molecules

MDPI

Transporter Protein-Guided Genome Mining for Head-to-Tail Cyclized Bacteriocins

Daniel Major [1], Lara Flanzbaum [1], Leah Lussier [2], Carly Davies [1], Kristian Mark P. Caldo [3] and Jeella Z. Acedo [2,*]

1   Department of Biology, Mount Royal University, Calgary, AB T3E 6K6, Canada; dmajo818@mtroyal.ca (D.M.); lflan365@mtroyal.ca (L.F.); cdavi671@mtroyal.ca (C.D.)
2   Department of Chemistry and Physics, Mount Royal University, Calgary, AB T3E 6K6, Canada; lluss039@mtroyal.ca
3   Department of Agriculture, Food and Nutritional Science, University of Alberta, Edmonton, AB T6G 2P5, Canada; caldo@ualberta.ca
*   Correspondence: jacedo@mtroyal.ca

Abstract: Head-to-tail cyclized bacteriocins are ribosomally synthesized antimicrobial peptides that are defined by peptide backbone cyclization involving the N- and C- terminal amino acids. Their cyclic nature and overall three-dimensional fold confer superior stability against extreme pH and temperature conditions, and protease degradation. Most of the characterized head-to-tail cyclized bacteriocins were discovered through a traditional approach that involved the screening of bacterial isolates for antimicrobial activity and subsequent isolation and characterization of the active molecule. In this study, we performed genome mining using transporter protein sequences associated with experimentally validated head-to-tail cyclized bacteriocins as driver sequences to search for novel bacteriocins. Biosynthetic gene cluster analysis was then performed to select the high probability functional gene clusters. A total of 387 producer strains that encode putative head-to-tail cyclized bacteriocins were identified. Sequence and phylogenetic analyses revealed that this class of bacteriocins is more diverse than previously thought. Furthermore, our genome mining strategy captured hits that were not identified in precursor-based bioprospecting, showcasing the utility of this approach to expanding the repertoire of head-to-tail cyclized bacteriocins. This work sets the stage for future isolation of novel head-to-tail cyclized bacteriocins to serve as possible alternatives to traditional antibiotics and potentially help address the increasing threat posed by resistant pathogens.

Keywords: antimicrobials; head-to-tail cyclized bacteriocins; circular bacteriocins; genome mining

1. Introduction

Bacteriocins refer to a highly diverse family of bacteria-derived ribosomally synthesized antimicrobial peptides [1]. They are mainly used as biopreservatives in food and animal industries and are utilized by adding the bacteriocin-producing organism, bacteriocin-containing fermentates, or purified peptides into food, animal feed, and other antimicrobial products [1–3]. Recent studies have shown that bacteriocins also play a crucial role in regulating gut microbiota [4,5].

A class of bacteriocins referred to as head-to-tail cyclized bacteriocins are characterized by a cyclic peptide backbone wherein an amide bond is formed between the N- to C-terminal amino acids [6–8]. Moreover, experimentally characterized head-to-tail cyclized bacteriocins display a saposin-fold composed of four to five helices in a compact globular architecture that is also observed for leaderless bacteriocins [9,10]. The compact saposin-fold and cyclic peptide backbone of head-to-tail cyclized bacteriocins confer increased stability against proteolytic digestion and denaturation typically caused by extreme temperature and pH conditions [9]. There are two suggested subgroups of head-to-tail cyclized bacteriocins. Subgroup i head-to-tail cyclized bacteriocins are highly cationic resulting in

high (generally >10) isoelectric point values, while subgroup ii bacteriocins are mainly neutral and have lower isoelectric points [6,8].

Several head-to-tail cyclized bacteriocins are active against a wide range of Gram-positive bacteria while activity against Gram-negative strains was observed when the bacterial membrane integrity was initially compromised with ethylenediaminetetraacetic acid (EDTA) or with the use of higher bacteriocin concentrations [7,11–14]. These antimicrobial peptides are reported to operate via diverse mechanisms of action that mostly involve pore formation in the target lipid membranes [15–17]. The most studied head-to-tail cyclized bacteriocin, enterocin AS-48, forms a dimer that non-specifically creates toroidal pores in lipid membranes [18,19]. Another head-to-tail cyclized bacteriocin, carnocyclin A, was reported to create anion-selective channels as a monomeric unit [16,20]. More recently, garvicin ML was shown to bind to a maltose ABC transporter complex that mediates activity particularly at low bacteriocin concentration [21].

Genes encoding proteins that are involved in the biosynthesis of head-to-tail cyclized bacteriocins are clustered together (i.e., in a biosynthetic gene cluster) either in the chromosomal DNA or plasmid DNA of the producer organism [7]. A typical biosynthetic gene cluster for head-to-tail cyclized bacteriocins consists of genes encoding the bacteriocin precursor peptide, transporter protein(s), a SpoIIM (stage II sporulation protein M) membrane protein (previously known as DUF95), an immunity protein, and one or more unknown hydrophobic proteins [6–8]. The inactive precursor peptide has an N-terminal leader sequence and C-terminal core peptide. During maturation, the leader peptide is cleaved, and a peptide bond is formed between the new N-terminal amino acid and the C-terminal residue, producing the active head-to-tail cyclized bacteriocin. The detailed mechanism by which this post-translational modification occurs has yet to be determined [7].

With the rapid advances in DNA sequencing technologies and the development of bioinformatic tools, a plethora of genomic data can now be mined for biosynthetic gene clusters that potentially encode novel natural products, both ribosomal and non-ribosomal in origin [22–27]. In 2017, the head-to-tail cyclized bacteriocin pumilarin from *Bacillus pumilus* B4107 was discovered through genome mining with the use of the bioinformatics tool BAGEL3 wherein the authors filtered their candidate biosynthetic gene clusters based on the presence of genes encoding a precursor peptide and a SpoIIM protein [28,29]. More recently, cerecyclin from *Bacillus cereus* DDD103 was identified through BLAST analysis using the precursor peptide sequences of known head-to-tail cyclized bacteriocins as search queries [30].

To further expand the repertoire of head-to-tail cyclized bacteriocins, we adopted a new genome mining strategy to search for novel head-to-tail cyclized bacteriocins. In particular, transporter protein sequences associated with 19 reported head-to-tail cyclized bacteriocins were used as query sequences in mining for putative head-to-tail cyclized bacteriocins. The Rapid ORF Description and Evaluation Online Tool (RODEO) was then employed to investigate neighboring genes and identify high probability functional biosynthetic gene clusters [31]. A total of 387 strains that potentially encode putative head-to-tail cyclized bacteriocins were identified, 127 of which are unique sequences. More than a hundred hits (i.e., 366 strains; Table S1) identified in this study were not detected in a previous precursor peptide-based genome mining study [32], highlighting the relevance of a transporter protein-based genome mining strategy to expand the collection of this family of circular antimicrobials.

2. Results and Discussion

2.1. Identification of Putative Head-to-Tail Cyclized Bacteriocins

The transporter protein sequences associated with characterized head-to-tail cyclized bacteriocins were used as driver sequences to search the NCBI genome database of non-redundant protein sequences (accessed 30 June 2020) for novel head-to-tail cyclized bacteriocins. The genomic sequences of the protein hits were then analyzed to determine which transporter proteins were clustered with genes encoding a bacteriocin precursor peptide

and a SpoIIM protein, since previous studies have shown that biosynthetic gene clusters of experimentally characterized head-to-tail cyclized bacteriocins contain at least these three genes (i.e., precursor peptide, SpoIIM, and transporter protein) [7,30,32]. In total, 387 hits were detected, 127 of which are unique peptide sequences (Table S1).

Several precursor peptides are encoded in multiple different organisms, while ~25% of the peptides are exclusively found in a single organism (Table S1). For example, the biosynthetic gene clusters for the characterized bacteriocins, amylocyclicin and pumilarin, are found in 61 and 43 different strains, respectively. On the other hand, carnocyclin A, aureocyclicin 4186, and garvicin ML, are produced exclusively by *Carnobacterium maltaromaticum* UAL307, *Staphylococcus aureus* 4185, and *Lactococcus garvieae* DCC43, respectively (Table S1).

The producer organisms of head-to-tail cyclized bacteriocins identified in this study span various bacterial genera, which are mainly under the Firmicutes phylum (Figure 1). The members of each subgroup are differentially distributed across different genera. Subgroup i head-to-tail cyclized bacteriocins are predominantly found in *Bacillus* followed by *Staphylococcus* and *Geobacillus*. The *Bacillus* genus is regarded as a gold mine of antibiotic candidates, not only for bacteriocins, but for lipopeptide antibiotics as well [33–37].
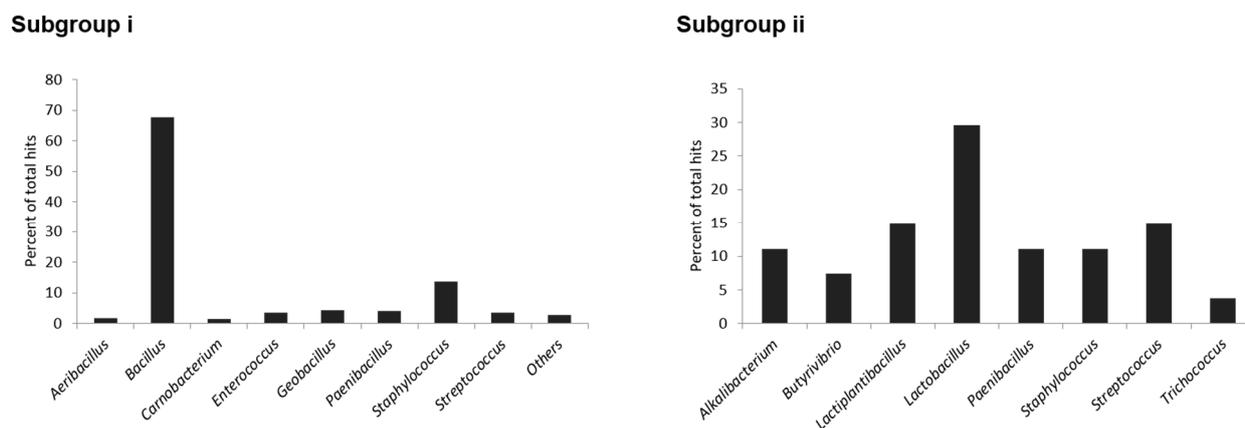

Figure 1. Genera distribution of subgroup **i** and subgroup **ii** head-to-tail cyclized bacteriocin producer organisms identified in this study.

Subgroup ii members were identified mainly in *Lactobacillus*, *Lactiplantibacillus*, and *Streptococcus* bacteria. *Paenibacillus*, *Staphylococcus* and *Streptococcus* were found to encode for biosynthetic genes for the two head-to-tail cyclized bacteriocin subgroups, whereas the other genera only encode genes from either one of the two subgroups.

2.2. Amino Acid Sequence Diversity of Identified Precursor Peptides

To investigate the diversity of the identified putative head-to-tail cyclized precursor peptide sequences, an all-by-all BLAST analysis was performed to create a sequence similarity network that grouped together the most related peptides (Figure 2). The results were then visualized using the Cytoscape program (version 3.9) [38]. In the sequence similarity network (Figure 2), each circle, referred to as a node, represents a unique peptide sequence. A line, referred to as an edge, is drawn between two nodes that share a similarity within the defined alignment score threshold.
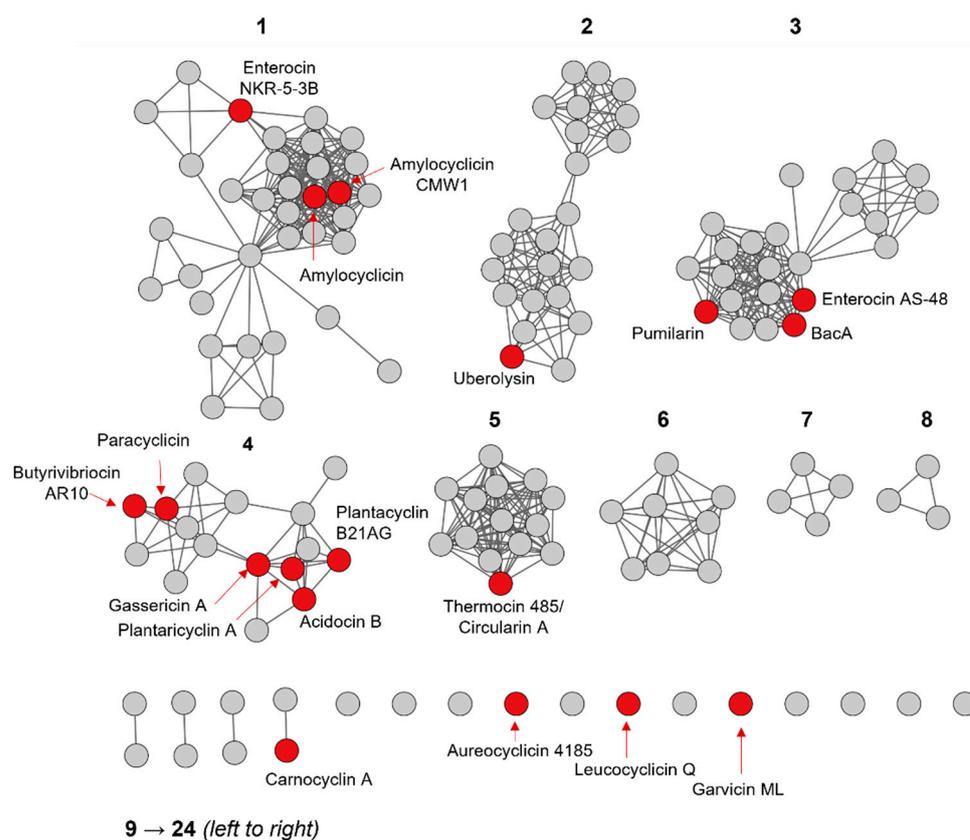
Figure 2. Amino acid sequence similarity network of head-to-tail cyclized bacteriocin precursor peptides resulting in 12 groups and 12 singletons. The group numbers are indicated in bold. Putative novel head-to-tail cyclized bacteriocins are shown in gray, while characterized bacteriocins are shown in red and are labelled.

The sequence similarity network analysis resulted in twelve groups and twelve singletons (Figure 2). The largest group is comprised of 34 unique peptide sequences that are produced by a total of 119 different bacterial strains (Table S1). Characterized members of this group include amylocyclicin, amylocyclicin CMW1, and enterocin NKR-5-3B. As mentioned earlier, amylocyclicin itself is produced by 61 different bacterial strains, and hence comprises ~50% of the organisms in this group. Amylocyclicin CMW1 is found exclusively in *Bacillus amyloliquefaciens* CMW1, while enterocin NKR-5-3B is produced by four different *Enterococcus* strains. Group 2 is comprised of 23 peptides produced by 39 different strains and are closely related to the characterized bacteriocin, uberolysin. The latter was detected in genomes of 13 different organisms. Group 3 has 3 characterized members (i.e., enterocin AS-48, pumilarin, and BacA) and 20 putative members. The sequences within this group are produced by 114 different strains, signifying that several members of this group are present in multiple organisms, including the characterized bacteriocin, pumilarin, which is specifically found in 43 different strains.

Interestingly, all the previously classified as subgroup ii head-to-tail cyclized bacteriocins (i.e., paracyclicin, butyrovibriocin AR10, acidocin B, gassericin A, plantaricyclin A, and plantacyclin B21AG) are found in a single group (group 4). This suggests that head-to-tail cyclized bacteriocins are more diverse than previously thought in terms of sequence homology and properties, and may likely not be restricted to the earlier proposed two subgroups. Subgroup i head-to-tail cyclized bacteriocins are defined by their cationic character and relatively high isoelectric point values; while subgroup ii bacteriocins are distinguished by their hydrophobicity, neutral property, and lower isoelectric points [6,8]. This classification scheme is in fact related to the amino acid sequence because the hydrophobic and electrostatic properties of peptides are dictated by the amino acid composition. This

highlights the importance of the sequence similarity network presented in this work that revealed a more diverse classification scheme for head-to-tail cyclized bacteriocins. It is also worth noting that several of the uncharacterized members of group 4 are produced by *Streptococcus pneumoniae* that differ at either the strain or sub strain level. *S. pneumoniae* strains are reported to produce a major family of bacteriocins known as pneumocins, including a potential circular bacteriocin provisionally named pneumocyclicin, which was found to be present in 34% of pneumococcal genomes [39]. Interestingly, all pneumococci express a pneumocin-capable ABC transporter ComAB, but only 25% express the more efficient ABC transporter BlpAB, suggesting there is diversity in how pneumocins are used by different strains [40]. Because *S. pneumoniae* are opportunistic pathogens of global renown, understanding how their bacteriocin production confers a competitive growth advantage in the nasopharyngeal environment could be greatly beneficial to medicine.

Group 5 includes the characterized bacteriocins, thermocin 485 and circularin A, which are identical peptides. Group 5 also has 13 putative bacteriocin members. Groups 6 to 11 do not have any characterized members, and thus, further investigation of these groups via isolation and characterization is desired. Group 7 consists of 4 unique peptide sequences that are found in a total of 43 strains. Each of these peptides come from the genus *Staphylococcus* and varies by species and strain, although the majority are from *Staphylococcus aureus* (Table S1). It is a promising group to investigate because none of the peptides are very well annotated on GenBank. *S. aureus* is another clinically relevant bacterium because some strains have developed methicillin resistance (MRSA). Understanding bacteriocin synthesis may be helpful for controlling these populations in pathogenic contexts.

Group 12 is comprised of two unique peptide sequences, one of which is the well-characterized bacteriocin, carnocyclin A. Lastly, among the 12 singletons (which share the lowest similarity to the rest of the identified peptides) are the experimentally validated peptides, aureocyclicin 4185, leucocyclicin Q, and garvicin ML.

2.3. Primary Sequence Analysis of Identified Groups in the Sequence Similarity Network

Each group in the sequence similarity network (Figure 2) was analyzed for primary sequence conservation. Amino acid sequences within a cluster that consists of three or more members (i.e., groups 1 to 8) were aligned using Clustal Omega [41], and results were visualized using WebLogo3 [42].

The sequence logos for groups 1 to 8 (Figure 3) show that the N-terminal region of the precursor peptides displays the most diversity, especially for groups 1 to 4. This region corresponds to the leader sequence that is cleaved during biosynthesis. On the other hand, the peptide sequences of the core peptides (i.e., C-terminus) are more conserved. The mechanism of the cyclization of head-to-tail cyclized bacteriocins is poorly understood, but it is worth noting that the majority of the sequence logos end with an aromatic amino acid (i.e., tryptophan or tyrosine), except for groups 4 and 8 that end in alanine. It appears that a hydrophobic C-terminal amino acid may be an essential requirement for core peptide cyclization.

The currently used classification scheme for head-to-tail cyclized bacteriocins is based on the hydrophobicity and the overall charge of the peptide sequences (i.e., subgroup i peptides are cationic, while subgroup ii peptides are neutral). The sequence logos show that groups 4 and 8 lack positively charged amino acids (i.e., lysine and arginine; colored in blue) and may be regarded as subgroup ii bacteriocins. In addition, members of group 11 and a singleton corresponding to a bacteriocin produced by *Lactobacillus nodensis* DSM 19682 (Table S1) also display hydrophobic core peptide sequences and may thus be classified as subgroup ii head-to-tail cyclized bacteriocins. The rest of the groups and singletons have highly conserved positively charged amino acids in the core peptide region, suggesting that they may be classified as subgroup i head-to-tail cyclized bacteriocins. These positively charge residues are proposed to introduce an overall cationic property to these bacteriocins, thereby facilitating their initial interactions with negatively charged phospholipids in bacterial membranes. Notably, there also appears to be several highly conserved tryptophan

residues throughout the core peptide sequences apart from the C-terminal amino acid mentioned earlier. Tryptophan residues are implicated in the anchoring of a bacteriocin to target bacterial membranes through hydrophobic interactions [43].

Figure 3. Amino acid sequence logos for precursor peptides of head-to-tail cyclized bacteriocins belonging to groups 1 to 8 of the sequence similarity network in Figure 2. Polar, neutral, basic, acidic, and hydrophobic amino acids are shown in green, purple, blue, red, and black, respectively [42]. The red broken lines indicate the predicted leader peptide cleavage site based on the characterized members of groups 1 to 5. Groups 6 to 8 do not have characterized members, and hence, the cleavage sites could not be proposed.

2.4. Identification of the Leader Peptide Cleavage Sites

Based on the experimentally validated head-to-tail cyclized bacteriocins, the lengths of the leader peptides can range from 2 to 48 amino acid residues [7]. The variation in length and composition of the leader sequences complicates our understanding of their potential roles in the biosynthesis of this family of bacteriocins. Out of the eight sequence logos presented in Figure 3, groups 1 to 5 have experimentally confirmed members. Hence, the leader peptide cleavage sites for these groups were deduced based on the amino acid

sequences of the characterized members. The cleavage sites are indicated with red broken lines in Figure 3. For other bacteriocin classes (e.g., linear two-peptide bacteriocins) [44], cleavage motifs are established, and hence, the leader peptide cleavage sites are readily identified. However, for the head-to-tail cyclized bacteriocins, our sequence alignment results highlight how challenging it is to postulate how the leader peptide cleavage and core peptide cyclization proceed as there appears to be no apparent sequence motif shared at the cleavage sites across the different groups. Members of group 1 have relatively longer leader sequences, and sequence conservation appears to start only at the first amino acid at the N-terminus of the core peptide. This contrasts with group 2 where there is an evident MFE motif prior to the cleavage site, followed by a series of hydrophobic amino acids at the N-terminus of the core peptide. As for group 3, there is a highly conserved proline residue four amino acids prior to the cleavage site, and similar to group 2, the first few amino acids at the N-terminus of the core peptide are highly conserved, especially the KEF motif for positions 3 to 5 of the core peptide. Group 4, which includes all the currently known subgroup ii head-to-tail cyclized bacteriocins, displays the previously noted highly conserved asparaginyl cleavage site [45]. The N-terminus of the core peptides consists of hydrophobic and aromatic residues. Lastly, group 5 consists of members with short leader peptides (i.e., up to six residues) that are more highly conserved compared to those observed in the earlier groups. The first few residues at the N-terminus of the core peptide are highly conserved.

2.5. Biosynthetic Gene Cluster Analysis

Representative biosynthetic gene clusters of the putative bacteriocins are presented in Figure 4. The gene clusters of all hits identified in this study consist of genes encoding a precursor peptide, a SpoIIM protein, and at least one transporter protein. Interestingly, some gene clusters include multiple copies of the precursor peptide gene, such as the three copies observed for some members of groups 1 and 2. Group 5 also has members with biosynthetic gene clusters comprised of two copies of the precursor peptide gene. The gene organization is mostly conserved (with few variations) among members of the same group except for group 6. Six different gene organizations were observed for this group (Figure 4) among the eight unique precursor peptide sequences that comprise group 6 (Table S1). Group 6 has no characterized member.

Genes encoding transposases that are associated with horizontal gene transfer can be found across the different groups. Another common occurrence is the presence of a Yip1 protein, which is a membrane protein that may be involved in transport. The presence of Yip1 protein is prevalent especially in group 1. Groups 2, 3, 5, and 11 also have members with Yip1 protein, particularly the gene clusters consisting of a set of three genes that are putative accessory ABC transporter components.

One of the longstanding questions about the biosynthesis of head-to-tail cyclized bacteriocins is how the leader peptide is cleaved and the core peptide is cyclized. To date, no enzyme has been identified to facilitate this transformation. It was postulated that genes encoding a peptidase/protease that catalyzes this reaction should be located elsewhere outside the gene cluster [30]. Intriguingly, a few gene clusters encountered in our study contained genes encoding peptidases. These clusters include two members of group 1 wherein a gene encoding a peptidase belonging to the M48 family of zinc peptidases is located three genes downstream from the bacteriocin precursor gene (Figure 4). Another zinc metallopeptidase of the M50 family is found in group 6, located in between a SpoIIM gene and a group of three genes encoding proteins putatively involved in transport (Figure 4). On the other hand, a number of putative peptidases are observed in some gene clusters belonging to group 4, which comprises peptides that are currently classified as subgroup ii head-to-tail cyclized bacteriocins. The possible involvement of these peptidases in leader peptide cleavage and core peptide cyclization remains to be determined experimentally.
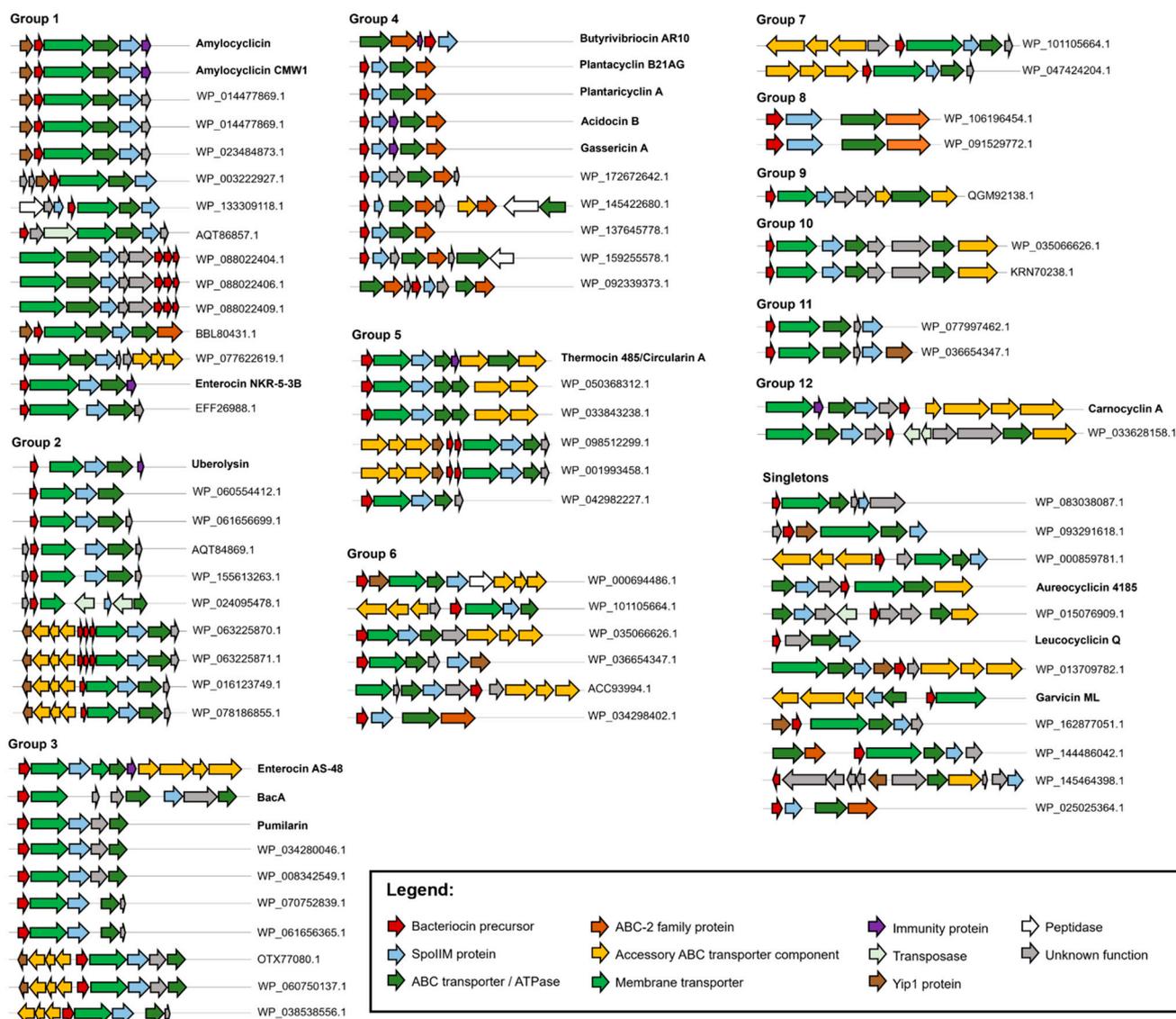
Figure 4. Representative biosynthetic gene clusters of head-to-tail cyclized bacteriocins. Characterized bacteriocins are labelled in bold with the bacteriocin name, while the putative bacteriocins are indicated with their respective protein accession number.

2.6. Phylogenetic Distribution of Identified Head-to-Tail Cyclized Bacteriocins

The evolutionary relationship of the head-to-tail cyclized bacteriocins was inferred through phylogenetic analysis. The five biggest groups identified in the sequence similarity network analysis (Figure 2) are shown in the phylogenetic tree in Figure 5.

In the phylogenetic tree, members of groups 1 to 5 are shown in blue, orange, violet, teal, and brown, respectively. The results revealed that group 1, which is the biggest group, evolved mainly from the earliest ancestors of head-to-tail cyclized bacteriocins. This group appears to be highly heterogeneous as the members are distributed into different subclades and exhibit varying degrees of hydrophobicity based on the grand average of hydropathicity index (GRAVY) data. Most of the other groups eventually evolved from a common ancestor that diverged from group 1 early in the evolution. As shown in the phylogenetic tree, the next three biggest groups belonging to subgroup i head-to-tail cyclized bacteriocins; namely groups 2, 3, and 5 form more tightly related subclades, and the hydrophobic character is more conserved among members of each group based on the GRAVY analysis.
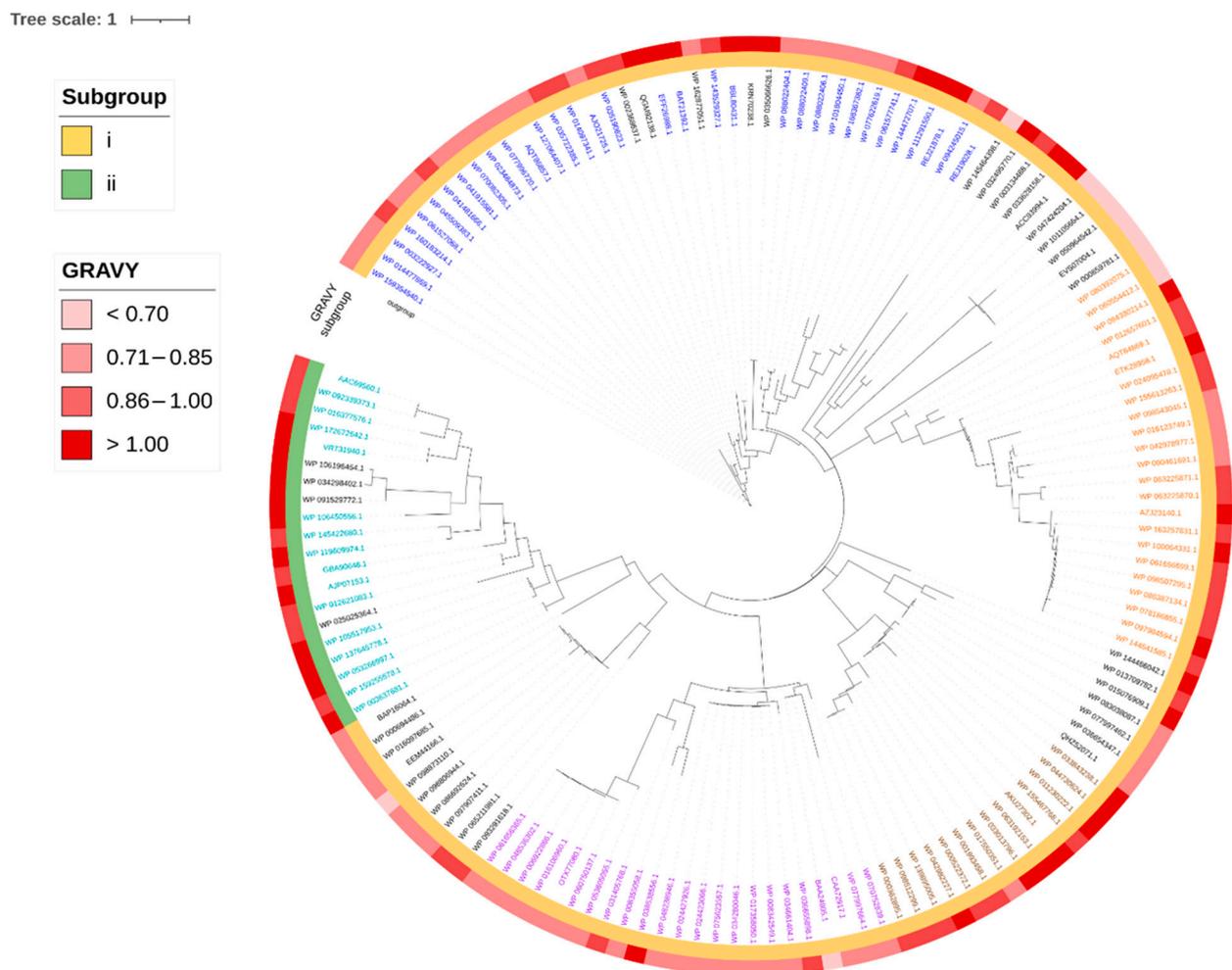
Figure 5. Phylogenetic relationships among putative head-to-tail cyclized bacteriocins identified using the transporter-guided genome mining. The maximum likelihood phylogenetic tree was generated using IQ-TREE [46] and annotated and visualized using iTOL [47]. Members of the five biggest groups in the sequence similarity network analysis are shown in different colors (Group 1—blue, 2—orange, 3—violet, 4—teal, 5—brown). The subgroup number of each accession number and their corresponding hydrophobicity based on GRAVY analysis are indicated in the annotation.

Subgroup ii head-to-tail cyclized bacteriocins, on the other hand, appear to diverge later from subgroup i members. Subgroup ii mainly include groups 4 and 8 members, which are both found to be more closely related to the singleton, leucocyclicin Q, and group 6 members. GRAVY analysis indicates that subgroup ii members appear to be consistently more hydrophobic compared to the members of subgroup i bacteriocins.

3. Materials and Methods

3.1. Genome Mining for Head-to-Tail Cyclized Bacteriocins

The amino acid sequences of transporter proteins associated with thirteen subgroup i (enterocin AS-48 [48], circularin A [49], uberolysin [50], carnocyclin A [51], enterocin NKR-5-3B [52], pumilarin [29], garvicin ML [53], aureocyclicin 4185 [54], leucocyclicin Q [13], amylocyclicin CMW1 [55], amylocyclicn [56], BacA [57], and thermocin 485 [58]), and six subgroup ii (acidocin B [45], gassericin A [59], butyrovibriocin AR10 [60], plantaricyclin A [61], plantacyclin B21AG [62], and paracyclicin [63]) head-to-tail cyclized bacteriocins were obtained and used as driver sequences to mine the NCBI database for putative transporter proteins associated with head-to-tail cyclized bacteriocins using protein BLAST with default search parameters. Duplicates were removed using Excel processing and the accession numbers were inputted into the RODEO Web Tool [31] to extract the genome

neighborhood for the various hits and allow for manual identification of high probability putative bacteriocin gene clusters characterized by the presence of genes encoding a precursor peptide, a SpoIIM protein, in addition to the transporter protein(s).

3.2. Precursor Peptide Sequence Similarity Analysis

A list of the precursor peptides from the RODEO analysis was retrieved and a sequence similarity network was generated using the EFI Enzyme Similarity Tool [64,65]. This allowed an all-by-all BLAST analysis (E-value 1×10^{-5}) and clustered together the most related proteins into groups. To generate the final sequence similarity network, different alignment scores were tested to select an appropriate clustering that grouped most hits with similar gene cluster organization together. This was obtained with a percent identity cut-off of at least 50%. Cytoscape (version 3.9) was then used to visualize the results [38]. The list of peptides with their associated producer organisms is presented in Table S1.

3.3. Generation of Precursor Peptide Sequence Logos

Groups from the sequence similarity network that consist of three or more members were subjected to further sequence analysis. The precursor peptide amino acid sequences belonging to each group were aligned using Clustal Omega [41], and the alignment results were used to create sequence logos using WebLogo3 [42]. The leader peptide cleavage sites were then inferred based on the cleavage sites of known head-to-tail cyclized bacteriocins.

3.4. Phylogenetic Analysis

Multiple sequence alignment of the putative head-to-tail cyclized precursor peptides was performed using the G-INS-i method implemented in the MAFFT web server (https://mafft.cbrc.jp/alignment/server/) [66]. IQ-TREE web server (http://iqtree.cibiv.univie.ac.at/, accessed on 17 October 2021) [46] was then used to determine the best-fit model for protein alignment (WAG + F + I + G4) based on the Bayesian information criterion score and to construct a phylogenetic tree using the maximum likelihood method. The phylogenetic tree was then visualized and annotated using iTOL v4 [47]. Hydrophobicity analysis was performed using the GRAVY Calculator tool (http://gravy-calculator.de).

4. Conclusions

In this study, we performed a new genome mining strategy to search for novel head-to-tail cyclized bacteriocins, specifically by using transporter protein sequences associated with 19 experimentally validated head-to-tail cyclized bacteriocins as query sequences, and the use of a set of bioinformatics tools. A total of 387 strains that potentially encode putative head-to-tail cyclized bacteriocins were identified, 127 of which are unique peptide sequences. Sequence similarity network, sequence analysis, and biosynthetic gene cluster analysis revealed that head-to-tail cyclized bacteriocins are widely distributed and are more diverse than previously thought. Phylogenetic analysis showed that members of group 1 in the sequence similarity network are the most evolutionarily diverse and that the other groups arose from a common ancestor that diverged from group 1. While this work identified putative novel head-to-tail cyclized bacteriocins, it does not predict whether these peptides are easily produced in the laboratory setting and to which pathogens will they be active against. Hence, future work must be directed towards peptide isolation and characterization. It is of particular interest to characterize representative members of groups in the sequence similarity network that do not currently have an experimentally characterized member. Upon the successful production of these bacteriocins, purified peptides can then be used in bioactivity assays against a range of bacterial indicator strains to determine the sensitivity and efficacy of their antimicrobial potential. This work sets the stage for isolation and characterization of diverse head-to-tail cyclized bacteriocins, with likely varying specificities, for their subsequent applications as biopreservatives and antimicrobial agents in various applications in food, animal, and health industries.

Supplementary Materials: The following are available online, Table S1: A complete list of head-tail cyclized bacteriocins identified in this study, classified based on the clustering result of the sequence similarity network analysis.

Author Contributions: Conceptualization, J.Z.A.; methodology, J.Z.A., D.M., L.F. and K.M.P.C.; validation, J.Z.A., D.M., L.F., L.L., C.D. and K.M.P.C.; formal analysis, J.Z.A., D.M., L.F. and K.M.P.C.; data curation, J.Z.A., D.M., L.F., L.L., C.D. and K.M.P.C.; writing—original draft, J.Z.A., D.M. and K.M.P.C.; writing—review and editing, J.Z.A., D.M., L.F. and K.M.P.C.; supervision, J.Z.A.; funding acquisition, J.Z.A. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in the supplementary material.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. O'Connor, P.M.; Kuniyoshi, T.M.; Oliveira, R.P.; Hill, C.; Ross, R.P.; Cotter, P.D. Antimicrobials for food and feed; a bacteriocin perspective. *Curr. Opin. Biotechnol.* **2020**, *61*, 160–167. [CrossRef]
2. Johnson, E.M.; Jung, D.Y.; Jin, D.Y.; Jayabalan, D.R.; Yang, D.S.H.; Suh, J.W. Bacteriocins as food preservatives: Challenges and emerging horizons. *Crit. Rev. Food Sci. Nutr.* **2018**, *58*, 2743–2767. [CrossRef]
3. Ben Lagha, A.; Haas, B.; Gottschalk, M.; Grenier, D. Antimicrobial potential of bacteriocins in poultry and swine production. *Vet. Res.* **2017**, *48*, 22. [CrossRef] [PubMed]
4. Hegarty, J.W.; Guinane, C.M.; Ross, R.P.; Hill, C.; Cotter, P.D. Bacteriocin production: A relatively unharnessed probiotic trait. *F1000Research* **2016**, *5*, 2587. [CrossRef]
5. Bäuerl, C.; Umu, Ö.C.O.; Hernandez, P.E.; Diep, D.B.; Pérez-Martínez, G. A method to assess bacteriocin effects on the gut microbiota of mice. *J. Vis. Exp.* **2017**, *125*, 56053. [CrossRef] [PubMed]
6. Van Belkum, M.J.; Martin-Visscher, L.A.; Vederas, J.C. Structure and genetics of circular bacteriocins. *Trends Microbiol.* **2011**, *19*, 411–418. [CrossRef] [PubMed]
7. Perez, R.H.; Zendo, T.; Sonomoto, K. Circular and leaderless bacteriocins: Biosynthesis, mode of action, applications, and prospects. *Front. Microbiol.* **2018**, *9*, 2085. [CrossRef]
8. Acedo, J.Z.; Chiorean, S.; Vederas, J.C.; van Belkum, M.J. The expanding structural variety among bacteriocins from Gram-positive bacteria. *FEMS Microbiol. Rev.* **2018**, *42*, 805–828. [CrossRef]
9. Towle, K.M.; Vederas, J.C. Structural features of many circular and leaderless bacteriocins are similar to those in saposins and saposin-like peptides. *MedChemComm* **2017**, *8*, 276–285. [CrossRef]
10. Acedo, J.Z.; van Belkum, M.J.; Lohans, C.T.; Towle, K.M.; Miskolzie, M.; Vederas, J.C. Nuclear magnetic resonance solution structures of lacticin Q and aureocin A53 reveal a structural motif conserved among leaderless bacteriocins with broad-spectrum activity. *Biochemistry* **2016**, *55*, 733–742. [CrossRef]
11. Martin-Visscher, L.A.; Yoganathan, S.; Sit, C.S.; Lohans, C.T.; Vederas, J.C. The activity of bacteriocins from *Carnobacterium maltaromaticum* UAL307 against Gram-negative bacteria in combination with EDTA treatment. *FEMS Microbiol. Lett.* **2011**, *317*, 152–159. [CrossRef] [PubMed]
12. Sawa, N.; Zendo, T.; Kiyofuji, J.; Fujita, K.; Himeno, K.; Nakayama, J.; Sonomoto, K. Identification and characterization of lactocyclicin Q, a novel cyclic bacteriocin produced by *Lactococcus* sp. strain QU 12. *Appl. Environ. Microbiol.* **2009**, *75*, 1552–1558. [CrossRef]
13. Masuda, Y.; Ono, H.; Kitagawa, H.; Ito, H.; Mu, F.; Sawa, N.; Zendo, T.; Sonomoto, K. Identification and characterization of leucocyclicin Q, a novel cyclic bacteriocin produced by *Leuconostoc mesenteroides* TK41401. *Appl. Environ. Microbiol.* **2011**, *77*, 8164–8170. [CrossRef]
14. Gálvez, A.; Maqueda, M.; Martínez-Bueno, M.; Valdivia, E. Bactericidal and bacteriolytic action of peptide antibiotic AS-48 against Gram-positive and Gram-negative bacteria and other organisms. *Res. Microbiol.* **1989**, *140*, 57–68. [CrossRef]
15. Gálvez, A.; Maqueda, M.; Martínez-Bueno, M.; Valdivia, E. Permeation of bacterial cells, permeation of cytoplasmic, and artificial membrane vesicles, and channel formation on lipid bilayers by peptide antibiotic AS-48. *J. Bacteriol.* **1991**, *173*, 886–892. [CrossRef]
16. Gong, X.; Martin-Visscher, L.A.; Nahirney, D.; Vederas, J.C.; Duszyk, M. The circular bacteriocin, carnocyclin A, forms anion-selective channels in lipid bilayers. *Biochim. Biophys. Acta* **2009**, *1788*, 1797–1803. [CrossRef]
17. Kawai, Y.; Ishii, Y.; Arakawa, K.; Uemura, K.; Saitoh, B.; Nishimura, J.; Kitazawa, H.; Yamazaki, Y.; Tateno, Y.; Itoh, T.; et al. Structural and functional differences in two cyclic bacteriocins with the same sequences produced by lactobacilli. *Appl. Environ. Microbiol.* **2004**, *70*, 2906–2911. [CrossRef] [PubMed]

18. Sánchez-Barrena, M.J.; Martínez-Ripoll, M.; Gálvez, A.; Valdivia, E.; Maqueda, M.; Cruz, V.; Albert, A. Structure of bacteriocin AS-48: From soluble state to membrane bound state. *J. Mol. Biol.* **2003**, *334*, 541–549. [CrossRef]

19. Cruz, V.L.; Ramos, J.; Melo, M.N.; Martinez-Salazar, J. Bacteriocin AS-48 binding to model membranes and pore formation as revealed by coarse-grained simulations. *Biochim. Biophys. Acta* **2013**, *1828*, 2524–2531. [CrossRef]

20. Martin-Visscher, L.A.; Gong, X.; Duszyk, M.; Vederas, J.C. The three-dimensional structure of carnocyclin A reveals that many circular bacteriocins share a common structural motif. *J. Biol. Chem.* **2009**, *284*, 28674–28681. [CrossRef]

21. Gabrielsen, C.; Brede, D.A.; Hernández, P.E.; Nes, I.F.; Diep, D.B. The maltose ABC transporter in *Lactococcus lactis* facilitates high-level sensitivity to the circular bacteriocin garvicin ML. *Antimicrob. Agents Chemother* **2012**, *56*, 2908–2915. [CrossRef] [PubMed]

22. Trivella, D.B.B.; de Felicio, R. The tripod for bacterial natural product discovery: Genome mining, silent pathway induction, and mass spectrometry-based molecular networking. *mSystems* **2018**, *3*. [CrossRef]

23. O'Neill, E.C.; Schorn, M.; Larson, C.B.; Millán-Aguiñaga, N. Targeted antibiotic discovery through biosynthesis-associated resistance determinants: Target directed genome mining. *Crit. Rev. Microbiol.* **2019**, *45*, 255–277. [CrossRef]

24. Sekurova, O.N.; Schneider, O.; Zotchev, S.B. Novel bioactive natural products from bacteria via bioprospecting, genome mining and metabolic engineering. *Microb. Biotechnol.* **2019**, *12*, 828–844. [CrossRef]

25. Poorinmohammad, N.; Bagheban-Shemirani, R.; Hamedi, J. Genome mining for ribosomally synthesised and post-translationally modified peptides (RiPPs) reveals undiscovered bioactive potentials of actinobacteria. *Antonie Van Leeuwenhoek* **2019**, *112*, 1477–1499. [CrossRef]

26. Huo, L.; Zhao, X.; Acedo, J.Z.; Estrada, P.; Nair, S.K.; van der Donk, W.A. Characterization of a dehydratase and methyltransferase in the biosynthesis of a ribosomally-synthesized and post-translationally modified peptide in Lachnospiraceae. *Chembiochem* **2019**, *21*, 190–199. [CrossRef]

27. Acedo, J.Z.; Bothwell, I.R.; An, L.; Trouth, A.; Frazier, C.; van der Donk, W.A. O-Methyltransferase-mediated incorporation of a β-amino acid in lanthipeptides. *J. Am. Chem. Soc.* **2019**, *141*, 16790–16801. [CrossRef] [PubMed]

28. Van Heel, A.J.; de Jong, A.; Montalbán-López, M.; Kok, J.; Kuipers, O.P. BAGEL3: Automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. *Nucleic Acids Res.* **2013**, *41*, W448–W453. [CrossRef] [PubMed]

29. Van Heel, A.J.; Montalbán-López, M.; Quentin, O.; Kuipers, O.P. Genome-guided identification of novel head-to-tail cyclized antimicrobial peptides, exemplified by the discovery of pumilarin. *Microb. Genom.* **2017**, *3*. [CrossRef]

30. Xin, B.; Liu, H.; Zheng, J.; Xie, C.; Gao, Y.; Dai, D.; Peng, D.; Ruan, L.; Chen, H.; Sun, M. *In silico* analysis highlights the diversity and novelty of circular bacteriocins in sequenced microbial genomes. *mSystems* **2020**, *5*, e00047-20. [CrossRef]

31. Tietz, J.I.; Schwalen, C.J.; Patel, P.S.; Maxson, T.; Blair, P.M.; Tai, H.C.; Zakai, U.I.; Mitchell, D.A. A new genome-mining tool redefines the lasso peptide biosynthetic landscape. *Nat. Chem. Biol.* **2017**, *13*, 470–478. [CrossRef] [PubMed]

32. Vezina, B.; Rehm, B.H.A.; Smith, A.T. Bioinformatic prospecting and phylogenetic analysis reveals 94 undescribed circular bacteriocins and key motifs. *BMC Microbiol.* **2020**, *20*, 77. [CrossRef]

33. Cochrane, S.A.; Vederas, J.C. Lipopeptides from *Bacillus* and *Paenibacillus* spp.: A gold mine of antibiotic candidates. *Med. Res. Rev.* **2016**, *36*, 4–31. [CrossRef]

34. Daas, M.S.; Rosana, A.R.R.; Acedo, J.Z.; Nateche, F.; Kebbouche-Gana, S.; Vederas, J.C.; Case, R. Draft genome sequences of *Bacillus cereus* E41 and *Bacillus anthracis* F34 isolated from Algerian salt lakes. *Genome Announc.* **2017**, *5*, e00383-17. [CrossRef]

35. Daas, M.S.; Acedo, J.Z.; Rosana, A.R.R.; Orata, F.D.; Reiz, B.; Zheng, J.; Natache, F.; Case, R.J.; Kebbouche-Gana, S.; Vederas, J.C. *Bacillus amyloliquefaciens* ssp. plantarum F11 isolated from Algerian salty lake as a source of biosurfactants and bioactive lipopeptides. *FEMS Microbiol. Lett.* **2018**, *365*. [CrossRef]

36. Daas, M.S.; Rosana, A.R.R.; Acedo, J.Z.; Douzane, M.; Nateche, F.; Kebbouche-Gana, S.; Vederas, J.C. Draft genome sequence of *Bacillus paralicheniformis* F47, isolated from an Algerian salty lake. *Genome Announc.* **2018**, *6*, e00190-18. [CrossRef] [PubMed]

37. Daas, M.S.; Rosana, A.R.R.; Acedo, J.Z.; Douzane, M.; Nateche, F.; Kebbouche-Gana, S.; Vederas, J.C. Insights into the draft genome sequence of bioactives-producing *Bacillus thuringiensis* DNG9 isolated from Algerian soil-oil slough. *Stand Genom. Sci.* **2018**, *13*, 25. [CrossRef]

38. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [CrossRef] [PubMed]

39. Bogaardt, C.; Tonder, A.J.; Brueggemann, A.B. Genomic analyses of pneumococci reveal a wide diversity of bacteriocins–including pneumocyclicin, a novel circular bacteriocin. *BMC Genom.* **2015**, *16*, 554. [CrossRef]

40. Wang, C.Y.; Patel, N.; Wholey, W.Y.; Dawid, S. ABC transporter content diversity in *Streptococcus pneumoniae* impacts competence regulation and bacteriocin production. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E5776–E5785. [CrossRef]

41. Sievers, F.; Higgins, D.G. Clustal Omega. *Curr. Protoc. Bioinform.* **2014**, *48*, 3.13.1–3.13.16. [CrossRef] [PubMed]

42. Crooks, G.E.; Hon, G.; Chandonia, J.M.; Brenner, S.E. WebLogo: A sequence logo generator. *Genome Res.* **2004**, *14*, 1188–1190. [CrossRef] [PubMed]

43. Kourie, J.I.; Shorthouse, A.A. Properties of cytotoxic peptide-formed ion channels. *Am. J. Physiol. Cell Physiol.* **2000**, *278*, C1063–C1087. [CrossRef]

44. Acedo, J.Z.; Towle, K.M.; Lohans, C.T.; Miskolzie, M.; McKay, R.T.; Doerksen, T.A.; Vederas, J.C.; Martin-Visscher, L.A. Identification and three-dimensional structure of carnobacteriocin XY, a class IIb bacteriocin produced by carnobacteria. *FEBS Lett.* **2017**, *591*, 1349–1359. [CrossRef]

45. Acedo, J.Z.; van Belkum, M.J.; Lohans, C.T.; McKay, R.T.; Miskolzie, M.; Vederas, J.C. Solution structure of acidocin B, a circular bacteriocin produced by *Lactobacillus acidophilus* M46. *Appl. Environ. Microbiol.* **2015**, *81*, 2910–2918. [CrossRef]

46. Trifinopoulos, J.; Nguyen, L.T.; von Haeseler, A.; Minh, B.Q. W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **2016**, *44*, W232–W235. [CrossRef] [PubMed]

47. Letunic, I.; Bork, P. Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **2016**, *44*, W242–W245. [CrossRef]

48. Samyn, B.; Martínez-Bueno, M.; Devreese, B.; Maqueda, M.; Gálvez, A.; Valdivia, E.; Coyette, J.; Beeumen, J.V. The cyclic structure of the enterococcal peptide antibiotic AS-48. *FEBS Lett.* **1994**, *352*, 87–90. [CrossRef]

49. Kemperman, R.; Kuipers, A.; Karsens, H.; Nauta, A.; Kuipers, O.; Kok, J. Identification and characterization of two novel clostridial bacteriocins, circularin A and closticin 574. *Appl. Environ. Microbiol.* **2003**, *69*, 1589–1597. [CrossRef]

50. Wirawan, R.E.; Swanson, K.M.; Kleffmann, T.; Jack, R.W.; Tagg, J.R. Uberolysin: A novel cyclic bacteriocin produced by *Streptococcus uberis*. *Microbiology* **2007**, *153*, 1619–1630. [CrossRef]

51. Martin-Visscher, L.A.; van Belkum, M.J.; Garneau-Tsodikova, S.; Whittal, R.M.; Zheng, J.; McMullen, L.M.; Vederas, J.C. Isolation and characterization of carnocyclin A, a novel circular bacteriocin produced by *Carnobacterium maltaromaticum* UAL307. *Appl. Environ. Microbiol.* **2008**, *74*, 4756–4763. [CrossRef]

52. Himeno, K.; Rosengren, K.J.; Inoue, T.; Perez, R.H.; Colgrave, M.L.; Lee, H.S.; Chan, L.Y.; Henriques, S.T.; Fujita, K.; Isibashi, N.; et al. Identification, characterization, and three-dimensional structure of the novel circular bacteriocin, enterocin NKR-5-3B, from *Enterococcus faecium*. *Biochemistry* **2015**, *54*, 4863–4876. [CrossRef]

53. Borrero, J.; Brede, D.A.; Skaugen, M.; Diep, D.B.; Herranz, C.; Nes, I.F.; Cintas, L.M.; Hernández, P.E. Characterization of garvicin ML, a novel circular bacteriocin produced by *Lactococcus garvieae* DCC43, isolated from mallard ducks (*Anas platyrhynchos*). *Appl. Environ. Microbiol.* **2011**, *77*, 369–373. [CrossRef]

54. Potter, A.; Ceotto, H.; Coelho, M.L.; Guimaráes, A.J.; Bastos, M.C. The gene cluster of aureocyclicin 4185: The first cyclic bacteriocin of *Staphylococcus aureus*. *Microbiology* **2014**, *160*, 917–928. [CrossRef]

55. Kurata, A.; Yamaguchi, T.; Kira, M.; Kishimoto, N. Characterization and heterologous expression of an antimicrobial peptide from *Bacillus amyloliquefaciens* CMW1. *Biotechnol. Biotechnol. Equip.* **2019**, *33*, 886–893. [CrossRef]

56. Scholz, R.; Vater, J.; Budiharjo, A.; Wang, Z.; He, Y.; Dietel, K.; Schwecke, T.; Herfort, S.; Lasch, P.; Borriss, R. Amylocyclicin, a novel circular bacteriocin produced by *Bacillus amyloliquefaciens* FZB42. *J. Bacteriol.* **2014**, *196*, 1842–1852. [CrossRef] [PubMed]

57. Tomita, H.; Fujimoto, S.; Tanimoto, K.; Ike, Y. Cloning and genetic and sequence analyses of the bacteriocin 21 determinant encoded on the *Enterococcus faecalis* pheromone-responsive conjugative plasmid pPD1. *J. Bacteriol.* **1997**, *179*, 7843–7855. [CrossRef]

58. Egan, K. Discovery and Evaluation of Novel and Characterised Bacteriocins for Future Applications. Ph.D. Thesis, School of Microbiology, University College Cork, Cork, Ireland, 2018.

59. Kawai, Y.; Saito, T.; Kitazawa, H.; Itoh, T. Gassericin A; an uncommon cyclic bacteriocin produced by *Lactobacillus gasseri* LA39 linked at N- and C-terminal ends. *Biosci. Biotechnol. Biochem.* **1998**, *62*, 2438–2440. [CrossRef] [PubMed]

60. Kalmokoff, M.L.; Cyr, T.D.; Hefford, M.A.; Whitford, M.F.; Teather, R.M. Butyrivibriocin AR10, a new cyclic bacteriocin produced by the ruminal anaerobe *Butyrivibrio fibrisolvens* AR10: Characterization of the gene and peptide. *Can. J. Microbiol.* **2003**, *49*, 763–773. [CrossRef] [PubMed]

61. Collins, F.W.J.; O'Connor, P.M.; O'Sullivan, O.; Gómez-Sala, B.; Rea, M.C.; Hill, C.; Ross, R.P. Bacteriocin gene-trait matching across the complete *Lactobacillus* pan-genome. *Sci. Rep.* **2017**, *7*, 3481. [CrossRef]

62. Golneshin, A. *Characterisation of Bacteriocin Genes and Proteins from Lactobacillus plantarum B21 as Potential New Antimicrobial Agents and Natural Food Preservatives*; School of Applied Sciences, RMIT University: Melbourne, Australia, 2014.

63. Borrero, J.; Kelly, E.; O'Connor, P.M.; Kelleher, P.; Scully, C.; Cotter, P.D.; Mahony, J.; van Sinderen, D. Purification, characterization and heterologous production of plantaricyclin A, a novel circular bacteriocin produced by *Lactobacillus plantarum* NI326. *Appl. Environ. Microbiol.* **2017**, *84*, e01801–e01817. [CrossRef]

64. Gerlt, J.A.; Bouvier, J.T.; Davidson, D.B.; Imker, H.J.; Sadkhin, B.; Slater, D.R.; Whalen, K.L. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim. Biophys. Acta* **2015**, *1854*, 1019–1037. [CrossRef] [PubMed]

65. Zallot, R.; Oberg, N.; Gerlt, J.A. The EFI web resource for genomic enzymology tools: Leveraging protein, genome, and metagenome databases to discover novel enzymes and metabolic pathways. *Biochemistry* **2019**, *58*, 4169–4182. [CrossRef] [PubMed]

66. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef] [PubMed]