

RESEARCH ARTICLE

Open Access



Precursor peptide-targeted mining of more than one hundred thousand genomes expands the lanthipeptide natural product family

Mark C. Walker^{1,2*} , Sara M. Eslami² , Kenton J. Hetrick² , Sarah E. Ackenhusen² , Douglas A. Mitchell^{2,3,4}  and Wilfred A. van der Donk^{2,3,5} 

Abstract

Background: Lanthipeptides belong to the ribosomally synthesized and post-translationally modified peptide group of natural products and have a variety of biological activities ranging from antibiotics to antinociceptives. These peptides are cyclized through thioether crosslinks and can bear other secondary post-translational modifications. While lanthipeptide biosynthetic gene clusters can be identified by the presence of genes encoding characteristic enzymes involved in the post-translational modification process, locating the precursor peptides encoded within these clusters is challenging due to their short length and high sequence variability, which limits the high-throughput exploration of lanthipeptide biosynthesis. To address this challenge, we enhanced the predictive capabilities of Rapid ORF Description & Evaluation Online (RODEO) to identify members of all four known classes of lanthipeptides.

Results: Using RODEO, we mined over 100,000 bacterial and archaeal genomes in the RefSeq database. We identified nearly 8500 lanthipeptide precursor peptides. These precursor peptides were identified in a broad range of bacterial phyla as well as the Euryarchaeota phylum of archaea. Bacteroidetes were found to encode a large number of these biosynthetic gene clusters, despite making up a relatively small portion of the genomes in this dataset. A number of these precursor peptides are similar to those of previously characterized lanthipeptides, but even more were not, including potential antibiotics. One such new antimicrobial lanthipeptide was purified and characterized. Additionally, examination of the biosynthetic gene clusters revealed that enzymes installing secondary post-translational modifications are more widespread than initially thought.

Conclusion: Lanthipeptide biosynthetic gene clusters are more widely distributed and the precursor peptides encoded within these clusters are more diverse than previously appreciated, demonstrating that the lanthipeptide sequence-function space remains largely underexplored.

Keywords: Lanthipeptide, RiPP, RODEO, Natural product, Genome-mining, Antibiotic

* Correspondence: markwalker@unm.edu

¹Department of Chemistry and Chemical Biology, University of New Mexico, 346 Clark Hall, 300 Terrace St. NE, Albuquerque, NM 87131, USA

²Department of Chemistry, University of Illinois at Urbana-Champaign, Roger Adams Laboratory, 600 S. Mathews Ave, Urbana, IL 61801, USA

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Ribosomally synthesized and post-translationally modified peptides (RiPPs) are an expanding group of natural products [1]. Lanthipeptides are among the most studied RiPPs and have a diverse array of structures and biological activities, including antibiotic [2–5], anti-fungal [6], anti-HIV [7, 8], and antinociceptive [9, 10] activities. Recent studies have demonstrated important roles of lanthipeptides produced by the human microbiome in disease and disease prevention [11, 12]. These peptidic natural products are characterized by the presence of macrocycles formed via thioether crosslinks between amino acid residue side chains, termed lanthionines or methyllanthionines [13]. Lanthipeptides are synthesized from a genetically encoded precursor peptide, generically named LanA, which can be divided into two portions; an N-terminal leader region, involved in recognition by the biosynthetic machinery, and a C-terminal core region, which is post-translationally modified. The essential enzymes in lanthipeptide biosynthesis dehydrate select serine and threonine residues in the core region to form dehydroalanine (Dha) and dehydrobutyrine (Dhb) residues, respectively, and then catalyze the conjugate addition of cysteine thiols onto the resulting alkenes to form the lanthionine or methyllanthionine crosslinks (Fig. 1a). Lanthipeptides can be divided into four classes based on the essential biosynthetic enzymes [13]. In class I lanthipeptides, separate proteins carry out the dehydration (LanB) [14] and cyclization (LanC) [15] reactions. LanB enzymes activate serine and threonine residues by glutamylation in a tRNA-dependent manner and produce the dehydrated residues through beta-elimination of glutamate. In classes II–IV, a single protein carries out dehydration and cyclization (LanM, LanKC, and LanL, respectively) (Fig. 1b) [13]. The C-terminal cyclization domains of LanMs, LanKCs, and LanLs are homologous to LanC cyclases; however, the LanKC cyclization domain lacks the zinc-binding residues that are conserved in the other cyclases [16]. The LanM dehydratase domain is related to lipid kinases and has acquired phosphate elimination activity in the kinase active site [17] whereas the LanKC and LanL proteins catalyze dehydration using dedicated kinase [18] and lyase domains. LanL proteins are related to OspF, a phosphothreonine lyase from certain pathogenic proteobacteria (InterPro entry IPR003519) [19]. Beyond dehydratases and cyclases, lanthipeptide biosynthetic gene clusters (BGCs) often encode transporters and proteases to remove the leader peptide (LanT/LanP) and sometimes additional enzymes that further decorate lanthipeptides with secondary modifications [13, 20].

Genome-mining studies based on these enzymes have revealed that lanthipeptide BGCs are distributed widely across bacterial phyla [21–32]. Despite the success in bioinformatically identifying likely lanthipeptide BGCs, it

has been an outstanding challenge to perform high-throughput analysis of the precursor peptides encoded in these gene clusters. Because of the short length of the genes encoding LanAs, they are often not annotated as genes and their variability renders identifying new precursors through homology searching challenging. To address this problem, we have expanded Rapid ORF Description & Evaluation Online (RODEO) [33] to predict lanthipeptide precursor peptides and mined the bacterial and archaeal genomes in the RefSeq database for new lanthipeptide natural products.

Results

Identification of potential lanthipeptide biosynthetic gene clusters

Potential lanthipeptide BGCs were identified by searching the non-redundant RefSeq database (release 93) with the LANC_like (PF05147) hidden Markov model (HMM) from the Protein family (Pfam) database [34], as this domain is shared among currently known classes of lanthipeptides (Fig. 1b). This search resulted in 12,705 proteins with LanC-like domains. The genomic context of these proteins was then examined to assign the clusters to the four separate lanthipeptide classes. If any of the proteins encoded in the seven genes upstream or downstream from the LanC-domain containing protein matched the Pfam HMMs for a LanB (PF04738 and PF14028), the cluster was categorized as class I. If the encoded proteins matched the Pfam HMM for the dehydratase domain of a LanM (PF13575), the cluster was categorized as class II. If the protein containing the LanC-like domain also matched the Pfam HMM for a protein kinase (PF00069), the cluster was categorized as class III or class IV. Classes III and IV were then separated using custom HMMs to distinguish LanKCs (class III) from LanLs (class IV). If none of the encoded proteins matched with these Pfam HMMs, the cluster was categorized as unclassified. This sorting resulted in 2753 putative class I lanthipeptide BGCs, 3708 class II BGCs, 2377 class III BGCs, 815 class IV BGCs, and 3052 unclassified sequences. With the exception of 33 putative class II BGCs from Archaea, lanthipeptide BGCs were exclusively identified in bacteria. Of the unclassified proteins, 1279 are likely not within lanthipeptide BGCs as these proteins are more similar to other proteins, such as endoglucanases [15]. In another 381 cases of unclassified proteins, the gene encoding the protein is within 3 kb of the beginning or end of a sequencing contig, suggesting incomplete data on the BGC. Intriguingly, a number of the remaining 1392 unclassified proteins are located within BGCs that encode proteins often associated with RiPP biosynthesis, such as ABC transporters and proteases, suggesting these clusters are potentially

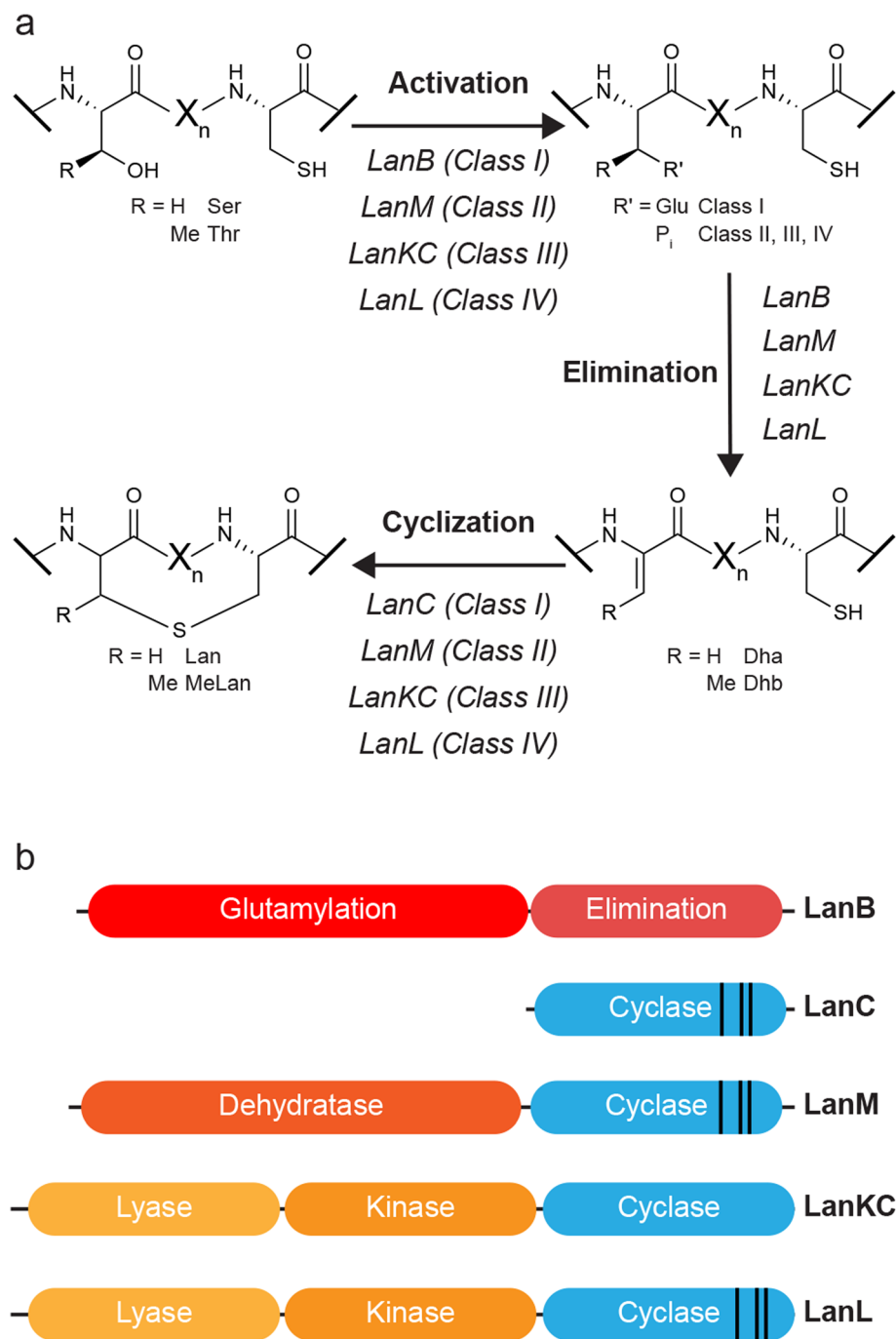


Fig. 1 Biosynthesis of lanthipeptides. **a** Installation of lanthionine or methyllanthionine thioether crosslinks in the four different classes of lanthipeptides. Dha: dehydroalanine, Dhb: dehydrobutyryne, Lan: lanthionine, MeLan: methyllanthionine. **b** Domain structure of the enzymes that install the thioether crosslinks in the different classes of lanthipeptides. The cyclase domains shared between the classes belong to the Pfam family PF05147. The black lines in the cyclase domains represent the location of the zinc-binding residues

involved in the biosynthesis of an as-of-yet uncharacterized class of RiPP.

Identification of precursor peptides

Having identified potential lanthipeptide BGCs, we set out to identify the cognate precursor peptide(s). The

DNA sequence encompassing seven genes downstream to seven genes upstream of the LanC-like domain-containing protein was searched for potential open reading frames (ORFs) beginning with an ATG, TTG, or GTG start codon. Potential ORFs that encoded peptides within the expected length range for LanAs (30–120

amino acids) and not located entirely within an annotated ORF were identified for scoring. A random subset consisting of 20% of the BGCs for each class were manually examined and the identified peptides were annotated as a precursor peptide or not based on characteristics such as similarity to lanthipeptide precursor Pfam families, being encoded immediately upstream or downstream and on the same strand as the class-defining modification enzyme, and the prevalence of Ser, Thr, and Cys residues at the C-terminus. If a precursor peptide could not be unambiguously identified in the BGC, all of the potential peptides from that cluster were set aside. Next, 2458 features were calculated for the peptides deemed to be lanthipeptide precursors (Supplementary Figure S1, Additional File 1) and ANOVA was used to identify the features that were most significantly different (p -value < 0.05) between high-confidence precursor peptides and likely peptides arising from translation of noncoding regions. These features were then calculated for the entire set of potential precursor peptides for each class, and the manually annotated peptides were used as a training set for support vector machine (SVM) classification of the peptides as precursor or not. The SVM classification, the presence of sequence motifs in the leader peptide, and other features were used in the RODEO framework to identify potential precursor peptides for the entire RefSeq database (Supplementary Tables S1, S2, S3, S4, Additional File 1). These improvements have been incorporated into the web tool and command line versions of RODEO and are publicly accessible (<http://ripp.rodeo>).

This approach resulted in the identification of 8405 precursor peptides (Additional File 2). Of these putative LanAs, 2698 (32% of the total) were from class I BGCs, 3002 (36%) were from class II BGCs, 2304 (27%) were from class III BGCs, and 401 (5%) were from class IV BGCs. Based on the number of times their cognate modifying enzymes are encoded in the genome data set, these precursors represent approximately 30,000 redundant lanthipeptides. Approximately 24% of class I precursors, 17% of class II precursors, 55% of class III precursors, and 86% of class IV precursors were not annotated as genes in the database. The majority of precursor peptides in class I (62%), class III (57%), and class IV (83%) BGCs are the only predicted precursor peptide in the cluster. Precursors in class II BGCs are roughly equally split between BGCs with a single precursor peptide (37%) and those with two precursor peptides (39%). Notable exceptions to this distribution are a class I BGC from *Tumebacillus flagellates* that encodes 10 distinct precursor peptides, a class II BGC from *Herbidospora mongoliensis* with 6 distinct precursor peptides, and a class III cluster from *Bacillus cereus* with 13 identical precursor peptides. The most abundant, ungapped sequence motifs from the leader and core regions of each class were identified using Multiple Em for

Motif Elicitation (MEME) (Supplementary Figure S2, Additional File 1) [35]. None of the leader peptide motifs were shared among the four lanthipeptide classes, which was expected given the differences in the respective lanthionine biosynthetic proteins. Interestingly, the most abundant core peptide motifs from each class were also restricted to that class. For example, the nisin/gallidermin lipid II-binding motif SxxxCTP(G/S) C [36] is only found in class I precursors and the mersacidin lipid II-binding motif TxTxEC [37, 38] is only found in class II precursors. Examining these sequence motifs also reveals that in addition to the long-recognized FxLD sequence motif in the leader peptides of class I LanAs [39], a number of class I LanAs from Bacteroidetes have an LxLxKx₅L motif instead. Many of the leader peptides that contain this motif end with a Gly-Gly sequence, and a C39-family Cys protease that removes leader peptides at GG sites [40, 41] is often encoded in the corresponding clusters. This GG leader motif has previously only been observed in class II [42] and III LanAs [43]. With the identification of these class I LanAs, approximately one third of all LanAs have a GG motif at the end of the leader peptide. Double-Gly motif leader peptides are also a common occurrence in other RiPP classes [1, 44]. Other frequently observed leader peptide sequence motifs are the (E/D - 8)(L/M - 7) motif in class II [45] and the LxLQ motif in class III lanthipeptide precursors (Supplementary Figure S2) [46]. Additional less frequent motifs that have not been experimentally investigated are depicted in Supplementary Figure S2.

Comparison with other genome mining tools

To explore the effectiveness of RODEO to predict precursor peptides, these results were compared to other genome mining packages. To achieve this comparison, 5240 genome records encoding identified lanthipeptide BGCs were submitted to antiSMASH 5.0 [47, 48]. AntiSMASH identified a similar number of BGCs as the above analysis, which is to be expected as both approaches utilize Pfam HMMs to identify the clusters, although antiSMASH does not distinguish between class III and class IV BGCs. AntiSMASH identified 55% of class I, 70% of class II, and 47% of class III or IV precursor peptides that were identified by RODEO. On the other hand, RODEO identified 93% of class I, 38% of class II, and 93% of class III or IV precursor peptides that were identified by antiSMASH. The majority of class II precursor peptides predicted by antiSMASH and not by RODEO appear to be false positives as 68% of those peptides are encoded in BGCs with at least one other precursor peptide identified by both tools. These putative false positive peptides have leader peptides that share neither similarity with the peptide identified by both

tools nor with each other (when three or more putative precursor peptides were predicted in a BGC). This lack of leader peptide homology calls into question whether these peptides would be modified by the same set of enzymes as most examples of experimentally verified BGCs that contain multiple precursor peptides show high sequence identity in their leader peptides [41, 49, 50]. Next, ten randomly selected genome records encoding each class of lanthipeptide BGC were selected and submitted to the web server BAGEL4 [29]. BAGEL4 identified 70, 90, 60, and 70% of class I-IV BGCs, respectively, and identified the precursor peptides as open reading frames, but typically did not predict which open reading frame in the BGC was the precursor peptide. Thus, the improvements in RODEO for lanthipeptide precursor peptide annotation described here provide both more information and higher confidence predictions.

A sequence similarity network analysis [51] (Fig. 2) reveals that the identified precursor peptides tend to cluster into families by lanthipeptide class and by taxonomic phylum (Fig. 2 and Supplementary Figure S3, Additional File 1). Even though a number of these families include lanthipeptides that have been characterized, as indicated by the representative lanthipeptides shown (Supplementary Table S5, Additional File 1), most families lack a characterized member, highlighting the scope of lanthipeptide sequence space that remains to be studied. In this work, we have labeled the precursor families by a Roman numeral indicating lanthipeptide class and an increasing Arabic number from left to right and top to bottom in the order generated by the Organic layout of Cytoscape [53]. Several of the uncharacterized families, including I 8, I 13, II 18, and II 32, appear to contain lipid II-binding motifs (Supplementary Figures S4, S5, S6, S7, Additional File 1) and are likely antibiotics. The four largest class I families (I 1–4) are from Actinobacteria and do not have a characterized member. Their core peptides contain a highly conserved Asp residue that is of particular note because the corresponding BGCs contain an *O*-methyltransferase (PF01135) and the conserved Asp is likely post-translationally modified [54]. A number of the class II families, such as II 2, II 13, II 17, and II 26 have conserved leader peptides and non-conserved core peptides. The leader peptides from families II 2, II 13, II 17, II 25, and II 29 belong to the nitrile hydratase leader peptide family of leader peptides, whereas the leader peptides from family II 26 belong to the Nif11 family of leader peptides [44]. The precursor peptides in family II 26 are from Cyanobacteria, however the prochlorosin lanthipeptides are not among them [49]. The prochlorosin precursor peptides (also in the Nif11 family) are located in a smaller cluster, which does not represent the actual size of this family of precursors as

many of them are encoded in genes located distantly from their cognate LanM in the genome [49, 55] and thus were not identified in our analysis that limited the distance between the LanC-domain containing protein and the precursor peptide to seven genes upstream or downstream. We suggest the name cyanotins for this family of RiPPs that are made from highly diverse core peptides, some of which lack Cys and hence cannot be precursors to lanthipeptides.

Other enzymes in lanthipeptide biosynthetic gene clusters

Very few of the BGCs with predicted precursors contained genes encoding class-defining enzymes from other lanthipeptide classes. For example, only six BGCs encoded a LanM and a LanB and LanC, and it is unclear if these encode a single biosynthetic pathway or two separate pathways encoded in close proximity. BGCs encoding a LanM and a LanKC have been identified previously [25]; however, the LanM-associated precursor peptides in those clusters lack Cys residues and therefore were not considered lanthipeptides in the current analysis. In contrast, enzymes that install secondary post-translational modifications are more broadly distributed. Other proteins present in the BGCs were characterized by searching the Pfam database of HMMs. Examining the most abundant proteins that hit at least one Pfam family reveals proteases, ABC transporters, and transcriptional regulators (Supplementary Tables S6, S7, S8, S9, Additional File 1). A number of class I BGCs contain split LanB enzymes that contain the glutamylation and elimination domains on separate polypeptides, as is seen in the biosynthesis of the lanthipeptide pinensin [6], as well as the thiopeptide family of RiPPs [56]. Other class I BGCs contain a full length LanB and an additional protein homologous to the LanB elimination domain. These proteins are also homologous to the enzyme in thiopeptide biosynthesis that catalyzes a formal [4 + 2] cycloaddition to install a substituted pyridine or (dehydro) piperidine moiety [14, 57–59]. Accordingly, it is an intriguing possibility that these domains catalyze a post-translational modification other than elimination. These standalone elimination domain proteins are also often fused to protein-L-isoaspartate *O*-methyltransferase (PCMT or PIMT, PF01135) family proteins and, in turn, many BGCs have these *O*-methyltransferases as standalone proteins. Notably, these elimination domain proteins and methyltransferases are nearly exclusively limited to class I BGCs (Supplementary Table S10, Additional File 1).

Enzymes that are among the most abundant in one class of lanthipeptide BGCs are generally also present in the other classes, if at lower abundance (Supplementary Table S10 and Figure S8, Additional File 1). For example, flavoprotein family enzymes, which have been shown to catalyze oxidative decarboxylation of the C-terminus of some lanthipeptides (LanDs) [60–64], halogenation of amino acid side chains [62], and oxidation of

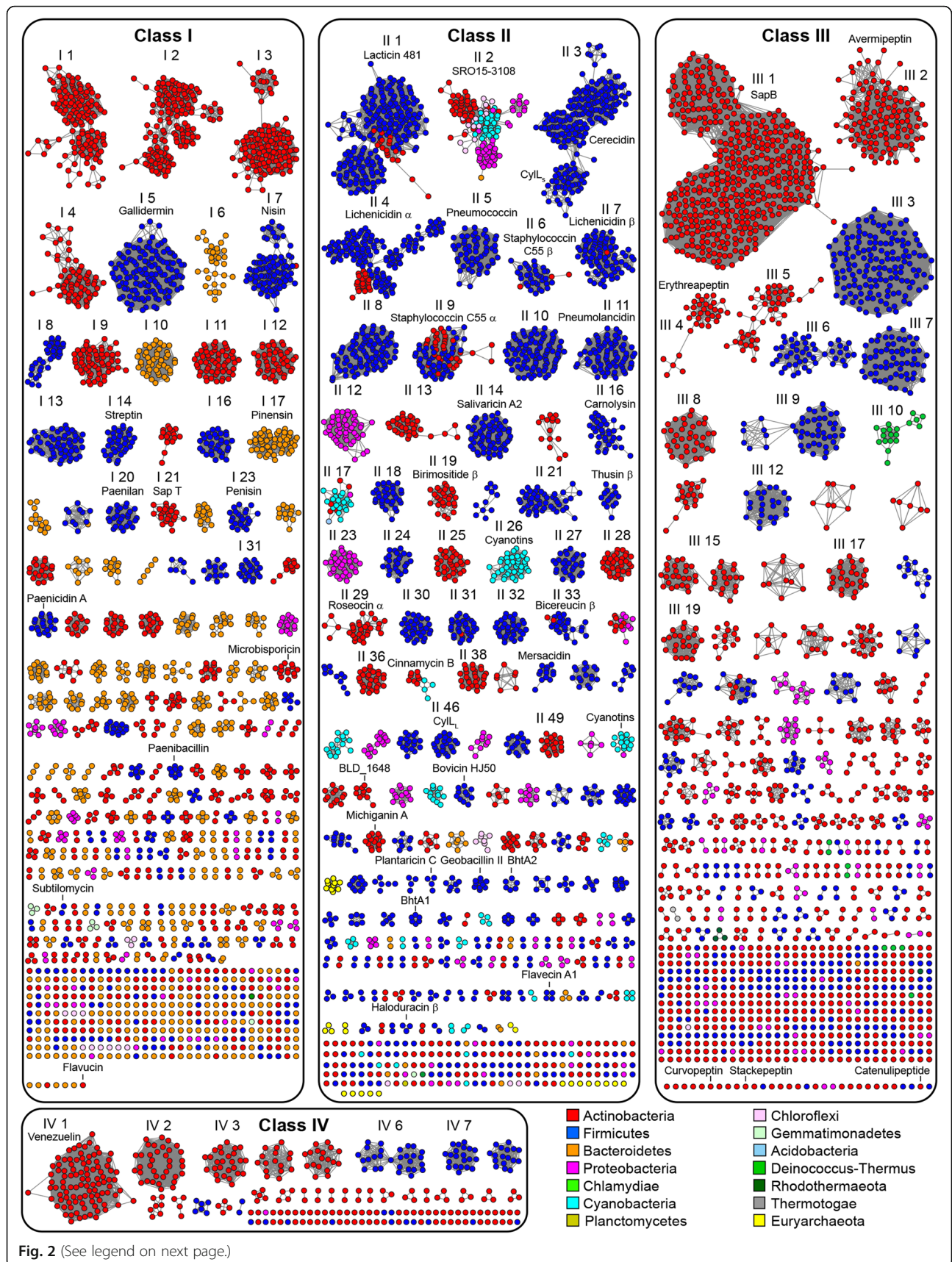
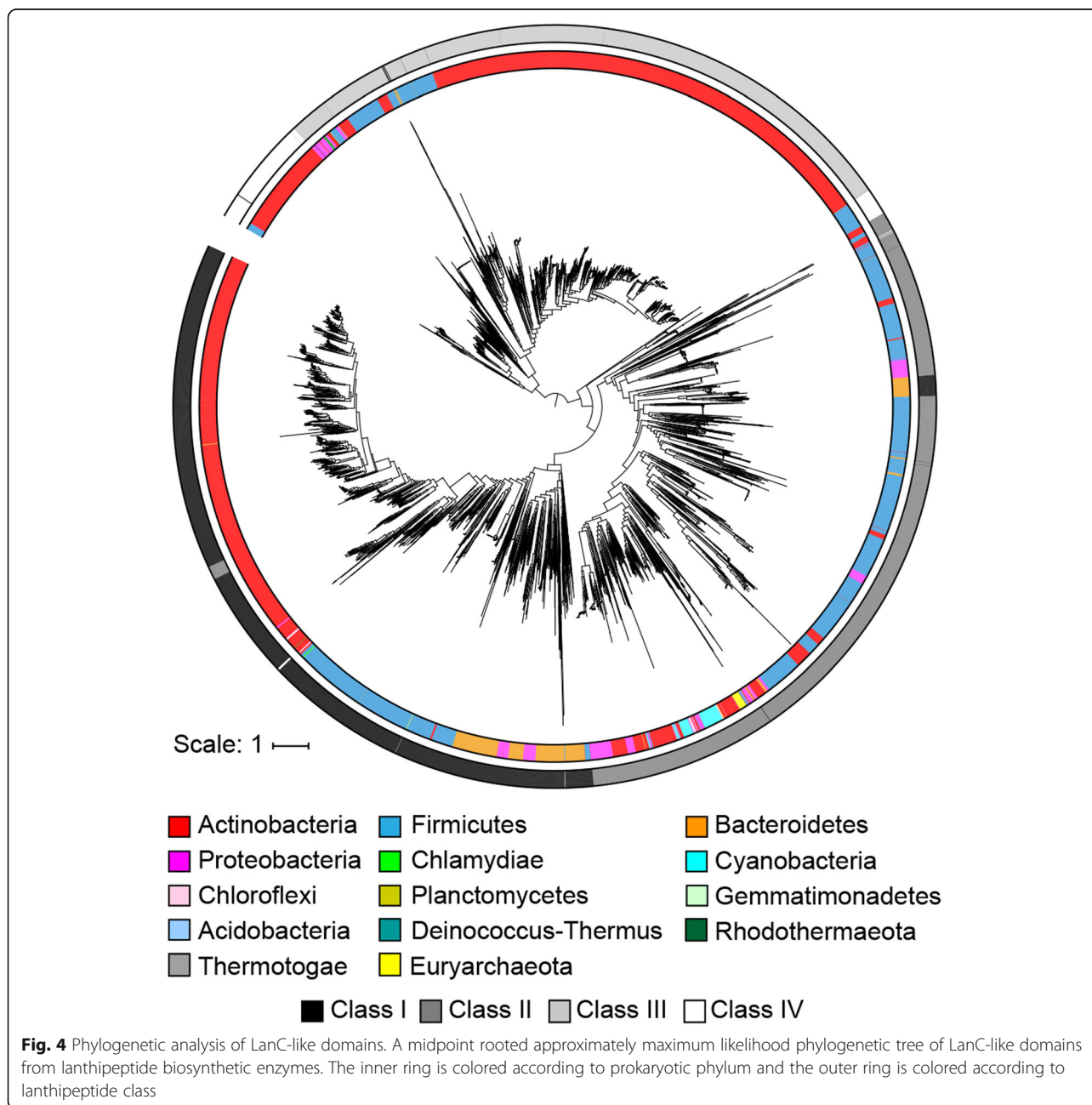


Fig. 2 (See legend on next page.)

The distribution of these proteins across phyla is inconsistent for the different classes of lanthipeptides (Fig. 3b). Nearly a quarter of the class I LanCs were identified in Bacteroidetes, despite their genomes making up a relatively small portion of those in the data set (Supplementary Figure S10, Additional File 1). This distribution suggests further genome sequencing efforts of Bacteroidetes may uncover additional novel lanthipeptide BGCs. At present, only the pinensins have been isolated from this phylum [6]. LanMs were the only lanthipeptide biosynthetic enzymes identified in Cyanobacteria. The majority of LanKCs and LanLs are from Actinobacteria and

Firmicutes (Fig. 2); however, no members of these class III or class IV lanthipeptides from Firmicutes have been characterized to date.

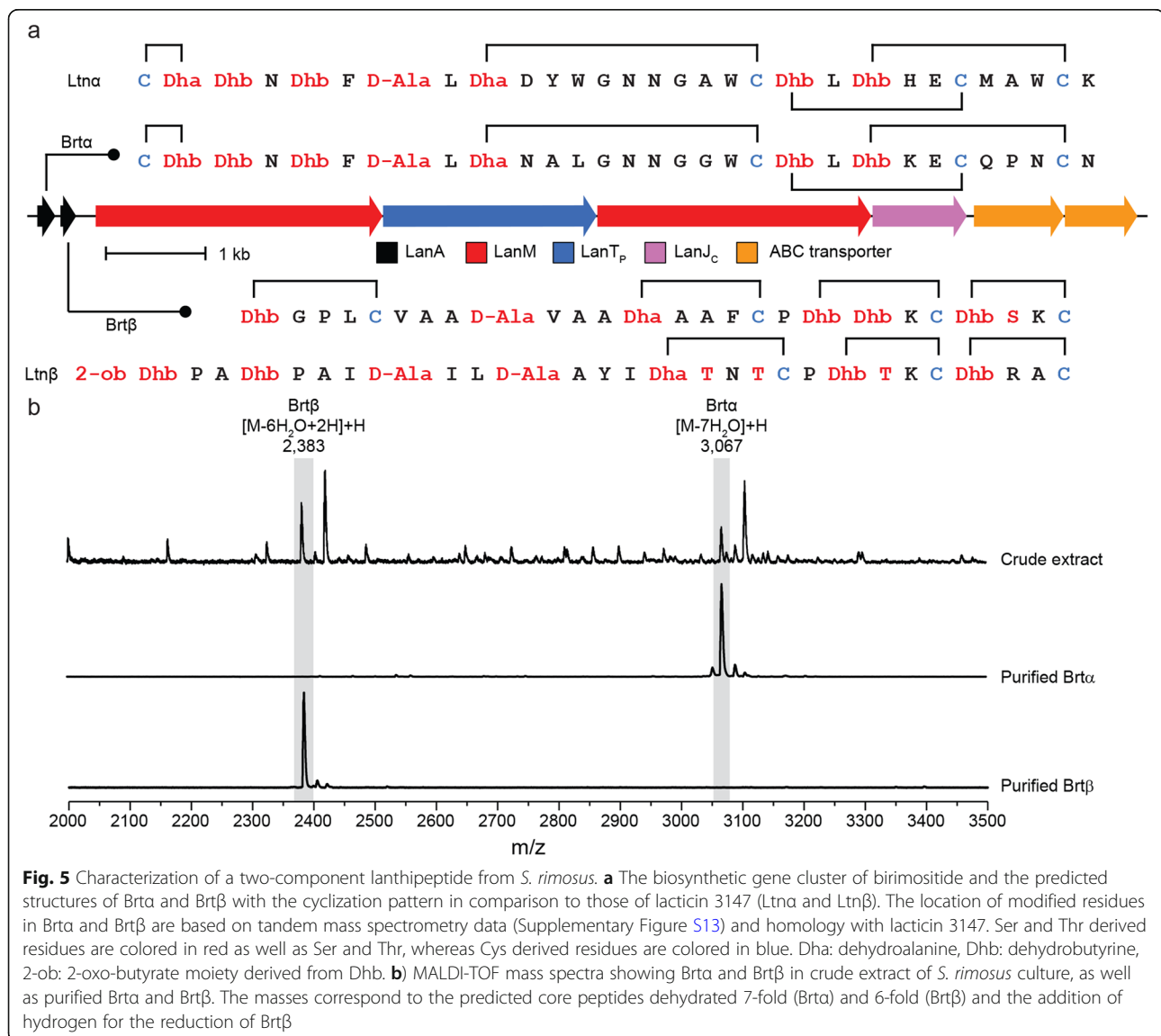
A phylogenetic tree of LanCs and LanC-like domains reveals clades corresponding to the class of lanthipeptide and then sub-clades of bacterial phyla (Fig. 4). This topology suggests the divergence of the lanthipeptide classes is ancient and supports the hypothesis that the lanthipeptide synthases that produce different classes may have arisen through convergent evolution [71]. Inclusion of human LanC-like proteins on the tree shows that they fall into the class IV clade, which is made up of proteins with LanC-



like domains linked to a kinase domain (Supplementary Figure S11, Additional File 1). Notably, human LanC-like proteins bind to kinases in various cell lines [72]. Some exceptions to grouping by class are observed, such as class I LanCs from Bacteroidetes that appear to be related to the LanC-like domains of class II LanMs from Firmicutes. The precursor peptides associated with these LanCs fall into family I 17, which includes the antifungal lanthipeptide pinensin [6]. Furthermore, a group of the LanC-like domains of LanMs from Actinobacteria are related to LanCs from the same phylum with the precursors associated with these LanMs falling in family II 28. Additionally, an analysis of the %GC content of the lanthipeptide BGCs versus the %GC content of the entire bacterial or archaeal genome was performed. Generally, these two values are in good agreement (Supplementary Figure S12, Additional File 1).

Identification of a two-component lanthipeptide BGC from *Streptomyces rimosus* subsp. *rimosus* WC3908

While several two-component lanthipeptides from Firmicutes have been characterized, only a single example, roseocin [73], has been reported from Actinobacteria. The roseocin α precursor peptide falls into family II 29 whereas the roseocin β precursor is a member of family II 2. To further explore two-component lanthipeptides from Actinobacteria, we focused on a BGC from *Streptomyces rimosus*, which encodes an α precursor peptide in family II 9 similar to lactacin 3147 α (Ltn α) [74, 75], and a β precursor peptide in family II 19 (Fig. 5a), the first example of a II 19 family member (Fig. 2). The BGC also encodes a flavin-dependent oxidoreductase belonging to the luciferase-like monooxygenase family (PF00296). The lactacin 3147 BGC encodes a LanJ_A that is involved in the conversion of Ser to D-Ala [68]. Curiously,



whereas the corresponding Ser residues appear conserved in the precursor peptides encoded in the *S. rimosus* BGC (Fig. 5a), a gene encoding LanJ_A is not present. Therefore, we hypothesized that the luciferase-like monooxygenase might carry out a similar reaction. Upon culturing *Streptomyces rimosus*, masses corresponding to the predicted two-component lanthipeptide were observed and the purified peptides were shown to display the synergistic activity characteristic of a two-component lantibiotic (Supplementary Figure S13, Additional File 1). We named the compound birimositide and designated the locus encoding its biosynthesis *brt*. High-resolution mass spectra of both peptides were consistent with the conversion of one Ser to a D-Ala (Fig. 5b). Therefore, we propose the generic name LanJ_C to luciferase-like monooxygenases that reduce dehydrated residues in lanthipeptides. Tandem mass spectrometry suggests that the α -peptide (Br α) is similar in structure to lactacin 3147 α including the position of the D-Ala, but that the β -peptide (Br β) is structurally more divergent from lactacin 3147 β (Ltn β) and contains a single D-Ala compared to two Ser to D-Ala conversions for Ltn β (Supplementary Figure S13 and Supplementary Table S12, Additional File 1).

Discussion

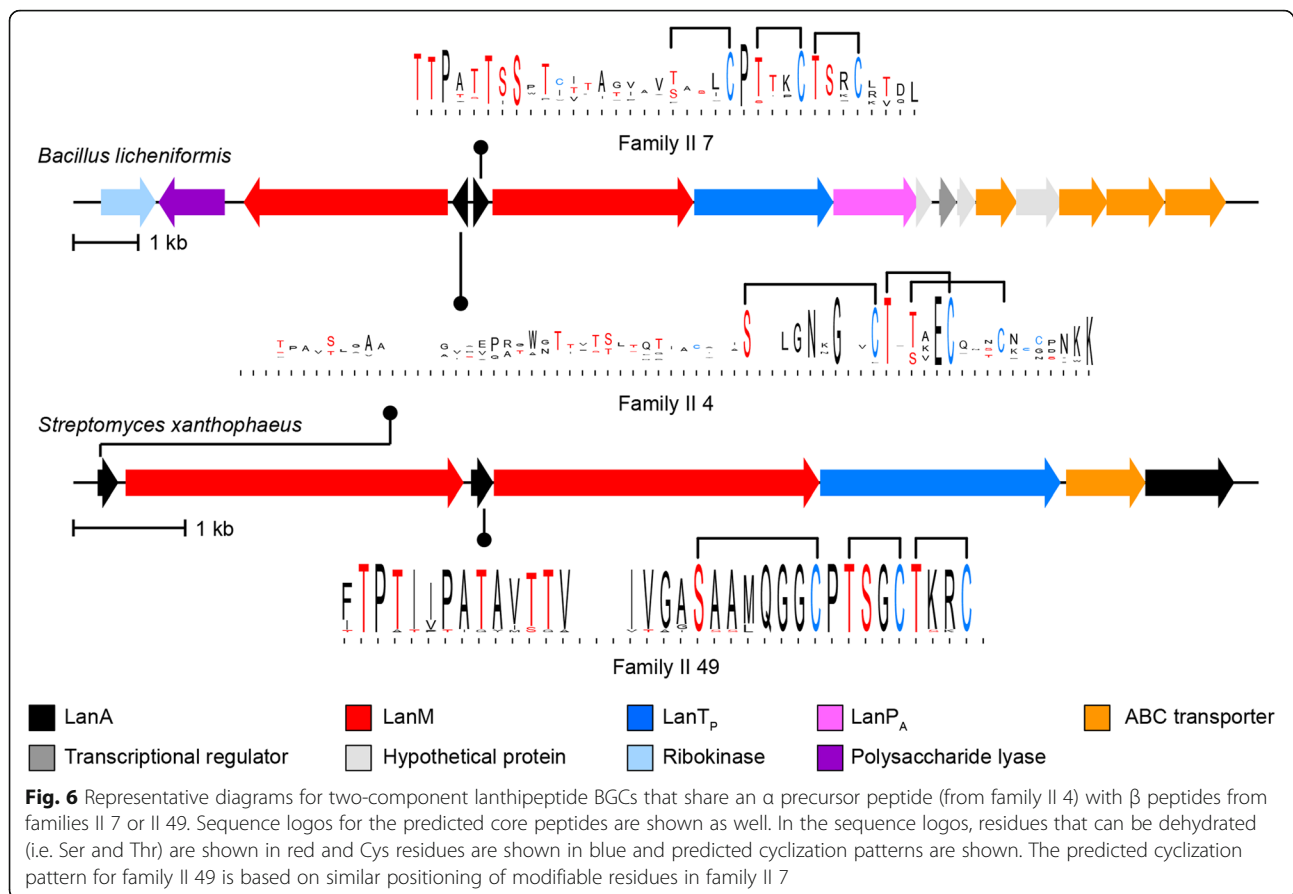
In this work, we improved the ability of RODEO to predict precursor peptides for all four classes of lanthipeptides. This expanded functionality facilitated the mining of more than one hundred thousand bacterial and archaeal genomes for the ability to produce lanthipeptides. These studies revealed that lanthipeptide BGCs are more broadly distributed than previously appreciated, with a large number of class I lanthipeptides in Bacteroidetes, the presence of class III and IV lanthipeptides in Firmicutes, and the detection of class II lanthipeptide BGCs in archaea. Examining the precursor peptides encoded in the gene clusters revealed that the majority of lanthipeptide natural product families have not been characterized, including a number that are likely antibiotics because of lipid II-binding motifs. As delineated below, several new insights have been revealed through this bioinformatics study.

As in a previous study that focused on Actinobacteria [25], the most common lanthipeptide precursor family when analyzing all currently available genomes from different phyla (III 1) is the morphogenic SapB peptide involved in sporulation [76]. The third- and fourth-most abundant precursor families (II 3 and II 4) comprise single- and two-component lanthipeptides in which two structurally dissimilar lanthipeptides exert synergistic bioactivity, with the individual peptides usually having low or no activity [77]. The fourth most abundant family

(II 4) includes the α -peptide of the two-component lanthipeptides lichenicidin [30, 78, 79], haloduracin [80], and thusin [81] (Supplementary Table S5, Additional File 1), and are primarily found in Firmicutes with some members from Actinobacteria. Unexpectedly, the precursors of the partner lanthipeptide that would make up the two-component systems are not in one family but are more diverse in structure. In Firmicutes, lichenicidin β is a member of family II 7, and haloduracin β and thusin β are in small families, whereas the putative partners of the actinobacterial lichenicidin α -like precursors are in family II 49 (Fig. 6). Precursors related to other two-component lanthipeptides such as staphylococcin C55 and the newly reported birimositide also share a similar bifurcated distribution. The precursors of staphylococcin C55 α and birimositide α are members of family II 9, which is comprised of peptides encoded in Firmicutes and Actinobacteria. However, the precursor of staphylococcin C55 β is part of family II 6, whereas birimositide β (and other putative partners of the Actinobacterial α -peptides) are in family II 19. The α peptides of currently investigated two-component lanthipeptides are involved in lipid II-binding. The resulting complex is believed to serve as a binding site for the β peptides, which results in pore formation in the bacterial membranes [78, 82, 83]. The more divergent structures of the β peptides may suggest that different features are required to form pores in the membranes of the target bacteria for lanthipeptide producers that live in different ecological niches.

The precursor peptides of the third most abundant family of peptides (II 3) have sequence homology with the CylL_S'' peptide that together with CylL_L'' makes up the enterococcal cytolysin. This two-component lantibiotic lyses bacterial and mammalian cells [84], and epidemiological studies have shown a clear correlation between the presence of the cytolysin biosynthetic gene locus and hospital-acquired infections [85]. The precursor peptides of the CylL_L'' peptides (II 30) are much less abundant than the precursor peptides for CylL_S'' (II 3) (Fig. 2). Genes encoding these CylL_S-like peptides are found mostly in species of *Bacillus* and sometimes in *Staphylococcus aureus*. Previous studies noted these peptides in *B. cereus* [23, 86] and several members have been isolated (cericidins) and shown to display antimicrobial activity without the need of a partner peptide, explaining why family II 3 is much larger than family II 30.

The most abundant class I precursor peptides in Actinobacteria (I 1–4) are encoded in BGCs that contain the previously mentioned ortholog of a protein isoaspartate methyltransferase as well as a fully conserved Asp in the core peptide (Fig. 7a). This methyltransferase is the highest co-occurring protein with lanthionine biosynthetic



enzymes (Additional File 1: Supplementary Table S9) even though it is only found in class I lanthipeptide BGCs. Recently, it was shown that one such enzyme methylated the conserved Asp in a precursor peptide of family I 2 encoded in *Streptomyces olivaceus* NRRL B-3009. Methylation led to the formation of a succinimide that was hydrolyzed to a mixture of aspartate and isoaspartate. The mature form of the natural product remains unknown [54].

The sixth most abundant precursor family (III 3) consists of class III lanthipeptides encoded by Firmicutes; currently class III peptides have only been isolated from Actinobacteria. At present, the organisms encoding family III 3 are restricted to *Streptococcus pneumoniae*. The precursor peptides display identical core sequences (Fig. 7b) that contain the characteristic Sx_2Sx_3C motif that in some class III lanthipeptides give rise to a lanthionine between the first Ser and last Cys of the motif, and in other peptides yield labionin [10, 43, 76, 87] (Fig. 7b). Unlike class III lanthipeptide gene clusters from Actinobacteria that usually do not contain a protease [88], the III 3 clusters encode for a prolyl oligopeptidase (we propose

the name LanP_B for these with the name LanP_A given to the previously characterized group of subtilisin-like leader peptide proteases [13]) as well as a CAAX family protease [89]. The smaller III 6 family is also found in Firmicutes but the phylogenetic distribution is more varied as these BGCs are present in different *Bacillus* and *Staphylococcus* species. Like class III BGCs from Actinobacteria, they do not encode a conserved protease in the BGC, but they do encode an ABC-type transporter.

Generally speaking, a particular family of lanthipeptides is usually only produced by a single phylum, but some exceptions are notable. The lacticin 481 family (II 1) is mostly produced by Firmicutes, but some Actinobacteria also encode members. Nisin A, a commercially used food preservative, and related lanthipeptides are almost exclusively encoded in Firmicutes genomes, but our analysis revealed a single Actinobacterium (*Nocardia vaccinii* NBRC 15922) encoding a member of this antimicrobial family. After finding an example of a nisin-family lanthipeptide in Actinobacteria we manually examined the genomic context of highly similar LanB and LanC enzymes. Nisin-like precursor peptides were

the cognate precursor peptides reveal the diversity of the lanthipeptide family of RiPPs and the extent to which a large portion of chemical space remains to be explored. The current study will facilitate prioritization of genome-mining studies for novel structures, new synergistic lanthipeptide pairs, or lanthipeptides from genera currently not known to produce such compounds.

Methods

Bioinformatic mining for lanthipeptides

The non-redundant protein, GenBank, and nucleotide records for bacteria and archaea in the RefSeq collection were downloaded from NCBI in May of 2019. The non-redundant protein records were searched with the LanC-Like Pfam HMM (PF05147.12) [90] using HMMER3 [91] with the default settings. The GenBank and nucleotide records were parsed using the list of LanC-Like proteins. Proteins encoded within seven ORFs upstream and downstream of the LanC-Like protein were annotated by searching the Pfam HMM database using HMMER with an E-value cutoff of 1×10^{-5} . Gene clusters that encoded proteins that matched with the Lant_dehydr_N Pfam HMM (PF04738.13) and the Lant_dehydr_C Pfam HMM (PF14028.6) were classified as class I clusters. Gene clusters that encoded a protein that matched with the DUF4135 Pfam HMM (PF13575.6) were classified as class II clusters. Gene clusters that encoded a protein that matched both the LANC_Like Pfam HMM and the Pkinase Pfam HMM (PF00069.25) were classified as class III or class IV clusters. Gene clusters that did not encode proteins that matched these HMMs were discarded as unclassified. Custom HMMs were developed for class III and class IV LanC-Like domains. The sequences of representative LanKCs and LanLs were aligned with Clustal-Omega [92], and these alignments were manually truncated to include only the LanC-like domain. HMMER was then used to generate HMMs from these alignments (Additional Files 3 and 4). The LanC-like proteins in the class III and IV gene clusters were then searched against these custom HMMs, and classified as class III if the E-value for the match with the class III HMM was lower than that for the class IV HMM, and as class IV if the E-value for the match with the class IV HMM was lower than that for the class III HMM. Additionally, the DNA sequence spanning the most upstream ORF and most downstream ORF was translated into all potential ORFs with ATG, GTG, or TTG start codons. Potential gene products that were longer than 120 amino acids or shorter than 30 amino acids were discarded, as were any that were encoded entirely within an annotated gene. Additionally, any ORF not encoding a Cys was discarded as it could not be a potential lanthipeptide precursor.

Finally, to reduce redundancy, the longest of the remaining ORFs with the same stop codon coordinates were retained for further analysis.

Scoring of potential precursor peptides

Leader peptide motifs were identified in the gene products identified above using the MEME bioinformatics application. The leader-core boundary was then estimated by searching the amino acids following the leader peptide motif for GG, GA, or S/T(x)₂₋₇C and setting the core region as starting immediately following a GG or GA motif or 1 residue before a S/T(x)₂₋₇C as long as that motif was more than 10 residues from the end of the peptide. If multiple of these motifs were identified, the one allowing the longest core region was used as the boundary. If none of these motifs were present or were present within 10 residues of the C-terminus, the C-terminal half of the ORF was used as the core. If a leader peptide motif was not identified, the same analysis was performed from the beginning of the ORF. Finally, if no Cys was present in the estimated core region, the ORF was discarded as not a lanthipeptide precursor. Features were then calculated for the potential core peptide, SVM classification was performed, and the potential precursor peptides were scored according to the rubrics in Supplementary Tables S1, S2, S3, S4.

Phylogenetic analysis of LanC and LanC-like domains

LanC and LanC-like domain containing proteins from clusters encoding likely precursor peptides per the analysis above, were retrieved. LanM, LanKC, and LanL enzymes were aligned separately using Clustal-Omega, manually truncated to their LanC-like domains, and then unaligned. Then an alignment of LanCs and LanC-like domains was constructed using Clustal-Omega and manually edited to remove large gaps. This alignment was used to calculate an approximately maximum likelihood phylogenetic tree using FastTree [93]. The tree was then visualized using the Interactive Tree of Life [94].

Precursor sequence logos

Likely precursor peptides were aligned using Clustal-Omega, and that alignment was used to generate sequence logos using WebLogo [95].

Purification of birimositide peptides

Streptomyces rimosus subsp. *rimosus* WC3908 was cultivated on GYM agar for 4 d at 30 °C. After growth, colonies were extracted with methanol. Solvent was removed under vacuum using rotary evaporation and the extract was resuspended in up to 10% MeCN:H₂O prior to high-performance liquid chromatography (HPLC) purification. HPLC purification was conducted using a C18 column (Macherey-Nagel, 100 Å, 250 × 10

mm, 5 μ m) connected to an Agilent Infinity II LC system with solvents A (0.1% trifluoroacetic acid in H₂O) and B (0.1% trifluoroacetic acid in MeCN) at a flow rate of 4 mL/min. The following method was used: stationary at 2% B for 10 min, followed by a linear increase from 2 to 98% B over 35 min, and holding stationary at 98% B for an additional 2 min. For LC coupled to mass spectrometry (MS), LC was performed on HPLC-purified samples using a C18 column (Acclaim PepMap RSLC, nanoViper, 75 μ m \times 15 cm) on a Dionex UltiMate 3000 LC. The following stepwise linear gradient was applied with solvents A (1% formic acid in H₂O) and B (1% formic acid in MeCN) at a flow rate of 0.3 μ L/min: starting at 4% B to 35% B over 45 min, 35 to 50% B over 6 min, 50 to 75% B over 6 min, and holding stationary at 75% B for 6 min. Hybrid Ion Trap-Orbitrap Mass Spectrometer (Thermo Fusion Tribrid Orbitrap, Thermo Scientific) Peptides were diluted 1:1 with ESI mix (80% MeCN, 19% H₂O, 1% formic acid) in H₂O. Calibration and tuning was conducted with Pierce LTQ Velos ESI Positive Ion Calibration (ThermoFisher). Data were obtained using the following parameters: resolution, 30,000; isolation window (MS/MS), 1.6 m/z; collision-induced energy (CID) (MS/MS), 35%; CID activation time, 10 ms; activation q value (MS/MS), 0.25. Data analysis was performed with the Qualbrowser application of Xcalibur software (Thermo-Fisher Scientific).

Birimositide BGC

The genome of *Streptomyces rimosus* subsp. *rimosus* WC3908 was sequenced as part of another genome mining effort focused on phosphonates [96]. The sequence of the birimositide BGC has been deposited under accession number MT037000.

Bioactivity assay

Brt α and Brt β were resuspended in 50% MeCN:H₂O to a concentration of 50 μ M. Each peptide was spotted at varying concentrations on 20 mL of Mueller-Hinton agar seeded with *Micrococcus luteus* ATCC 4698 or *Lactococcus lactis* sp. *cremoris* at a final OD₆₀₀ of 0.05. Inhibition was assessed after overnight growth at 37 °C for *M. luteus* and 30 °C for *L. lactis*. Images were obtained using a Gel-Doc XR+ molecule imager (Bio-Rad).

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12864-020-06785-7>.

Additional file 1: Figure S1. Features calculated to score precursor peptides; **Table S1.** Features and scoring for class I precursors; **Table S2.** Features and scoring for class II precursors; **Table S3.** Features and scoring for class III precursors; **Table S4.** Features and scoring for class IV precursors; **Figure S2.** Sequence motifs present in more than 100 lanthipeptide precursor peptides; **Figure S3.** Sequence similarity

network of predicted precursor peptides with permissive similarity cutoff; **Table S5.** Location of top BLAST hits of known lanthipeptides from the MIBiG and BAGEL databases in the sequence similarity networks presented in Fig. 2; **Figure S4.** Sequence logos generated from alignments of class I precursor peptides in clusters with 20 or more members; **Figure S5.** Sequence logos generated from alignments of class II precursor peptides in clusters with 20 or more members; **Figure S6.** Sequence logos generated from alignments of class III precursor peptides in clusters with 20 or more members; **Figure S7.** Sequence logos generated from alignments of class IV precursor peptides in clusters with 20 or more members; **Table S6.** Twenty most abundant proteins in class I BGCs that belong to at least one Pfam; **Table S7.** Twenty most abundant proteins in class II BGCs that belong to at least one Pfam; **Table S8.** Twenty most abundant proteins in class III BGCs that belong to at least one Pfam; **Table S9.** Twenty most abundant proteins in class IV BGCs that belong to at least one Pfam; **Table S10.** Distribution of Pfams that are in the 20 most abundant protein families in one class among the other three classes; **Figure S8.** Example biosynthetic gene clusters encoding the enzymes in Table S10; **Table S11.** Distribution of select Pfam protein families from BGCs; **Figure S9.** Example biosynthetic gene clusters encoding the enzymes in Table S11; **Figure S10.** Phylogenetic distribution of genomes in the dataset; **Figure S11.** A approximately maximum likelihood midpoint rooted phylogenetic tree of LanC and LanC-like domains including human LanC-like proteins; **Figure S12.** GC content of clusters versus genomes; **Figure S13.** ESI MS/MS and bioactivity of birimositide α and β ; **Table S12.** Expected and observed monoisotopic masses for Brt α and Brt β using ESI MS.

Additional file 2. Excel File containing precursor peptides identified in this study.

Additional file 3. Custom HMM for LanKC LanC-like domains.

Additional file 4. Custom HMM for LanL LanC-like domains.

Abbreviations

BGC: Biosynthetic gene cluster; Dha: Dehydroalanine; Dhb: Dehydrobutyryne; HMM: Hidden Markov model; LanA: Lanthipeptide precursor peptide; LanB: Class I lanthipeptide dehydratase; LanC: Class I lanthipeptide cyclase; LanKC: Class III lanthipeptide synthetase; LanL: Class IV lanthipeptide synthetase; LanM: Class II lanthipeptide synthetase; MEME: Multiple em for motif elicitation; ORF: Open reading frame; Pfam: Protein family; RODEO: Rapid ORF description & evaluation online; RiPP: Ribosomally synthesized and post-translationally modified peptide; SVM: Support vector machine

Acknowledgements

We acknowledge technical assistance from Bryce Kille in implementing the lanthipeptide precursor peptide identification modules in RODEO.

Authors' contributions

M.C.W., D.A.M., and W.A.V. conceived of the study. M.C.W. performed the genome mining. S.M.E., K.J.H. and S.E.A. isolated and characterized birimositide. M.C.W., D.A.M. and W.A.V. drafted the manuscript and all authors have read and approved the final manuscript.

Funding

This work was supported in part by the National Institutes of Health (GM058822 to W.A.V., GM123998 to D.A.M., F32 GM0112284 to M.C.W., F31 GM113486 to K.J.H.) and the David and Lucile Packard Fellowship for Science and Engineering (D.A.M.). The funders had no role in the design of the study, the collection, analysis, and interpretation of the data, nor the decision to publish and preparation of the manuscript.

Availability of data and materials

All biosynthetic gene clusters and precursor peptides identified in this study are available in Additional file 2. Genomes used in this study are available from NCBI in RefSeq release 93. The software used to perform the genome-mining studies is available at <https://github.com/mcwalker-group/reimagined-octo-funicular>. The improved lanthipeptide precursor prediction has

been incorporated into the RODEO web tool at <http://ripp.rodeo> and the command line version at <https://github.com/the-mitchell-lab/rodeo2>.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Chemistry and Chemical Biology, University of New Mexico, 346 Clark Hall, 300 Terrace St. NE, Albuquerque, NM 87131, USA.

²Department of Chemistry, University of Illinois at Urbana-Champaign, Roger Adams Laboratory, 600 S. Mathews Ave, Urbana, IL 61801, USA. ³Carl R.

Woese Institute for Genomic Biology, University of Illinois, Urbana, IL 61801, USA. ⁴Department of Microbiology, University of Illinois at

Urbana-Champaign, 601 S. Goodwin Ave, Urbana, IL 61801, USA. ⁵Howard

Hughes Medical Institute, University of Illinois at Urbana-Champaign, 600 S.

Mathews Ave, Urbana, IL 61801, USA.

Received: 25 November 2019 Accepted: 18 May 2020

Published online: 03 June 2020

References

- Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, Camarero JA, Campopiano DJ, Challis GL, Clardy J, et al. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep*. 2013;30(1):108–60.
- Jabés D, Brunati C, Candiani G, Riva S, Romanó G, Donadio S. Efficacy of the new lantibiotic NAI-107 in experimental infections induced by multidrug-resistant gram-positive pathogens. *Antimicrob Agents Chemother*. 2011;55(4):1671–6.
- Mathur H, O'Connor PM, Hill C, Cotter PD, Ross RP. Analysis of anti-*Clostridium difficile* activity of thuricin CD, vancomycin, metronidazole, ramoplanin, and actagardine, both singly and in paired combinations. *Antimicrob Agents Chemother*. 2013;57(6):2882–6.
- Goldstein BP, Wei J, Greenberg K, Novick R. Activity of nisin against *Streptococcus pneumoniae*, in vitro, and in a mouse infection model. *J Antimicrob Chemother*. 1998;42(2):277–8.
- Hirsch A. The assay of the antibiotic nisin. *J Gen Microbiol*. 1950;4:70–83.
- Mohr KI, Volz C, Jansen R, Wray V, Hoffmann J, Bernecker S, Wink J, Gerth K, Stadler M, Müller R. Pinensins: the first antifungal lantibiotics. *Angew Chem Int Ed*. 2015;54(38):11254–8.
- Féris G, Petrova MI, Andrei G, Huskens D, Hoorelbeke B, Snoeck R, Vanderleyden J, Balzarini J, Bartoschek S, Brönstrup M, et al. The lantibiotic peptide labyrinthopeptin A1 demonstrates broad anti-HIV and anti-HSV activity with potential for microbicidal applications. *PLoS One*. 2013;8(5):e64010.
- Smith TE, Pond CD, Pierce E, Harmer ZP, Kwan J, Zachariah MM, Harper MK, Wyche TP, Maitainaho TK, Bugni TS, et al. Accessing chemical diversity from the uncultivated symbionts of small marine animals. *Nat Chem Biol*. 2018;14(2):179–85.
- Iorio M, Sasso O, Maffioli SI, Bertorelli R, Monciardini P, Sosio M, Bonezzi F, Summa M, Brunati C, Bordoni R, et al. A glycosylated, labionin-containing lantipeptide with marked antinociceptive activity. *ACS Chem Biol*. 2014;9(2):398–404.
- Meindl K, Schmiederer T, Schneider K, Reicke A, Butz D, Keller S, Guhring H, Vertesy L, Wink J, Hoffmann H, et al. Labyrinthopeptins: a new class of carbacyclic lantibiotics. *Angew Chem Int Ed*. 2010;49(6):1151–4.
- Kim SG, Becattini S, Moody TU, Shliha PV, Littmann ER, Seok R, Gjonbalaj M, Eaton V, Fontana E, Amoretti L, et al. Microbiota-derived lantibiotic restores resistance against vancomycin-resistant *Enterococcus*. *Nature*. 2019;572(7771):665–9.
- Duan Y, Llorente C, Lang S, Brandl K, Chu H, Jiang L, White RC, Clarke TH, Nguyen K, Torralba M, et al. Bacteriophage targeting of gut bacterium attenuates alcoholic liver disease. *Nature*. 2019;575(7783):505–11.
- Repka LM, Chekan JR, Nair SK, van der Donk WA. Mechanistic understanding of lantipeptide biosynthetic enzymes. *Chem Rev*. 2017;117:5457–520.
- Ortega MA, Hao Y, Zhang Q, Walker MC, van der Donk WA, Nair SK. Structure and mechanism of the tRNA-dependent lantibiotic dehydratase NisB. *Nature*. 2015;517(7535):509–12.
- Li B, Yu JP, Brunzelle JS, Moll GN, van der Donk WA, Nair SK. Structure and mechanism of the lantibiotic cyclase involved in nisin biosynthesis. *Science*. 2006;311(5766):1464–7.
- Wang H, van der Donk WA. Biosynthesis of the class III lantipeptide catenulipeptin. *ACS Chem Biol*. 2012;7:1529–35.
- Dong SH, Tang W, Lukk T, Yu Y, Nair SK, van der Donk WA. The enterococcal cytolysin synthetase has an unanticipated lipid kinase fold. *eLife*. 2015;4:e07607.
- Müller WM, Schmiederer T, Enslé P, Süßmuth RD. In vitro biosynthesis of the prepeptide of type-III lantibiotic labyrinthopeptin A2 including formation of a C-C bond as a post-translational modification. *Angew Chem Int Ed*. 2010;49(13):2436–40.
- Goto Y, Li B, Claesen J, Shi Y, Bibb MJ, van der Donk WA. Discovery of unique lantionine synthetases reveals new mechanistic and evolutionary insights. *PLoS Biol*. 2010;8(3):e1000339.
- Funk MA, van der Donk WA. Ribosomal natural products, tailored to fit. *Acc Chem Res*. 2017;50(7):1577–86.
- Skinnider MA, Johnston CW, Edgar RE, Dejong CA, Merwin NJ, Rees PN, Magarvey NA. Genomic charting of ribosomally synthesized natural product chemical space facilitates targeted mining. *Proc Natl Acad Sci U S A*. 2016;113(42):E6343–51.
- Letzel AC, Pidot SJ, Hertweck C. Genome mining for ribosomally synthesized and post-translationally modified peptides (RiPPs) in anaerobic bacteria. *BMC Genomics*. 2014;15:983.
- Singh M, Sareen D. Novel LanT associated lantibiotic clusters identified by genome database mining. *PLoS One*. 2014;9(3):e91352.
- Begley M, Cotter PD, Hill C, Ross RP. Identification of a novel two-peptide lantibiotic, lichenicidin, following rational genome mining for LanM proteins. *Appl Environ Microbiol*. 2009;75(17):5451–60.
- Zhang Q, Doroghazi JR, Zhao X, Walker MC, van der Donk WA. Expanded natural product diversity revealed by analysis of lantipeptide-like gene clusters in actinobacteria. *Appl Environ Microbiol*. 2015;81(13):4339–50.
- Azevedo AC, Bento CB, Ruiz JC, Queiroz MV, Mantovani HC. Distribution and genetic diversity of bacteriocin gene clusters in rumen microbial genomes. *Appl Environ Microbiol*. 2015;81(20):7290–304.
- Marsh AJ, O'Sullivan O, Ross RP, Cotter PD, Hill C. In silico analysis highlights the frequency and diversity of type 1 lantibiotic gene clusters in genome sequenced bacteria. *BMC Genomics*. 2010;11(679):1–21.
- Poorinmohammad N, Bagheban-Shemirani R, Hamed J. Genome mining for ribosomally synthesized and post-translationally modified peptides (RiPPs) reveals undiscovered bioactive potentials of actinobacteria. *Antonie Van Leeuwenhoek*. 2019;112(10):1477–99.
- van Heel AJ, de Jong A, Song C, Viel JH, Kok J, Kuipers OP. BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Res*. 2018;46(W1):W278–w281.
- Dischinger J, Josten M, Szekat C, Sahl HG, Bierbaum G. Production of the novel two-peptide lantibiotic lichenicidin by *Bacillus licheniformis* DSM 13. *PLoS One*. 2009;4(8):e6788.
- Tracanna V, de Jong A, Medema MH, Kuipers OP. Mining prokaryotes for antimicrobial compounds: from diversity to function. *FEMS Microbiol Rev*. 2017;41(3):417–29.
- Merwin NJ, Mousa WK, Dejong CA, Skinnider MA, Cannon MJ, Li HX, Dial K, Gunabalasingam M, Johnston C, Magarvey NA. DeepRiPP integrates multiomics data to automate discovery of novel ribosomally synthesized natural products. *Proc Natl Acad Sci U S A*. 2020;117(1):371–80.
- Tietz JI, Schwalen CJ, Patel PS, Maxson T, Blair PM, Tai HC, Zakai UI, Mitchell DA. A new genome-mining tool redefines the lasso peptide biosynthetic landscape. *Nat Chem Biol*. 2017;13(5):470–8.
- Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 2016;44(D1):D279–85.
- Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren JY, Li WW, Noble WS. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*. 2009;37:W202–8.

36. Hsu ST, Breukink E, Tischenko E, Lutters MA, De Kruijff B, Kaptein R, Bonvin AM, Van Nuland NA. The nisin-lipid II complex reveals a pyrophosphate cage that provides a blueprint for novel antibiotics. *Nat Struct Mol Biol*. 2004;11(10):963–7.
37. Hsu ST, Breukink E, Bierbaum G, Sahl HG, de Kruijff B, Kaptein R, van Nuland NA, Bonvin AM. NMR study of mersacidin and lipid II interaction in dodecylphosphocholine micelles. Conformational changes are a key to antimicrobial activity. *J Biol Chem*. 2003;278(15):13110–7.
38. Szekat C, Jack RW, Skutlarek D, Farber H, Bierbaum G. Construction of an expression system for site-directed mutagenesis of the lantibiotic mersacidin. *Appl Environ Microbiol*. 2003;69(7):3777–83.
39. van der Meer JR, Rollema HS, Siezen RJ, Beerthuyzen MM, Kuipers OP, de Vos WM. Influence of amino acid substitutions in the nisin leader peptide on biosynthesis and secretion of nisin by *Lactococcus lactis*. *J Biol Chem*. 1994;269(5):3555–62.
40. Håvarstein LS, Diep DB, Nes IF. A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. *Mol Microbiol*. 1995;16(2):229–40.
41. Bobeica SC, Dong SH, Huo L, Mazo N, McLaughlin MI, Jimenez-Oses G, Nair SK, van der Donk WA. Insights into AMS/PCAT transporters from biochemical and structural characterization of a double Glycine motif protease. *eLife*. 2019;8.
42. Dirix G, Monsieurs P, Marchal K, Vanderleyden J, Michiels J. Screening genomes of gram-positive bacteria for double-glycine-motif-containing peptides. *Microbiology*. 2004;150(Pt 5):1121–6.
43. Völler GH, Krawczyk JM, Pesic A, Krawczyk B, Nachtigall J, Süßmuth RD. Characterization of new class III lantibiotics-erythreapeptin, avermipeptin and griseopeptin from *Saccharopolyspora erythraea*, *Streptomyces avermitilis* and *Streptomyces griseus* demonstrates stepwise N-terminal leader processing. *ChemBioChem*. 2012;13:1174–83.
44. Haft DH, Basu MK, Mitchell DA. Expansion of ribosomally produced natural products: a nitrile hydratase- and Nif11-related precursor family. *BMC Biol*. 2010;8:70.
45. Patton GC, Paul M, Cooper LE, Chatterjee C, van der Donk WA. The importance of the leader sequence for directing lanthionine formation in lactacin 481. *Biochemistry*. 2008;47(28):7342–51.
46. Müller WM, Enslé P, Krawczyk B, Süßmuth RD. Leader peptide-directed processing of labyrinthopeptin A2 precursor peptide by the modifying enzyme LabKC. *Biochemistry*. 2011;50(39):8362–73.
47. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res*. 2019;47(W1):W81–7.
48. Blin K, Kazempour D, Wohlleben W, Weber T. Improved lanthipeptide detection and prediction for antiSMASH. *PLoS One*. 2014;9(7):e103665.
49. Li B, Sher D, Kelly L, Shi Y, Huang K, Knerr PJ, Joewono I, Rusch D, Chisholm SW, van der Donk WA. Catalytic promiscuity in the biosynthesis of cyclic peptide secondary metabolites in planktonic marine cyanobacteria. *Proc Natl Acad Sci U S A*. 2010;107(23):10430–5.
50. Zhao X, van der Donk WA. Structural characterization and bioactivity analysis of the two-component lantibiotic Flv system from a ruminant bacterium. *Cell Chem Biol*. 2016;23(2):246–56.
51. Gerlt JA, Bouvier JT, Davidson DB, Imker HJ, Sadkhin B, Slater DR, Whalen KL. Enzyme function initiative-enzyme similarity tool (EFI-EST): a web tool for generating protein sequence similarity networks. *Biochim Biophys Acta*. 2015;1854(8):1019–37.
52. Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, Blin K, de Bruijn I, Chooi YH, Claesen J, Coates RC, et al. Minimum information about a biosynthetic gene cluster. *Nat Chem Biol*. 2015;11(9):625–31.
53. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–504.
54. Acedo JZ, Bothwell IR, An L, Trouth A, Frazier C, van der Donk WA. O-methyltransferase-mediated incorporation of a β -amino acid in Lanthipeptides. *J Am Chem Soc*. 2019;141(42):16790–801.
55. Cubillos-Ruiz A, Berta-Thompson JW, Becker JW, van der Donk WA, Chisholm SW. Evolutionary radiation of lanthipeptides in marine cyanobacteria. *Proc Natl Acad Sci U S A*. 2017;114(27):E5424–33.
56. Hudson GA, Zhang Z, Tietz JI, Mitchell DA, van der Donk WA. In vitro biosynthesis of the core scaffold of the thiopeptide thiomuracin. *J Am Chem Soc*. 2015;137(51):16012–5.
57. Wever WJ, Bogart JW, Baccile JA, Chan AN, Schroeder FC, Bowers AA. Chemoenzymatic synthesis of thiazolyl peptide natural products featuring an enzyme-catalyzed formal [4 + 2] cycloaddition. *J Am Chem Soc*. 2015;137(10):3494–7.
58. Zhang Z, Hudson GA, Mahanta N, Tietz JI, van der Donk WA, Mitchell DA. Biosynthetic timing and substrate specificity for the Thiopeptide Thiomuracin. *J Am Chem Soc*. 2016;138(48):15511–4.
59. Cogan DP, Hudson GA, Zhang Z, Pogorelov TV, van der Donk WA, Mitchell DA, Nair SK. Structural insights into enzymatic [4+2] aza-cycloaddition in thiopeptide antibiotic biosynthesis. *Proc Natl Acad Sci U S A*. 2017;114(49):12928–33.
60. Kupke T, Stevanovic S, Sahl HG, Götz F. Purification and characterization of EpiD, a flavoprotein involved in the biosynthesis of the lantibiotic epidermin. *J Bacteriol*. 1992;174(16):5354–61.
61. Majer F, Schmid DG, Altena K, Bierbaum G, Kupke T. The flavoprotein MrsD catalyzes the oxidative decarboxylation reaction involved in formation of the peptidoglycan biosynthesis inhibitor mersacidin. *J Bacteriol*. 2002;184(5):1234–43.
62. Ortega MA, Cogan DP, Mukherjee S, Garg N, Li B, Thibodeaux GN, Maffioli S, Donadio S, Sosio M, Escano J, et al. Two flavoenzymes install 5-chlorotryptophan and 2-aminovinyl cysteine during the biosynthesis of the lantibiotic NAI-107. *ACS Chem Biol*. 2017;12(2):548–57.
63. Wiebach V, Mainz A, Siegert MJ, Jungmann NA, Lesquame G, Tirat S, Dreux-Zigha A, Aszodi J, Le Beller D, Süßmuth RD. The anti-staphylococcal lipolanthines are ribosomally synthesized lipopeptides. *Nat Chem Biol*. 2018;14(7):652–4.
64. Mo T, Yuan H, Wang F, Ma S, Wang J, Li T, Liu G, Yu S, Tan X, Ding W, et al. Convergent evolution of the Cys decarboxylases involved in aminovinyl-cysteine (AviCys) biosynthesis. *FEBS Lett*. 2019;593(6):573–80.
65. Shi Y, Bueno A, van der Donk WA. Heterologous production of the lantibiotic Ala (0) actagardine in *Escherichia coli*. *Chem Commun*. 2012;48(89):10966–8.
66. Lohans CT, Li JL, Vederas JC. Structure and biosynthesis of carnolysin, a homologue of enterococcal cytolsin with D-amino acids. *J Am Chem Soc*. 2014;136:13150–3.
67. Huo L, van der Donk WA. Discovery and characterization of bicereucin, an unusual D-amino acid-containing mixed two-component lantibiotic. *J Am Chem Soc*. 2016;138(16):5254–7.
68. Cotter PD, O'Connor PM, Draper LA, Lawton EM, Deegan LH, Hill C, Ross RP. Posttranslational conversion of L-serines to D-alanines is vital for optimal production and activity of the lantibiotic lactacin 3147. *Proc Natl Acad Sci U S A*. 2005;102(51):18584–9.
69. Velásquez JE, Zhang X, van der Donk WA. Biosynthesis of the antimicrobial peptide epilancin 15X and its N-terminal lactate moiety. *Chem Biol*. 2011;18:857–67.
70. Burkhart BJ, Schwalen C, Mann G, Naismith JH, Mitchell DA. YcaO-dependent posttranslational amide activation: biosynthesis, structure, and function. *Chem Rev*. 2017;117(8):5389–456.
71. Zhang Q, Yu Y, Velásquez JE, van der Donk WA. Evolution of lanthipeptide synthetases. *Proc Natl Acad Sci U S A*. 2012;109(45):18361–6.
72. Zeng M, van der Donk WA, Chen J. Lanthionine synthetase C-like protein 2 (LanCL2) is a novel regulator of Akt. *Mol Biol Cell*. 2014;25(24):3954–61.
73. Singh M, Chaudhary S, Sareen D. Roseocin, a novel two-component lantibiotic from an actinomycete. *Mol Microbiol*. 2020;113(2):326–37.
74. Ryan MP, Jack RW, Josten M, Sahl HG, Jung G, Ross RP, Hill C. Extensive post-translational modification, including serine to D-alanine conversion, in the two-component lantibiotic, lactacin 3147. *J Biol Chem*. 1999;274(53):37544–50.
75. Martin NI, Sprules T, Carpenter MR, Cotter PD, Hill C, Ross RP, Vederas JC. Structural characterization of lactacin 3147, a two-peptide lantibiotic with synergistic activity. *Biochemistry*. 2004;43(11):3049–56.
76. Kodani S, Hudson ME, Durrant MC, Buttner MJ, Nodwell JR, Willey JM. The SapB morphogen is a lantibiotic-like peptide derived from the product of the developmental gene *ramS* in *Streptomyces coelicolor*. *Proc Natl Acad Sci U S A*. 2004;101(31):11448–53.
77. Garneau S, Martin NI, Vederas JC. Two-peptide bacteriocins produced by lactic acid bacteria. *Biochimie*. 2002;84(5–6):577–92.
78. Shenkarev ZO, Finkina EI, Nurmukhamedova EK, Balandin SV, Mineev KS, Nadezhdin KD, Yakimenko ZA, Tagaev AA, Temirov YV, Arseniev AS, et al. Isolation, structure elucidation, and synergistic antibacterial activity of a novel two-component lantibiotic lichenicidin from *Bacillus licheniformis* VK21. *Biochemistry*. 2010;49(30):6462–72.

79. Caetano T, Krawczyk JM, Mosker E, Süßmuth RD, Mendo S. Heterologous expression, biosynthesis, and mutagenesis of type II lantibiotics from *Bacillus licheniformis* in *Escherichia coli*. *Chem Biol*. 2011;18(1):90–100.
80. McClarren AL, Cooper LE, Quan C, Thomas PM, Kelleher NL, van der Donk WA. Discovery and in vitro biosynthesis of haloduracin, a two-component lantibiotic. *Proc Natl Acad Sci U S A*. 2006;103(46):17243–8.
81. Xin BY, Zheng JS, Liu HL, Li JH, Ruan LF, Peng DH, Sajid M, Sun M. Thusin, a novel two-component lantibiotic with potent antimicrobial activity against several gram-positive pathogens. *Front Microbiol*. 2016;7.
82. Wiedemann I, Bottiger T, Bonelli RR, Wiese A, Hagge SO, Gutsmann T, Seydel U, Deegan L, Hill C, Ross P, et al. The mode of action of the lantibiotic lactacin 3147—a complex mechanism involving specific interaction of two peptides and the cell wall precursor lipid II. *Mol Microbiol*. 2006; 61(2):285–96.
83. Oman TJ, Lupoli TJ, Wang T-SA, Kahne D, Walker S, van der Donk WA. Haloduracin binds the peptidoglycan precursor lipid II with 2:1 stoichiometry. *J Am Chem Soc*. 2011;133:17544–7.
84. Coburn PS, Gilmore MS. The *Enterococcus faecalis* cytolysin: a novel toxin active against eukaryotic and prokaryotic cells. *Cell Microbiol*. 2003;5(10):661–9.
85. Cox CR, Coburn PS, Gilmore MS. Enterococcal cytolysin: a novel two component peptide system that serves as a bacterial defense against eukaryotic and prokaryotic cells. *Curr Protein Pept Sci*. 2005;6(1):77–84.
86. Wang J, Zhang L, Teng K, Sun S, Sun Z, Zhong J. Cerecidins, novel lantibiotics from *Bacillus cereus* with potent antimicrobial activity. *Appl Environ Microbiol*. 2014;80(8):2633–43.
87. Krawczyk B, Völler GH, Völler J, Enslé P, Süßmuth RD. Curvopeptin: a new lanthionine-containing class III lantibiotic and its co-substrate promiscuous synthetase. *ChemBioChem*. 2012;13(14):2065–71.
88. Chen S, Xu B, Chen E, Wang J, Lu J, Donadio S, Ge H, Wang H. Zn-dependent bifunctional proteases are responsible for leader peptide processing of class III lanthipeptides. *Proc Natl Acad Sci U S A*. 2019;116(7): 2533–8.
89. Maxson T, Deane CD, Molloy EM, Cox CL, Markley AL, Lee SW, Mitchell DA. HIV protease inhibitors block streptolysin S production. *ACS Chem Biol*. 2015;10(5):1217–26.
90. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, et al. The Pfam protein families database in 2019. *Nuc Acids Res*. 2019;47(D1):D427–32.
91. Eddy SR. Accelerated profile HMM searches. *PLoS Comput Biol*. 2011;7(10): e1002195.
92. Sievers F, Higgins DG. Clustal omega for making accurate alignments of many protein sequences. *Protein Sci*. 2018;27(1):135–45.
93. Price MN, Dehal PS, Arkin AP. FastTree 2—Approximately maximum-likelihood trees for large alignments. *PLoS One*. 2010;5(3):e9490.
94. Letunic I, Bork P. Interactive tree of life (iTOL) v4: recent updates and new developments. *Nuc Acids Res*. 2019;47(W1):W256–9.
95. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res*. 2004;14(6):1188–90.
96. Ju KS, Gao J, Doroghazi JR, Wang KK, Thibodeaux CJ, Li S, Metzger E, Fudala J, Su J, Zhang JK, et al. Discovery of phosphonic acid natural products by mining the genomes of 10,000 actinomycetes. *Proc Natl Acad Sci U S A*. 2015;112(39):12175–80.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

