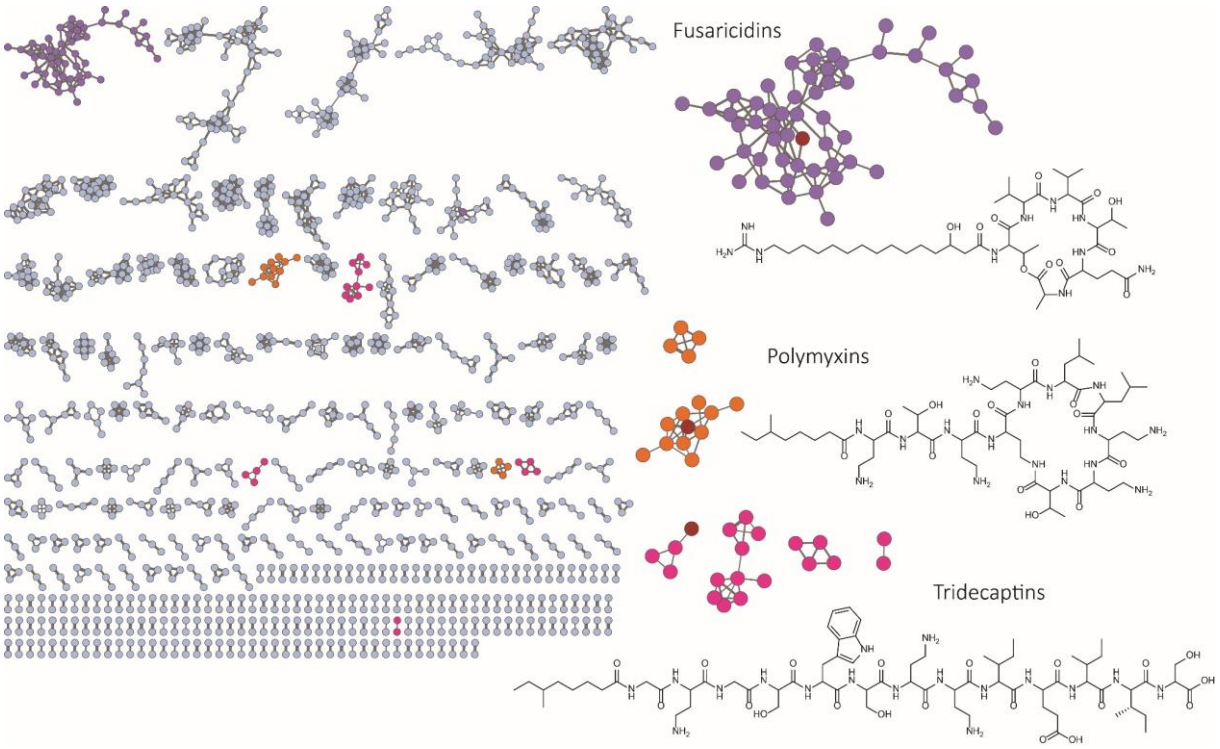
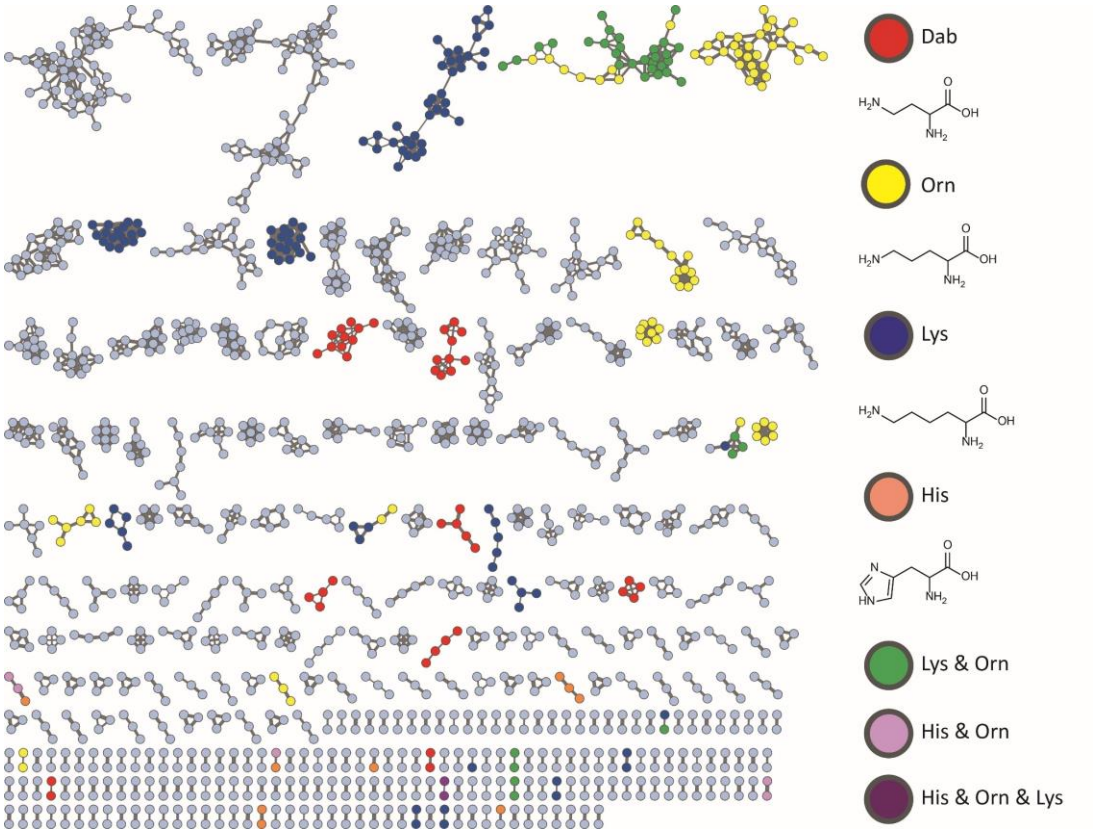


Figure 2. Molecular network of mass features detected in the bacterial extracts from the *Paenibacillus* isolates. Each node represents a parent ion detected in one or more strains. Different node colors represent distinct molecular families. The red nodes highlight the specific annotated compound in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

750 

751

752 **Figure 3. MassQL-annotated molecular network of the ions detected in the bacterial**

753 **extracts from *Paenibacillus* isolates.** Colors of the nodes, as indicated in the figure,

754 highlight parent ions that contain specific basic amino acid residues (Dab, Orn, His, Lys

755 or combinations thereof) in the structures.

756

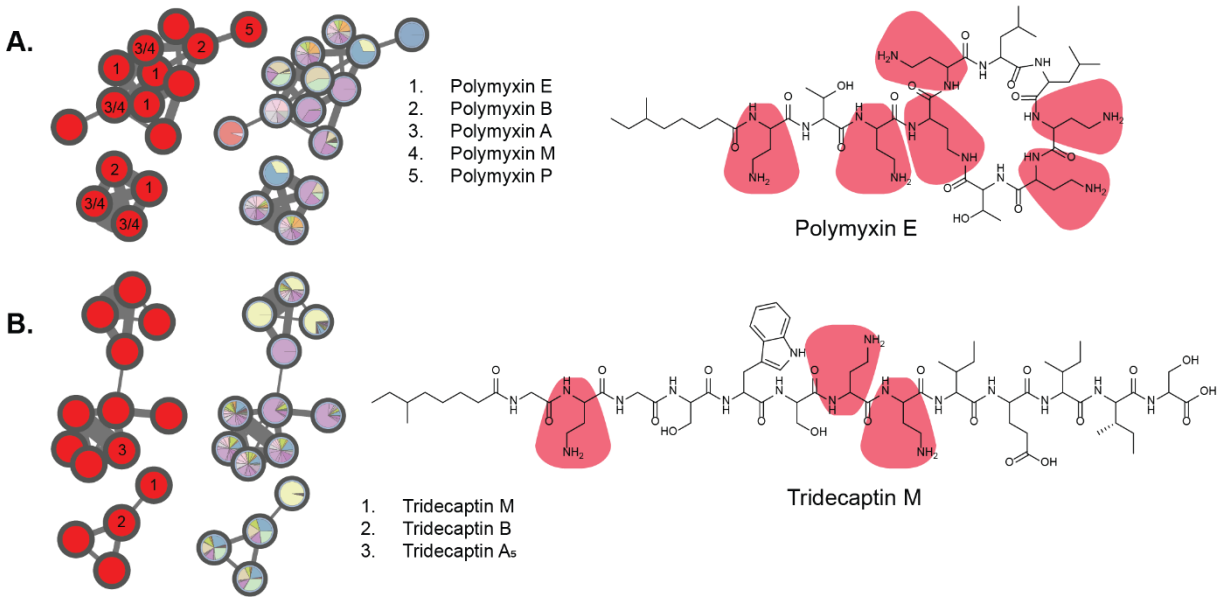
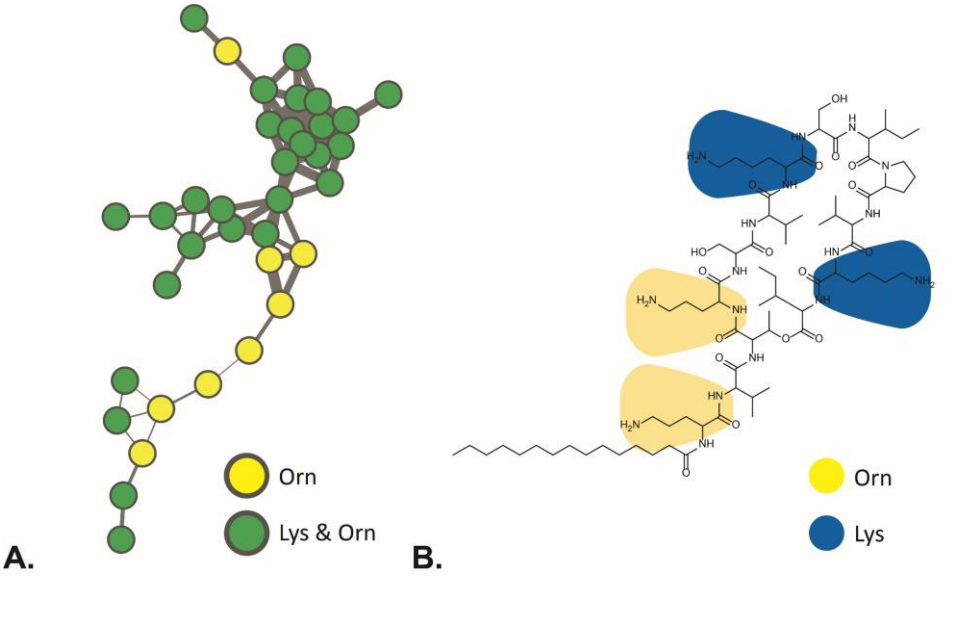


Figure 4. Dab-containing lipopeptide families highlighted by MassQL search. Molecular families of polymyxins (A) and tridecaptins (B) consist of multiple structural analogues, reflecting high chemical diversity. Pie charts were mapped to the nodes to represent the relative precursor ion intensities within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



765

766 **Figure 5. MassQL-based detection of paenibacterins. A.** MassQL-annotated

767 molecular family of paenibacterins. **B.** Chemical structure of paenibacterin A.

768

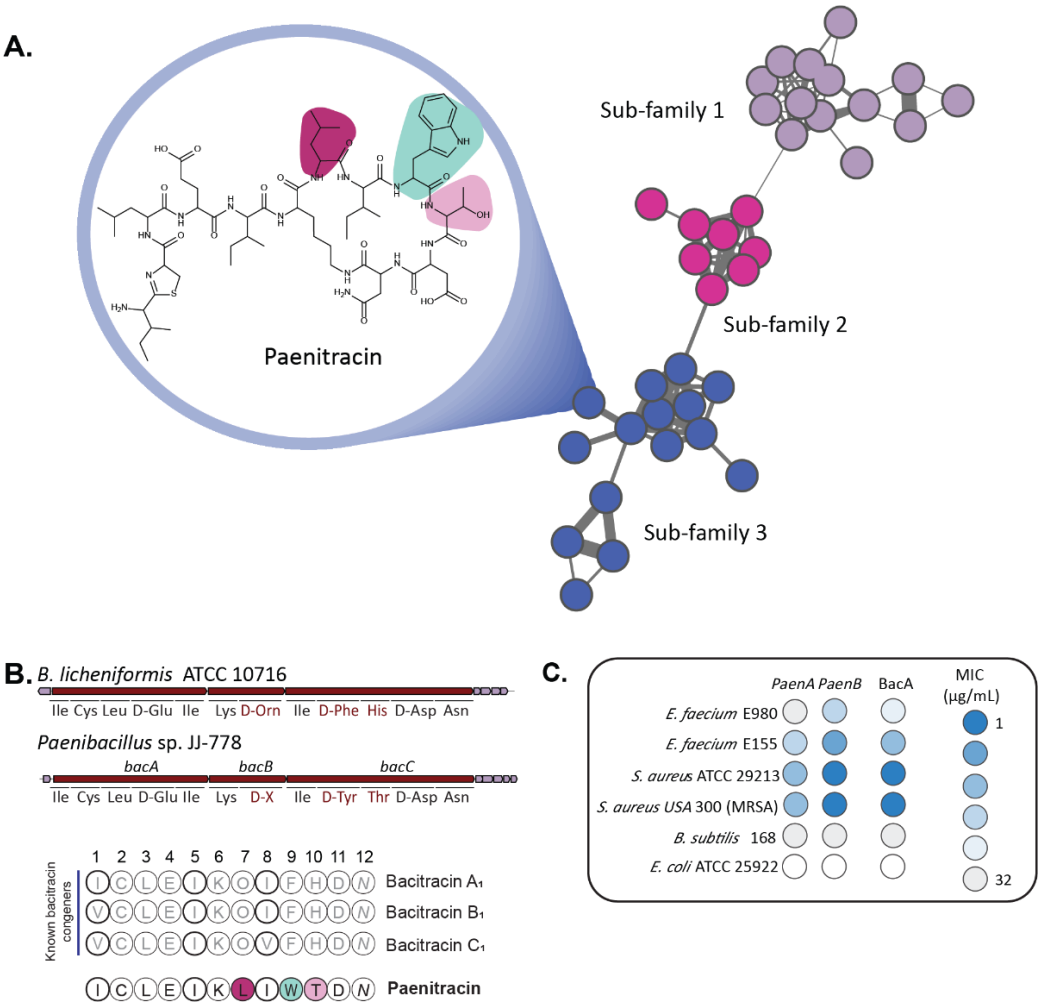
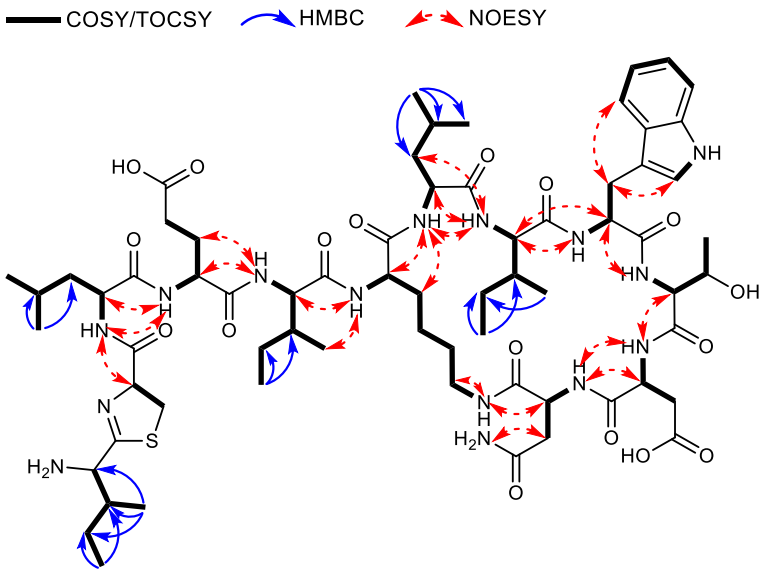


Figure 6. Molecular family of Lys-containing NRPs highlighted by MassQL search. **A.** Exploring the mass features from sub-family 3 led to the discovery of new bacitracin analog, designated as paenitracin. **B.** The BGC and its predicted amino acid substrates of new paenitracin from the genome of *Paenibacillus* 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.



780

781 **Figure 7. Correlations obtained by COSY, HMBC, and NOESY measurements in**

782 **NMR of paenitracin A (1).**