

# Bioinformatic prospecting identified 99 novel, misannotated and unnoticed putative circular bacteriocins

**Ben Vezina**

Griffith University Eskitis Institute for Cell and Molecular Therapies <https://orcid.org/0000-0003-4224-2537>

**Bernd H.A. Rehm**

Griffith University Eskitis Institute for Cell and Molecular Therapies

**Andrew T. Smith** (✉ [andrew.smith@griffith.edu.au](mailto:andrew.smith@griffith.edu.au))

---

## Research article

**Keywords:** circular, bacteriocin, antimicrobial, bioinformatics, prospecting, clusters, phylogenetics

**Posted Date:** October 1st, 2019

**DOI:** <https://doi.org/10.21203/rs.2.15348/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

## Abstract

Circular bacteriocins are antimicrobial peptides produced by bacteria with a N and C termini ligation. They have desirable properties such as activity at low concentrations along with thermal, pH and proteolytic resistance. There are nineteen experimentally confirmed circular bacteriocins as part of bacteriocin gene clusters, with transport, membrane and immunity proteins. Traditionally, novel antimicrobials are found by testing large numbers of isolates against indicator strains, with no promise of corresponding novel sequence. Through bioprospecting publicly available sequence databases, we identified ninety-nine circular bacteriocins across a variety of bacteria bringing the total to 118. They were grouped into two families within class IIc (IIc i and ii) and further divided into subfamilies based on similarity to experimentally confirmed circular bacteriocins. Within subfamilies, sequences overwhelmingly shared similar characteristics such as sequence length, presence of a polybasic region, conserved locations of aromatic residues, C and N termini, gene clusters similarity, translational coupling and hydrophobicity profiles. At least ninety were predicted to be putatively functional based on gene clusters. Furthermore, bacteriocins identified from *Enterococcus*, *Staphylococcus* and *Streptococcus* species may have activity against clinically relevant strains, due to the presence of putative immunity genes required for expression in a toxin-antitoxin system. Some strains such as *Paenibacillus larvae* subsp. *pulvifaciens* SAG 10367 contained multiple circular bacteriocin gene clusters from different subfamilies, while some strains such as *Bacillus cereus* BCE-01 contained clusters with multiple circular bacteriocin structural genes. Sequence analysis provided rapid insight into identification of novel, putative circular bacteriocins, as well as conserved genes likely essential for circularisation. This represents an expanded library of putative antimicrobial proteins which are potentially active against human, plant and animal pathogens.

## Background

Circular bacteriocins are a class of ribosomally produced antimicrobial peptides with a covalent peptide bond between the N and C termini (Samyn, Martínez-Bueno et al. 1994, Kawai, Saito et al. 1998). The circularisation of the molecule improves thermostability, pH tolerance and proteolytic resistance (Borrero, Kelly et al. 2018), under which conditions most other proteins would be denatured or inactivated. Linearising or nicking circular bacteriocins hampers these intrinsic properties as well as causing a significant reduction in anti-microbial potency (Kawai, Kemperman et al. 2004, Montalbán-López, Spolaore et al. 2008, Sánchez-Hidalgo, Montalbán-López et al. 2011). They have been shown to work by binding to the cell membrane and creating pores, which act as non-selective ion channels causing cell death (Gálvez, Maqueda et al. 1991, Kawai, Ishii et al. 2004, Gong, Martin-Visscher et al. 2009, Himeno, Rosengren et al. 2015). Receptor molecules binding circular bacteriocins may also be involved, as demonstrated by garvicin ML targeting the maltose ABC transporter (Gabrielsen, Brede et al. 2012).

Bacteriocins have many advantages over traditional antimicrobials such as antibiotics. There have been no major reports of bacteriocin resistance, possibly due to their strong activity at low concentrations (Perez, Zendo et al. 2014). Because bacteriocins are encoded, it means they can be genetically engineered and targeted towards specific organisms (Perez, Zendo et al. 2014, Jiménez, Diep et al. 2015). Due to these characteristics, there is also considerable scope for use in anti-spoilage and food-safety applications.

Circular bacteriocins are class IIc (or class V) bacteriocins which can be divided into two families, IIc i and IIc ii based on sequence identity (Cotter, Hill et al. 2005, Gabrielsen, Brede et al. 2014). Table 1 shows the list of experimentally confirmed circular bacteriocins and their characteristics. Class IIc circular bacteriocins are short sequences (58–70 amino acids in length), four (five in the case of AS-48 and BacA) helical segments that enclose a tightly packed hydrophobic core, a saposin fold, no cysteine pairs, and all (except butyriovibriocin AR10) contain a polybasic region involved in binding to target cell membranes (González, Langdon et al. 2000, Martin-Visscher, Gong et al. 2009, Acedo, van Belkum et al. 2015, Himeno, Rosengren et al. 2015).

Circular bacteriocins are usually produced by a gene cluster or operon consisting of 4–10 genes. The mechanism of circularisation and roles of each gene within clusters have not yet been completely elucidated (Maqueda, Sánchez-Hidalgo et al. 2008, Gabrielsen, Brede et al. 2014), though annotation and mutagenesis studies have provided insight into this (Cebrián, Maqueda et al. 2010, Sánchez-Hidalgo, Montalbán-López et al. 2011). A pre-peptide encoded by the bacteriocin structural gene is produced, followed by signal sequence/leader peptide cleavage. This mature peptide is then able to be either circularised within the cell then secreted which has been shown for leucocyclin Q (Mu, Masuda et al. 2014), or secreted and then circularised (Perez, Zendo et al. 2018). The genes involved and the process are not well understood, and it's possible that different pathways exist for different circular bacteriocins. Circularisation appears contingent on hydrophobic N and C termini residues along with the signal sequence, which is required for correct mature peptide processing (Perez, Sugino et al. 2017).

Circular bacteriocin gene clusters are often constituted of overlapping genes, demonstrating a tight organisational structure or genes which depend upon the ribosomal binding site of upstream genes. This indicates expression is regulated by translational coupling (Perez, Ishibashi et al. 2016). All of the currently identified circular bacteriocin gene clusters contain at least two genes that are translationally-coupled (Table 1).

There are nineteen experimentally confirmed circular bacteriocins. Evolutionary-based approaches such as sequence alignments, phylogenetics and gene cluster analysis can provide insight and allow novel identification. This study has identified many new and unmentioned putative circular bacteriocins based on sequence similarity from publicly available sequence data. These putative circular bacteriocins were analysed for characteristics commonly found in circular bacteriocins.

## Results

### Identification and characteristics of putative circular bacteriocins

This study has identified ninety-nine putative circular bacteriocins within a range of microorganisms, bringing the total known circular bacteriocins to 118 (Fig S1, Figure 1). Fig S1 contains detailed information about each identified circular bacteriocin, characteristics, strain information and accession numbers. As

signal sequences can be highly species specific (von Heijne and Abrahmsen 1989, Himeno, Rosengren et al. 2015), they were not used for identification of putative circular bacteriocins. Although they are essential for correct folding, circularisation and bioactivity of circular bacteriocins (Perez, Sugino et al. 2017), removing them from database mining allowed identification of distantly-related putative circular bacteriocins. While some putative circular bacteriocins were annotated correctly, many were unannotated or annotated as branched-chain amino acid aminotransferases which are involved in amino acid catabolism (Thage, Rattray et al. 2004), despite having high similarity and sequence motifs to the mature sequences of known circular bacteriocins.

None contained disulphide bonds. Cysteine residues existed only as single residues in 10/118 of the putative and experimentally confirmed sequences, indicating they are not present for disulphide bond formation (Fig S1).

Almost every putative and experimentally-confirmed circular bacteriocin contained a polybasic region (Fig S1, Figure 1). The paracyclicin and butyrivibriocin AR10 subfamilies were distinct from this trend, as they contained 1–2 basic residues. Both paracyclicin and butyrivibriocin AR10 have been experimentally confirmed. The circular bacteriocin from *Alkalibacterium* AK22 (NZ\_JANL01000003.1) (from the paracyclicin subfamily) did not contain any basic residues. 97.5% of the sequences identified contained aromatic residues, which were locationally-conserved (Fig S1, Table 3). Only *Bacillus krulwichiae* AM31D, *Virgibacillus proomii* V-P and *Alkalibacterium* AK22 contained circular bacteriocin sequences without.

Two *Paenibacillus larvae* strains each harboured two independent putative circular bacteriocins clusters. *Paenibacillus larvae* subsp. *pulvificiens* SAG 10367 (NZ\_CP020557) contained amylocyclin-like and uberolysin-like clusters, while *Paenibacillus larvae* subsp. *larvae* ERIC\_I (NZ\_CP019651.1) harboured AS–48-like and uberolysin-like clusters.

Table 2 shows the list of bacteriocins identified which may be active against the WHO's global priority list of antibiotic resistant bacteria due to the presence of putative immunity genes within the gene clusters (WHO 2017).

## Phylogenetics of circular bacteriocins

Based on the sequence analysis of bacteriocins, there appears to be two different families of class IIc circular bacteriocins, family IIc i and IIc ii, each cluster with bootstrap values of 100 (Cotter, Hill et al. 2005, Gabrielsen, Brede et al. 2014) (Fig S2). Out of the 118 sequences, 89 (75.4%) are part of family IIc i 29 (24.6%) are from IIc ii (Fig S2, Table 3). However, there is considerable sequence divergence within these families, with family IIc i demonstrating a wide variety of sequence lengths and compositions. Therefore, the most appropriate way to classify these sequences was to separate them based on their most closely-related experimentally confirmed circular bacteriocin. In some cases such as streptocyclin, divergence was considered too high (based on bootstrap values) and new subfamilies were coined using the 'cyclin' suffix.

Due to phylogenetic ambiguity and divergence of the identified circular bacteriocin sequences, it was inappropriate to classify each putative circular bacteriocin into currently identified/characterised subfamilies. To remedy this, new circular bacteriocin subfamilies were proposed and named including streptocyclin, akalicyclin, krulwicyclin, bacillocyclin and venezuelacyclin (Fig S1).

Family IIc i was composed of the circularin, lactocyclin/leucocyclin, bacillocyclin, AS–48, amylocyclin, enterocin NKR–5–3B, uberolysin, aureocyclin 4185/garvicin ML, venezuelacyclin, krulwicyclin and carnocyclin A subfamilies. IIc ii were composed of the paracyclicin, akalicyclin, streptocyclin, butyrivibriocin AR10, gassericin/acidocin and plantaricyclin/plantacyclin subfamilies. Due to sequence similarity and phylogenetic branch position, several experimentally confirmed circular bacteriocins were classified within the same subfamily. They included aureocyclin 4185 and garvicin ML (61.4% similarity), lactocyclin and leucocyclin (82% similarity), gassericin and acidocin (100% similarity), plantaricyclin and plantacyclin (94.8% similarity). Some of these subfamilies will most likely fracture into clearer, distinct subfamilies as more sequences become available. Several putative circular bacteriocins were found on lone phylogenetic branches and did not fit into subfamilies and were not classified beyond the familial level (Figure 1). Hydrophobicity of mature circular bacteriocins

Analysis of hydrophobicity profiles suggested two major profiles (Figure 2), with a few exceptions (Fig S3). This gave further evidence that the putative sequences identified were most likely circular bacteriocins. The two major hydrophobicity profiles of the circular bacteriocins matched the phylogenetic family classifications of IIc i and IIc ii (Figure 2). It appears, despite sequence divergence within families, residues are mutating to residues which maintain the hydrophobic profile of the protein. In general, the N terminus of class IIc i tended to have a variable hydrophobic profile, reflecting the sequence divergence and residue length differences within the family.

Both families have similar regions within the hydrophobicity profiles, despite the sequence variability within and between them. In general they are considerably hydrophobic. The C and N termini of every sequence was also found to be hydrophobic (Fig S3). Both families also have a notable polybasic region (residues 52–65 in IIc i and 14–19 in IIc ii) which produces two hydrophilic troughs.

Despite not fitting into any direct phylogenetic subfamilies within family IIc ii, *Bacillus pumilus* GM3FR, *Paeniclostridium sordellii* R26833 and *Bacillus thuringiensis* serovar *indiana* HD521 all match the hydrophobic profile of family IIc ii. Sequence logos (Fig S4) showed high levels of conservation within the IIc ii family, while IIc i had high levels of conservation at the N and C termini. The conserved termini may be implicated as a ligation motif, allowing circularisation of the C and N termini.

## Gene cluster analysis

To determine the number of putatively functional circular bacteriocins, each putative cluster was compared to the cluster of its most closely related experimentally confirmed circular bacteriocin (Figure 1). Table 3 shows a summary of this analysis.

Though there was high cluster divergence between families, similar genes were found in clusters in almost every case, but not limited to: ABC transporters, putative immunity gene/s, transmembrane proteins, SpoIIIM proteins, permeases etc. Table S1 shows a general summary across the identified subfamilies. This provided more evidence that most of these putative sequences were circular bacteriocins, in line with the sequence similarity and hydrophobic profile results. Different gene clusters showed different degrees of similarity, with many having gene rearrangements, inversions, insertions and sharing low sequence similarities between homologues.

Several bacteriocin clusters appeared incomplete (Fig S5) and it is probable that some of these clusters were vestigial or pseudogenes. Of the total 118 circular bacteriocin clusters, a conservative estimate of 90 (76.3%) were putatively functional (Fig S1), though the number is likely higher due to the percentage of gene clusters which contain translational coupling (92.4%). As this analysis was restricted by limited sequence data and assemblies, other genes outside the clusters required for circular bacteriocin production may be present elsewhere in the genome. These would be functional but would be scored as non-functional via this analysis. ABC transporters were seen in every single experimentally confirmed circular bacteriocin cluster, as well 95/99 of putatively identified clusters (Fig S5). This indicates that these 5 circular bacteriocins without ABC transporters were either inactive vestigial remnants or exported via another ABC transporter. Circular bacteriocin ABC transporters are highly similar to ABC transporters within the genomes. It was unclear if non-cluster transporters would be involved in production of circular bacteriocins and were thus considered putatively non-functional. HyID/efflux RND transporters were only present in a few clusters within subfamilies and were not indicative of a putatively functional cluster, as previously demonstrated (Martínez-Bueno, Valdivia et al. 1998). The clusters from *C. polysaccharolyticum* DSM 1801 and *L. bacterium* 3-1 acPFp are examples of unambiguously disrupted gene clusters which would most likely be non-functional. The summary of the cluster analysis for each putative circular bacteriocin (functional/non-functional) is found in the table section of (Figure 1).

21.2% of the clusters were found on plasmids, 65.3% were chromosomally located, and the remaining 13.6% were considered unknown (Table 3, Figure 1). 20.3% were associated with mobile genetic elements such as insertion sequences (Fig S5).

In the AS-48 subfamily (Figure 3), six genes as-48ABCC1DD1 have been shown to be essential for AS-48 production (Martínez-Bueno, Valdivia et al. 1998). This consists of the bacteriocin structural gene, a short and long putative membrane protein/stage II sporulation protein M, another putative transmembrane protein, an ABC transporter and an immunity gene (Martínez-Bueno, Valdivia et al. 1998). All six genes were found in most clusters, though putative immunity genes were not identified in 3/10 clusters. This analysis revealed stage II sporulation protein M domains were commonly found in the putative membrane proteins of the identified circular bacteriocin clusters. Other times, they were found encoded by two separate genes (Figure 3). Therefore, they were treated as similar genes.

Immunity genes from clusters of experimentally confirmed circular bacteriocins appear to have two to three transmembrane domains (Figure S3). They also contain large hydrophilic region/s which occur between these domains. Acidic residues were also found outside these transmembrane domains in 10/15 experimentally-confirmed circular bacteriocin immunity genes. There were no cysteine pairs found in the immunity genes except for in the atypical lycD sequence from leucocyclin Q.

To demonstrate the identification of putative circular bacteriocin subfamilies, which were most likely functional, cluster analysis of the putative bacillocyclin subfamily is shown in Fig S6. Five of the six gene clusters match the gene cluster profile of the AS-48 subfamily (closest phylogenetic relative) and appear to be intact.

Another previously undescribed observation was that some strains contained multiple structural bacteriocin genes within the same cluster (Fig S7). *Bacillus cereus* BCE-01 (NZ\_MVPV01000042.1) contained two different circularin-like circular structural bacteriocin genes with 82.89% identity. 80% identity was found between the signal sequences of these two structural genes. *Bacillus thuringiensis* AFS079576 (NZ\_NUXU01000032.1) also contained two circularin-like structural genes with 81.58% identity within the same cluster. 80% identity was found between the signal sequences of these two structural genes. *Bacillus weihenstephanensis* SDA\_NFFE664 (NZ\_FMBF01000026.1) contained three uberolysin-like circular structural genes with 100% identity and 92% identity, respectively. Each circular bacteriocin structural gene from *B. weihenstephanensis* SDA\_NFFE664 had identical signal sequences to the others in the cluster.

Each structural gene within these multi-structural gene clusters had independent putative promoters. Another observation is that a single putative immunity gene was found within these clusters, indicating it is most likely the single immunity factor for each circular bacteriocin variant.

## Discussion

### Putatively functional circular bacteriocins

This study shows that circular bacteriocins are much more prevalent than originally expected (Perez, Zendo et al. 2018). A literature search revealed a circular bacteriocin was likely isolated from *Lactobacillus acidophilus* IBB 801, though this was not confirmed nor was sequence data available (Zamfir, Callewaert et al. 1999). Some circular bacteriocins identified here have 100% similarity to other circular bacteriocins despite being present in different species. It has been shown that identical circular bacteriocins can have different structures and activities based on the presence of D-amino acids, as is the case with gassericin A and reuterin 6 (Kawai, Ishii et al. 2004). This study has shown bacteria from a wide range of sources including milk, soil, urine, plant cores, honeybee larvae, deep sea water and more (Figure 1) contain putatively functional circular bacteriocin clusters. This indicates a potentially large reservoir of circular bacteriocin-producing strains and circular bacteriocins which could be used as therapeutics, food preservatives (Perez, Zendo et al. 2018), or in other applications such as use as vector proteins to stabilise bioactive proteins (Iwai and Plückthun 1999). There are many bioactive peptides which report low stability (Espinosa-Hernández, Morales-Camacho et al. 2019), which could be stabilised with C-N terminal ligation (Clark, Fischer et al. 2005) found in circular bacteriocins. During the process of this manuscript being written, circular bacteriocin amylocyclin CMW1 was discovered (Kurata, Yamaguchi et al. 2019).

This sequence was successfully predicted as a circular bacteriocin from this dataset, appearing in *Bacillus amyloliquefaciens* LL3. This co-occurrence provides more evidence that the predicted circular bacteriocins are likely correctly identified.

*Bacillus* spp. also contained the largest range of putative circular bacteriocins in this dataset. They contained clusters from Ilc I subfamilies: AS-48, amylocyclicin, enterocin NKR-5-3B, uberolysin, lactocyclin/leucocyclin, circularin, bacillocyclin, krulwicyclin. They also contained Ilc ii circular bacteriocins which were not assigned subfamilies. However, this may have been due to their phylogenetic heterogeneity, some of which has been remediated through reclassifications based on next generation sequencing rather than phenotype (Nazina, Tourova et al. 2001, Logan, Berge et al. 2009).

The percentage of gene clusters which contain translational coupling (92.4%) is most likely a better representation of functional clusters than the conservative prediction based on gene presence (76.3%) found in Table 3. Translational coupling indicates a high level of cluster structure conservation (van de Guchte, Kok et al. 1991) and it would be highly unusual for these genes to be asserting such a high degree of organisational structure if they were not positively-selected for, that is if they were not functional/expressed. Mutations in these tightly-packed clusters will not only alter the ends of particular gene products, but also impact transcription of downstream genes in alternative reading frames.

Presence of polybasic and aromatic residues were locationally conserved, found in 94.1% and 97.5% of identified circular bacteriocins, respectively (Table 3, Fig S1). Aromatic residues are often found flanking transmembrane-associated helices, allowing penetration into membranes (Braun and von Heijne 1999, Gleason, Greathouse et al. 2013). Trp24 has been shown to be essential in the biological activity of AS-48, as it is located in a hydrophobic region that interacts with the membrane (Sanchez-Hidalgo, Fernández-Escamilla et al. 2010).

It has been previously pointed out that circular bacteriocins have similar hydrophobicity profiles (Kawai, Kemperman et al. 2004). Analysis of hydrophobicity profiles allowed increased confidence in the identification of putative circular bacteriocins discerned through sequence similarity. Hydrophobic profiles were maintained within subfamilies, as well as more generally within the families Ilc i and ii. By comparing profiles of putatively identified sequences to the average profile of each family, it can be determined which family they belong to. This could also be used to screen out non-circular bacteriocins. Though the hydrophobicity profiles are different between the families Ilc i and ii, if the profile of Ilc ii is flipped, the profile is surprisingly similar to Ilc i (Fig S3). There are particular sequences which show divergence to the profiles, such as *L. mesenteroides* TK41401 (leucocyclicin Q) and *Lactococcus* sp. QU 12 (lactocyclicin Q) from Ilc i, and *Trichococcus alkaliphilus* B5 (paracyclincin subfamily) and *Alkalibacterium* AK22 (akalicyclin subfamily) from Ilc ii.

A hydrophilic region was found in every putative and experimentally confirmed circular bacteriocin (Figure 2, Fig S3). This usually overlapped with the uni/polybasic region and implied a conserved functional region. There is strong evidence for a similar mechanism of action for this region, given the high levels of evolutionary conservation. This region is most likely involved with cell membrane interaction and binding based on the positively-charged basic residues and the negatively-charged cell membrane (Kim, Mosior et al. 1991, Jiménez, C Barrachi-Saccilotto et al. 2005). In experiments, the positively-charged (and polybasic) region of AS-48 (residues 49-69) showed no killing activity, but showed competitive binding to the negatively-charged membrane against the wild type AS-48 bacteriocin (Jiménez, C Barrachi-Saccilotto et al. 2005), indicating the role this region plays in the bactericidal activity of circular bacteriocins. Butyrivibriocin AR10 uncharacteristically does not contain a polybasic region (only a single basic residue), yet is functional as a circular bacteriocin against other *B. fibrisolvens* isolates (Kalmokoff and Teather 1997). It has a hydrophobic profile with a hydrophilic region which is consistent with family Ilc ii. This indicates polybasic regions aren't necessarily required for antimicrobial activity, but the hydrophilic region is.

## Phylogenetics

Phylogenetic classification has resolution trade-offs. A higher number of families (reasonably up to 6) could have been attributed, though due to the similar proposed modes of action (Gabrielsen, Brede et al. 2014, Perez, Zendo et al. 2018) and conserved structural motifs (González, Langdon et al. 2000, Martin-Visscher, Gong et al. 2009, Acedo, van Belkum et al. 2015, Himeno, Rosengren et al. 2015), further familial division would result in diminished returns. By classifying circular bacteriocins into groups with higher resolution such as subfamilies, experimentally-confirmed circular bacteriocins can be used as type-sequences and accurate sequence analysis and comparisons can be performed. This reduces the background noise of distantly-related circular bacteriocins within the immediate sequence family. It's highly probable that the putative circular bacteriocins within each subfamily share a similar mechanism of action but have their own distinct spectrum of activity. The phylogenetic classifications were further enforced by cluster analysis. For example, uberolysin and amylocyclicin circular bacteriocin subfamilies are distinct at the cluster level, have different hydrophobicity profiles at their C termini (Fig S3), yet are not divergent regarding structural gene homology despite a size difference of 6 residues.

## Conserved genes within circular bacteriocin clusters

Cluster analysis proved to be informative for determining putative functional circular bacteriocins, as well as phylogenetic classification. Recently-diverged structural genes would most likely have similar associated genes within their bacteriocin clusters. The drawback of this type of analysis was the associated genes essential for circular bacteriocin product may not be present within the same cluster but elsewhere within the genomic material. However, given a conservative 77.6% estimate of putative functionality, a number of potentially useful antimicrobial peptides have been highlighted (Figure 1). It is probable that some of these clusters contain non-functional pseudogenes, but given that most clusters were 'intact' upon comparison to experimentally confirmed clusters, the genes are considered conserved for circular bacteriocin production (Tomita, Fujimoto et al. 1997, Kawai, Saito et al. 1998, Martínez-Bueno, Valdivia et al. 1998, Kalmokoff, Cyr et al. 2003, Kemperman, Kuipers et al. 2003, Wirawan, Swanson et al. 2007, Martin-Visscher, van Belkum et al. 2008, Sawa, Zendo et al. 2009, Borrero, Brede et al. 2011, Masuda, Ono et al. 2011, Golneshin 2014, Potter, Ceotto et al. 2014, Scholz, Vater et al. 2014, Acedo, van Belkum et al. 2015, Perez, Ishibashi et al. 2016, Collins, O'Connor et al. 2017, Borrero, Kelly et al. 2018).

Strains with multi-structural gene clusters are an undescribed phenomenon until now. Given their high sequence identities to each other, it is clear they are a result of duplication events in which slight variants with independent promoters have been selected for. It is most likely that these strains swap or co-express variable circular bacteriocins via response regulators and quorum sensing (Kalmokoff, Cyr et al. 2003, Wirawan, Swanson et al. 2007, Bartholomae, Buivydas et al. 2017), allowing expression of different circular bacteriocins with a slightly different spectrum of activity/microbial targets. These multi-structural gene clusters can also give us insights into the putative immunity genes. It appears one putative immunity gene is enough to provide protection against each circular bacteriocin variant within the cluster. This indicates immunity genes may provide broader immunity than once thought and may possibly provide immunity to similar circular bacteriocins with as low as ~80% similarity. Based on the presence of two (sometimes three) putative transmembrane domains, as well as the central hydrophilic region and presence of acidic residues at the termini, the mechanism of immunity can be proposed. Immunity proteins may function as transmembrane proteins and competitively bind positively-charged/polybasic regions of corresponding circular bacteriocins, thus reducing pore formation within the cell membrane. Acidic residues found in the immunity proteins may compete with the negatively-charged cell membrane. Though, further experimental analysis is required, as immunity has been shown to be a cumulative effect with other genes within the cluster demonstrating a role in immunity (Mercedes, Antonio et al. 2004, Mu, Masuda et al. 2014, Perez, Ishibashi et al. 2016). More broadly, the observation that immunity genes are present in most gene clusters indicates these bacteria are susceptible to their own bacteriocins. Therefore, related species may also prove susceptible if lacking the corresponding immunity gene. This is hopeful as circular bacteriocins identified here were found in *Enterococcus*, *Staphylococcus* and *Streptococcus* species, which are currently regarded by the WHO as priority organisms for discovery of new antimicrobials (WHO 2017).

## Selfish genetic elements

Although providing fitness to the cell, circular bacteriocins and their associated clusters can be thought of as selfish genetic elements. Given the high stability of circular bacteriocins, if at any time the cluster is mutated or plasmid is lost, the immunity factors associated with the cluster may also be lost. The ex-producer would then be susceptible to the bacteriocin, and therefore this phenotype will be selected against. Also, given the high temporal stability of circular bacteriocins, they would also be more stable than the immunity genes which would be more susceptible to proteases, heat, pH etc, and would require continual renewal via gene expression. By nature, it is a toxin-antitoxin system which locks the producing strain into a long-term partnership. It has been demonstrated by removing the circular bacteriocin gassericin A from a plasmid, segregational stability of that plasmid drops (Ito, Kawai et al. 2009). This explains why so many of the circular bacteriocin clusters identified were putatively intact (Figure 1, Table 3), regardless if they are chromosomally associated or plasmid-borne. As previously described, the spectrum of antimicrobial activity (usually to closely related species) of circular bacteriocins provides further evidence of the toxin-antitoxin relationship (Martin-Visscher, van Belkum et al. 2008, Ito, Kawai et al. 2009, Egan 2018). Coincidentally, the circular bacteriocin from *L. nodensis* DSM 19682 was previously highlighted by a similar genome-mining study and the strain was not found to demonstrate antimicrobial activity against a range of bacteria including *Enterococci* and *Lactobacilli* (Collins, O'Connor et al. 2017). Given that the gene cluster was identified as intact (Figure 1, Fig S5), it is possible the bacteriocin was not tested against closely-related strains (including *L. nodensis*) which may demonstrate susceptibility.

Based on the selfish gene hypothesis, any clusters currently identified as missing a putative immunity gene should realistically have one that is just yet to be identified. Production of the circular bacteriocin without immunity factors will result in death of the producer, therefore sequence data from pure cultures would not exist, indicating the immunity gene is currently unidentified. Alternatively, if the antimicrobial activity mode of action relies on specific target receptors not found in the producer strain, immunity genes would not be needed. Another alternative explanation is recent inactivation of the entire immunity-gene-lacking cluster, which given enough time will eventually be reduced to pseudogenes and vestigial fragments. Though, given the general intactness of the clusters, the former hypothesis seems most likely. Being associated with conjugative plasmids or mobile genetic elements (Table 3) such as transposons allows wider dissemination of these genes within populations.

## Methods

### Identification of putative circular bacteriocins

NCBI was mined (date accessed: 20/2/19) against the 17 known mature circular bacteriocin sequences minus signal sequences.

### Phylogenetic analysis of putative circular bacteriocins

Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) (date accessed 25/2/19) (Chojnacki, Cowley et al. 2017) was used for alignments and exported to fasta format, which was used as input for RAXML (raxmlHPC-PTHREADS-SSE3 version 8.2.10) (Stamatakis 2014) using the following parameters for ML + rapid bootstrap analysis with 100 replicates:

```
-T 2 -f a -x 285 -m PROTGAMMABLOSUM62 -p 639 -N 100
```

The bipartitions output file was used in FigTree version 1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>) for viewing/manipulation. Microsoft Excel version 1902 was used to compile the table of putative circular bacteriocin and characteristics, which was then manipulated using Inkscape version 0.92 (<https://www.inkscape.org>).

### Circular bacteriocin characteristics

Polybasic residues were identified in the mature bacteriocin via the 'Mark' function in Notepad++ version 7.5.9 searching for the string "R|K|H" using the following search modifiers: 'Regular expression' and 'Match case'.

## Hydrophobicity analysis

Hydrophobicity profiles were generated using the protscale website <https://web.expasy.org/protscale/> with a sliding window of 9 (Kyte and Doolittle 1982). 95% confidence intervals were calculated using the Descriptive Statistics module from the Data Analysis ToolPak in Microsoft Excel. As C and N termini would be joined in the mature circular bacteriocin form, the first four residues were copied to the end of the sequence and the final four residues were copied to the beginning of the sequence to account for the sliding window of 9. This was performed by searching the amino acid fasta file for: `^([A-Z]{4})([A-Z]*)([A-Z]{4})$` and replacing with `$4$1$2` with search modifiers: 'Regular expression' and 'Match case' in Notepad++.

## Transmembrane domain analysis

Sequences were submitted to Phobius (Krogh, Sonnhammer et al. 2007) (date accessed: 1/3/19).

## Sequence logos

Skyline (date accessed: 9/4/19) was used after Clustal Omega alignments using the 'Observed Counts', 'Alignment sequences are full length', and 'Information Content—All' parameters (Wheeler, Clements et al. 2014).

## Gene cluster analysis

To determine if the circular bacteriocin structural gene and associated gene clusters were present on plasmids or chromosome, tBlastn and BLASTn (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) (NCBI 2017) was used to see if there were significant nucleotide hits to plasmids or chromosomes on NCBI. Size was also considered; if a gene cluster was on a contig >100 kb, it was considered most likely chromosomal. Functional domains were determined using HMMER version 3.2.1 (<http://hmmer.org/>) (Luciani, Lopez et al. 2018), along with NCBI annotations to infer gene function. Presence of plasmid-determinants such as repA/B and mobilisation genes were used to determine presence of cluster on plasmid. Presence of chromosomal determinants such as the 16s and tRNA genes were used to infer chromosomal localisation. If location was unclear, they were determined as 'Unknown'.

For gene clusters broken up amongst multiple contigs, contigs containing cluster elements were first joined with 5 N's, and then used for cluster alignments and analysis.

Easyfig version 2.2.3 (Sullivan, Petty et al. 2011) was used to align and visualise gene clusters using the tblastx function with an e-value cut-off of 0.001. *Lactococcus* sp. QU 12 was excluded from cluster analysis as only the structural gene sequence data is publicly available.

## Declarations

## Availability of data and materials

All data generated or analysed during this study are included in this published article [and its supplementary information files].

## Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Funding

This work was funded by Griffith University.

## Author Contributions

Analysis performed by BV. Experimental design by BV, BHAR and ATS. All authors read and approved the final manuscript.

## Acknowledgments

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors

## References

Acedo, J. Z., M. J. van Belkum, C. T. Lohans, R. T. McKay, M. Miskolzie and J. C. Vederas (2015). "Solution Structure of Acidocin B, a Circular Bacteriocin Produced by *Lactobacillus acidophilus* M46."

*Applied and Environmental Microbiology* 81(8): 2910.

Bartholomae, M., A. Buivydas, J. H. Viel, M. Montalbán-López and O. P. Kuipers (2017). "Major gene-regulatory mechanisms operating in ribosomally synthesized and post-translationally modified peptide (RiPP) biosynthesis." *Molecular Microbiology* 106(2): 186–206.

Borrero, J., D. A. Brede, M. Skaugen, D. B. Diep, C. Herranz, I. F. Nes, L. M. Cintas and P. E. Hernández (2011). "Characterization of Garvicin ML, a Novel Circular Bacteriocin Produced by *Lactococcus garvieae* DCC43, Isolated from Mallard Ducks (*Anas platyrhynchos*)." *Applied and Environmental Microbiology* 77(1): 369.

Borrero, J., E. Kelly, P.M. Connor, P. Kelleher, C. Scully, P. D. Cotter, J. Mahony and D. van Sinderen (2018). "Plantaricyclin A, a Novel Circular Bacteriocin Produced by *Lactobacillus plantarum* NI326: Purification, Characterization, and Heterologous Production." *Applied and Environmental Microbiology* 84(1): e01801–01817.

Braun, P. and G. von Heijne (1999). "The Aromatic Residues Trp and Phe Have Different Effects on the Positioning of a Transmembrane Helix in the Microsomal Membrane." *Biochemistry* 38(30): 9778–9782.

Cebrián, R., M. Maqueda, J. L. Neira, E. Valdivia, M. Martínez-Bueno and M. Montalbán-López (2010). "Insights into the Functionality of the Putative Residues Involved in Enterocin AS-48 Maturation." *Applied and Environmental Microbiology* 76(21): 7268.

Ceotto, H., D. Brede, Z. Salehian, J. dos Santos Nascimento, P. C. Fagundes, I. F. Nes and M. do Carmo de Freire Bastos (2010). "Aureocins 4185, Bacteriocins Produced by *Staphylococcus aureus* 4185: Potential Application in Food Preservation." *Foodborne Pathogens and Disease* 7(10): 1255–1262.

Ceotto, H., J. d. S. Nascimento, M. A. V. d. P. Brito and M. d. C. d. F. Bastos (2009). "Bacteriocin production by *Staphylococcus aureus* involved in bovine mastitis in Brazil." *Research in microbiology* 160(8): 592–599.

Chojnacki, S., A. Cowley, J. Lee, A. Foix and R. Lopez (2017). "Programmatic access to bioinformatics tools from EMBL-EBI update: 2017." *Nucleic acids research* 45(W1): W550-W553.

Clark, R. J., H. Fischer, L. Dempster, N. L. Daly, K. J. Rosengren, S. T. Nevin, F. A. Meunier, D. J. Adams and D. J. Craik (2005). "Engineering stable peptide toxins by means of backbone cyclization: Stabilization of the  $\alpha$ -conotoxin MII." *Proceedings of the National Academy of Sciences of the United States of America* 102(39): 13767.

Collins, F. W. J., P.M. O'Connor, O. O'Sullivan, B. Gómez-Sala, M. C. Rea, C. Hill and R. P. Ross (2017). "Bacteriocin Gene-Trait matching across the complete *Lactobacillus* Pan-genome." *Scientific Reports* 7(1): 3481.

Cotter, P. D., C. Hill and R. P. Ross (2005). "Bacteriocins: developing innate immunity for food." *Nature Reviews Microbiology* 3: 777.

Egan, K. (2018). *Discovery and evaluation of novel and characterised bacteriocins for future applications*, University College Cork.

Espinosa-Hernández, E., J. I. Morales-Camacho, D. A. Fernández-Velasco, C. G. Benítez-Cardoza, F. d. F. Rosas-Cárdenas and S. Luna-Suárez (2019). "The insertion of bioactive peptides at the C-terminal end of an 11S globulin changes the structural stability and improves the antihypertensive activity." *Electronic Journal of Biotechnology* 37: 18–24.

Fujimoto, S., H. Tomita, E. Wakamatsu, K. Tanimoto and Y. Ike (1995). *Physical mapping of the conjugative bacteriocin plasmid pPD1 of Enterococcus faecalis and identification of the determinant related to the pheromon response*.

Gabrielsen, C., D. A. Brede, P. E. Hernández, I. F. Nes and D. B. Diep (2012). "The Maltose ABC Transporter in *Lactococcus lactis* Facilitates High-Level Sensitivity to the Circular Bacteriocin Garvicin ML." *Antimicrobial Agents and Chemotherapy* 56(6): 2908.

Gabrielsen, C., D. A. Brede, I. F. Nes and D. B. Diep (2014). "Circular Bacteriocins: Biosynthesis and Mode of Action." *Applied and Environmental Microbiology* 80(22): 6854.

Gálvez, A., M. Maqueda, M. Martínez-Bueno and E. Valdivia (1991). "Permeation of bacterial cells, permeation of cytoplasmic and artificial membrane vesicles, and channel formation on lipid bilayers by peptide antibiotic AS-48." *Journal of Bacteriology* 173(2): 886.

Gálvez, A., M. Maqueda, E. Valdivia, A. Quesada and E. Montoya (1986). "Characterization and partial purification of a broad spectrum antibiotic AS-48 produced by *Streptococcus faecalis*." *Canadian Journal of Microbiology* 32(10): 765–771.

Gleason, N. J., D. V. Greathouse, C. V. Grant, S. J. Opella and R. E. Koeppe (2013). "Single Tryptophan and Tyrosine Comparisons in the N-Terminal and C-Terminal Interface Regions of Transmembrane GWALP Peptides." *The Journal of Physical Chemistry B* 117(44): 13786–13794.

Golneshin, A. (2014). *Characterisation of bacteriocin genes and proteins from Lactobacillus plantarum B21 as potential new antimicrobial agents and natural food preservatives* Doctor of Philosophy, RMIT University.

Gong, X., L. A. Martin-Visscher, D. Nahirney, J. C. Vederas and M. Duszyk (2009). "The circular bacteriocin, carnocyclin A, forms anion-selective channels in lipid bilayers." *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1788(9): 1797–1803.

González, C., G. M. Langdon, M. Bruix, A. Gálvez, E. Valdivia, M. Maqueda and M. Rico (2000). "Bacteriocin AS-48, a microbial cyclic polypeptide structurally and functionally related to mammalian NK-lysin." *Proceedings of the National Academy of Sciences* 97(21): 11221.

Himeno, K., K. J. Rosengren, T. Inoue, R. H. Perez, M. L. Colgrave, H. S. Lee, L. Y. Chan, S. T. Henriques, K. Fujita, N. Ishibashi, T. Zendo, P. Wilaipun, J. Nakayama, V. Leelawatcharamas, H. Jikuya, D. J. Craik and K. Sonomoto (2015). "Identification, Characterization, and Three-Dimensional Structure of the Novel Circular Bacteriocin, Enterocin NKR-5-3B, from *Enterococcus faecium*." *Biochemistry* 54(31): 4863-4876.

Ishibashi, N., K. Himeno, K. Fujita, Y. Masuda, R. H. Perez, T. Zendo, P. Wilaipun, V. Leelawatcharamas, J. Nakayama and K. Sonomoto (2012). "Purification and Characterization of Multiple Bacteriocins and an Inducing Peptide Produced by *Enterococcus faecium* NKR-5-3 from Thai Fermented Fish." *Bioscience, Biotechnology, and Biochemistry* 76(5): 947-953.

Ito, Y., Y. Kawai, K. Arakawa, Y. Honme, T. Sasaki and T. Saito (2009). "Conjugative Plasmid from *Lactobacillus gasseri* LA39 That Carries Genes for Production of and Immunity to the Circular Bacteriocin Gassericin A." *Applied and Environmental Microbiology* 75(19): 6340.

Iwai, H. and A. Plückthun (1999). "Circular  $\beta$ -lactamase: stability enhancement by cyclizing the backbone." *FEBS Letters* 459(2): 166-172.

Jiménez, J. J., D. B. Diép, J. Borrero, L. Gútiéz, S. Arbulu, I. F. Nes, C. Herranz, L. M. Cintas and P. E. Hernández (2015). "Cloning strategies for heterologous expression of the bacteriocin enterocin A by *Lactobacillus sakei* Lb790, *Lb. plantarum* NC8 and *Lb. casei* CECT475." *Microbial Cell Factories* 14(1): 166.

Jiménez, M. A., A. C Barrachi-Saccilotto, E. Valdivia, M. Maqueda and M. Rico (2005). *Design, NMR characterization and activity of a 21-residue peptide fragment of bacteriocin AS-48 containing its putative membrane interacting region.*

Kalmokoff, M. L., T. D. Cyr, M. A. Hefford, M. F. Whitford and R. M. Teather (2003). "Butyriovibriocin AR10, a new cyclic bacteriocin produced by the ruminal anaerobe *Butyriovibrio fibrisolvens* AR10: characterization of the gene and peptide." *Canadian Journal of Microbiology* 49(12): 763-773.

Kalmokoff, M. L. and R. M. Teather (1997). "Isolation and characterization of a bacteriocin (Butyriovibriocin AR10) from the ruminal anaerobe *Butyriovibrio fibrisolvens* AR10: evidence in support of the widespread occurrence of bacteriocin-like activity among ruminal isolates of *B. fibrisolvens*." *Applied and Environmental Microbiology* 63: 394-402.

Kawai, Y., Y. Ishii, K. Arakawa, K. Uemura, B. Saitoh, J. Nishimura, H. Kitazawa, Y. Yamazaki, Y. Tateno, T. Itoh and T. Saito (2004). "Structural and Functional Differences in Two Cyclic Bacteriocins with the Same Sequences Produced by *Lactobacilli*." *Applied and Environmental Microbiology* 70(5): 2906.

Kawai, Y., R. Kemperman, J. Kok and T. Saito (2004). "The Circular Bacteriocins Gassericin A and Circularin A." *Current protein & peptide science* 5: 393-398.

Kawai, Y., T. Saito, H. Kitazawa and T. Itoh (1998). "Gassericin A; an Uncommon Cyclic Bacteriocin Produced by *Lactobacillus gasseri* LA39 Linked at N- and C-Terminal Ends." *Bioscience, Biotechnology, and Biochemistry* 62(12): 2438-2440.

Kemperman, R., A. Kuipers, H. Karsens, A. Nauta, O. Kuipers and J. Kok (2003). "Identification and Characterization of Two Novel Clostridial Bacteriocins, Circularin A and Closticin 574." *Applied and Environmental Microbiology* 69(3): 1589.

Kim, J., M. Mosior, L. A. Chung, H. Wu and S. McLaughlin (1991). "Binding of peptides with basic residues to membranes containing acidic phospholipids." *Biophysical Journal* 60(1): 135-148.

Krogh, A., E. L. L. Sonnhammer and L. Käll (2007). "Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server." *Nucleic Acids Research* 35(suppl\_2): W429-W432.

Kurata, A., T. Yamaguchi, M. Kira and N. Kishimoto (2019). "Characterization and heterologous expression of an antimicrobial peptide from *Bacillus amyloliquefaciens* CMW1." *Biotechnology & Biotechnological Equipment* 33(1): 886-893.

Kyte, J. and R. F. Doolittle (1982). "A simple method for displaying the hydropathic character of a protein." *Journal of Molecular Biology* 157(1): 105-132.

Logan, N. A., O. Berge, A. H. Bishop, H.-J. Busse, P. De Vos, D. Fritze, M. Heyndrickx, P. Kämpfer, L. Rabinovitch, M. S. Salkinoja-Salonen, L. Seldin and A. Ventosa (2009). "Proposed minimal standards for describing new taxa of aerobic, endospore-forming bacteria." *International Journal of Systematic and Evolutionary Microbiology* 59(8): 2114-2121.

Luciani, A., R. Lopez, S. C. Potter, Y. Park, R. D. Finn and S. R. Eddy (2018). "HMMER web server: 2018 update." *Nucleic Acids Research* 46(W1): W200-W204.

Maqueda, M., M. Sánchez-Hidalgo, M. Fernández, M. Montalbán-López, E. Valdivia and M. Martínez-Bueno (2008). "Genetic features of circular bacteriocins produced by Gram-positive bacteria." *FEMS Microbiology Reviews* 32(1): 2-22.

Martin-Visscher, L. A., X. Gong, M. Duszyk and J. C. Vederas (2009). "The Three-dimensional Structure of Carnocyclin A Reveals That Many Circular Bacteriocins Share a Common Structural Motif." *Journal of Biological Chemistry* 284(42): 28674-28681.

Martin-Visscher, L. A., M. J. van Belkum, S. Gameau-Tsodikova, R. M. Whittall, J. Zheng, L. M. McMullen and J. C. Vederas (2008). "Isolation and characterization of carnocyclin a, a novel circular bacteriocin produced by *Carnobacterium maltaromaticum* UAL307." *Applied and environmental microbiology* 74(15): 4756-4763.

- Martínez-Bueno, M., A. Gálvez, E. Valdivia and M. Maqueda (1990). "A transferable plasmid associated with AS-48 production in *Enterococcus faecalis*." *Journal of Bacteriology* 172(5): 2817.
- Martínez-Bueno, M., E. Valdivia, A. Gálvez, J. Coyette and M. Maqueda (1998). "Analysis of the gene cluster involved in production and immunity of the peptide antibiotic AS-48 in *Enterococcus faecalis*." *Molecular Microbiology* 27(2): 347-358.
- Masuda, Y., H. Ono, H. Kitagawa, H. Ito, F. Mu, N. Sawa, T. Zendo and K. Sonomoto (2011). "Identification and Characterization of Leucocyclin Q, a Novel Cyclic Bacteriocin Produced by *Leuconostoc mesenteroides* TK41401." *Applied and Environmental Microbiology* 77(22): 8164.
- Mercedes, M., G. Antonio, B. Manuel Martinez, S.-B. Maria Jose, G. Carlos, A. Armando, R. Manuel and V. Eva (2004). "Peptide AS-48: Prototype of a New Class of Cyclic Bacteriocins." *Current Protein & Peptide Science* 5(5): 399-416.
- Montalbán-López, M., B. Spolaore, O. Pinato, M. Martínez-Bueno, E. Valdivia, M. Maqueda and A. Fontana (2008). "Characterization of linear forms of the circular enterocin AS-48 obtained by limited proteolysis." *FEBS Letters* 582(21-22): 3237-3242.
- Mu, F., Y. Masuda, T. Zendo, H. Ono, H. Kitagawa, H. Ito, J. Nakayama and K. Sonomoto (2014). "Biological function of a DUF95 superfamily protein involved in the biosynthesis of a circular bacteriocin, leucocyclin Q." *Journal of Bioscience and Bioengineering* 117(2): 158-164.
- Nazina, T. N., T. P. Tourova, A. B. Poltarus, E. V. Novikova, A. A. Grigoryan, A. E. Ivanova, A. M. Lysenko, V. V. Petrunyaka, G. A. Osipov, S. S. Belyaev and M. V. Ivanov (2001). "Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzonensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*, *G. th.*" *International Journal of Systematic and Evolutionary Microbiology* 51(2): 433-446.
- NCBI (2017). "Database resources of the National Center for Biotechnology Information." *Nucleic Acids Research* 46(D1): D8-D13.
- Perez, R. H., N. Ishibashi, T. Inoue, K. Himeno, Y. Masuda, N. Sawa, T. Zendo, P. Wilaipun, V. Leelawatcharamas, J. Nakayama and K. Sonomoto (2016). "Functional Analysis of Genes Involved in the Biosynthesis of Enterocin NKR-5-3B, a Novel Circular Bacteriocin." *Journal of Bacteriology* 198(2): 291.
- Perez, R. H., H. Sugino, N. Ishibashi, T. Zendo, P. Wilaipun, V. Leelawatcharamas, J. Nakayama and K. Sonomoto (2017). "Mutations near the cleavage site of enterocin NKR-5-3B prepeptide reveal new insights into its biosynthesis." *Microbiology* 163(4): 431-441.
- Perez, R. H., T. Zendo and K. Sonomoto (2014). "Novel bacteriocins from lactic acid bacteria (LAB): various structures and applications." *Microbial Cell Factories* 13(1): S3.
- Perez, R. H., T. Zendo and K. Sonomoto (2018). "Circular and Leaderless Bacteriocins: Biosynthesis, Mode of Action, Applications, and Prospects." *Frontiers in Microbiology* 9(2085).
- Potter, A., H. Ceotto, M. L. V. Coelho, A. J. Guimarães and M. d. C. d. F. Bastos (2014). "The gene cluster of aureocyclin 4185: the first cyclic bacteriocin of *Staphylococcus aureus*." *Microbiology* 160(5): 917-928.
- Samyn, B., M. Martínez-Bueno, B. Devreese, M. Maqueda, A. Gálvez, E. Valdivia, J. Coyette and J. Van Beeumen (1994). "The cyclic structure of the enterococcal peptide antibiotic AS-48." *FEBS Letters* 352(1): 87-90.
- Sanchez-Hidalgo, M., A.M. Fernández-Escamilla, M. Martínez-Bueno, E. Valdivia, L. Serrano and M. Maqueda (2010). *Conformational Stability and Activity of Circular Enterocin AS-48 Derivatives*.
- Sánchez-Hidalgo, M., M. Montalbán-López, R. Cebrián, E. Valdivia, M. Martínez-Bueno and M. Maqueda (2011). "AS-48 bacteriocin: close to perfection." *Cellular and Molecular Life Sciences* 68(17): 2845-2857.
- Sawa, N., T. Zendo, J. Kiyofuji, K. Fujita, K. Himeno, J. Nakayama and K. Sonomoto (2009). "Identification and Characterization of Lactocyclin Q, a Novel Cyclic Bacteriocin Produced by *Lactococcus* sp. Strain QU 12
- " *Applied and Environmental Microbiology* 75: 1552-1558.
- Scholz, R., J. Vater, A. Budiharjo, Z. Wang, Y. He, K. Dietel, T. Schwecke, S. Herfort, P. Lasch and R. Borriss (2014). "Amylocyclin, a Novel Circular Bacteriocin Produced by *Bacillus amyloliquefaciens* FZB42." *Journal of Bacteriology* 196(10): 1842.
- Stamatakis, A. (2014). "RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies." *Bioinformatics* 30(9): 1312-1313.
- Sullivan, M. J., N. K. Petty and S. A. Beatson (2011). "Easyfig: a genome comparison visualizer." *Bioinformatics* 27(7): 1009-1010.
- Thage, B. V., F. P. Rattray, M. W. Laustsen, Y. Ardö, V. Barkholt and U. Houlberg (2004). "Purification and characterization of a branched-chain amino acid aminotransferase from *Lactobacillus paracasei* subsp. *paracasei* CHCC 2115." *Journal of Applied Microbiology* 96(3): 593-602.

Tomita, H., S. Fujimoto, K. Tanimoto and Y. Ike (1997). "Cloning and genetic and sequence analyses of the bacteriocin 21 determinant encoded on the *Enterococcus faecalis* pheromone-responsive conjugative plasmid pPD1." *Journal of Bacteriology* 179(24): 7843.

van de Guchte, M., J. Kok and G. Venema (1991). "Distance-dependent translational coupling and interference in *Lactococcus lactis*." *Molecular and General Genetics MGG* 227(1): 65–71.

von Heijne, G. and L. Abrahmsen (1989). "Species-specific variation in signal peptide design Implications for protein secretion in foreign hosts." *FEBS Letters* 244(2): 439–446.

Wheeler, T. J., J. Clements and R. D. Finn (2014). "Skyline: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models." *BMC Bioinformatics* 15(1): 7.

WHO (2017). WHO priority pathogens list for R&D of new antibiotics. W. H. Organisation. <https://www.who.int>, Organisation.

Wirawan, R. E., K. M. Swanson, T. Kleffmann, R. W. Jack and J. R. Tagg (2007). "Uberolysin: a novel cyclic bacteriocin produced by *Streptococcus uberis*." *Microbiology* 153(5): 1619–1630.

Zamfir, M., R. Callewaert, P. C. Cornea, L. Savu, I. Vatafu and L. De Vuyst (1999). "Purification and characterization of a bacteriocin produced by *Lactobacillus acidophilus* IBB 801." *Journal of Applied Microbiology* 87(6): 923–931.

## Tables

Table 1: Information about the 19 experimentally confirmed circular bacteriocins and their producer strains. <sup>P</sup> denotes the characteristic is putative based on the shared characteristics with the other circular bacteriocins.

Class	Circular bacteriocin	Bacterial producer	3D structure	$\alpha$ helices	aa size of mature peptide (Mass in Da)	Cluster location	Conjugative plasmid (Y/N/Unknown)	Genes in cluster	Translational coupling	A
IIC i	Aureocyclicin 4185	<i>Staphylococcus aureus</i> 4185	Model based on carnocyclin A	4 <sup>P</sup>	60 (5,607 Da)	pRJ101	Unknown (data unavailable)	11	Y	K
IIC i	Enterocin NKR-5-3B	<i>Enterococcus faecium</i> NKR-5-3B	NMR solution structure	4	64 (6316.4 Da)	Unknown (data unavailable)	N	5	Y	L
IIC i	Amylocyclicin	<i>Bacillus amyloliquefaciens</i> FZB42	N/A	4 <sup>P</sup>	64 (6,382 Da)	Chromosome	N	6	Y	C
IIC i	Amylocyclicin CMW1	<i>Bacillus amyloliquefaciens</i> CMW1	N/A	4 <sup>P</sup>	64 (6351.59 Da)	Chromosome	N	6	Y	D
IIC i	AS-48	<i>Enterococcus faecalis</i> subsp. <i>liquefaciens</i> S-48	NMR solution structure	5	70 (7,150.17 Da)	pMB2	Y	10	Y	A
IIC i	bacA	<i>Enterococcus faecalis</i> 39-5	N/A	5 <sup>P</sup>	70 (7,150.17 Da)	pPD1	Y	9	Y	D
IIC i	Carnocyclin A	<i>Carnobacterium maltaromaticum</i> UAL307	NMR	4	60 (5,862 Da)	Unknown (data unavailable)	Unknown (data unavailable)	10	Y	E
IIC i	Circularin A	<i>Clostridium beijerinckii</i> ATCC 25752	N/A	5 <sup>P</sup>	69 (6789.05 Da)	Chromosome	N	8	Y	A
IIC i	Thermocin 458	<i>Geobacillus stearothermophilus</i> DSM 458	N/A	5 <sup>P</sup>	70 (6933.25 Da)	Chromosome	N	8	Y	N
IIC i	Garvicin ML	<i>Lactococcus garvieae</i> DCC43	PSIPRED and Jpred3 model	4 <sup>P</sup>	60 (6,007.2 Da)	Chromosome	N	8 <sup>P</sup>	Y	N
IIC i	Lactocyclicin Q	<i>Lactococcus</i> sp. QU 12	Model using PSIPRED	4 <sup>P</sup>	61 (6,062.8 Da)	Unknown (data unavailable)	Unknown (data unavailable)	Unknown (data unavailable)	Y	A
IIC i	Leucocyclicin Q	<i>Leuconostoc mesenteroides</i> TK41401	N/A	4 <sup>P</sup>	61 (6133.23 Da)	Unknown (data unavailable)	Unknown (data unavailable)	5	Y	A
IIC i	Uberolysin	<i>Streptococcus uberis</i> 42	N/A	5 <sup>P</sup>	70 (7043.94 Da)	Chromosome	N	6 <sup>P</sup>	Y	D
IIC ii	Acidocin B	<i>Lactobacillus acidophilus</i> M46	NMR solution structure	4	58 (5,621.5 Da)	pCV461	Unknown (data unavailable)	7	Y	K
IIC ii	Butyrivibriocin AR10	<i>Butyrivibrio fibrisolvens</i> AR10	N/A	4 <sup>P</sup>	58 (5981.5 Da)	Chromosome	N	5	Y	A
IIC ii	Paracyclicin	<i>Lactobacillus paracasei</i> subsp. <i>paracasei</i> JCM 8130/DSM 5622	N/A	4 <sup>P</sup>	58 (5905.75 Da)	pLBPC-1	Y	7 <sup>P</sup>	Y	N
IIC ii	Gasserin A/ Reuterin 6	<i>Lactobacillus gasseri</i> LA39/ <i>Lactobacillus reuteri</i> LA6	N/A	4 <sup>P</sup>	58 (5,653.6 Da)	pLgLA39	Y	7	Y	A
IIC ii	Plantacyclin B21AG	<i>Lactobacillus plantarum</i> B21	N/A	4 <sup>P</sup>	58 (5,668 Da)	pB21AG01	Y <sup>P</sup>	7	Y	N
IIC ii	Plantaricyclin A	<i>Lactobacillus plantarum</i> NI326	N/A	4 <sup>P</sup>	58 (5,572 Da)	pB21AG01-like plasmid	Y <sup>P</sup>	7	Y	N

Table 2: List of circular bacteriocins identified with potential activity against clinically relevant isolates as part of the WHO's Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics (WHO 2017).

Table 3: Summary results of the mature bacteriocin sequence and gene cluster analysis.

Pathogen potentially susceptible	Bacteriocin producer	Experimental anti-bacterial activity against clinically relevant bacteria	Source
<i>Enterococcus faecium</i>	<i>Enterococcus faecalis</i> 39-5	<i>S. aureus</i> , <i>E. faecalis</i> , <i>E. faecium</i> , <i>S. agalactiae</i> , <i>S. sanguis</i>	(Fujimoto, Tomita et al. 1995)
<i>Enterococcus faecium</i>	<i>Enterococcus faecalis</i> subsp. <i>liquefaciens</i> S-48	<i>S. faecalis</i> S-47, <i>Escherichia coli</i> U-9, <i>E. faecalis</i> OGIX	(Gálvez, Maqueda et al. 1986, Martínez-Bueno, Gálvez et al. 1990)
<i>Enterococcus faecium</i>	<i>Enterococcus faecium</i> NKR-5-3B	<i>B. coagulans</i> , <i>B. subtilis</i> , <i>L. lactis</i> ssp. <i>lactis</i> , <i>L. sakei</i> ssp. <i>sakei</i>	(Ishibashi, Himeno et al. 2012)
<i>Enterococcus faecium</i>	<i>Enterococcus villorum</i> F1129D	N/A	This study
<i>Enterococcus faecium</i>	<i>Enterococcus thailandicus</i> a523	N/A	This study
<i>Enterococcus faecium</i>	<i>Enterococcus faecium</i> E8172	N/A	This study
<i>Enterococcus faecium</i>	<i>Enterococcus faecium</i> IHC105	N/A	This study
<i>Enterococcus faecium</i>	<i>Enterococcus faecalis</i> AZ19	N/A	This study
<i>Staphylococcus aureus</i>	<i>Staphylococcus aureus</i> 4185	<i>B. cereus</i> , <i>B. coagulans</i> , <i>B. licheniformis</i> , <i>L. monocytogenes</i> , <i>M. luteus</i>	(Ceotto, Nascimento et al. 2009, Ceotto, Brede et al. 2010)
<i>Staphylococcus aureus</i>	<i>Staphylococcus felis</i> F9	N/A	This study
<i>Staphylococcus aureus</i>	<i>Staphylococcus felis</i> F21	N/A	This study
<i>Staphylococcus aureus</i>	<i>Staphylococcus xylosum</i> INIFAP 004-15	N/A	This study
<i>Staphylococcus aureus</i>	<i>Staphylococcus kloosii</i> SNUC 4696	N/A	This study
<i>Staphylococcus aureus</i>	<i>Staphylococcus aureus</i> 3688STDY6124959	N/A	This study
<i>Staphylococcus aureus</i>	<i>Staphylococcus epidermidis</i> AU23	N/A	This study
<i>Streptococcus pneumoniae</i> /Group A/Group B	<i>Streptococcus uberis</i> 42	<i>E. faecalis</i> , <i>E. hirae</i> , <i>M. luteus</i> , <i>S. aureus</i> , <i>S. agalactiae</i> , <i>S. salivarius</i> , <i>S. pyogenes</i> , <i>S. equisimilis</i> , <i>S. dysgalactiae</i> , <i>S. anginosus</i>	(Wirawan, Swanson et al. 2007)
<i>Streptococcus pneumoniae</i> /Group A/Group B	<i>Streptococcus uberis</i> NCTC 4674	N/A	This study
<i>Streptococcus pneumoniae</i> /Group A/Group B	<i>Streptococcus suis</i> YS111	N/A	This study
<i>Streptococcus pneumoniae</i> /Group A/Group B	<i>Streptococcus</i> sp. OBRC6	N/A	This study
<i>Streptococcus pneumoniae</i> /Group A/Group B	<i>Streptococcus gordonii</i> BCA7	N/A	This study
<i>Streptococcus pneumoniae</i> /Group A/Group B	<i>Streptococcus cristatus</i> 787_SOLI	N/A	This study
<i>Streptococcus pneumoniae</i> /Group A/Group B	<i>Streptococcus</i> sp. CCH8-G7	N/A	This study
<i>Streptococcus pneumoniae</i> /Group A/Group B	<i>Streptococcus mitis</i> KCOM 1350	N/A	This study
<i>Streptococcus pneumoniae</i> /Group A/Group B	<i>Streptococcus pneumoniae</i> UoS2029	N/A	This study
<i>Streptococcus pneumoniae</i> /Group A/Group B	<i>Streptococcus thermophilus</i> NCTC 12958	N/A	This study
<i>Streptococcus pneumoniae</i> /Group A/Group B	<i>Streptococcus pneumoniae</i> Zo 18m39-2	N/A	This study

	Characteristic	No. of circular bacteriocins	Percentage of bacteriocins (%)
Mature bacteriocin sequence	Total circular bacteriocins	118	100.0
	Experimentally-confirmed	19	16.1
	Total identified	99	83.9
	Containing polybasic region	111	94.1
	Containing aromatic residues	115	97.5
	Family IIc i	89	75.4
	Family IIc ii	29	24.6
Bacteriocin cluster	Predicted functional (conservative)	90	76.3
	Translational coupling	109	92.4
	Chromosomally-located	77	65.3
	Plasmid-associated	25	21.2
	Unknown location	16	13.6
	Mobile genetic element-associated	24	20.3

## Figures

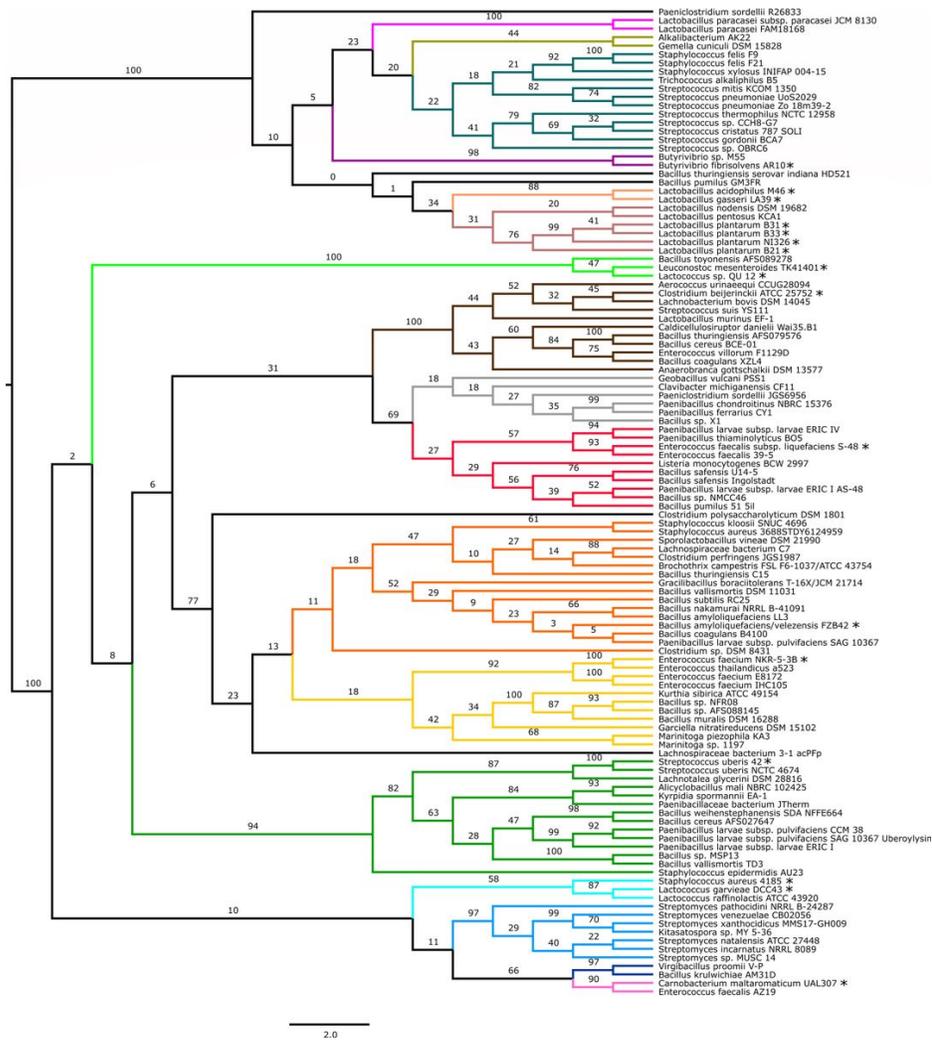
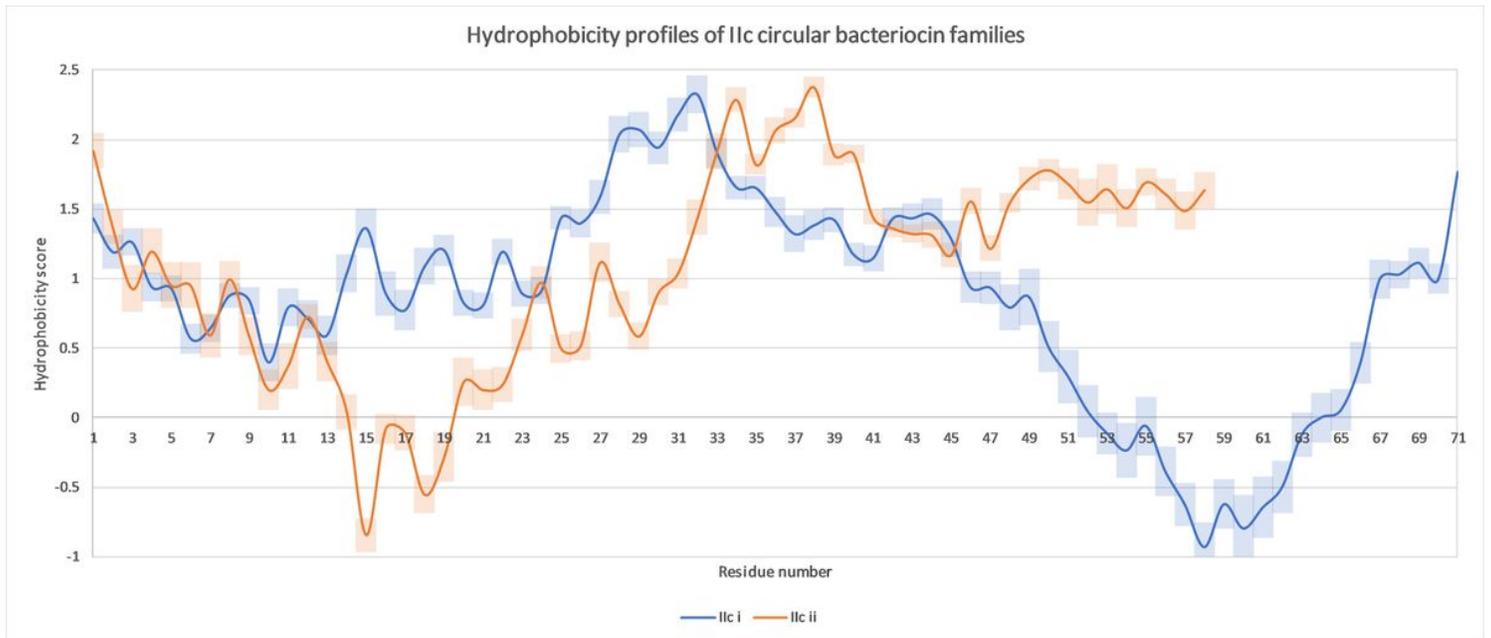
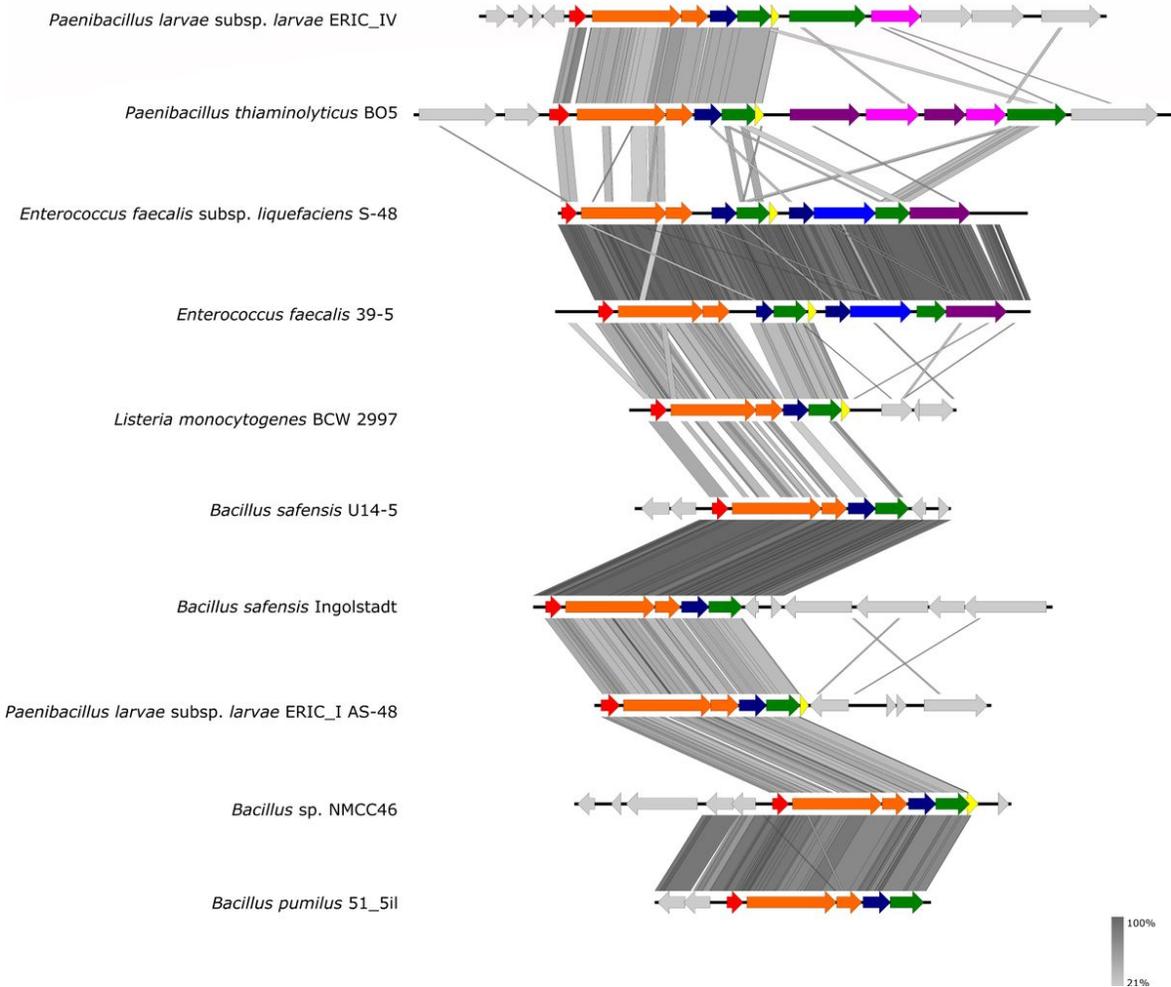


Figure 1

Phylogenetic tree showing the subfamilies of circular experimentally confirmed and putative bacteriocins. \*Experimentally-confirmed. Red: AS-48 subfamily. Orange: amylocyclin subfamily. Yellow: enterocin NKR-5-3B subfamily. Green: uberolysin subfamily. Light: aureocyclin 4185/garvicin ML subfamily. Blue: putative venezuelacyclin subfamily. Dark blue: putative krulwicyclin subfamily. Pink: carnocyclin subfamily. Fuschia: paracyclin subfamily. Gold: akalicyclin subfamily. Teal: streptocyclin subfamily. Purple: butyriovibriocin AR10 subfamily. Peach: gassericin/acidocin subfamily. Brown: plantaricyclin/plantacyclin subfamily. Light green: lactocyclin/leucocyclin subfamily. Dark brown: circularin family. Grey: putative bacillocyclin subfamily. Black shows any putative circular bacteriocins which are too distinct to fit within the proposed subfamilies.



**Figure 2**  
 Hydrophobicity profiles of two circular bacteriocin families. The blue lines shows family Ilc i and the orange line shows Ilc ii. The semi-transparent bars show the 95% confidence intervals. Hydrophobicity scores were calculated based on Kyte and Doolittle with a sliding window of 9 (Kyte and Doolittle 1982).  
**Conserved putative circular bacteriocin gene clusters within the AS-48 subfamily**



**Figure 3**

Diagram showing the conserved genes within the AS-48 subfamily. Arrows show putative genes within the gene cluster. Red shows the bacteriocin structural gene. Orange shows putative membrane proteins/stage II sporulation protein M. Yellow shows putative immunity genes such as as-48D1. Green shows putative ABC-transporter proteins such as as-48D and as-48G. Blue shows HylD/efflux RND transporters such as as-48F. Dark blue shows other transmembrane proteins. Pink shows putative binding-protein-dependent proteins and extracellular solute-binding proteins. Purple shows ABC-II/FtsX permeases such as-48H, based on work done with AS-48 (Mercedes, Antonio et al. 2004). Grey shows genes of unknown function which may or may not be related to circular bacteriocin production. Figure produced in Easyfig and Inkscape.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigS5.docx](#)
- [FigS2.docx](#)
- [FigS4.docx](#)
- [FigS7.docx](#)
- [FigS6.docx](#)
- [FigS1.xlsx](#)
- [FigS3.xlsx](#)
- [TableS1.docx](#)