Universiteit
Leiden
The Netherlands

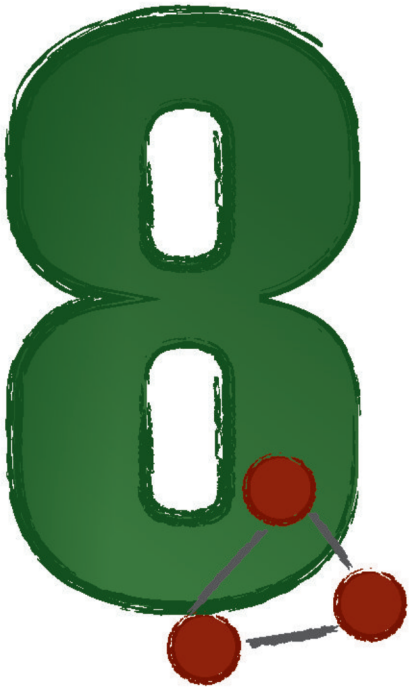
# Expanding the chemical space of antibiotics produced by Paenibacillus and Streptomyces

Machushynets, N.V.

**Citation**
Machushynets, N. V. (2024, September 5). *Expanding the chemical space of antibiotics produced by Paenibacillus and Streptomyces*. Retrieved from https://hdl.handle.net/1887/4082475

| | |
|---|---|
| Version: | Publisher's Version |
| License: | Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden |
| Downloaded from: | https://hdl.handle.net/1887/4082475 |

**Note:** To cite this publication please use the final published version (if applicable).



# Chapter 8

## Chimeric biosynthesis unveils diverse tridecaptin variants in the *Paenibacillus* genus

Nataliia V Machushynets, Barbara R. Terlouw, Chao Du, Karol Al Ayed, Somayah S Elsayed, Julian Schill, Vincent Trebosc, Michel Pieren, Mark R. Liles, Marnix H Medema, Nathaniel I Martin, Gilles P van Wezel

# Abstract

Modern sequencing technologies offer deep insight into the vast potential of *Paenibacillus* species to produce an extensive range of nonribosomal peptides (NRPs). Despite the expanding knowledge of the diverse microbial strategies used to produce these natural products, translating genomic information into bioactive lead compounds remains a complex challenge. Here, we show that chimeric biosynthesis occurs within the family of tridecaptin antibiotics. Through genome mining, numerous new tridecaptin biosynthetic gene clusters (BGCs) were identified in *Paenibacillus* genomes. Surprisingly, 15 strains harbor both a full tridecaptin BGC and a second BGC that specifies a truncated decamer. The encoded NRPS domain architectures suggested the capability of these strains to produce multiple tridecaptin variants through collaborative action between this tridecaptin-like NRPS and a shorter NRPS encoded within the 'main' tridecaptin BGC. Indeed, *Paenibacillus* sp. JJ-1683 was found to produce both tridecaptin $A_5$ and tridecaptin B, providing an unusual example of chimeric biosynthesis of NRPS lipopeptide antibiotics. These findings reveal chimeric biosynthesis as a strategy enabling strains to dynamically produce similar compounds with distinct chemistry and bioactivity profiles.

# Introduction

Demand for new antibiotics is at an all-time high due to the emergence of antimicrobial resistance among human pathogens that poses a serious but largely unchallenged global health threat (WHO, 2019). With the advent of increasingly affordable DNA sequencing technologies and the development of advanced bioinformatic tools, bacterial genome mining for novel natural products is poised to play an important role in addressing this threat (Baltz, 2021, Palazzotto & Weber, 2018). In recent years, genome mining has revealed evidence for an expansive trove of yet-undiscovered drug-like molecules hidden in the genomes of biosynthetically talented bacterial genera (Katz & Baltz, 2016, Medema *et al.*, 2021, Scherlach & Hertweck, 2021). One such genus is *Paenibacillus*, which includes many bacterial species commonly found in soil that are often associated with plant roots. Members of this genus can offer protection against insects, herbivores, and phytopathogens, including other bacteria, fungi, nematodes, and viruses, by producing a wealth of natural products (Olishevska *et al.*, 2019, Grady *et al.*, 2016).

Genes encoding biosynthetic pathways for natural products are frequently clustered in bacterial genomes (Medema & Fischbach, 2015). With the help of genome mining tools such as antiSMASH, such biosynthetic gene clusters (BGCs) can be identified in microbial (meta) genomes and classified based on the type of molecule they are predicted to produce, such as polyketides, ribosomally synthesized and post-translationally modified peptides (RiPPs), terpenes, and nonribosomal peptides (Blin *et al.*, 2023). For some natural product classes, it is well-understood how the encoded biosynthetic machinery builds a natural product, and thus, the core scaffold structure of the produced natural product can be (partially) predicted. This is especially true for nonribosomal peptides (NRPs): short peptides that are oligomerized in an mRNA- and ribosome-independent manner. Members of the *Paenibacillus* genus produce a plethora of NRPs, including polymyxins, octapeptins, paenibacterins, fusaricidins, and tridecaptins (Cochrane & Vederas, 2016).

NRP synthetase (NRPS) assembly lines select and condense amino acids iteratively to assemble peptides (Sieber & Marahiel, 2005). To achieve nonribosomal peptide synthesis, minimal enzymatic complexes require four distinct domains called core or essential domains: condensation (C), adenylation (A), peptidyl carrier protein (PCP) and a terminal thioesterase (TE) domain (Marahiel, 2016). The C, A, and PCP domains make up functional modules, with C domains catalyzing condensation, A domains mediating building block selection through adenylation and subsequent thiolation, and non-catalytic PCP domains functioning as mechanical arms to which intermediate scaffolds are tethered between reactions. An NRPS assembly line only requires a single TE domain, typically at the end of an NRPS enzyme, which is responsible for chain release. As A domains govern building block selection, many NRP structure prediction algorithms attempt to predict the selected substrate from the A

domain sequence. Examples of such machine learning algorithms include NRPSpredictor2 (Röttig *et al.*, 2011), AdenylPred (Mongia *et al.*, 2023), SANDPUMA (Chevrette *et al.*, 2017), and AdenPredictor (Mongia *et al.*, 2023), which were trained on A domain sequences of known specificity (Jenke-Kodama & Dittmann, 2009).

However, translating genomic information into predicted molecular structures remains challenging (Udwary *et al.*, 2021). The diversity of NRPS-derived peptides is not only determined by the incorporation of different amino acid building blocks but also by an ever-increasing number of characterized tailoring reactions and enzymatic domains (Wenski *et al.*, 2022). In addition, a constantly expanding library of scientific literature describes the unraveling of new examples of nonribosomal peptide synthetases exhibiting rare domains and noncanonical domain and module organizations, leading to unique microbial strategies to synthesize novel natural products (Duban *et al.*, 2022).

One notable example entails collaborative natural product biosynthesis, involving crosstalk between two separate BGCs. Such pairing of biosynthetic enzymes from distinct BGCs within the same biosynthetic pathway can give rise to increased chemical diversity and an expanded scope of biological activity. Over the past few years, this phenomenon of collaborative biosynthesis has received increased attention, and specialized metabolites produced through such strategies have been referred to in the literature as 'hybrid molecules' or 'chimeras' (Mevers *et al.*, 2019, Yin *et al.*, 2022). Chimeric biosynthesis has previously been observed in various NRPS and PKS (hybrid) systems, including the biosynthesis of leucanicidin and endophenaside C (Wu *et al.*, 2016b), actinomycin L (Machushynets *et al.*, 2022), azaphilone (Huang *et al.*, 2020), penilactones (Fan *et al.*, 2019), spirotryprostatin A (Yan *et al.*, 2019), and echinocandin B (Cacho *et al.*, 2012).

In a prior study, we applied global genome mining for an in-depth examination of the diversity of NRPS BGCs across 785 *Paenibacillus* genomes. BiG-SCAPE analysis revealed numerous uncharacterized gene cluster families (GCFs), which led to the discovery of the new tridecaptin variants $A_5$ and D. Tridecaptins are linear nonribosomal antibacterial peptides produced by various *Paenibacillus* spp., with potent activity against multidrug-resistant Gram-negative bacteria (Cochrane *et al.*, 2014, Cochrane *et al.*, 2015, Jangra *et al.*, 2019b). Tridecaptins are open-chain tridecapeptides with a β-hydroxy fatty acid tail at the N terminus. Numerous natural tridecaptin variants (tridecaptins A, B, C, M, and G) differ from each other in the fatty acid tail and/or amino acid composition (Bann *et al.*, 2021). Typically, tridecaptin BGCs comprise five genes encoding different proteins that are responsible for the synthesis and secretion of the active peptide: TriA (a putative thioesterase), TriB and TriC (ABC transporters), TriD (a 10-module NRPS) and TriE (a 3-module NRPS) (da Costa *et al.*, 2022). The peptide scaffold is assembled by the cooperative action of TriD and TriE, with TriD incorporating the first ten amino acids and TriE incorporating the last three (Bann *et al.*, 2021).

Here, we present evidence for the widespread chimeric biosynthesis of tridecaptin variants in the genus *Paenibacillus* through a combination of genome mining and experimental validation. We found that 15 different *Paenibacillus* genomes harbored both a full-length tridecaptin BGC and a separate, tridecaptin-like NRPS at a different locus. We hypothesized that these strains are capable of producing multiple tridecaptin variants through chimeric biosynthesis. Specifically, we propose that one variant is produced by the full-length BGC and the second by collaborative action between the tridecaptin-like NRPS at the secondary locus and the shorter NRPS in the full-length BGC. We verified this hypothesis by sequencing the genome of *Paenibacillus* JJ-1683, a strain which we experimentally showed to produce both tridecaptin $A_5$ and tridecaptin $B_1$. We demonstrated that *Paenibacillus* JJ-1683 contains both loci that match the bioinformatics predictions for the NRPS enzymes involved in the speculated biosynthetic routes for both tridecaptins. As such, this work provides an example of chimeric biosynthesis of NRPS lipopeptide antibiotics, and hints at chimeric biosynthesis as a strategy to swap between the production of two similar compounds with slightly different chemistry and bioactivity profiles.

# Results and Discussion

## Bioinformatics prediction of novel tridecaptin BGCs

We previously demonstrated that members of the genus *Paenibacillus* have great biosynthetic potential to produce NRP-like natural products. BiG-SCAPE (Navarro-Muñoz *et al.*, 2020) networking analysis revealed gene cluster families (GCFs) of numerous unknown NRPS-like BGCs alongside previously discovered polymyxins, paenilipoheptins, fusaricidins, paenibacterins, octapeptins, bacitracins, cilagicins, and tridecaptins (**Chapter 7**). To characterize the putative structures of the tridecaptins produced by *Paenibacillus* sp. strains, we first identified all tridecaptin-associated GCFs from our BiG-SCAPE analysis (Figure 1A). This yielded 7 GCFs that comprise tridecaptin-associated BGCs across 117 strains. Notably, some of these BGCs were incomplete, as they were found on the contigs of heavily fragmented genomes, and full tridecaptin clusters could not always be derived.

To maximize the discovery of tridecaptin BGCs, including those that did not cluster into GCFs, we identified all adenylation-domain-containing proteins encoded by the BGCs of the 785 publicly available *Paenibacillus* genomes. We then extracted the A domains and their active site signatures and predicted their substrate selectivity with the adenylation domain substrate prediction algorithm PARAS (Table S1). BGCs with ten or more modules were retained, and from these, putative tridecaptin assembly lines with associated core structure predictions were manually assigned based on building block predictions and the position of modules containing epimerization domains. This yielded 79 putative tridecaptin assembly lines (Table S2). Of these, 53 were full-length clusters containing all 13 modules, while the other 26 had one or more missing modules. Among the 53 full-length clusters, 10 were found to be unique.

Next, we computationally cross-referenced the predicted structures with all known tridecaptin peptide scaffolds (Table 1) and found full-length predicted scaffolds corresponding to tridecaptins $G_4$, $A_1/A_3$, $A_5$ and D (Table S2). Additionally, we predicted the structures of nine potentially novel tridecaptins, seven of which were full-length (Figure 1B; Table S2). This included two putative variants of tridecaptin A encoded in three *Paenibacillus kribbensis* strains, which are predicted to contain an alanine at position 3 as only previously observed in tridecaptin G (da Costa *et al.*, 2022). Notably, the second variant encoded by two of the strains was predicted to incorporate a glycine moiety in module 6, which recognizes a serine in all known tridecaptin variants. Other predicted novel tridecaptins include a tridecaptin $B_\gamma/C_{\beta_1}$ variant in *Paenibacillus peoriae* KCTC 3763, and a tridecaptin $A_5$ variant encoded in seven strains of *Paenibacillus polymyxa*. Finally, we found a putative tridecaptin M variant encoded in *Paenibacillus* sp. M-152 and *Paenibacillus* sp. lzh-N1, for which our predictions matched the six characterized amino acids of tridecaptin $M_{11}$. Tridecaptin $M_{11}$ was originally isolated from *Paenibacillus* sp. M-152 in a very low yield, and its amino acid sequence was only partially characterized through MS/MS analysis (Jangra *et al.*, 2019a). We here propose the sequence of its full amino acid scaffold as D-Val–D-Dab–D-Ser–D-Gly–D-Trp–L-Ser–L-Dab–D-Dab–L-Phe–L-Glu-L-Ile–D-Ile–L-Ser, based on our genome mining for the producing strain.

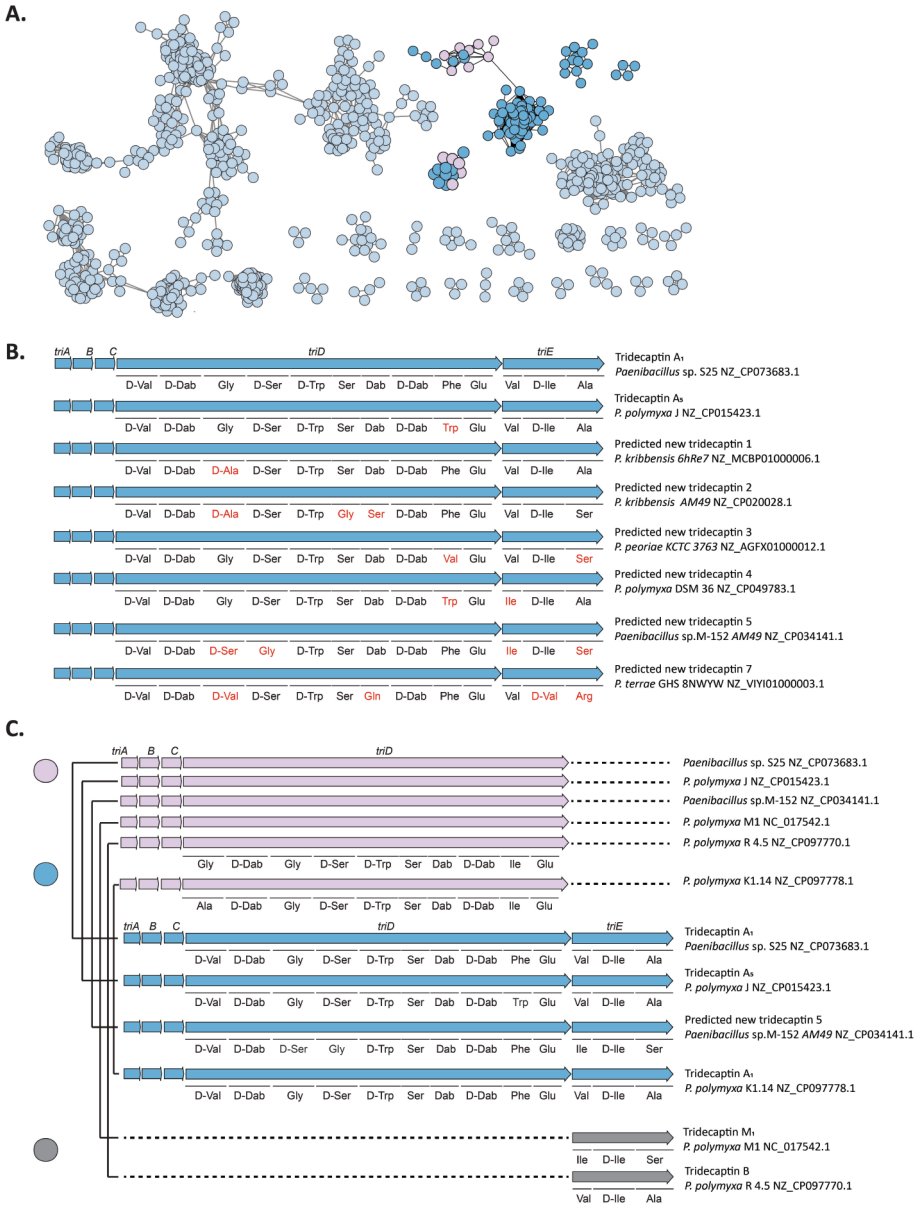



**Figure 1. Sequence similarity network of NRPS BGCs and tridecaptin BGC analysis. A.** BiG-SCAPE sequence similarity network (SSN) and tridecaptin-like GCFs visualized in Cytoscape (c 0.25). Each node represents one tridecaptin BGC identified by antiSMASH 6.0. Blue: full-length tridecaptin BGCs. Pink: truncated tridecaptin BGCs. **B.** New tridecaptin variants predicted in this study. The A domains and their active site signatures were extracted, and their substrate selectivity was predicted using the adenylation domain substrate prediction algorithm PARAS. BGCs with ten or more modules were retained, and from these, putative tridecaptin clusters assembly lines with associated core structure predictions were manually assigned. The amino acids highlighted in red differ from those in the structures of known tridecaptins. **C.** Gene architectures and amino acid scaffolds of full-length (blue) and truncated ten-(pink) and three-modular (grey) tridecaptin BGCs. Black connectors indicate BGCs that co-occur in the same genome. Collaboration between the two clusters, where the truncated cluster provides the first 10 amino acids and the full-length cluster provides the last three, leads to tridecaptin B production.

**Table 1. Amino acid scaffolds of known tridecaptins.**

| Compound | 1[D] | 2[D] | 3[D] | 4[D] | 5[D] | 6 | 7 | 8[D] | 9 | 10 | 11 | 12[D] | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $TriA_1/A_3$ | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Phe | Glu | Val | Ile | Ala |
| $TriA_2$ | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Phe | Glu | Val | Val | Ala |
| $TriA_4$ | Val | Dab | Gly | Ser | Phe | Ser | Dab | Dab | Phe | Glu | Val | Ile | Ala |
| $TriA_5$ | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Val | Ile | Ala |
| $TriB_1$ | Gly | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Ile | Glu | Val | Ile | Ala |
| TriBα | Gly | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Ile | Glu | Val | Ile | Ser |
| TriBβ / $M_2$ | Gly | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Ile | Glu | Val | Val | Ser |
| TriBγ | Gly | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Val | Glu | Val | Val | Ser |
| TriBδ | Gly | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Val | Glu | Val | Ile | Ser |
| TriCα1/Cα2 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Phe | Glu | Val | Val | Ser |
| TriCβ1 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Phe | Glu | Val | Ile | Ser |
| TriD | Phe | Dab | Ile | Ser | Trp | Ser | Ser | Dab | Trp | Ser | Val | Ile | Dab |
| TriE | Ala | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Val | Ile | Ala |
| $TriG_1$ | Ile | Dab | Ala | Dab | Trp | Ser | Asp | Dab | Phe | Asp | Val | Val | Arg |
| $TriG_1$ | Leu | Dab | Ala | Dab | Trp | Ser | Asp | Dab | Phe | Asp | Val | Val | Arg |
| $TriG_2$ | Val | Dab | Ala | Dab | Trp | Ser | Asp | Dab | Phe | Asp | Val | Val | Arg |
| $TriG_3$ | Ile | Dab | Ala | Dab | Trp | Ser | Asp | Dab | Trp | Asp | Val | Val | Arg |
| $TriG_3$ | Leu | Dab | Ala | Dab | Trp | Ser | Asp | Dab | Trp | Asp | Val | Val | Arg |
| $TriG_4$ | Val | Dab | Ala | Dab | Trp | Ser | Asp | Dab | Trp | Asp | Val | Val | Arg |
| $TriM_1$ | Gly | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Ile | Glu | Ile | Ile | Ser |
| $TriM_5$ | Gly | Dab | Gly | Ser | Phe | Ser | Dab | Dab | Ile | Glu | Ile | Val | Ser |
| $TriM_6$ | Gly | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Ile | Glu | Ile | Val | Ser |
| $TriM_7$ | Gly | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Val | Glu | Ile | Ile | Ser |
| $TriM_8$ | Gly | Dab | Gly | Ser | Phe | Ser | Dab | Dab | Ile | Glu | Ile | Ile | Ser |
| $TriM_{11}$ | ND | ND | ND | ND | ND | ND | ND | Dab | Phe | Glu | Ile | Ile | Ser |

[D] Modules incorporate a D-amino acid. ND, not determined.

## Discovery of tridecaptin BGCs with unusual architecture

Some of the detected BGCs differed from the reference tridecaptin $A_1$ not only by the predicted amino acid composition but also by their architecture. These truncated BGCs contained the ten-modular NRPS gene corresponding to the *triD* gene but lacked the terminal 3-module *triE* gene (Figure 1C). Importantly, these truncated tridecaptin clusters were not predicted to encode the TE domain that is usually located in the TriE synthase and ensures the release of the final product from its carrier protein. Besides the genome sequences being of high quality, the truncated ten-module tridecaptin clusters always co-occurred with a full-length BGC in the same genome that was predicted to encode a different tridecaptin (Figure 1C). The seven remaining ten-modular tridecaptin BGCs, which were not flagged by BiG-SCAPE as part of this family, were also co-localized either with full-length tridecaptin BGCs or with another truncated tridecaptin cluster that only contained a predicted *triE* ortholog.

Interestingly, the predicted amino acid scaffolds for some of the truncated tridecaptins perfectly matched the N-terminal parts of the scaffolds of previously isolated tridecaptins $B_1$/$M_1$. These truncated BGCs always co-localized with the full-length BGCs predicted to produce tridecaptin $A_1$ (*Paenibacillus* sp. S25, *P. peoriae* ZBSF16, *P. polymyxa* R 4.24), tridecaptin $A_5$ (*P. polymyxa* J, *P. polymyxa* EB4 G3, P. *polymyxa* MEZ6, *P. polymyxa* R 6.14, *P. polymyxa* HY96-2, *P. polymyxa* SQR-21) or tridecaptin $M_{11}$ (*Paenibacillus* sp. M-152, *Paenibacillus* sp. lzh-N1, *P. polymyxa* CJX518). The authors who previously isolated these compounds ascribed the production of tridecaptins $B_1$/$M_1$ to full-length tridecaptin BGCs within the genomes of their producing strains, *P. polymyxa* NRRL B-30507 and *Paenibacillus* sp. M-152, respectively (Cochrane *et al.*, 2015, Jangra *et al.*, 2019b). However, both groups noticed discrepancies between the adenylation domain predictions and the observed scaffolds for these full-length clusters, which instead corresponded to tridecaptin $A_1$ (Table S3) and the putative tridecaptin $M_{11}$ (Table S4), respectively. Specifically, in *P. polymyxa* NRRL B-30507, the greatest discrepancies were observed for adenylation domains 1 and 9, which the authors predicted to select valine and phenylalanine, respectively, rather than the glycine and isoleucine, which are found in the peptide scaffolds of tridecaptin $B_1$ and $M_1$ (Tables S3, S4). Similarly, glycine (position 1) and isoleucine (position 9)-recognizing adenylation domains are found in the TriD peptide synthase of the truncated tridecaptin cluster in the genome of *Paenibacillus* sp. M-152 (Table S2).

We hypothesized that the truncated BGC in *Paenibacillus* sp. M-152 might be involved in the chimeric biosynthesis of tridecaptin $M_1$ by biosynthesizing the first ten amino acids, upon which the last three are incorporated by the TriE synthase encoded by the full-length tridecaptin $M_{11}$ BGC (Figure 1C; Table S4). This seems particularly likely considering that both tridecaptins $M_1$ and $M_{11}$ were isolated from this strain. At this point, we strongly suspected that the tridecaptin $B_1$ biosynthesis in *P. polymyxa* NRRL B-30507 was also chimeric, especially considering the wide-spread co-occurrence of full-length and

truncated BGCs among *Paenibacillus* genomes that would encode for tridecaptin $A_1$ and $B_1$, respectively. Unfortunately, the genome of this strain is too heavily fragmented to confirm that it also harbors a truncated tridecaptin BGC that might be responsible for the chimeric biosynthesis of tridecaptin $B_1$.

## Screening of tridecaptin B producers using large-scale metabolomics

To strengthen our hypothesis that chimeric biosynthesis also occurs for tridecaptin A/$B_1$-producing strains, we set out to identify and sequence a *Paenibacillus* strain that produced both tridecaptin A and B. For this purpose, we screened a collection of 227 plant-associated *Paenibacillus* isolates (see Materials and Methods) for producers of tridecaptin B. To accomplish this, the isolates were grown in 96 well plates, and extracts of the cultures were analyzed using liquid chromatography coupled with mass spectrometry (LC-MS). A mass features having an *m/z* of 1458.8315, 729.9194 or 486.9487 corresponding to an $[M+H]^+$, $[M+2H]^{2+}$ or $[M+3H]^{3+}$ ions of tridecaptin $B_1$, respectively; were searched in the obtained LC-MS dataset. Eventually, a triply charged ion peak having an *m/z* of 486.9487, eluting at a retention time of 5.09 min, was detected and annotated as tridecaptin $B_1$ based on its MS/MS spectrum assignment (Figure 2). We found that tridecaptin $B_1$ was produced by seven *Paenibacillus* isolates, namely *Paenibacillus* sp. JJ-195, *Paenibacillus* sp. JJ-1650, *Paenibacillus* sp. JJ-333, *Paenibacillus* sp. JJ-1683, *Paenibacillus* sp. JJ-1604, *Paenibacillus* sp. JJ-1667, and *Paenibacillus* sp. JJ-1669. Importantly, all tridecaptin $B_1$ producers also produced the previously discovered tridecaptin $A_5$, which was characterized by the MS/MS analysis of its triply charged ion peak having an *m/z* of 539.9715 (Figure 2). Interestingly, a total of 28 tridecaptin $A_5$ producers were detected in the collection; therefore, 21 strains exclusively produced tridecaptin $A_5$.

Of the isolates that produced both tridecaptin $B_1$ and tridecaptin $A_5$, *Paenibacillus* sp. JJ-1683 was selected for further analysis to assess whether the production of tridecaptin $B_1$ could potentially be encoded by two collaborating BGCs involved in chimeric biosynthesis. The genome of *Paenibacillus* sp. JJ-1683 was sequenced using the PacBio platform. Assembly of the PacBio reads with Falcon (version 1.8.1) (Lohans *et al.*, 2014) resulted in a single contig of 5.8 Mb. The *Paenibacillus* sp. JJ-1683 genome was analyzed using antiSMASH 6.0.1 (Blin *et al.*, 2021) to obtain an overview of the predicted BGCs. Indeed, two tridecaptin BGCs were found in the genome of *Paenibacillus* sp. JJ-1683 (Table S2) that produced tridecaptin $B_1$, together with tridecaptin $A_5$. The first, full-length, BGC encoded both ten-modular TriD and three-modular TriE synthases, while the second BGC was truncated and encoded exclusively a ten-modular TriD ortholog. Conversely, *Paenibacillus* sp. JJ-21 harbors solely the full-length tridecaptin BGC predicted to encode the production of tridecaptin $A_5$ (Table S2). Notably, tridecaptin $A_5$ was detected in the extract of this strain, whereas tridecaptin $B_1$ was not found (Figure 2A).

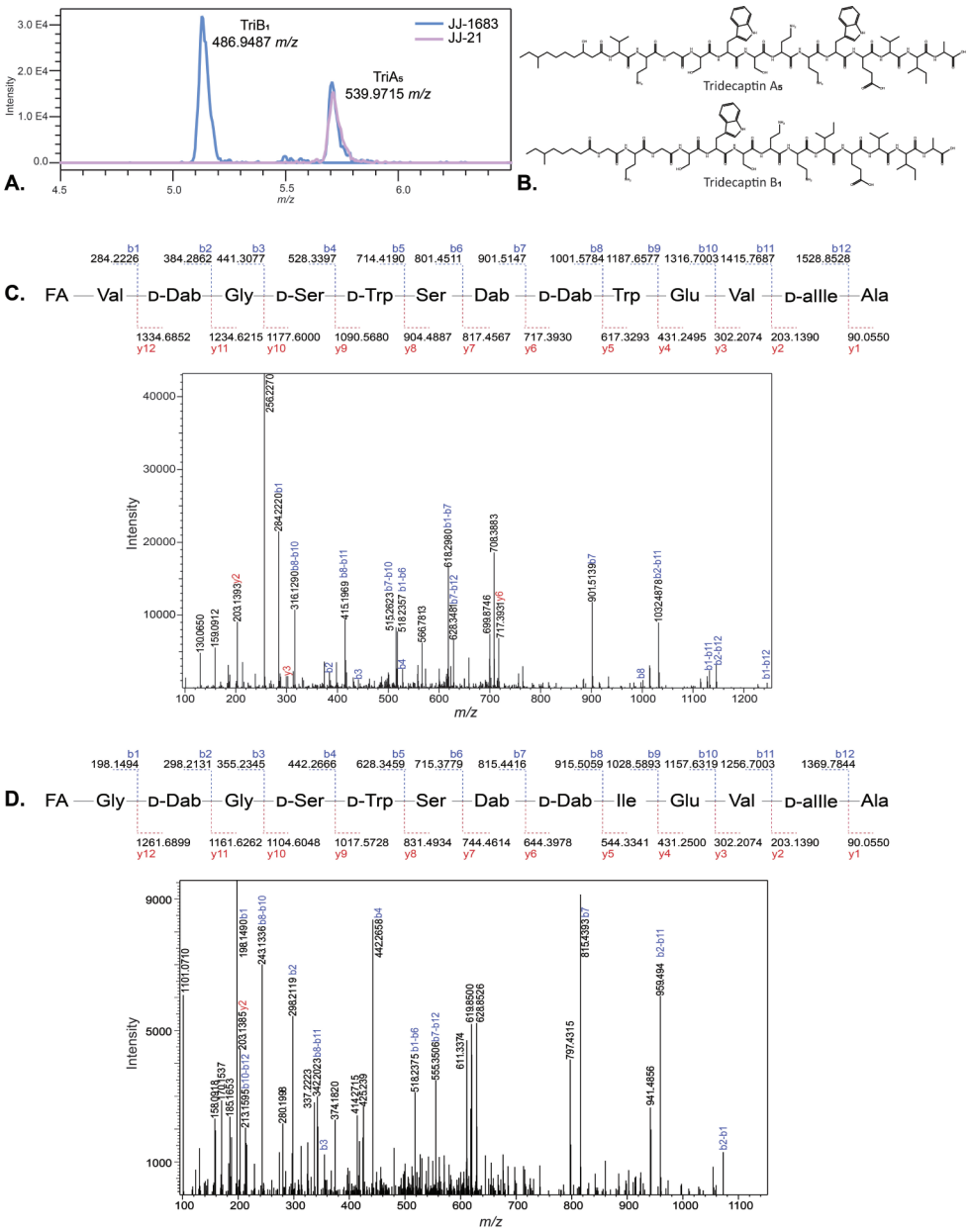



**Figure 2. Extracted ion chromatograms (EICs) (A) and chemical structures (B) of tridecaptin $A_5$ (539.9715 *m/z*) and $B_1$ (486.9487 *m/z*).** Note that, *Paenibacillus* sp. JJ-1683 produced both tridecaptin $B_1$ and tridecaptin $A_5$, while *Paenibacillus* sp. JJ-21 produced solely tridecaptin $A_5$. The assignment of the sequence of amino acid residues of tridecaptin $A_5$ **(C)** and tridecaptin $B_1$ **(D)** was conducted based on the mass differences between the consecutive *y* and *b* ions.

Production of tridecaptin $B_1$ exclusively by the strain that harbors two tridecaptin BGCs strongly suggests that both the truncated and full-length clusters are functional and required for the biosynthesis of this hybrid tridecaptin (Figure 3). This is further supported by the widespread co-occurrence of full-length and truncated tridecaptin BGCs: we found evidence of chimeric tridecaptin biosynthesis in 18 *Paenibacillus* strains in total (Table S2). Out of all 61 genomes in which tridecaptin BGCs were detected, this represents around 25% of all putative tridecaptin producers. A partially related mechanism has been shown previously for the biosynthesis of fungal NRPs, in which a single NRPS is involved in two separate biosynthesis pathways (Zhao *et al.*, 2016). Some families of *Pseudomonas* lipopeptides display a "split" BGC configuration, with the initiator NRPS gene at a genomic location distant from the operon with the other NRPS genes (Cesa-Luna *et al.*, 2023).

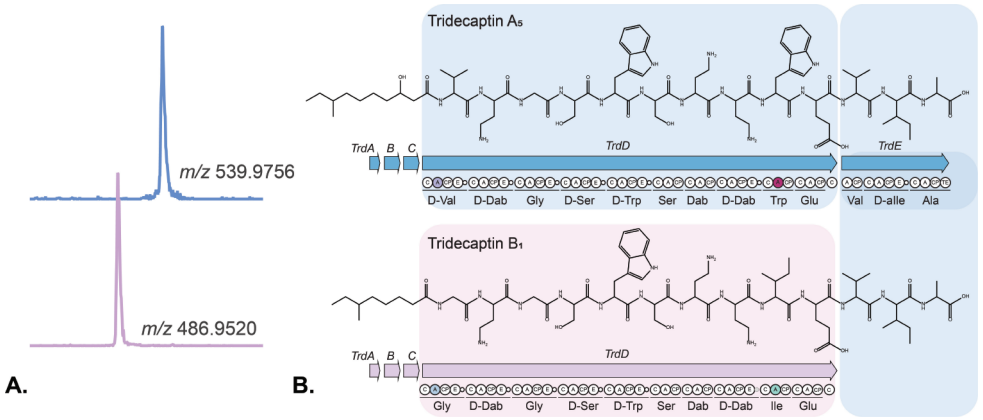



**Figure 3. EICs of tridecaptin $A_5$ (blue) and tridecaptin $B_1$ (pink) co-produced by *Paenibacillus* sp. JJ-1683 (A) and proposed biosynthetic pathway of tridecaptin $B_1$ (B).**

## Bioactivity profile of synthetic tridecaptins $A_5$ and $B_1$ analogs

Chimeric biosynthesis is a meaningful mechanism of bacterial and fungal secondary metabolism, which has the potential to further increase the structural diversity of encoded molecules and subsequent biological activity. To assess whether tridecaptin diversity caused by chimeric biosynthesis also translates to a strain's potential for functional diversity, we were also motivated to test tridecaptins $A_5$ and $B_1$ for bioactivity against a panel of common human pathogens.

Due to the limited production of tridecaptins $A_5$ and $B_1$ by *Paenibacillus* sp. JJ-1683, we developed synthetic analogs for both lipopeptides. Previous studies have shown that the natural N-terminal lipid tail of tridecaptin $A_1$ can be replaced with octanoic acid without impacting the antimicrobial properties of the peptide (Lohans *et al.*, 2014). Thus, to simplify the synthetic procedure, the chiral lipid tails found in the natural lipopeptides were replaced with commercially available octanoic acid, yielding the synthetic tridecaptin analogs Oct-Tri$A_5$ and Oct-Tri$B_1$ (Table S5, Figures S1, S2). As we previously reported, the synthetic

analogs of tridecaptin $A_5$, Oct-Tri$A_5$, have unusual broad-spectrum activity against Gram-positive and Gram-negative ESKAPE pathogens (**Chapter 7**). On the other hand, Oct-Tri$B_1$ displays moderate activity against Gram-negative strains and is not active against Gram-positive bacteria (Table 2). The selective activity observed for Oct-Tri$B_1$ is consistent with the antimicrobial spectrum known for several natural tridecaptins (i.e., tridecaptins A, C, and M) (Bann *et al.*, 2021, Jangra *et al.*, 2019b).

**Table 2. *In vitro* minimum inhibitory concentrations (MICs) of Oct-Tri$B_1$ and Oct-Tri$A_5$.**

| Strain | | Species | MIC[a] | | |
|---|---|---|---|---|---|
| | | | Oct-Tri$B_1$ | Oct-Tri$A_5$ | Colistin |
| BV18 | UAMS-1625 | *S. aureus* | > 64 | 8 | > 64 |
| BV249 | ATCC 29213 | *S. aureus* | > 64 | 8 | > 64 |
| BV1402 | 1126387 | *S. aureus* | > 64 | 8 | > 64 |
| BV29 | ATCC 29212 | *E. faecalis* | > 64 | 8 | > 64 |
| BV99 | ATCC 25922 | *E. faecalis* | > 64 | 8 | > 64 |
| BV144 | IHMA 890472 | *E. faecium* | > 64 | 8 | > 64 |
| BV159 | ATCC 51559 | *E. faecium* | > 64 | 8 | > 64 |
| BV94 | IHMA 777621 | *A. baumannii* | > 64 | 4 | 64 |
| BV160 | ATCC 17978 | *A. baumannii* | > 64 | 4 | 1 |
| BV374 | HUMC1 | *A. baumannii* | > 64 | 8 | 1 |
| BV23 | ATCC 25922 | *E. coli* | 16 | 2 | 1 |
| BV1475 | 1220120 | *E. coli* | 16 | 4 | 0.5 |
| BV1469 | 1214245 | *E. coli* | 16 | 4 | 32 |
| BV306 | ATCC 13883 | *K. pneumoniae* | 8 | 4 | 1 |
| BV1445 | 1227947 | *K. pneumoniae* | 8 | 4 | 0.5 |
| BV1447 | 1228586 | *K. pneumoniae* | 4 | 4 | 64 |
| BV34 | PAO1 | *P. aeruginosa* | >64 | 16 | 2 |
| BV1551 | 1226072 | *P. aeruginosa* | 16 | 8 | 0.5 |
| BV1544 | 1218019 | *P. aeruginosa* | >64 | 4 | > 64 |
| Hemolysis EC50 [µg/mL] | | | >128 | 6.9 | n/a |
| IC50 on HepG2 w/o FCS [µg/mL] | | | >128 | 11.4 | n/a |

[a] MIC values reported in units of µg/mL

In conclusion, this study is the first to reveal the biosynthetic potential of the *Paenibacillus* genus to produce hybrid tridecaptin antibiotics. Through BiG-SCAPE analysis we gained insight into the diversity of tridecaptin BGCs from *Paenibacillus* spp. and discovered new

tridecaptin BGCs including molecules with unusual biosynthetic architectures. Using LC-MS analysis and genome mining studies, we revised the biosynthetic pathway of tridecaptin $B_1$ and demonstrated that the biosynthesis of tridecaptin $B_1$ is hybrid and involves two separate BGCs. Tridecaptin $B_1$ assembly line provides an unusual example of chimeric biosynthesis of NRPS lipopeptide antibiotics.

# Materials and Methods

## Global genome mining of NRPSs

Genomes of all *Paenibacillus* spp. available from RefSeq (release 213) (O'Leary *et al.*, 2016) were downloaded from the NCBI FTP site. All genomes with more than 400 contigs were considered low-quality assemblies and were removed from the collection, leaving a set of 785 *Paenibacillus* genomes. All genomes were analyzed using AntiSMASH (version 6.0.1) (Blin *et al.*, 2021) to obtain biosynthetic gene cluster predictions. These predictions were then used as input for BiG-SCAPE (version 1.1.4) (Navarro-Muñoz *et al.*, 2020) analysis with a distance matrix cutoff set to 0.25. The resulting network of gene cluster families was visualized using Cytoscape (3.9.1) (Shannon *et al.*, 2003).

To detect tridecaptin BGCs in our dataset, we used the adenylation domain predictor PARAS (v0.0.1, available at https://github.com/BTheDragonMaster/parasect) to extract adenylation domains and predict their substrate specificity. PARAS is an adenylation domain predictor that uses an HMM to extract the A-domain active site and was trained on over 2800 adenylation domains from fungi and bacteria. Importantly, PARAS also reports the confidence score of a prediction. The *triD* and *triE* NRPS genes are always positioned on the same DNA strand, with typically only 165 bp between them. Keeping this in mind, we identified putative tridecaptin NRPS enzymes by sorting A-domain containing proteins into 'protein blocks': proteins that are encoded side by side on the genome on the same strand, with a maximum of 1000 bp between them. Then, we counted the adenylation domains for each protein block, and only considered those protein blocks that had ten or more adenylation domains. We chose a cutoff of ten instead of thirteen to also capture potentially truncated clusters, some of which we had detected with our BiG-SCAPE analysis. Then, we predicted the adenylation domains of all remaining protein blocks with PARAS (v0.0.1), and manually selected putative tridecaptin-producing BGCs based on predicted substrate selectivity and epimerization domain presence in modules 1-5, 8, and 12 as predicted by antiSMASH (Blin *et al.*, 2023).

Finally, for the three genomes, that contained a truncated ten-modular tridecaptin BGC but not a full-length tridecaptin BGC, we manually inspected the antiSMASH results to search for NRPS BGCs encoding fewer than ten NRPS modules to see if these might contribute to tridecaptin biosynthesis as well. We found three three-modular BGCs with an epimerization domain in their second module (corresponding to module 12 in full-length tridecaptin BGCs) and predicted their adenylation domain specificity as previously described.

## Growth of *Paenibacillus* spp. and specialized metabolite extraction

Our method for cultivation and extraction of specialized metabolites was adapted from previously described procedures (Nguyen *et al.*, 2016)2016. Briefly, 227 *Paenibacillus* spp. from the Auburn University Plant-Associated Microbial strain collection were inoculated into 600 µL of Tryptic Soy Agar (TSA, Bacto Soybean-Casein Digest Medium, 30 g/L) in 2.0 mL 96-deep well plates (Thermo Scientific, Nunc 2.0 mL DeepWell Plate) and incubated at

30 °C for 72 h. The cultures were extracted with 300 µL 100% isopropanol acidified with 0.1% formic acid. The plates were resealed with the same 96 Well-Cap Mats, sonicated for 10 min, and extracted for an additional 50 min. 200 µL of these crude extracts were transferred into a pre-washed 96 well plate (Agilent Technologies, 96 well plates, 0.5 mL, polypropylene) and lyophilized to dryness. The extraction protocol was repeated once more for a total extract volume of 400 µL. Dried samples were redissolved in 160 µL of methanol: water (1:1 v/v).

## LC-MS/MS analysis

For LC-MS analyses 1 µL of crude extract was injected into a Shimadzu Nexera X2 UHPLC system coupled to a Shimadzu 9030 QTOF mass spectrometer, and data acquisition was performed as previously described (van Bergeijk *et al.*, 2022). Briefly, all the samples were analyzed in positive polarity, using data-dependent acquisition mode. In this regard, full scan MS spectra (*m/z* 100–1700, scan rate 10 Hz, ID enabled) were followed by two data-dependent MS/MS spectra (*m/z* 100–1700, scan rate 10 Hz, ID disabled) for the two most intense ions per scan. The ions were fragmented using collision-induced dissociation (CID) with fixed collision energy (CE 20 eV) and excluded for 1 s before being re-selected for fragmentation. The parameters used for the ESI source were: interface voltage 4 kV, interface temperature 300 °C, nebulizing gas flow 3 L/min, and drying gas flow 10 L/min.

## Genome sequencing, assembly and annotation of *Paenibacillus* sp. JJ-1683

*Paenibacillus* sp. JJ-1683 and *Paenibacillus* sp. JJ-21 were grown in TSB (Bacto Soybean-Casein Digest Medium, 30 g/L) at 30 °C and 220 rpm for 24 h. DNA was extracted from *Paenibacillus* sp. JJ-1683 and *Paenibacillus* sp. JJ-21 as described (Kieser, 2000). DNA quality was verified by agarose gel electrophoresis. PacBio sequencing and assembly was performed by Novogene (UK). Generally, the library was prepared using a 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) (Chin *et al.*, 2014). BGCs in this genome were annotated using AntiSMASH (version 6.0.1) (Blin *et al.*, 2021).

## Solid phase peptide synthesis

Peptides were prepared on 0.25 mmol scale on either preloaded Wang resin or CTC resin. The synthesis of Oct-TriB was performed on preloaded Fmoc-Ala-Wang resin (0.29 mmol/g loading). All couplings with the exception of Fmoc-d-*allo*-Ile-OH and Fmoc-d-Ile-OH were performed using 4 eq. of amino acid or fatty acid, benzotriazol-1-yloxytris(dimethylamino) phosphonium hexafluorophosphate (BOP) (4 eq.) and N,N-diisopropylethylamine (DiPEA) (8 eq.) in 10 mL of DMF for 1 h at RT, under nitrogen flow. Fmoc-d-allo-Ile and Fmoc-d-Ile-OH were coupled by treating the resin with 2 eq. of the amino acid, 2 eq. of BOP and 4 eq. of DiPEA in 10 mL of DMF overnight at RT. Fmoc group removal was performed by treating the resin with 10 mL of piperidine : DMF (1 : 4, v/v) for 5 min and then again for 15 min.

Final sidechain deprotection and cleavage from the resin was carried out by treating the resin with 10 mL of TFA : TIPS : $H_2O$ (95 : 2.5 : 2.5, v/v) for 90 min. The reaction mixture was filtered through cotton, the filtrate was precipitated in MTBE : petroleum ether (1 : 1, v/v) and centrifuged (4500 rpm, 5 min). The pellet was then resuspended in MTBE : petroleum ether (1 : 1, v/v) and centrifuged again (4500 rpm, 5 min). Finally the pellet containing the crude lipopeptide was dissolved in tBuOH:$H_2O$ (1 : 1, v/v) and lyophilized overnight. The crude mixtures were subsequently purified by RP-HPLC. Fractions were assessed by HPLC and LC-MS and product containing fractions were pooled, frozen and lyophilized to yield the pure lipopeptides in >95% purity (determined by HPLC).

## Peptide synthesis

Fmoc-L-Dab(Boc)-OH, Fmoc-D-Dab(Boc)-OH, Fmoc-D-Ile-OH, Fmoc-allo-Ile-OH and 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate (HATU) were purchased from Combi-Blocks. All other Fmoc-amino acids, the Fmoc-Ala-Wang resin and Fmoc-Glu(OtBu)-Wang resin were purchased from P3 BioSystems. 2-Chlorotrityl chloride resin (CTC) was purchased from Iris Biotech. Octanoic acid was purchased from Alfa Aesar. ((1H-Benzo[d][1,2,3]triazol-1-yl)oxy)tris(dimethylamino) phosphonium hexafluorophosphate (BOP), N,N-Diisopropylcarbodiimide (DIC) and triisopropylsilane (TIPS) were purchased from Manchester Organics. Diisopropylethylamine (DIPEA), piperidine, trifluoroacetic acid (TFA) and dimethyl sulfoxide (DMSO) were purchased from Carl Roth. Dichloromethane ($CH_2Cl_2$) and petroleum ether were purchased from VWR Chemicals. Acetonitrile (MeCN), dimethylformamide (DMF) and methyl tertiary-butyl ether (MTBE) were purchased from Biosolve.



*Resin swelling*

The resin was swollen in 10 mL of DMF for 300 s prior to the first coupling.

*Automated coupling protocol*

| Step | Function | Duration/Temperature |
|---|---|---|
| 1 | Initial Deprotection N-terminus | 15 s at 60°C then 30 s at 70°C |
| 2 | Deprotection N-terminus | 15 s at 60°C then 180 s at 70°C |
| 3 | Wash (DMF) | RT |
| 4 | Wash (DMF) | RT |
| 5 | Wash (DMF) | RT |
| 6 | Coupling amino acid | 15 s at 60°C then 300 s at 70°C |

After coupling of the final residue on the synthesizer, the resin was washed with DCM, filtered and treated with 3 mL of TFA : TIPS : $H_2O$ (95 : 2.5 : 2.5, v/v) for 90 min. The reaction mixture was filtered through cotton, the filtrate was precipitated in MTBE : petroleum ether (1 : 1, v/v) and centrifuged (4500 rpm, 5 min). The pellet was then resuspended in MTBE : petroleum

ether (1 : 1, v/v) and centrifuged again (4500 rpm, 5 min). Finally the pellet containing the crude lipopeptide was dissolved in tBuOH : $H_2O$ (1 : 1, v/v) and lyophilized overnight. The crude mixtures were subsequently purified by RP-HPLC. Fractions were assessed by HPLC and LC-MS and product containing fractions were pooled, frozen and lyophilized to yield the pure lipopeptides in >95% purity (determined by HPLC).

*Prep RP-HPLC purification*

Peptides were purified using a BESTA-Technik system with a Dr. Maisch Reprosil Gold 120 $C_{18}$ column (25 × 250 mm, 10 µm) and equipped with a ECOM Flash UV detector monitoring at 214 nm and 254 nm. The following solvent system, at a flow rate of 12 mL/min, was used: solvent A, 0.1 % TFA in water/acetonitrile 95/5; solvent B, 0.1 % TFA in water/acetonitrile 5/95. Gradient elution was as follows: 100:0 (A/B) for 5 min, 100:0 to 50:50 (A/B) over 50 min, 50:50 to 0:100 (A/B) for 3min, then reversion back to 100:0 (A/B) over 1 min, 100:0 (A/B) for 5 min.

*Analytical RP-HPLC*

Shimadzu Prominence-i LC-2030 system with a Dr. Maisch ReproSil Gold 120 $C_{18}$ column (4.6 × 250 mm, 5 µm) at 30 °C and equipped with a UV detector monitoring 214 nm and 254 nm. The following solvent system, at a flow rate of 1 mL/min, was used: solvent A, 0.1 % TFA in water/acetonitrile 95/5; solvent B, 0.1 % TFA in water/acetonitrile 5/95. Method A: Gradient elution was as follows: 100:0 (A/B) for 2 min, 100:0 to 0:100 (A/B) over 23 min, 0:100 (A/B) for 1 min, 0:100 (A/B) then reversion back to 100:0 (A/B) over 1 min, 100:0 (A/B) for 3 min. Method B: Gradient elution was as follows: 100:0 (A/B) for 2 min, 100:0 to 50:50 (A/B) over 45 min, 50:50 (A/B) to 0:100 (A/B) over 1 min, 0:100 (A/B) for 6 min then reversion back to 100:0 (A/B) over 1min, 100:0 (A/B) for 5 min.

*HRMS analysis*

HRMS analyses were performed on a Thermo Scientific Dionex UltiMate 3000 HPLC system with a Phenomenex Kinetex $C_{18}$ (2.1 x 150 mm, 2.6 µm) column at 35 °C and equipped with a diode array detector. The following solvent system, at a flow rate of 0.3 mL/min, was used: solvent A, 0.1 % formic acid in water; solvent B, 0.1% formic acid in acetonitrile. Gradient elution was as follows: 95:5 (A/B) for 1 min, 95:5 to 5:95 (A/B) over 9 min, 5:95 to 2:98 (A/B) over 1 min, 2:98 (A/B) for 1 min, then reversion back to 95:5 (A/B) over 2 min, 95:5 (A/B) for 1 min. This system was connected to a Bruker micrOTOF-Q II mass spectrometer (ESI ionization) calibrated internally with sodium formate.

### Antimicrobial testing

All minimum inhibitory concentrations were determined according to Clinical and Standards Laboratory Institute (CLSI) guidelines. In brief, agar plates were inoculated from glycerol stocks and incubated overnight at 35 °C. Peptides DMSO stocks were dispensed into 96-well round bottom plates using a TECAN D300e dispenser. Colistin sulfate (Sigma-Aldrich) was dissolved in water and serially diluted in cation-adjusted Mueller Hinton broth (10 µL per well). The final concentrations of the test compounds were 64–0.06 µg/mL. Inocula were prepared by direct colony suspension in NaCl at 0.5 MacFarland, which was further diluted 200-fold in cation-adjusted Mueller Hinton broth for a target inoculum of 2 x $10^5$ colony forming units/mL. 100 µL of the prepared inoculum was added to each well of the prepared plates containing the test compounds. Plates were sealed with parafilm and incubated for 20 h at 35 °C. Images of all plates were recorded and MIC was assessed visually.

## Acknowledgments

The work was supported by the NACTAR program of The Netherlands Organization for Scientific Research (NWO), Grant 16440 to G.P.vW., M.H.M. and N.I.M.

# Supplementary information

**Table S1. Active site conservation among *Paenibacillus* tridecaptin adenylation domains.** Active sites were extracted using PARAS. Residues that deviate from the most commonly found active site for that substrate are indicated in bold. Confidence shows the average confidence of PARAS predictions for that specific active site. The number of domains with that active site among tridecaptin BGCs is also given. Rows shown in color correspond to active sites found in the BGCs of *Paenibacillus sp.* M-152 and *Paenibacillus polymyxa* NRRL B-30507.

| (Predicted) substrate | Active site signature | Confidence | Number of domains |
|---|---|---|---|
| Valine | DAFWLGGTFK | 1.00 | 95 |
| | DAFWIGGTFK | 0.97 | 4 |
| | DAFWLGGAFK | 0.77 | 1 |
| | DAFFIGATFK | 0.36 | 1 |
| 2,4-diaminobutyric acid | DVGEISSIDK | 1.00 | 200 |
| | DVGEISAIDK | 0.96 | 10 |
| | DVGETGTIDK | 0.43 | 1 |
| Alanine | DVFWLGGTFK | 0.77 | 48 |
| | DVTNFAIIYK | 0.81 | 2 |
| Tryptophan | DAWAFAGVAK | 0.95 | 83 |
| | DAWAVAGVAK | 0.80 | 22 |
| | DAWAFAGIAK | 0.80 | 7 |
| Serine | DVWHFSLVDK | 1.00 | 147 |
| | DVWHYSLVDK | 0.99 | 2 |
| | DVWHFSMVEK | 0.55 | 1 |
| Aspartic acid | DMKDVGVVDK | 0.34 | 4 |
| Arginine | DPEDIGHIDK | 0.33 | 3 |
| Glycine | DILQMGMVWK | 1.00 | 81 |
| Glutamic acid | DAKDLGVVDK | 0.95 | 67 |
| Isoleucine | DAFFLGITFK | 0.95 | 47 |
| | DAFFLGVTFK | 0.91 | 25 |
| | DAFFLGATFK | 0.42 | 4 |
| | DGFFLGVVYK | 0.81 | 2 |
| | DGFFLGVVFK | 0.85 | 1 |

| | | | |
|---|---|---|---|
| Phenylalanine | DAWTFAGVAK | 0.73 | 8 |
| | DAWTVAGVAK | 0.78 | 2 |
| | DAWTIAAICK | 0.92 | 2 |
| | DAWTIAGVAK | 0.75 | 2 |
| | DAWTVAAVAK | 0.87 | 1 |
| | DAWTFAAVAK | 0.88 | 1 |
| Glutamine | DAQDLGVVDK | 0.59 | 1 |
| Leucine | DAWLLGAIVK | 0.91 | 1 |

**Table S2. Predicted *Paenibacillus* tridecaptin biosynthetic gene clusters.**

In fifteen genomes, both a full-length (13-modular) and a truncated (10-modular) tridecaptin BGC can be found. In three genomes, two truncated (10-modular and 3-modular) tridecaptin BGCs can be found. Novel tridecaptins are highlighted in green. The tridecaptin BGC from the genomes of *Paenibacillus* sp. JJ-1683 and *Paenibacillus* sp. JJ-21 are highlighted in bold.

| Strain | Accession number | 1[D] | 2[D] | 3[D] | 4[D] | 5[D] | 6 | 7 | 8[D] | 9 | 10 | 11 | 12[D] | 13 | Compound |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *P. elgii* MER 157 | NZ_JAMAVM010000004.1. region001 | Val | Dab | Ala | Dab | Trp | Ser | Asp | Dab | Trp | Asp | Val | Val | Arg | Tridecaptin $G_4$ |
| *P. elgii* SWL-W8 | NZ_JAAIVH010000001.1. region006 | Val | Dab | Ala | Dab | Trp | Ser | Asp | Dab | Trp | Asp | Val | Val | Arg | Tridecaptin $G_4$ |
| *P. jamilae* KACC 10925 | NZ_QVPU01000010.1. region001 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Val | Ile | Ala | Tridecaptin $A_5$ |
| *P. kribbensis* 6hRe76 | NZ_MCBP01000006.1. region002 | Val | Dab | Ala | Ser | Trp | Ser | Dab | Dab | Phe | Glu | Val | Ile? | Ala | New Tridecaptin 1 |
| *P. kribbensis* AM49 | NZ_CP020028.1.region007 | Val | Dab | Ala | Ser | Trp | Gly | Ser | Dab | Phe | Glu | Val | Ile? | Ser | New Tridecaptin 2 |
| *P. kribbensis* PS04 | NZ_CP041731.1.region012 | Val | Dab | Ala | Ser | Trp | Gly | Ser | Dab | Phe | Glu | Val | Ile? | Ser | New Tridecaptin 2 |
| *P. mucilaginosus* K02 | NC_017672.3.region011 | Phe | Dab | Ile | Ser | Trp | Ser | Ser | Dab | Trp | Ser | Val | Ile | Dab | Tridecaptin D |
| *P. mucilaginosus* KNP414 | NC_015690.1.region011 | Phe | Dab | Ile | Ser | Trp | Ser | Ser | Dab | Trp | Ser | Val | Ile | Dab | Tridecaptin D |
| *P. peoriae* HS311 | NZ_CP011512.1.region010 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Val | Ile | Ala | Tridecaptin $A_5$ |
| *P. peoriae* KCTC 3763 | NZ_AGFX01000012.1. region001 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Val | Glu | Val | Ile | Ser | New Tridecaptin 3 |
| *P. peoriae* ZBSF16 | NZ_CP092831.1.region009 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Phe | Glu | Val | Ile | Ala | Tridecaptin $A_1/A_3$ |
| *P. peoriae* ZBSF16 | NZ_CP092831.1.region010 | Gly | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Ile | Glu | --- | --- | --- | Tridecaptin $B_1$ |
| *P. peoriae* ZF390 | NZ_CP061172.1.region008 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Phe | Glu | Val | Ile | Ala | Tridecaptin $A_1/A_3$ |
| *P. polymyxa* 2020 | NZ_CP049598.1.region019 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Ile | Ile | Ala | New Tridecaptin 4 |
| *P. polymyxa* A18 | NZ_JWJJ01000007.1.region001 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Val | Ile | Ala | Tridecaptin $A_5$ |
| *P. polymyxa* ATCC 15970 | NZ_CP011420.1.region010 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Val | Ile | --- | Tridecaptin $A_5$ |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *P. polymyxa* C12 | NZ_CP023711.1.region001 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Val | Ile | Ala | Tridecaptin $A_5$ |
| *P. polymyxa* CICC 10580 | NZ_JNCB01000011.1. region006 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Val | Ile | Ala | Tridecaptin $A_5$ |
| *P. polymyxa* CJX518 | NZ_CP029370.1.region007 | Val | Dab | Ser | Gly | Trp | Ser | Dab | Dab | Phe | Glu | Ile | Ile | Ser | New Tridecaptin 5 |
| *P. polymyxa* CJX518 | NZ_CP029370.1.region008 | Gly | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Ile | Glu | --- | --- | --- | Tridecaptin $M_1$ |
| *P. polymyxa* CR1 | NC_023037.2.region009 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Val | Ile | Ala | Tridecaptin $A_5$ |
| *P. polymyxa* DCB2-2 | NZ_RXYF01000012.1. region001 | --- | --- | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Val | Ile | Ala | Tridecaptin $A_5$ |
| *P. polymyxa* DSM 292 | NZ_OXKC02000014.1. region001 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Val | Ile | Ala | Tridecaptin $A_5$ |
| *P. polymyxa* DSM 36 | NZ_CP049783.1.region018 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Ile | Ile | Ala | New Tridecaptin 4 |
| *P. polymyxa* E681 | NZ_CP048793.1.region008 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Phe | Glu | Val | Ile | Ala | Tridecaptin $A_1/A_3$ |
| *P. polymyxa* EB4 G3 | NZ_CP084033.1.region008 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Val | Ile | Ala | Tridecaptin $A_5$ |
| *P. polymyxa* EB4 G3 | NZ_CP084033.1.region009 | Gly | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Ile | Glu | --- | --- | --- | Tridecaptin $B_1$ |
| *P. polymyxa* G25-124 | NZ_LDIE01000050.1. region001 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Ile | Ile | Ala | New Tridecaptin 4 |
| *P. polymyxa* HK4 | NZ_JAACZY010000021.1. region001 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Ile | Ile | Ala | New Tridecaptin 4 |
| *P. polymyxa* HY96-2 | NZ_CP025957.1.region010 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | --- | Val | Ile | Ala | Tridecaptin $A_5$ |
| *P. polymyxa* HY96-2 | NZ_CP025957.1.region011 | Gly | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Ile | Glu | --- | --- | --- | Tridecaptin $B_1$ |
| *P. polymyxa* J | NZ_CP015423.1.region015 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Val | Ile | Ala | Tridecaptin $A_5$ |
| *P. polymyxa* J | NZ_CP015423.1.region017 | Gly | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Ile | Glu | --- | --- | --- | Tridecaptin $B_1$ |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *P. polymyxa* KCCM 40454 | NZ_POVT01000002.1. region001 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Ile | Ile | Ala | New Tridecaptin 4 |
| *P. polymyxa* K 1.14 | NZ_CP097778.1.region009 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Phe | Glu | Val | Ile | Ala | Tridecaptin $A_1/A_3$ |
| *P. polymyxa* K 1.14 | NZ_CP097778.1.region011 | Ala | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Ile | Glu | --- | --- | --- | New Tridecaptin 6 |
| *P. polymyxa* LY214 | NZ_NOLA01000018.1. region001 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Val | Ile | Ala | Tridecaptin $A_5$ |
| *P. polymyxa* M1 | NC_017542.1.region008 | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | Ile | Ile | Ser | Tridecaptin $M_1$ |
| *P. polymyxa* M1 | NC_017542.1.region009 | Gly | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Ile | Glu | --- | --- | --- | Tridecaptin $M_1$ |
| *P. polymyxa* MEZ6 | NZ_CP086373.1.region009 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Val | Ile | Ala | Tridecaptin $A_5$ |
| *P. polymyxa* MEZ6 | NZ_CP086373.1.region010 | Gly | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Ile | Glu | --- | --- | --- | Tridecaptin $B_1$ |
| *P. polymyxa* NCTC10343 | NZ_UGSC01000001.1. region012 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Ile | Ile | Ala | New Tridecaptin 4 |
| *P. polymyxa* NRRL B-30509 | NZ_JTHO01000008.1. region001 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Val | Ile | Ala | Tridecaptin $A_5$ |
| *P. polymyxa* R 3.13 | NZ_CP097771.1.region013 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Val | Ile | Ala | Tridecaptin $A_5$ |
| *P. polymyxa* R 4.24 | NZ_CP097769.1.region005 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Phe | Glu | Val | Ile | Ala | Tridecaptin $A_1/A_3$ |
| *P. polymyxa* R 4.24 | NZ_CP097769.1.region006 | Gly | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Ile | Glu | --- | --- | --- | Tridecaptin $B_1$ |
| *P. polymyxa* R 4.5 | NZ_CP097770.1.region005 | Gly | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Ile | Glu | --- | --- | --- | Tridecaptin $B_1$ |
| *P. polymyxa* R 4.5 | NZ_CP097770.1.region007 | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | Val | Ile | Ala | Tridecaptin $B_1$ |
| *P. polymyxa* R 5.31 | NZ_CP097767.1.region001 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Val | Ile | Ala | Tridecaptin $A_5$ |
| *P. polymyxa* R 6.14 | NZ_CP097766.1.region009 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Val | Ile | Ala | Tridecaptin $A_5$ |
| *P. polymyxa* R 6.14 | NZ_CP097766.1.region010 | Gly | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Ile | Glu | --- | --- | --- | Tridecaptin $B_1$ |
| *P. polymyxa* SC2 | NC_014622.2.region008 | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | Ile | Ile | Ser | Tridecaptin $M_1$ |
| *P. polymyxa* SC2 | NC_014622.2.region009 | Gly | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Ile | Glu | --- | --- | --- | Tridecaptin $M_1$ |
| *P. polymyxa* SQR-21 | NZ_CP006872.1.region009 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Val | Ile | Ala | Tridecaptin $A_5$ |
| *P. polymyxa* SQR-21 | NZ_CP006872.1.region010 | Gly | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Ile | Glu | --- | --- | --- | Tridecaptin $B_1$ |
| *P. polymyxa* YC0573 | NZ_CP017968.3.region011 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Val | Ile | Ala | Tridecaptin $A_5$ |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *P. polymyxa* ZF129 | NZ_CP040829.1.region014 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Val | Ile | Ala | Tridecaptin $A_5$ |
| *P. polymyxa* ZJ-9 | NZ_CP092779.1.region010 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Ile | Ile | Ala | New Tridecaptin 4 |
| *Paenibacillus* sp. EKM211P | NZ_JAAMNS010000007.1. region001 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Ile | Ile | Ala | New Tridecaptin 4 |
| *Paenibacillus* sp. F4 | NZ_POVS01000007.1. region001 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Val | Ile | Ala | Tridecaptin $A_5$ |
| *Paenibacillus* sp. ICGEB2008 | NZ_AMQU01000029.1. region001 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Val | Ile | Ala | Tridecaptin $A_5$ |
| ***Paenibacillus* sp. JJ-1683** | **Paenibacillus sp._MBT_ JJ-1683.region008** | **Val** | **Dab** | **Gly** | **Ser** | **Trp** | **Ser** | **Dab** | **Dab** | **Trp** | **Glu** | **Val** | **Ile** | **Ala** | **Tridecaptin $A_5$** |
| ***Paenibacillus* sp. JJ-1683** | **Paenibacillus sp._MBT_ JJ-1683.region009** | **Gly** | **Dab** | **Gly** | **Ser** | **Trp** | **Ser** | **Dab** | **Dab** | **Ile** | **Glu** | --- | --- | --- | **Tridecaptin $B_1$** |
| ***Paenibacillus* sp. JJ-21** | **Paenibacillus sp._MBT_ JJ-21.region011** | **Val** | **Dab** | **Gly** | **Ser** | **Trp** | **Ser** | **Dab** | **Dab** | **Trp** | **Glu** | **Val** | **Ile** | **Ala** | **Tridecaptin $A_5$** |
| *Paenibacillus* sp. M-152 | NZ_CP034141.1.region008 | Val | Dab | Ser | Gly | Trp | Ser | Dab | Dab | Phe | Glu | Ile | Ile | Ser | New Tridecaptin 5 |
| *Paenibacillus* sp. M-152 | NZ_CP034141.1.region009 | Gly | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Ile | Glu | --- | --- | --- | Tridecaptin $M_1$ |
| *Paenibacillus* sp. PvR133 | NZ_JAFIBS010000001.1. region012 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Val | Ile | Ala | Tridecaptin $A_5$ |
| *Paenibacillus* sp. S02 | NZ_CP073682.1.region011 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Val | Ile | Ala | Tridecaptin $A_5$ |
| *Paenibacillus* sp. S25 | NZ_CP073683.1.region012 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Phe | Glu | Val | Ile | Ala | Tridecaptin $A_1/A_3$ |
| *Paenibacillus* sp. S25 | NZ_CP073683.1.region014 | Gly | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Ile | Glu | --- | --- | --- | Tridecaptin $B_1$ |
| *Paenibacillus* sp. UKAQ 18 | NZ_JAKFSF010000002.1. region001 | --- | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Phe | Glu | Val | --- | --- | Tridecaptin A/B/C |
| *Paenibacillus* sp. UNCCL52 | NZ_JMLR01000007.1. region003 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Val | Ile | Ala | Tridecaptin $A_5$ |
| *Paenibacillus* sp. cl130 CL130 | NZ_FOYG01000003.1. region001 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Trp | Glu | Val | Ile | Ala | Tridecaptin $A_5$ |
| *Paenibacillus* sp. lzh-N1 | NZ_CP025696.1.region011 | Gly | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Ile | Glu | --- | --- | --- | Tridecaptin $M_1$ |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Paenibacillus* sp. lzh-N1 | NZ_CP025696.1.region012 | Val | Dab | Ser | Gly | Trp | Ser | Dab | Dab | Phe | Glu | Ile | Ile | Ser | New Tridecaptin 5 |
| *P. terrae* sp. GHS.8NWYW.5 | NZ_VIYI01000003.1.region001 | Val | Dab | Val | Ser | Trp | Ser | Gln | Dab | Phe | Glu | Val | Val | Arg | New Tridecaptin 7 |
| *P. terrae* sp. HPL-003 | NC_016641.1.region014 | --- | Dab | Gly | Trp | Leu | Ser | Dab | Dab | Phe | Glu | Val | Ile | Ala | New Tridecaptin 8 |
| *P. terrae* sp. HPL-003 | NC_016641.1.region010 | Val | Dab | Gly | Ser | Trp | Ser | Val | Gln | Ala | --- | --- | --- | --- | New Tridecaptin 9 |
| *P. terrae* sp. NRRL B-30644 | NZ_JTHP01000234.1. region001 | Val | Dab | Gly | Ser | Trp | Ser | Dab | Dab | Phe | Glu | Val | Ile | Ala | Tridecaptin $A_1/A_3$ |

[D] Modules incorporate a D-amino acid. ND, not determined.

**Table S3. Bioinformatic analysis of tridecaptin BGC of *Paenibacillus polymyxa* NRRL B-30507.**

| Module number | Tridecaptin $B_1$ | Active site signature | Tridecaptin $A_5$ | Predicted aa (PARAS) | Predicted aa (Cochrane *et al.*, 2015) |
|---|---|---|---|---|---|
| 1 | Gly | DAFWLGGTFK | Val | Val (1.00) | Val |
| 2 | Dab | DVGEISSIDK | Dab | Dab (1.00) | Orn |
| 3 | Gly | DILQMGMVWK | Gly | Gly (1.00) | Gly |
| 4 | Ser | DVWHFSLVDK | Ser | Ser (1.00) | Ser |
| 5 | Trp | DAWAFAGVAK | Trp | Trp (0.94) | Phe |
| 6 | Ser | DVWHFSLVDK | Ser | Ser (1.00) | Ser |
| 7 | Dab | DVGEISSIDK | Dab | Dab (1.00) | Orn |
| 8 | Dab | DVGEISSIDK | Dab | Dab (1.00) | Orn |
| 9 | Ile | DAWAFAGVAK | Trp | Trp (0.94) | Phe |
| 10 | Glu | DAKDLGVVDK | Glu | Glu (0.96) | Glu |
| 11 | Val | DAFWLGGTFK | Val | Val (1.00) | Val |
| 12 | Ile | DAFFLGITFK | Ile | Ile (0.93) | Ile |
| 13 | Ala | DVFWLGGTFK | Ala | Ala (0.73) | Val |

The tridecaptin BGC in *P. polymyxa* NRRL B-30507 matches the structure of tridecaptin $A_5$, not that of tridecaptin $B_1$. For PARAS predictions, prediction confidence (scale from 0.0-1.0) is shown between brackets.

**Table S4. Bioinformatic analysis of tridecaptin BGC of *Paenibacillus* sp. M-152.**

The predicted scaffold for truncated tridecaptin $M_1$ matches the isolated structure from *Paenibacillus sp*. M-152. The postulated structure of tridecaptin $M_{11}$ is also shown, which matches with the full-length BGC. For PARAS predictions, prediction confidence (scale from 0.0-1.0) is shown between brackets.

| Module number | $M_1$ | Predicted aa truncated (PARAS) | Active site signature truncated | $M_{11}$ | Predicted aa full (PARAS) | Predicted aa full (Jangra *et al.*, 2019b) | Active site signature full |
|---|---|---|---|---|---|---|---|
| 1 | Gly | Gly (1.00) | DILQMGMVWK | ND | Val (1.00) | Ile/Val | DAFWLGGTFK |
| 2 | Dab | Dab (1.00) | DVGEISSIDK | ND | Dab (1.00) | Dab | DVGEISSIDK |
| 3 | Gly | Gly (1.00) | DILQMGMVWK | ND | Ser (1.00) | Ser | DVWHFSLVDK |
| 4 | Ser | Ser (1.00) | DVWHFSLVDK | ND | Gly (1.00) | Gly | DILQMGMVWK |
| 5 | Trp | Trp (0.95) | DAWAFAGVAK | ND | Trp (0.95) | Trp | DAWAFAGVAK |
| 6 | Ser | Ser (1.00) | DVWHFSLVDK | ND | Ser (1.00) | Ser | DVWHFSLVDK |
| 7 | Dab | Dab (1.00) | DVGEISSIDK | ND | Dab (1.00) | Dab | DVGEISSIDK |
| 8 | Dab | Dab (1.00) | DVGEISSIDK | Dab | Dab (1.00) | Dab | DVGEISSIDK |
| 9 | Ile | Ile (0.98) | DAFFLGITFK | Phe | Phe (0.75) | Phe | DAWTIAGVAK |
| 10 | Glu | Glu (0.95) | DAKDLGVVDK | Glu | Glu (0.95) | Glu | DAKDLGVVDK |
| 11 | Ile | | | Ile | Ile (0.95) | Ile/aIle | DAFFLGITFK |
| 12 | Ile | | | Ile | Ile (0.94) | Ile/aIle | DAFFLGVTFK |
| 13 | Ser | | | Ser | Ser (1.00) | Ser | DVWHFSLVDK |

**Table S5. HPLC and HRMS analysis of the synthetic peptides prepared in this study.**

| Peptide | Name | Chemical formula | Calcd. exact mass | Mass found | Calcd. | Overall yield [%] |
|---|---|---|---|---|---|---|
| 1 | Oct-Tri$B_1$ | $C_{66}H_{109}N_{17}O_{19}$ | 1443.8086 | 722.9122 $[M+2H]^{2+}$ | 722.9116 | 27 |
| 2 | Oct-Tri$A_5$ | $C_{74}H_{114}N_{18}O_{19}$ | 1558.8508 | 780.4332 $[M+2H]^{2+}$ | 780.4327 | 23 |

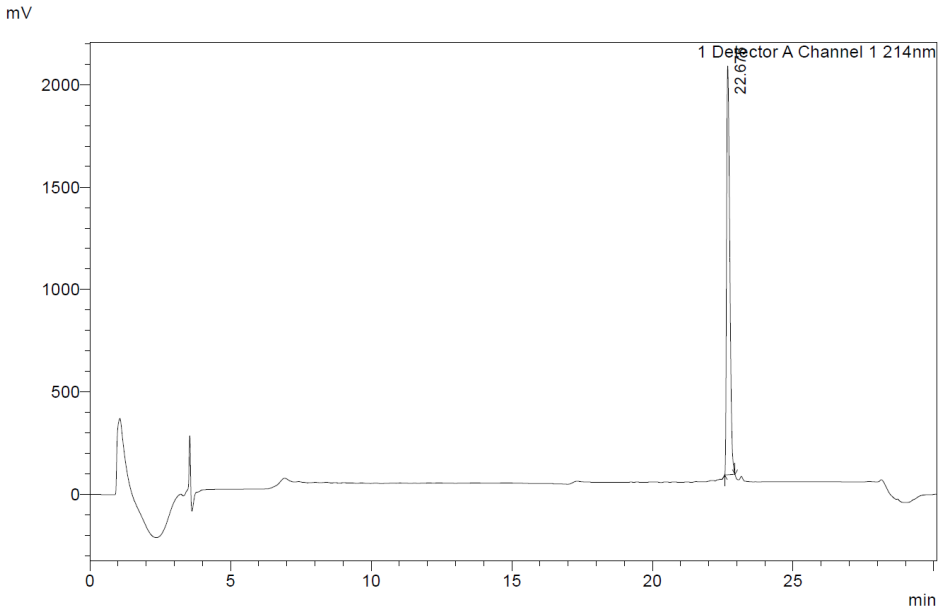



**Figure S1. HPLC trace showing the reinjection of purified Oct-TriB$_1$.** The peptide eluted as a single peak at 22.675 min using the HPLC method B.

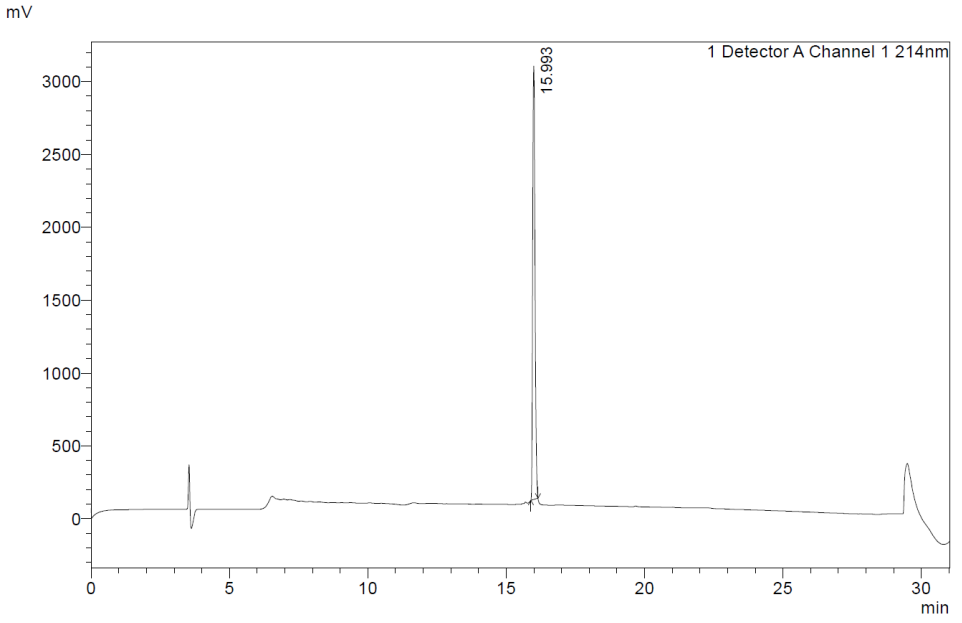



**Figure S2. HPLC trace showing the reinjection of purified Oct-TriA$_5$.** The peptide eluted as a single peak at 15.993 min using the HPLC method A.