WESLEY ELIAS BHERING BARRIOS

METAGENOME MINING OF PEDIOCIN-LIKE BACTERIOCINS AND CREATION OF pET-pedABCD, A CUSTOMIZABLE PEDIOCIN EXPRESSION VECTOR

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Bioquímica Aplicada, para obtenção do título de *Magister Scientiae*.

Orientador: Tiago Antônio de O. Mendes

Coorientador: Hilário Cuquetto Mantovani

VIÇOSA – MINAS GERAIS

2020

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T
Barrios, Wesley Elias Bhering, 1993-
B276m Metagenome mining of *pediocin-like* bacteriocins and
2020 creation of pET-pedABCD, a customizable pediocin expression vector / Wesley Elias Bhering Barrios. – Viçosa, MG, 2020.
57 f. : il. (algumas color.) ; 29 cm.

Orientador: Tiago Antônio de Oliveira Mendes.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Referências bibliográficas: f. 50-57.

1. Pediocinas. 2. Proteínas - Síntese. 3. Expressão gênica. 4. Mineração de dados (Computação). I. Universidade Federal de Viçosa. Departamento de Bioquímica e Biologia Molecular. Programa de Pós-Graduação em Bioquímica Aplicada. II. Título.

CDD 22. ed. 572.6

WESLEY ELIAS BHERING BARRIOS

METAGENOME MINING OF PEDIOCIN-LIKE BACTERIOCINS AND CREATION OF pET-pedABCD, A CUSTOMIZABLE PEDIOCIN EXPRESSION VECTOR

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Bioquímica Aplicada, para obtenção do título de *Magister Scientiae*.

APROVADA: 30 de setembro de 2020.

Assentimento:

Wesley Elias Bhering Barrios

Wesley Elias Bhering Barrios
Autor

Tiago Antônio de Oliveira Mendes

Tiago Antônio de Oliveira Mendes
Orientador

AGRADECIMENTOS

Agradeço a Deus, por me fortificar a cada nova luta e me dar a coragem, saúde e resiliência necessária para chegar até aqui e ir além.

A esta Instituição pelo ambiente criativo que proporciona e por ter proporcionado o caminho que me trouxe até aqui.

Ao meu orientador Prof. Dr. Tiago Antônio de Oliveira Mendes, por ter me dado a chance de recomeçar. Agradeço-o pelo suporte constante no pouco tempo que lhe coube, por suas correções e confiança depositada em todas as fases deste trabalho.

Ao meu coorientador Dr. Hilário Cuquetto Mantovani, por ter contribuído de forma concisa com seus conhecimentos compartilhados na área deste trabalho.

À Dra. Ramila Cristiane Rodrigues, pela ajuda, discussões, conversas esclarecedoras e disponibilidade em participar da banca.

Ao Dr. Luciano Gomes Fietto, pela disponibilidade em participar da banca ,por todos os conhecimentos transferidos durantes as disciplinas do Bacharelado em Bioquímica, e as valiosas contribuições em meu TCC.

À Dra. Maria Raquel Venturim Cosate, pela cordialidade e disponibilidade em participar da banca

Aos integrantes do LBM por terem me recebido com carinho; à Dra. Ananda Aguilar por estar sempre disponível para ajudar na resolução de problemas, e ao Msc. Renato Senra pelas valiosas dicas na análise de expressão proteica por SDS-PAGE.

Aos professores e colaboradores dos mais diferentes níveis da Universidade Federal de Viçosa, os quais além de me ensinarem, contribuíram para manutenção e funcionamento dessa instituição.

A minha família, luz da minha vida, por seu amor incondicional. Agradeço a vocês por apoiar todas as minhas decisões e acreditar que o melhor sempre está por vir.

A minha amada filha Laura que deu um novo sentido a minha existência e me fez enxergar que devo procurar todos os dias, a melhor versão de mim.

À minha esposa Débora por seu infinito amor. Agradeço-a por me apoiar, aconselhar, e por tudo de positivo que acrescentou em minha vida pessoal e profissional. Serei sempre grato!

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela concessão da bolsa de estudos.

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), pela concessão da bolsa de estudos.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão da bolsa de estudos.

Agradeço também a todos que de alguma forma, contribuíram para a minha formação.

RESUMO

BARRIOS, Wesley Elias Bhering, M.Sc., Universidade Federal de Viçosa, setembro de 2020. **Metagenome mining of pediocin-like bacteriocins and creation of pET-pedABCD, a customizable pediocin expression vector**. Orientador: Tiago Antônio de Oliveira Mendes. Coorientador: Hilário Cuquetto Mantovani.

Produtos lácteos fermentados, o intestino humano e o rúmen bovino são alguns ambientes que contém bactérias do ácido láctico (LAB). Nesses ambientes competitivos, a síntese de compostos antimicrobianos como bacteriocinas atua no controle da população de bactérias gram-positivas por meio de diversos mecanismos, favorecendo o microrganismo produtor. As bacteriocinas são classificadas em 3 classes, sendo a classe II, subclasse A, bacteriocinas análogas à pediocina, peptídeos com ampla atividade contra bactérias Gram-positivas e negativas. A pediocina PA-1, precursora desta subclasse, age através da inserção na membrana plasmática da bactéria alvo ligando-se ao transportador de manose amplamente conservado em bactérias gram-positivas ou por interações eletrostáticas, formando um poro que leva à morte celular. O elevado aumento de bactérias com resistência aos antibióticos clássicos evidencia a importância da descoberta de novos antimicrobianos. Uma alternativa para este problema pode estar nos metagenomas, que correspondem ao sequenciamento do DNA de comunidades microbianas, fornecendo acesso ao potencial genético de cepas não-cultiváveis, possibilitando a descoberta de novas bacteriocinas. Neste âmbito, este trabalho teve como objetivo principal a mineração de metagenomas de produtos lácteos fermentados utilizando o algoritmo Memi. Para tal, um novo padrão de aminoácidos característico da classe IIa foi desenvolvido para aumentar a precisão dos resultados e diminuição de falsos positivos. Inicialmente, novos padrões de aminoácidos característicos da classe IIa foram propostos a partir do alinhamento de sequências depositadas. Foi possível definir o memi-9 como novo padrão de aminoácidos característico das bacteriocinas classe IIa, apresentando 92,2% de sensibilidade e 100% de precisão. A mineração de metagenomas resultou no mapeamento de 52 bacteriocinas da classe IIa já conhecidas e outros possíveis candidatos para compor a classe. Ademais, produzimos o vetor pET-pedABCD por síntese química, para expressão heteróloga de bacteriocinas utilizando toda maquinaria de biossíntese necessária para produção do peptídeo ativo no meio de cultura extracelular / fração celular solúvel. Este é o primeiro plasmídeo modular, teoricamente capaz de expressar diversas bacteriocinas da classe IIa por Golden Gate Assembly. pET-pedABCD foi transformado em *E. coli* BL21(DE3) Arctic Express sob seleção de Canamicina e a presença do cassete de expressão foi confirmada em dois clones por PCR com primers T7. Análise do SDS-PAGE correspondente à fração insolúvel, preliminarmente demonstrou a expressão da principal proteína para processamento e transporte do peptídeo precursor (PedD). Dada a importância da resistência antimicrobiana na esfera da saúde única, desenvolveu-se um método de mineração de bacteriocinas da classe IIa em metagenomas e criação de um sistema de baixo custo para produção modular de bacteriocinas, somando-se ao arsenal de ferramentas no combate a bactérias resistentes a antimicrobianos.

Palavras-chave: Mineração de Metagenomas. Bacteriocinas classe IIa. Pediocin-like. Padrão de Aminoácidos.

ABSTRACT

BARRIOS, Wesley Elias Bhering, M.Sc., Universidade Federal de Viçosa, September, 2020. **Metagenome mining of pediocin-like bacteriocins and creation of pET-pedABCD, a customizable pediocin expression vector**. Adviser: Tiago Antônio de Oliveira Mendes. Co-Adviser: Hilário Cuquetto Mantovani.

Fermented dairy products, the human intestine, and the bovine rumen are some environments that contain lactic acid bacteria (LAB). In these competitive environments, the synthesis of antimicrobial compounds such as bacteriocins acts to control the population of gram-positive bacteria through various mechanisms, favoring the producer microorganism. The bacteriocins are classified into 3 classes, being class II, subclass A, bacteriocins analogous to pediocin, peptides with ample activity against Gram-positive and negative bacteria. Pediocin PA-1, the precursor of this subclass, acts through the insertion in the plasma membrane of the target bacterium by binding to the mannose carrier widely conserved in gram-positive bacteria or by electrostatic interactions, forming a pore that leads to cell death. The high increase of bacteria with resistance to classical antibiotics highlights the importance of the discovery of new antimicrobials. An alternative to this problem may be in metagenomes, which correspond to the DNA sequencing of microbial communities, providing access to the genetic potential of non-cultivable strains, enabling the discovery of new bacteriocins. In this context, the main objective of this work was the mining of metagenomes of fermented dairy products using the Memi algorithm. To this end, a new standard of amino acids characteristic of class IIa was developed to increase the accuracy of results and decrease false positives. Initially, new characteristic class IIa amino acid standards were proposed from the alignment of deposited sequences. It was possible to define memi-9 as a new standard of amino acids characteristic of class IIa bacteriocins, presenting 92.2% sensitivity and 100% accuracy. The mining of metagenomes resulted in the mapping of 52 already known class IIa bacteriocins and other possible candidates to compose the class. Also, we produced the pET-pedABCD vector by chemical synthesis, for heterologous expression of bacteriocins using all biosynthesis machinery necessary to produce the active peptide in the extracellular culture medium / soluble cell fraction. This is the first modular plasmid theoretically capable of expressing several class IIa bacteriocins by Golden Gate Assembly. pET-pedABCD was transformed into E. coli BL21(DE3) Arctic Express under Kanamicin selection and the presence of the expression cassette was confirmed in two clones by PCR with T7 primers. Analysis of SDS-PAGE corresponding to the insoluble fraction preliminarily demonstrated the expression of the main protein for processing and transport of the precursor peptide (PedD). Given the importance of antimicrobial resistance in the one health sphere, a class IIa bacteriocin mining method was developed in metagenomes and a low-cost system for the modular production of bacteriocins was created, adding to the arsenal of tools to combat antimicrobial-resistant bacteria.

Keywords: Metagenome Mining. Class IIa bacteriocins. Pediocin-like. Amino acid pattern.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

SUMMARY

1. INTRODUCTION

The mammalian gut microbiome is a diverse environment modulated by several factors; and the host well-being is directly linked to the composition of its gut microbial communities (JAMI et al., 2013). A classic example of these relationships is found in the first compartment of the digestive tract of ruminants - the rumen, an environment that concentrates several species of bacteria, that play a relevant role in the regulation of food digestion, intake and the uptake of energy and nutrition (LOZUPONE et al., 2012; UMU et al., 2013). Another example in the human gut, is the beneficial ingestion of fermented dairy food harboring Lactic Acid Bacteria (LAB) (PASOLLI et al., 2020).

The microorganisms from this community compete with each other for substrates and ecological niches for colonization (O'TOOLE; COONEY, 2008). To explain the coexistence of competing species, the syntrophy of the microbial community must be understood as a multifactorial phenomenon that is still being characterized (LIBBY et al., 2019). The colonization and survival in the bacterial competitive environment are favored by the synthesis of antimicrobial compounds named bacteriocins, which are antimicrobial peptides that act in the population control of homologous species. Bacteriocin production result in the adaptive advantage that enables improved access to limited resources and advantageous features for the producing microorganism (MAJEED et al., 2011; UMU et al., 2013).

Originally classified as antibiotics, bacteriocins are Ribosomally synthesized and Post-translationally modified Peptides (RiPPs) with antimicrobial activity (AHMAD et al., 2017). However, the classification as antibiotics was abandoned due to distinct biosynthetic pathways of these compounds (MANTOVANI; RUSSELL, 2008). The first description was reported 80 years ago when antagonism between strains of *Escherichia coli* was discovered (GRATIA, 1925).

Secreted by various Gram-positive and Gram-negative bacteria, and some members of Archaea, pediocin-like bacteriocins are one of these antimicrobial peptides. Pediocin-like peptides are synthesized as biologically inactive precursor due to the presence of the leader peptide in its N-terminal region. Subsequently, the precursor peptide is transported and cleaved in a GA/GG motif at the end of the leader sequence to generate the mature form of bacteriocin. The active peptide is then secreted to the cell exterior via a dedicated bacteriocin transport system (ABC transporter) or a general secretion system (e.g. Sec pathway) (KUMARIYA et al., 2019; PORTO et al., 2017).

At the molecular level, at least four genes organized as an *operon* (usually organized in one or two operons) are required for expression of the active bacteriocin: (1) *pedA*, encondes the pre-peptide; (2) *pedB* encodes the immunity protein that confers immunity to the host; (3) *pedD* encodes the ABC transporter, the bacteriocin carrier anchored to the membrane, and (4) *pedC*, that encodes an accessory protein responsible for disulfide bond formation; but their specific role is not fully elucidated (DIEP et al., 2007; PORTO et al., 2017).

The pre-peptide leader sequence may play a dual role; in the cytosol, the precursor sequence keeps the bacteriocin inactive, preventing self-injury, and inside the cell, as a cleavage motif during the exportation of the mature bacteriocin (DRIDER et al., 2006).

Versatile compounds, pediocin-like peptides act across various mechanisms that can be broadly divided into cell envelope recognition, pore formation/membrane biosynthesis inhibition, modulation of gene expression and protein production into the cell (COTTER; ROSS; HILL, 2013). According to its primary structure, functional similarity, thermal stability, and molecular mass (COTTER; HILL; ROSS, 2005a), at least three main groups (Table 1) have been proposed to classify these antimicrobial agents (AZEVEDO et al., 2015).

The class I are composed of post-translationally modified antimicrobial peptides, such as lanthipeptides, sactipeptides, and lasso peptides, which differ in molecular structures, mechanisms of action, and the enzymatic apparatuses involved in modification and maturation/export of the precursor peptides (AZEVEDO et al., 2015). Class II bacteriocins consist of small peptides that undergo little or no post-translationally modification of residues and can be sub-divided into class IIa (pediocin-like bacteriocins), class IIb (two-component bacteriocins), class IIc (circular bacteriocins) and class IId (linear, non-pediocin-like bacteriocins) (KUMARIYA et al., 2019). Class III consists of thermolabile proteins also called bacteriolysins, with a molecular mass greater than 10 kDa and homologous to endopeptidases (COTTER; HILL; ROSS, 2005b; NILSEN; NES; HOLO, 2003).

Table 1 Classification of Bacteriocins. * Some Examples. Adapted from KUMARIYA et al., 2019.

Class		**Features**	**Example**	**Mechanism of action**	**Receptors**	**Producers***	**Bacterial spectrum**
I	Ia	Lantibiotics (<5 kDa peptides containing lanthionine and β-methyl lanthionine)	Nisin	Membrane permeabilization by pore formation	Lipid II	*S. aureus, L. lactis*	Gram-positive bacteria
	Ib	Carbacyclic lantibiotics containing labyrinthin and labionin	Labyrinthopeptin A1	Not known	Not known	*Actinomadura namibiensis*	HIV, HSV
	Ic	Sactibiotics (sulphur to alpha carbon-containing antibiotics)	Thuricin CD	Not known	Not known	*B. thuringiensis*	Gram-positive bacteria
II	IIa	Small heat-stable peptides, synthesized in a form of precursor which is processed after two glycine residues, active against *Listeria*, have a consensus sequence of YGNGV-C in the N-terminal	Pediocin PA-1, sakacins A and P, leucocin A.	Membrane permeabilization by pore formation	Mannose permease	*P. pentosaceus, P. acidilactici, Lactobacillus sakei*	Gram-positive and Gram-negative bacteria
	IIb	Two component systems: two different peptides required to form an active pore complex	Lactococcins G, plantaricin EF and plantaricin JK	Membrane permeabilization by pore formation	UppP (undecaprenyl pyrophosphate phosphatise)	*L. lactis* subsp. *cremoris*, *Lb. plantarum*	Gram-positive bacteria
	IIc	Circular bacteriocins	Gassericin A, enterocin AS-48, garvicin ML	Membrane permeabilization by pore formation	ABC transporter	*L. gasseri*, *E. faecalis, L. garvieae*	Gram-positive bacteria
	IId	Unmodified, linear, leaderless, non pediocin-like bacteriocins	Bactofencin A, LsbB	Membrane permeabilization by pore formation	Metallopeptidase	*L. salivarius, L. lactis* subsp. *Lactis*	Gram-positive bacteria
III		Large molecules sensitive to heat	Helveticin M, helveticin J and enterolysin A	Membrane permeabilization by pore formation	Not known	*Lb. crispatus, L. helveticus, E. faecalis*	Gram-positive and Gram-negative bacteria

Pediocin-like bacteriocins constitute the most dominant and perhaps well-studied group of class IIa bacteriocins. Produced by some lactic acid bacteria (LAB), they are cationic peptides containing 25 to 50 amino acids with a less conserved C-terminal region and two well-conserved structural regions: the highly conserved YGNGV/L "pediocin box" motif in the amino-terminal portion, and well known spaced cysteine residues that form a disulfide bridge (OPPEGÅRD et al., 2015). Located after a hinge region, the C-terminal domain has been used as a basis for subgrouping these bacteriocins into IIa, IIb, and IIc (COTTER; HILL; ROSS, 2005b; JOHNSEN; FIMLAND; NISSEN-MEYER, 2005). Interestingly, the C-terminal region was confirmed as responsible for the target specificity of class IIa bacteriocins (JOHNSEN; FIMLAND; NISSEN-MEYER, 2005).

Since the initial identification of these peptides in the early nineties, more than 20 different pediocin-like bacteriocins have now been characterized. Despite their high structural similarity, pediocins can markedly differ in their antimicrobial spectrum of activity (DRIDER et al., 2006), mainly because of the broad spectrum of antimicrobial activity against Gram-positive bacteria justifying the considerable industrial interest in this class of bacteriocins (PORTO et al., 2017). For instance, listeriosis outbreaks are recurrent and the increasing discovery of strains of *Listeria monocytogenes* resistant to bacteriocins has focused the attention of bacteriocin researchers in the pursuit of new food preservation methods. This resulted in the isolation of a large number of class IIa bacteriocins, all of which are highly active against the food pathogen *L. monocytogenes* (INGHAM et al., 2003; RILEY; WERTZ, 2002).

The pediocin-like bacteriocins function as a membrane interacting ligand upon specifically binding to a target cell membrane protein (i.e., Man-PTS). In general, they form an N-terminal three-stranded antiparallel beta-sheet and hairpin spaced amphiphilic helical C-terminal structure, which allows them to be inserted into the membrane of the target cell, leading to depolarization by dissipating the proton motive force (PMF) and consequently, an imbalance of intracellular content which induce cell death (FIMLAND et al., 2005; RÍOS COLOMBO et al., 2018).

Chemically synthesized pediocin PA-11 showed antimicrobial activity against many pathogens, such as *Listeria monocytogenes* (MIC = 6.8 nM), *Carnobacterium divergens* ATCC 35677 (MIC = 1.9 nM), *Leuconostoc mesenteroides* ATCC 23386 (MIC = 1.9 nM), *Listeria seeligeri* ATCC 35967 MIC = 4.7 nM), *Clostridium perfringens* AAC 1–222 (MIC = 37.8 nM), *Clostridium perfringens* AAC 1–223 (MIC = 75.7 nM), *Listeria murrayi* ATCC 25401 (MIC =

151.4 nM), and *Lactobacillus plantarum* ATCC 8014 (MIC = 605.5 nM) (BÉDARD et al., 2018).

Also, antitumoral activity was demonstrated for pediocin PA-1, inhibiting human lung carcinoma cells, colorectal adenocarcinoma, human colon adenocarcinoma HT29, and HeLa cells (KUMAR et al., 2011; VILLARANTE et al., 2011).

In the producer host, pediocin expression is regulated by a pediocin operon independent three-component system, which includes the Inducer Factor (IF: pheromone or inducer factor), the membrane histidine protein kinase MHK (pheromone receptor), and the transcription factor called response regulator (RR) (Figure 1). At constant basal levels, IF is actively exported by the ABC transporter system. Once its concentration rises, IF interacts with MHK which auto-phosphorylate a conserved histidine residue located in the cytoplasmic side. Then, the response regulator is phosphorylated and becomes active, which triggers the expression of the pediocin-like peptide operon (biosynthesis, IF, RR and MHK).

In this route, IF act a positive-feedback molecule which promotes pediocin expression autoregulation. There is some correlation between IF, bacteriocin production, and cell growth that needs to be clarified as other environmental signals that might affect class II bacteriocin regulation (PORTO et al., 2017).

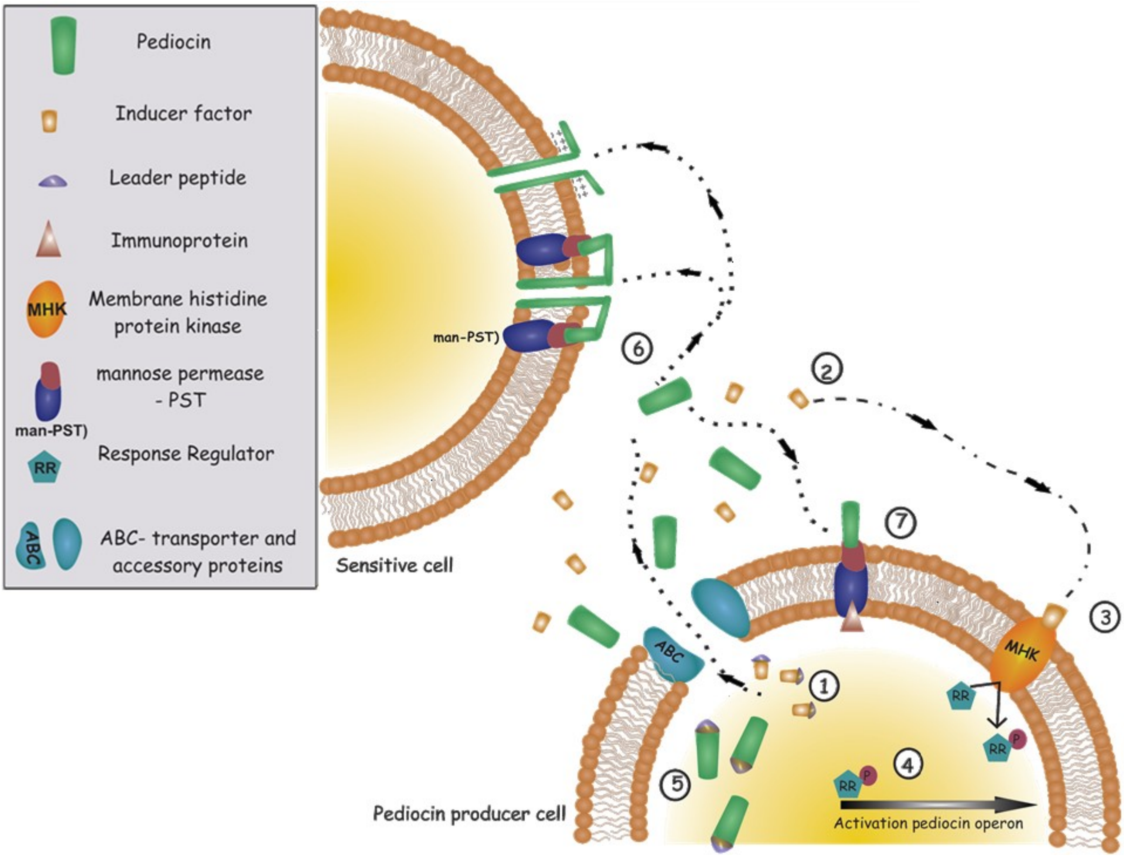


Figure 1 - Modulation of pediocin production, signaling, and mode of action (PORTO et al., 2017). IF is constantly produced (1) and when IF concentration rises outside the cell (2), it binds to MKH (3), promoting its autophosphorylation. RR is phosphorylated, inducing the expression of the pediocin-operon (4). The pediocin structural gene is expressed with the leader peptide, which is recognized and cleaved by accessory proteins in the ABC transporter and exported (5). Mature pediocin then binds to sensitive cell membranes through man-PST or electrostatic interactions (6), forming pores that dissipate partial or total ΔΨ and ΔpH (7).

Inside the class IIa bacteriocins, minor structural variation alters target specificity, and slight changes in membrane composition/transporters in target cells may lead to considered altered susceptibility to a pediocin. Thus, the elucidation of what manages the specificity of these peptides at a molecular level and the susceptibility of target cells is of great importance for the optimal and rational exploitation of bacteriocins as antimicrobial agents (FIMLAND et al., 2005; NISSEN-MEYER et al., 2009).

The identification of new antimicrobial peptide gene clusters has been based on the use of classical microbiology methods, starting in the screening of large collections of strains, based on their ability to produce novel antimicrobials. The bacteria selection is followed by the subsequent identification of the responsible genes through subcloning, mutagenesis, reverse genetics, or, more recently, sequencing of the corresponding genome. Despite constant improvements in culturing techniques, it is still estimated that 99% of microbial life does not have the optimal cultivation conditions known at the time (SULLIVAN et al., 2013).

Metagenomes are the whole DNA sequencing of microbial communities from the most diverse environments. They have great sequence richness, offering access to genomes of ‘unculturable’ bacteria, circumventing the need to culture them in the laboratory. In this case, metagenomic DNA sequencing provides access to the genetic potential of ‘unculturable’ unknown strains, which can help in understanding their physiology and role in the ecosystem, contributing to the creation of new metabolic pathways for the production of metabolites of industrial and medical interest (SHARMA et al., 2005).

Thus, besides the promising results achieved by wet-lab based screenings, the large amount of genetic information enclosed in metagenomic DNA databases suggests that *in silico* approaches for bacteriocin gene clusters could be more successful in the identification of novel bacteriocin gene clusters in microbial communities (ZOU et al., 2018).

In this panorama, we propose the creation of characteristic amino acid patterns for pediocin-like bacteriocins, mining of bovine ruminal and dairy food metagenomes using the new Memi software, and the design/validation of a new modular pET-28 a(+) based protein expression vector, aiming the expression of custom pediocin-like bacteriocins in *Escherichia coli* BL21(DE3) Arctic Express culture medium.

2. IMPORTANCE

Antibiotic therapy has revolutionized the human health system, significantly increasing life expectancy and decreasing mortality from bacterial infections. However, several problems are associated with the exacerbated use of antibiotics, and their harms affect the whole sphere of one-health, generating a great alert to support the continued use of these antimicrobials (COTTER; ROSS; HILL, 2013).

Although antimicrobial resistance is not a new feature, the scale of multidrug-resistant microorganisms found globally is extraordinarily high (LEVY; BONNIE, 2004). This problem has increasing adversities in medical treatments worldwide. According to a recent report of the World Health Organization published in 2019 (FUTURE; DRUG-RESISTANT, 2019), if neglected, the multi-resistant bacteria may cause 10 million deaths per year in 2050, due to infections unresponsive to antibiotic treatment.

There are numerous molecular mechanisms that bacteria use to overlap susceptibility to antibiotics, resulting in decreased activity or concentration in the cell (Figure 2) . But in addition to these mechanisms, other problems such as mutation rate, frequency of horizontal gene transfer, indiscriminate use of antibiotics as animal growth promoters and in agriculture; makes antimicrobial resistance a one-health problem, affecting all ecosystems in the environment (HUGHES, 2003). Thus, there is an urgent need to identify and develop alternative therapeutic strategies for clinical applications in human and animal health.

The figure below shows some of the molecular mechanisms of acquired antimicrobial resistance.

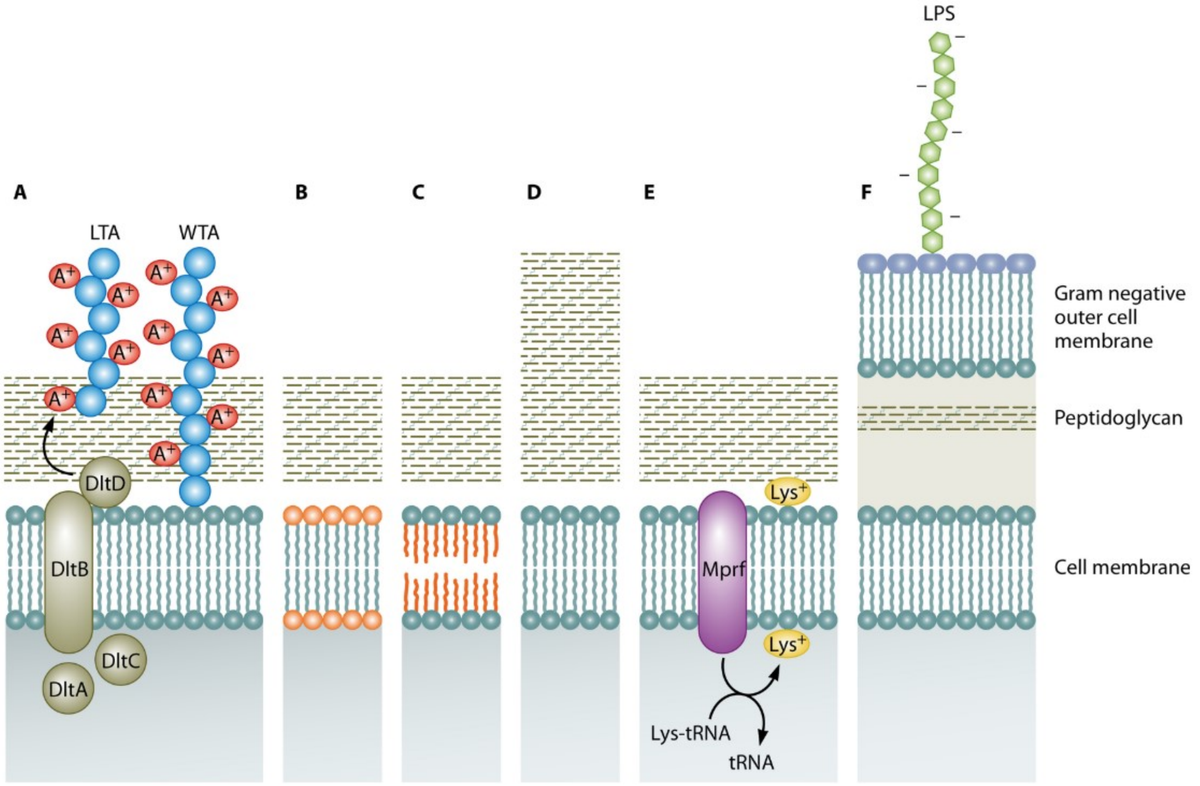


Figure 2 – Acquired antimicrobial resistance. Several factors confer antimicrobial resistance, some of the cell wall and membrane-associated factors are (A) D-Alanylation of lipoteichoic acids (LTA) and wall teichoic acids (WTA) by the *dltABCD* operon, which confers a positive charge; (B) Alterations in phospholipid composition; (C) Changes in membrane fatty acid composition; (D) cell wall thickening; (E) Lysine esterification of one of the two hydroxyl groups of phosphatidylglycerol (PG) by Mprf; and (F) Gram-negative outer cell membrane containing lipopolysaccharide (LPS) - from DRAPER et al., 2015.

Bacteriocins have emerged as one of the promising molecules to tackle antimicrobial resistance (BÉDARD et al., 2018). Produced by a wide variety of bacteria to combat other microorganisms in their competitive environments, these peptides have been regarded as a very promising source of antimicrobials. Considered “designer drugs,” they have many properties that include their potency (10^3 to 10^6 times more than several other antimicrobials including conventional antibiotics), their variation in the spectrum of action, the possibility of being bioengineered and, their low toxicity (BÉDARD; BIRON, 2018; COTTER; ROSS; HILL, 2013). An interesting fact is the consumption of lactic acid bacteria dating back to millennia ago as fermented foods (COTTER; ROSS; HILL, 2013).

They can be used both in their purified form and in the form of probiotics due to the GRAS (generally regarded as safe) classification of most of their producing microorganisms (PORTO et al., 2017). Chemical synthesis has been proposed for the large-scale production of active bacteriocins, nevertheless, the use of purified bacteriocins is not practical since very few bacteriocins have been successfully prepared in satisfactory yields using such means, not being economically viable (BÉDARD et al., 2018; DESRIAC et al., 2010).

To date, only nisin (e.g., Nisaplin®) and pediocin PA-1 (e.g., ALTA®-2431) have been commercialized as food preservatives, pointing the importance of the discovery of novel broad-spectrum and safe bacteriocins to add to the arsenal of available antimicrobials for food preservation, and formulation of cocktails with higher antimicrobial efficiency.

In this way, the importance of the search for new pediocin-like bacteriocins is justified by its potential contribution in the fight against antimicrobial drug resistance. It presents antimicrobial action in small concentrations, and the existence of a large data set of metagenomes possibly rich in this class of antimicrobials, could markedly alter the current scenario against antimicrobial resistance in food related pathogens.

3. GOALS

3.1. AIM

Prospection of pediocin-like bacteriocins in metagenomes from dairy food and the bovine rumen and the design and validation of pET-pedABCD – a modular pediocin-like bacteriocin expression vector.

3.2. OBJECTIVES

To employ the Memi algorithm to discover new pediocin-like bacteriocins in metagenomes from dairy food and rumen of cattle, using canonical amino acid standards;

To develop new amino acid standards to pediocin-like bacteriocins;

To design and produce a new plasmid for pediocin-like bacteriocin heterologous expression in *E. coli* BL21(DE3) Artic Express;

To validate the expression of proteins encoded in the plasmid , SDS-PAGE analysis was employed.

4. MATERIALS AND METHODS

4.1. The algorithm

Data mining was performed using the Memi software developed by de Souza, I.A.F (Master of Science dissertation in preparation). The algorithm is conceptually simple: its proposed problem is to receive a query pattern and short reads from metagenomic sequencing raw data, in either FASTA or FASTQ format. The query pattern is modeled as non-deterministic finite automata (NDFA) represented by the 5-tuple (Q, ∑, δ, q0, F) where Q is a finite set of states, ∑ is the alphabet, composed by a set of finite symbols representing the 20 amino acids and the stop codon, δ it the transition function, q0 is the initial state from where the read is processed, and F is a set of final states, in this case, the only accepted final state is the stop codon. Graphically the NDFA is represented by a directed graph (Figure 3).

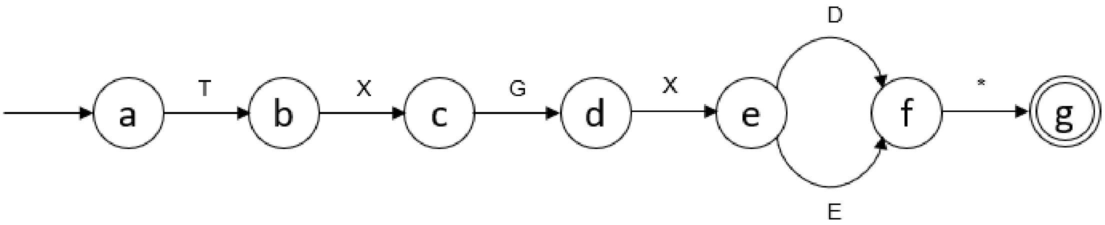


Figure 3 - Algorithm functionality. In the representation, the vertices are the states, the edges represent a transition consuming symbol from the alphabet (a to g), whereas the initial state is denoted by an empty incoming edge and the final state is indicated by a double circle. The representation is the NDFA for a query pattern T-X-G-X-[DE]-*, where the symbol X

indicates the transition happening using any symbol from the alphabet, the amino acids inside the square brackets ([DE]) indicate that there are two possible transitions – for Glu or Asp – at that position, and the * symbol is the stop codon, the only acceptable transition from a final state. If the final state (stop codon) is not achieved, the read is discarded.

The algorithm then, read each short sequence (termed word) from the raw metagenomic data and check if this word belongs to the NDFA vocabulary.

The algorithm's asymptotic complexity is O(mn), where m is a constant and n is the number of reads in the metagenomic data. The m constant value is the product of the query pattern length (QPL) divided by 3 (i.e. the number of amino acid residues) and the length of the read, also divided by 3.

The algorithm receives a pattern of amino acids in the PROSITE format. In the end, the sequences found in amino acid residues are generated. It is customizable concerning the choice of how many mismatches are allowed, and flexible about the standard, since unconserved positions can be considered as either amino acid.

The raw output files were cleaned, deduplicated, and amino acids after stop codons removed along with the last line (referring to the script's execution time). After this step, all the program's output files were cured manually in the search for new pediocin-like bacteriocins. For this, the strings were submitted to alignment with the blastp algorithm (ALTSCHUL et al., 1990), and positive matches were identified as possible existing bacteriocins present in the analyzed dataset.

4.2. Metagenomes

All metagenomes from dairy food (PASOLLI et al., 2020) and a small data-set of bovine rumen metagenomes (STEWART et al., 2019) were selected to search for new pediocin-like bacteriocins.

4.3. Selection of pediocin-like bacteriocins

The amino acid sequences of the pediocin-like bacteriocins were retrieved from the databases of online platforms Bagel 4 (VAN HEEL et al., 2018), BACTIBASE (ZOUHIR et al., 2010), CAMP$_{R3}$ (WAGHU et al., 2016), LAMP2 (ZHAO et al., 2013), MiBig (MEDEMA et al., 2015) InterPro (MITCHELL et al., 2019) - under motifs: IPR002633, IPR023384, IPR023388- and UniProt (BATEMAN, 2019) - domain PS60030. Then, due to the greater variability of the leader peptide, only the core peptide part of the sequence was maintained (removal of the leader peptide - all sequence before the GG / GA cleavage motif). The file containing all the core peptides of the selected pediocin-like bacteriocins was subjected to global multiple alignments by the MUSCLE algorithm (Edgar, 2004) using the MEGA X software (Kumar, Stecher, Li, Knyaz, & Tamura, 2018). Subsequently, the alignment was analyzed, and the most divergent bacteriocins or with variations in the pediocin-box were eliminated to carry out the position-to-position amino acid conservation study.

4.4. Investigation of the profile of conserved amino acids

The conserved amino acid profiles were generated in the PROSITE format (SIGRIST et al., 2002) using the PRATT tool (JONASSEN; COLLINS; HIGGINS, 1995), version 2.1 of the ExPaSy suite (GASTEIGER et al., 2003). The generated PROSITE standard was then submitted to the ScanProsite tool (DE CASTRO et al., 2006) on the ExPaSy website, where the standards were submitted against the PROSITE motif database. This analysis allowed for the selection of the standard with a wider range of pediocin-like bacteriocins, lower false positives, and e-values.

4.5. Validation of amino acid standards for pediocin-like bacteriocins

To validate the generated amino acid patterns for pediocin-like bacteriocins, the UniprotKB database of class IIa peptides (https://www.uniprot.org/uniprot/?query=family:%22Bacteriocin+class+IIA/YGNGV+family%22&sort=score) was analyzed with the generated amino acid patterns using option 3 of the ScanProsite tool (https://prosite.expasy.org/scanprosite/). Three patterns from literature and three new patterns was tested against all UniprotKB and IPR023384 (https://www.ebi.ac.uk/interpro/entry/InterPro/IPR023384/) peptides with parameters “greedy, no overlaps” and allowing 1 X character in a scanned sequence. The best standards were designed by calculating sensibility and precision (BHADRA et al., 2018). The pattern with the best precision and better coverage was selected to carry metagenome mining analysis.

4.6. Metagenome mining

To mine the metagenomes of dairy food and bovine rumen, all reads were downloaded to Jupiter, the computer cluster of UFV (Universidade Federal de Viçosa). All reads were trimmed to the Phred quality value of 20 using the Trimmomatic 0.36 algorithm (BOLGER; LOHSE; USADEL, 2014) set with the following options: “HEADCROP:7, ILLUMINACLIP: ../adapters.fasta:2:30:10, LEADING:3, TRAILING:3, SLIDINGWINDOW:4:20, MINLEN:36”. Then, to optimize the resources of the cluster, all reads were split into files containing 200000 lines. Moreover, the files were submitted to the Memi algorithm loaded with the literature patterns and discovered amino acid patterns.

The algorithm output was cleaned to remove amino acids after stop codons, repeated sequences, and the final line, referent to the algorithm time of execution. Each output file holds the origin file name as a prefix to ensure the traceability of results. Preliminarily, all cleaned output files of each data set were joined in one fasta file, where * (indicating stop codon) was removed, and a fasta header was designated for each peptide. This file was then edited for removal of duplicated sequences and further submitted to CAMP Sign (http://www.campsign.bicnirrh.res.in/) (WAGHU; BARAI; IDICULA-THOMAS, 2016), an online tool for the identification of AMPs using AMP family signatures, whether being amino acid patterns, or HHM profiles for the bacteriocin family.

Besides, all sequences were submitted to BLASTP, non-redundant protein sequences database, to search for homologous sequences. Duplicated hits were removed and unique ones recorded.

4.7. Design of the pediocin-like bacteriocin expression cassette

The expression cassette was assembled carrying out the genetic engineering of the precursor gene of Pediocin PA-1 (*pedA*), produced by *Pediococcus acidilactici* (RODRÍGUEZ; MARTÍNEZ; KOK, 2002) and all the protein machinery necessary for its coding (*pedA*) – UniProtKB accession: P29430, immunity (*pedB*) - UniProtKB accession: P36496, maturation (*pedC*) - UniProtKB accession: P37249, cleavage and transport (*pedD*) - UniProtKB accession: P36497.

With the use of the Benchling platform (https://www.benchling.com/), the complete expression cassette was submitted to codon optimization for *Escherichia coli* (O157:H7

EDL933) with medium (0.33 to 0.66) GC content and avoiding hairpins options marked, also, NcoI, XhoI, BsaI were avoided.

The precursor gene has two regions (leader and core). The leader sequence of the pediocin PA-1 was fixed and two restriction sites for the BsaI enzyme were inserted upstream of the leader, for insertion of customized core peptides through Golden Gate Assembly technique (ENGLER; KANDZIA; MARILLONNET, 2008). BsaI sites flank an internal region, which can be amplified by PCR with M13 primers, and will serve as an indication of the non-insertion of the desired core sequence (band of 120 bp indicates the non-insertion).

After the complete *in silico* bioengineering of *pedA, pedB, pedC,* and *pedD* genes, the sequence was joined in a pediocin-like bacteriocin expression cassette, controlled by the T7 promoter, and finally submitted to chemical synthesis by the company FastBio Ltda and inserted in the pET-28 a(+) plasmid. The plasmid containing all the above-mentioned features was named pET-pedABCD.

The core sequences derived from the mining of pediocin-like bacteriocins from dairy food and bovine rumen metagenomes were flanked by the same BsaI sites and were chemically synthesized. Thus, it was expected that the protein machinery for functional expression of Pediocin PA-1 could recognize the leader sequence and were able to process, mature, and export the new pediocin-like bacteriocins inserted in the engineered precursor gene, here called *pro-pedA*. Precursor *pro-pedA* was engineered to be inducible due to the possibility of toxicity in the cell and its expression was under the control of the T7 lac operon (T7 promoter, lac operator, and RBS), whereas the *pedB, pedC,* and *pedD* genes display constitutive expression conferred by the T7 promoter. None of the genes had histidine tags due to the possibility of altering the function/recognition of bacteriocins.

4.8. Transformation and Selection

Competent *E. coli* BL21(DE3) cells were transformed with 200 µg of pET-pedABCD plasmid DNA and plated in 1,5% agar-LB plates with 50 µg/mL Kanamycin (selection marker) according to manufacturer manual (AGILENT TECHNOLOGIES, 2015). Transforming colonies were selected and transferred to the LB liquid medium containing 50 µg/mL Kan. Glycerol stocks of cells were made and stored at -80 °C. Moreover, each transformant was grown in 5 mL LB plus 50 µg/mL of Kan to perform the extraction of plasmid DNA.

The plasmid DNA of transformants was extracted with StrataPrep Plasmid Miniprep Kit from Agilent Technologies, catalog #400761 according to the manufacturer's manual.

The purified plasmid DNAs, plus synthetized pET-pedABCD (positive control), were loaded in a 1% agarose gel to verify plasmid integrity. Moreover, the concentration of plasmid DNA was measured by spectrophotometry and fluorometry (Qubit 3.0), $A^{260/280,}$ and $A^{260/230}$.

4.9. Expression Cassette Confirmation

To verify the presence of the pediocin expression cassette, a PCR using GoTaq® Flexi DNA Polymerase from Promega Corporation, catalog: M8295 with T7 primers (

Table 2) was carried, using the purified plasmid DNA as template. PCR cycling conditions are presented below (Table 3).

Table 2 - Sequences of T7 primers used in this study.

Primer	Sequence (5' - 3')
T7_F_LBM	TAATACGACTCACTATAGGG
T7_R_LBM	GCTAGTTATTGCTCAGCGG

Table 3 – PCR cycling conditions of the expression cassette with T7 primers. The green cells represent the 35 amplification cycles.

Step	Temperature (°C)	Time
Initial Denaturation	94	10'
Denaturation	94	45"
Annealing	53	40"
Extension	72	3'36"
Final Extension	72	5'
Hold	4	∞

4.10. Protein Expression in *Escherichia coli* BL21(DE3) Arctic Express

Protein expression was performed using the strain *Escherichia coli* BL21(DE3) Artic Express of Agilent Technologies – catalog #230192. This strain was used due to the ability of toxic protein expression and by the presence of mesophilic chaperonins that assist protein folding at low temperatures, a condition that promotes a greater amount of soluble proteins expressed. The entire transformation and protein expression protocol was performed according to the manufacturer's manual plus adjustments (AGILENT TECHNOLOGIES, 2015). Induction of proteins was done in 50 mL falcon tubes containing 20 ml of transformant cell culture in LB broth, 50 µg/mL of Kan, and 0.8 mM of IPTG.

4.11. Identification of pET-pedABCD expressed proteins

At the time post-induction (TPI) of 0 and 18 hours, transformed cell cultures and negative control were centrifugated at 12.000 RPM for 10 minutes. Supernatant and cell pellets were placed in individual 50 mL Falcon tubes, where the cell pellet fractions were resuspended in 5 mL of 1X PBS and the supernatant fractions stored at -20 °C. Furthermore, the resuspended cell pellets were sonicated in a SONIC, USA 130W, 20 kHz sonicator. All samples were placed on the ice during the procedure, and settings were defined at 50% of amplitude, 5-minute run with cycles of 30s of pulse, and 30s without pulse.

Thereafter, samples were centrifuged at 12000 RPM for 10 min at 4 ºC; the supernatant corresponding to soluble cellular fraction was collected and aliquoted, and the insoluble fraction was resuspended in 4 mL of filtered cold 1X PBS and aliquoted, all samples were kept on ice during all steps.

To confirm the expression of proteins encoded by the pET-pedABCD vector, 30 µL of each sample was mixed with 10 µL of 4X SDS unstained loading buffer and incubated at 95 ºC for 10 minutes. Then, 30 µL of each sample was loaded on 12,5% SDS-PAGE and submitted to the source settings: 110V (fixed), 400 mA, and 30W for 2 hours and 15 minutes. SDS-PAGE's of the soluble cellular fraction and insoluble fraction was stained with Coomassie Brilliant Blue R-250 dye for 30 minutes, and then, unstained with Unstaining Buffer (25% methanol, 7,5% acetic acid) until bands were clear and well defined.

Mature pediocins are expected to be released directly into the culture medium or soluble cellular fraction, which promotes and simplifies downstream applications. PedB and PedC proteins are expected to be found in the soluble fraction, and PedD in the insoluble fraction, due to is membrane localization.

5. RESULTS AND DISCUSSION

Through time, many approaches were developed to address the genome mining of RiPPs (Figure 4Figure 4 - Timeline of the release of RiPP genome mining tools and methods of search. Several tools were developed to search for new RiPP's in genomic data, however, metagenomics has been loading large amounts of data-rich in unprecedented RiPP genetic information. Thus, new tools must be developed to optimize the metagenomic mining of such compounds. From ZHONG et al., 2020.). There are four genome mining methods employed by the available tools at the moment: search for PTM (post-translational modification) enzymes and precursors; comparison of BGCs (biosynthesis gene clusters) and clustering by similarity network, search for conserved RiPP tailoring enzymes, or the search for precursor peptides using machine learning. However, these methods are not applied to metagenomes due to the fragmentation present in this type of data, not being able to find PTM enzymes, BCG's and RiPP tailoring enzymes efficiently.

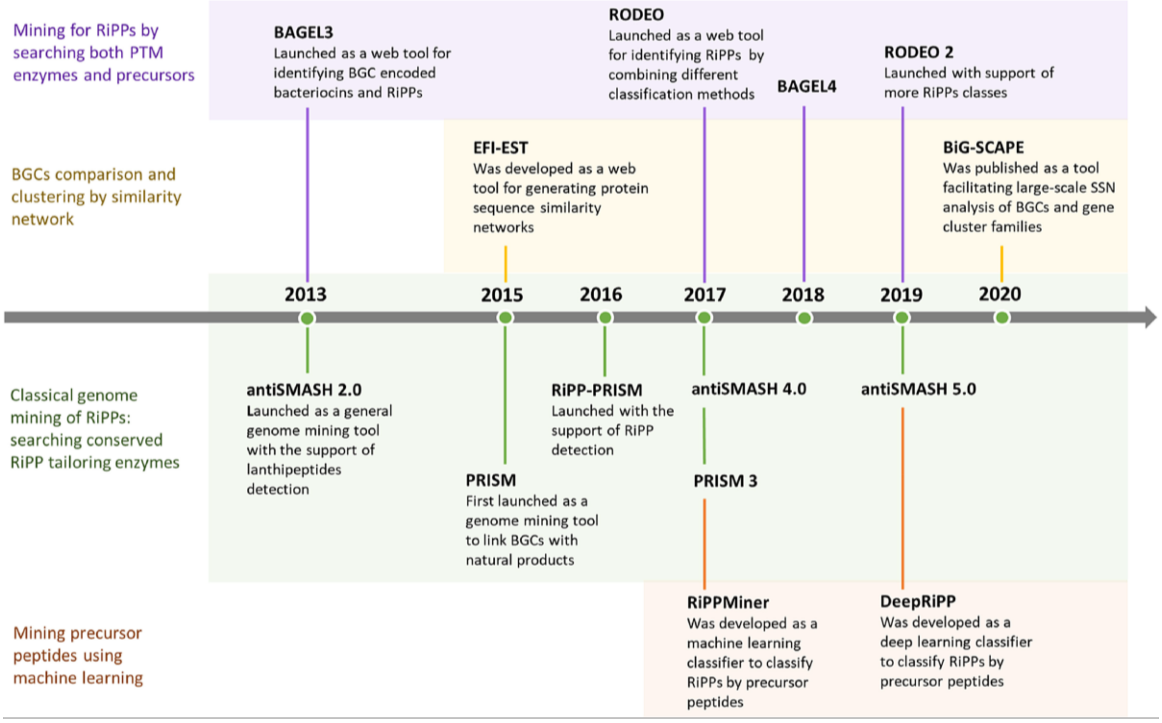


Figure 4 - Timeline of the release of RiPP genome mining tools and methods of search. Several tools were developed to search for new RiPP's in genomic data, however, metagenomics has been loading large amounts of data-rich in unprecedented RiPP genetic information. Thus, new tools must be developed to optimize the metagenomic mining of such compounds. From ZHONG et al., 2020.

To tackle this obstacle, the search of the Memi algorithm is driven by a pattern of amino acids in the PROSITE format from a given class of proteins; in this case, class IIa bacteriocins. this class is suitable because the class IIa bacteriocins are 35 to 58 amino acids long, and 10 to 20 fewer amino acids removing the leader peptide (variable); then, the algorithm can access all the information of class IIa bacteriocins present in metagenomes, which are composed of reads of 150 base pairs.

Also, Memi is not dependent on the gene context, and very sensible if loaded with well-designed amino acid patterns (Table 4). The relation between false positives and true positives can be modulated by adjusting the number of wild cards in the algorithm, and through this same approach, novel sequences can be discovered by relaxing the number of wild cards allowed in the search parameters.

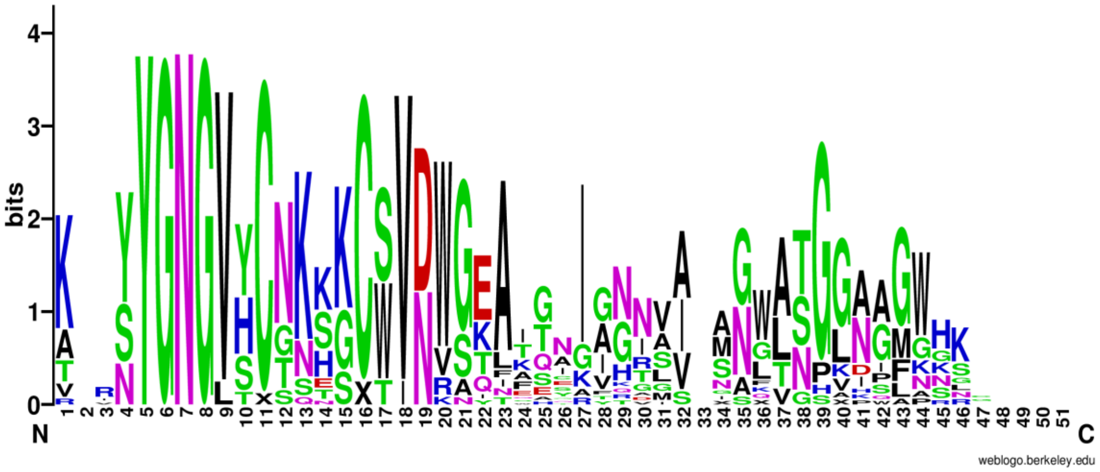
Table 4 - Comparison of genome mining tools with the Memi algorithm used in this work. Memi is currently the best option for the search of class IIa bacteriocins in metagenomes due to its independence of the gene context, the capability of prediction of virtually any class of peptides in the range of 50 amino acids in length. Adapted from ZHONG et al., 2020.

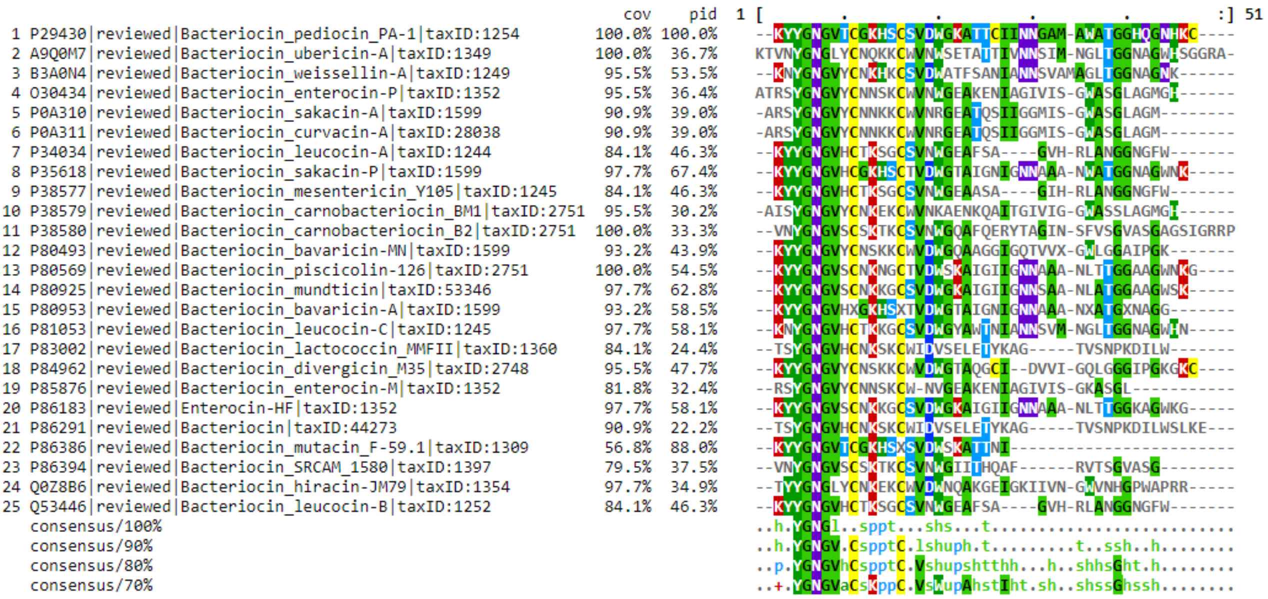
Mining tools	**Description**	**Advantages**	**Limitations**	**Methods**	**Ref**
Memi	**Read miner algorithm - metagenome or genome**	**Prediction of any family of proteins based on conserved amino acid patterns,** **Gene context-independent** **Graphical Interface**	**Read size**	**Pattern-based,**	**This work**
antiSMASH 5	Integrated platform for analyzing BGCs and metabolites including RiPPs	Integration of many other bioinformatics tools. Able to analyze and classify more than 50 classes of BGCs.	Gene context-dependent.	Rule-based, Hidden Markov model (HMM)	https://doi.org/10.1093/nar/gkz310
PRISM 3	Integrated platform for analyzing BGCs and metabolites including RiPPs	Able to analyze and classify more than 20 classes of BGCs. Able to predict the natural product structures of some types of BGCs.	Gene context-dependent.	Rule-based, HMM	https://doi.org/10.1093/nar/gkx320
BAGEL4	Combination of direct precursor peptides mining and indirect rule-based BGCs detection of RiPPs.	Able to predict both RiPPs BGCs and their precursor peptides.	Direct mining is searching precursor peptides against Bacteriocins and RiPPs databases, unable to mine for novel RiPPs precursor peptides.	Rule-based, HMM	https://doi.org/10.1093/nar/gky383
RiPPMiner	Machine learning classifier to predict RiPPs structural features based on precursor peptides	Able to classify different classes of RiPPs and predict their cleavage sites by precursor peptides sequences in a gene context-independent manner.	Prediction accuracy is poor for small classes of RiPPs (e.g. Sactipeptides, Linaridins).	Support vector machine (SVM)	https://doi.org/10.1093/nar/gkx408

RODEO	Mining for RiPPs BGCs and predicting precursor peptides by the combination of HMM, heuristic scoring, and machine learning.	Accurate prediction of RiPPs precursor peptides and their cleavage sites by context genes.	The current version is limited to some RiPPs classes (e.g. Lanthipeptides, Lasso peptides, etc.). Gene context-dependent.	Rule-based heuristic scoring, HMM, SVM	https://doi.org/10.1038/nchembio.2319
RiPPER	Family-independent identification of RiPPs precursor peptides	Able to identify novel precursor peptides of small RiPP class.	No user-friendly web tool is available. Requires prior knowledge of RiPP class to set the parameters. Gene context-dependent.	Prodigal scoring, HMM	https://doi.org/10.1093/nar/gkz192
DeepRiPP	Deep learning-based genome mining	Able to classify RiPPs and predict the cleavage sites by precursor peptides sequences in a gene context-independent manner.	Limited training sets and thus low accuracy for small RiPP class.	Deep neural network	https://doi.org/10.1073/pnas.1901493116

5.1. Multiple sequence alignment analysis

To determine the core peptides of reviewed UniprotKB pediocin-like bacteriocins (https://www.uniprot.org/uniprot/?query=family:%22bacteriocin%20class%20iia%20yngngv%20family%22&fil=reviewed%3Ayes&sort=score) and all Interprot class IIa bacteriocins, a aligment using the muscle algorithm was done (EDGAR, 2004) and the MSA's (Figure 5 and Figure 7) were submitted to the web logo tool (CROOKS, 2004), generating a bitmap with conserved amino acid positions for reviewed (Figure 6) and all pediocin-like bacteriocins (Figure 8).





```
                                                               cov    pid  1 [         .         .         .         .         :] 51
 1 P29430|reviewed|Bacteriocin_pediocin_PA-1|taxID:1254      100.0% 100.0%   --KYYGNGVTCGKHSCSVDWGKATTCIINNGAM-AWATGGHQGNHKC----
 2 A9Q0M7|reviewed|Bacteriocin_ubericin-A|taxID:1349         100.0%  36.7%   KTVNYGNGLYCNQKKCWVNWSETATTIVNNSIM-NGLTGGNAGWHSGGRA-
 3 B3A0N4|reviewed|Bacteriocin_weissellin-A|taxID:1249        95.5%  53.5%   --KNYGNGVYCNKHKCSVDWATFSANIANNSVAMAGLTGGNAGNK------
 4 O30434|reviewed|Bacteriocin_enterocin-P|taxID:1352         95.5%  36.4%   ATRSYGNGVYCNNSKCWVNWGEAKENIAGIVIS-GWASGLAGMGH------
 5 P0A310|reviewed|Bacteriocin_sakacin-A|taxID:1599           90.9%  39.0%   -ARSYGNGVYCNNKKCWVNRGEATQSIIGGMIS-GWASGLAGM--------
 6 P0A311|reviewed|Bacteriocin_curvacin-A|taxID:28038         90.9%  39.0%   -ARSYGNGVYCNNKKCWVNRGEATQSIIGGMIS-GWASGLAGM--------
 7 P34034|reviewed|Bacteriocin_leucocin-A|taxID:1244          84.1%  46.3%   --KYYGNGVHCTKSGCSVNWGEAFSA----GVH-RLANGGNGFW-------
 8 P35618|reviewed|Bacteriocin_sakacin-P|taxID:1599           97.7%  67.4%   --KYYGNGVHCGKHSCTVDWGTAIGNIGNNAAA-NWATGGNAGWNK-----
 9 P38577|reviewed|Bacteriocin_mesentericin_Y105|taxID:1245   84.1%  46.3%   --KYYGNGVHCTKSGCSVNWGEAASA----GIH-RLANGGNGFW-------
10 P38579|reviewed|Bacteriocin_carnobacteriocin_BM1|taxID:2751 95.5%  30.2%  -AISYGNGVYCNKEKCWVNKAENKQAITGIVIG-GWASSLAGMGH------
11 P38580|reviewed|Bacteriocin_carnobacteriocin_B2|taxID:2751 100.0%  33.3%   --VNYGNGVSCSKTKCSVNWGQAFQERYTAGIN-SFVSGVASGAGSIGRRP
12 P80493|reviewed|Bacteriocin_bavaricin-MN|taxID:1599        93.2%  43.9%   --KYYGNGVYCNSKKCWVDWGQAAGGIGQTVVX-GWLGGAIPGK-------
13 P80569|reviewed|Bacteriocin_piscicolin-126|taxID:2751     100.0%  54.5%   --KYYGNGVSCNKNGCTVDWSKAIGIIGNNAAA-NLTTGGAAGWNKG----
14 P80925|reviewed|Bacteriocin_mundticin|taxID:53346          97.7%  62.8%   --KYYGNGVSCNKKGCSVDWGKAIGIIGNNSAA-NLATGGAAGWSK-----
15 P80953|reviewed|Bacteriocin_bavaricin-A|taxID:1599         93.2%  58.5%   --KYYGNGVHXGKHSXTVDWGTAIGNIGNNAAA-NXATGXNAGG-------
16 P81053|reviewed|Bacteriocin_leucocin-C|taxID:1245          97.7%  58.1%   --KNYGNGVHCTKKGCSVDWGYAWTNIANNSVM-NGLTGGNAGWHN-----
17 P83002|reviewed|Bacteriocin_lactococcin_MMFII|taxID:1360   84.1%  24.4%   --TSYGNGVHCNKSKCWIDVSELETYKAG-----TVSNPKDILW-------
18 P84962|reviewed|Bacteriocin_divergicin_M35|taxID:2748      95.5%  47.7%   --KYYGNGVYCNSKKCWVDWGTAQGCI--DVVI-GQLGGGIPGKGKC----
19 P85876|reviewed|Bacteriocin_enterocin-M|taxID:1352         81.8%  32.4%   --RSYGNGVYCNNSKCW-NVGEAKENIAGIVIS-GKASGL-----------
20 P86183|reviewed|Enterocin-HF|taxID:1352                    97.7%  58.1%   --KYYGNGVSCNKKGCSVDWGKAIGIIGNNAAA-NLTTGGKAGWKG-----
21 P86291|reviewed|Bacteriocin|taxID:44273                    90.9%  22.2%   --TSYGNGVHCNKSKCWIDVSELETYKAG-----TVSNPKDILWSLKE---
22 P86386|reviewed|Bacteriocin_mutacin_F-59.1|taxID:1309      56.8%  88.0%   --KYYGNGVTCGKHSXSVDWSKATTNI------------------------
23 P86394|reviewed|Bacteriocin_SRCAM_1580|taxID:1397          79.5%  37.5%   --VNYGNGVSCSKTKCSVNWGIITHQAF------RVTSGVASG--------
24 Q0Z8B6|reviewed|Bacteriocin_hiracin-JM79|taxID:1354        97.7%  34.9%   --TYYGNGLYCNKEKCWVDWNQAKGEIGKIIVN-GWVNHGPWAPRR-----
25 Q53446|reviewed|Bacteriocin_leucocin-B|taxID:1252          84.1%  46.3%   --KYYGNGVHCTKSGCSVNWGEAFSA----GVH-RLANGGNGFW-------
   consensus/100%                                                            ..h.YGNGl..sppt...shs...t..........................
   consensus/90%                                                             ..h.YGNGV.CspptC.lshuph.t.........t..ssh..h........
   consensus/80%                                                             ..p.YGNGVhCspptC.Vshupshtthh...h..shhsGht.h........
   consensus/70%                                                             ..+.YGNGVaCsKppC.VsWupAhstIht.sh..shssGhssh........
```

Figure 5 – MUSCLE (EDGAR, 2004) MSA of 25 reviewed class IIa bacteriocins. Made with the EMBL (MADEIRA et al., 2019) tool Mview (BROWN; LEROY; SANDER, 1998).

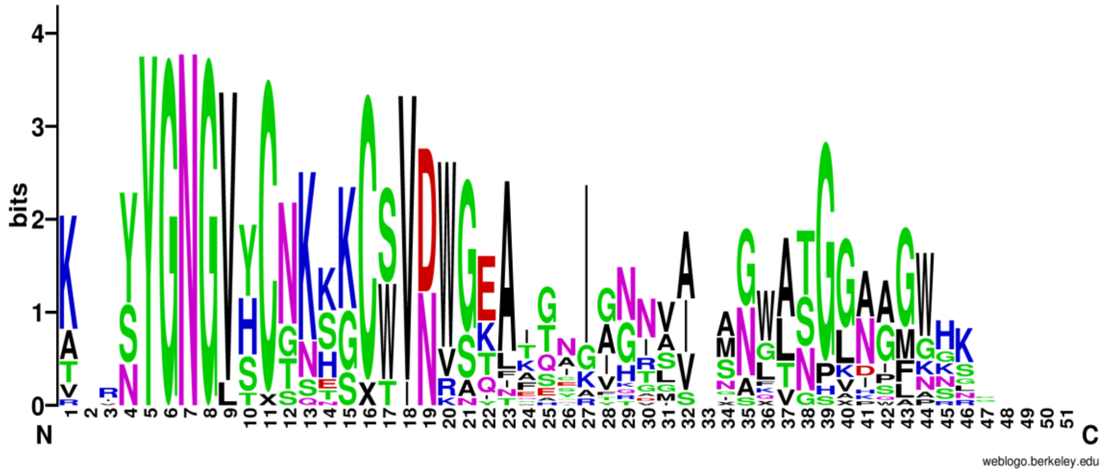
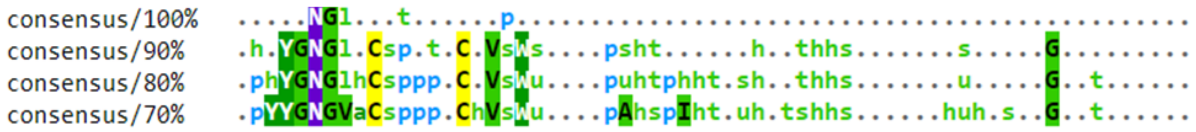


Figure 6 - Web logo of conserved amino acid positions generated from the MSA of reviewed pediocin-like bacteriocins (core peptides). Retrieved from UniProtKB (IPR002633:reviewed).



```
consensus/100%   .....NGl...t......p..............................................
consensus/90%    .h.YGNGl.Csp.t.C.VsWs....psht......h..thhs.......s.....G.........
consensus/80%    .phYGNGlhCsppp.C.VsWu....puhtphht.sh..thhs.......u.....G..t......
consensus/70%    .pYYGNGVaCsppp.ChVsWu....pAhspIht.uh.tshhs......huh.s..G..t......
```

Figure 7 - Conserved positions at 100%, 90%, 80%, and 70% of similarity of the MUSCLE MSA of all UniProtKB Class IIa bacteriocins. Made with Mview.

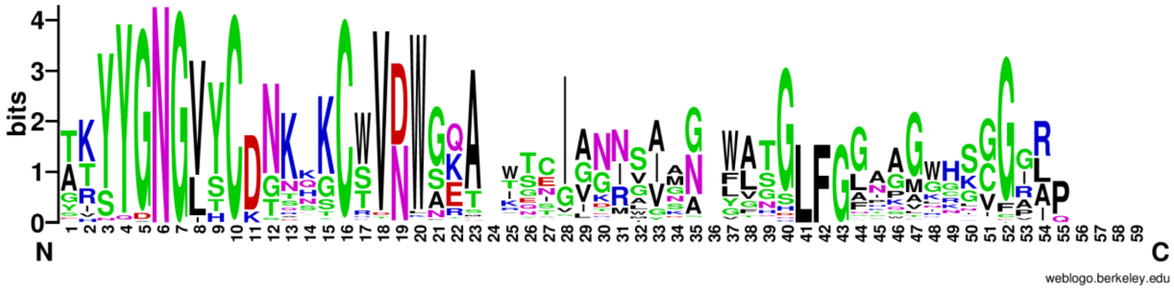


Figure 8 - Web logo of conserved amino acid positions generated from the MSA of all Interprot (IPR002633: both) pediocin-like bacteriocins (core peptides).

As extensivelly reported in literature (BALANDIN; SHEREMETEVA; OVCHINNIKOVA, 2019; COLLINS et al., 2018; DIEP et al., 2007; DRIDER et al., 2006; KUMAR et al., 2011; RÍOS COLOMBO et al., 2018; RODRÍGUEZ; MARTÍNEZ; KOK, 2002; SUSHIDA et al., 2018; UMU et al., 2016) and evidenced in figures 5, 6, 7 and 8, the pediocin-box is the characteristic region of pediocin-like bacteriocins, with the YGNG[VLI] well-conserved pediocin-box (inside the french braces in Figure 9 - 3D structure of Pediocin PA-1/AcH. Yellow arrows represent the cationic hydrophobic N-terminal triple anti-parallel β-folded sheet stabilized by a single disulfide bond and several hydrogen bonds; the red rectangle represents the amphiphilic helical conformation. Disulfide bonds are represented by linked cysteines. Adapted from BALANDIN; SHEREMETEVA; OVCHINNIKOVA, 2019.).

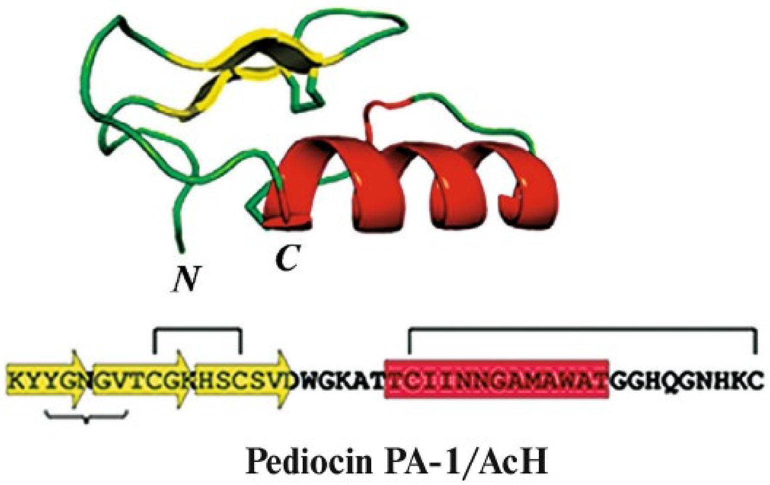


Figure 9 - 3D structure of Pediocin PA-1/AcH. Yellow arrows represent the cationic hydrophobic N-terminal triple anti-parallel β-folded sheet stabilized by a single disulfide bond and several hydrogen bonds; the red rectangle represents the amphiphilic helical conformation. Disulfide bonds are represented by linked cysteines. Adapted from BALANDIN; SHEREMETEVA; OVCHINNIKOVA, 2019.

The N-terminal triple anti-parallel β-folded sheet has a net positive charge and is fundamental for the pediocin-PA1 antimicrobial activity; it exerts initial electrostatic interactions with the plasmatic membrane of target bacteria (MILLER et al., 1998). While the C-terminal part of the bacteriocin confers its spectrum of activity (DRIDER et al., 2006). Therefore, the amino acid patterns designed focused on the conservation of the N-terminal region to preserve initial interaction and further pore formation.

5.2. Amino Acid Pattern Quality Control Parameters

Based on the abovementioned alignments, nine conserved amino acid patterns were designed, but only the best 3 (memi-7, memi-8, and memi-9) (Table 5) were used for further applications due to its better overall parameters. Alongside with developed patterns, another 3 amino acid patterns were considered to compare quality parameters (Table 5).

Table 5 - Amino acid patterns proposed and present in literature

Pattern	**PROSITE Sequence**	**Reference**
memi-9	[KTRVIA]-[YSN]-[YP]-G-N-G-[VLI]-x-C-x(4)-C-x-[VLI]	This work
memi-8	[YP]-G-N-G-[VLI]-x-C-x(4)-C-[SWTY]-[VI]-[DN]	This work
memi-7	[KTRVIA]-[YSN]-[YP]-G-N-G-[VLI]-[YHSTQ]-C-[NGST]-[KSND]-x-[KGST]-C-[SWTY]-[VI]-[DN]-[WKRV]-[GAQNS]-x-[ANL]-x(3)-[IKG]	This work
ahmad3a	Y-G-N-G-V-Y-C-N-x(2)-K-C-W-V-x(8)-I	(AHMAD et al., 2017)
ahmad3b	K-Y-Y-G-N-G-V-x-C-x-K-x(2)-C-x-V-x-W	(AHMAD et al., 2017)
PS60030	Y-G-N-G-[VL]-x-C-x(4)-C	PROSITE

Table 6 – Quality parameters of tested patterns. The memi-9 pattern was the best one. It demonstrated the highest sensibility amongst 100% precision patterns. The most important factor to mine metagenomes is precision, given the large dataset and reverse translation codon bias.

Pattern	**True positives**	**False Positives**	**Sensitivity(%)**	**Precision(%)**
memi-9	213	0	92,20779221	100
memi8	200	0	86,58008658	100
memi7	146	0	63,2034632	100
ahmad3a	42	0	18,18181818	100
ahmad3b	77	0	33,33333333	100
PS60030	229	7	99,13419913	97,03389831

Hence, the quality parameters of all amino acid patterns were evaluated against the same database of class IIa bacteriocins in Interprot. Comparing with other databases (Table 7), the memi-9 pattern was chosen as the best overall pattern. Memi-9 showed higher precision and optimal sensitivity when compared with literature ones.

Table 7 - Comparison of the number of hits by database.

	Number of hits per pattern/database							
database	Proteins	Positives	memi-9	memi8	memi7	ahmad3a	ahmad3b	PS60030
bagel4	811	51	51	51	47	11	22	50
UniprotKB	563082	229	213	200	146	42	77	236
PDB	168889	11	11	11	9	1	8	11
APD3	3167	31	31	31	28	6	11	30

5.3. Metagenome Mining

The mining of metagenomes yielded the occurrence of many existent class IIa bacteriocins (Table 8), but some newly discovered sequences (data not shown) are under investigation to decide what is the best one to carry the heterologous expression and evaluation of bacteriocin activity.

Table 8 - Metagenome mining of dairy food using the memi-9 pattern with 1 wild card allowed. Analyzing 222 paired reads, Memi loaded with memi-9 pattern could identify, uniquely, 52 pediocin-like bacteriocins with expected e-value ranging from 8,0E-04 to 2,00E-20, which names and Ids are listed.

Project Accession	Type	Type of food	Number of reads	Unique Hits	Names	Identifiers	e-values (range)	
							From	to
PRJEB15423	Paired	Surface ripened cheese	222	52	AvicinA, Bac43, Bacillocin1580, Bacillocin602, Bacteriocin31, Bacteriocin43, BacteriocinT8, bavaricin-MN, CarnobacteriocinB2, CarnobacteriocinBM1, Coagulin, CurvacinA/sakacinA, Divercin_RV41, DivercinV41, DivergicinM35, Eenterocin_P-like, EntA, Enterocin_CRL35, Enterocin_P-like, Enterocin_SE-K4, EnterocinA, Enterocin-HF, Enterocin-M, EnterocinP, EnterocinSE-K4, LactococcinMMFII, Leucocin_B_, LeucocinA, LeucocinC, Listeriocin743A, Mesentericin, MesentericinY105, MundticinATO6, MundticinKS, MundticinL, Pediocin, Pediocin_PA, PediocinPA-1/AcH, PenocinA, Piscicolin126/PiscicocinVia, Plantaricin423, PlantaricinC19, putative_bacteriocin, sakacin_G_skgA1, sakacin_G_skgA2, Sakacin_P, Sakacin_P_, Sakacin5X, SakacinG, SakacinP/Sakacin674, ubericin-A, WeissellinA	Bactibase_180|BAC180, Bagel_II_15|15.2, Bactibase_206|BAC206, Bactibase_204|BAC204, APD_2169|AP00854, APD_206|AP01183, Bactibase_167|BAC167, Bactibase_67|BAC067, APD_502|AP00145, APD_1927|AP00636, APD_2158|AP00844, APD_1928|AP00637, Bagel_II_65|68.2, APD_2156|AP00842, APD_2157|AP00843, Bagel_II_75|78.2, Bagel_II_77|80.2, Bagel_II_83|86.2, Bagel_II_92|95.2, Bagel_II_95|98.2, APD_2155|AP00841, Bactibase_175|BAC175, Bagel_II_99|102.2, APD_2168|AP00853, APD_201|AP01179, APD_2163|AP00849, Bagel_II_134|139.2, APD_1101|AP00199, APD_2162|AP00848, APD_2159|AP00845, Bagel_II_142|148.2, APD_1926|AP00635, APD_1922|AP00631, APD_2160|AP00846, Bactibase_184|BAC184, Bagel_II_156|163.2, Bagel_II_159|166.2, APD_1925|AP00634, APD_223|AP01199, APD_1923|AP00632, APD_2166|AP00851, APD_2167|AP00852, Bagel_II_198|205.2, Bagel_II_202|209.2, Bagel_II_203|210.2, Bagel_II_204|211.2, Bagel_II_205|212.2, APD_2161|AP00847, APD_2165|AP00850, APD_1924|AP00633, Bagel_II_220|227.2, Bactibase_200|BAC200	8,0E-04	2,00E-20
PRJEB20873		Nunu						
PRJEB32768		Cheese						
PRJNA603575		Yogurt and dietary supplement						
PRJNA286900		Cotija cheese						

Among these 52 unique pediocin-like bacteriocins, additional new candidate sequences (data not shown) were found, harboring amino acid point mutations and other features that will be further exploited.

5.4. pET-pedABCD expression vector modularity and versatility

The pET-pedABCD (Figure 10) is the functional plasmid that encodes all proteins necessary to maturation (PedC), cleavage/transport (PedD) of the precursor protein (pro-PedA), and to confer immunity (PedB).

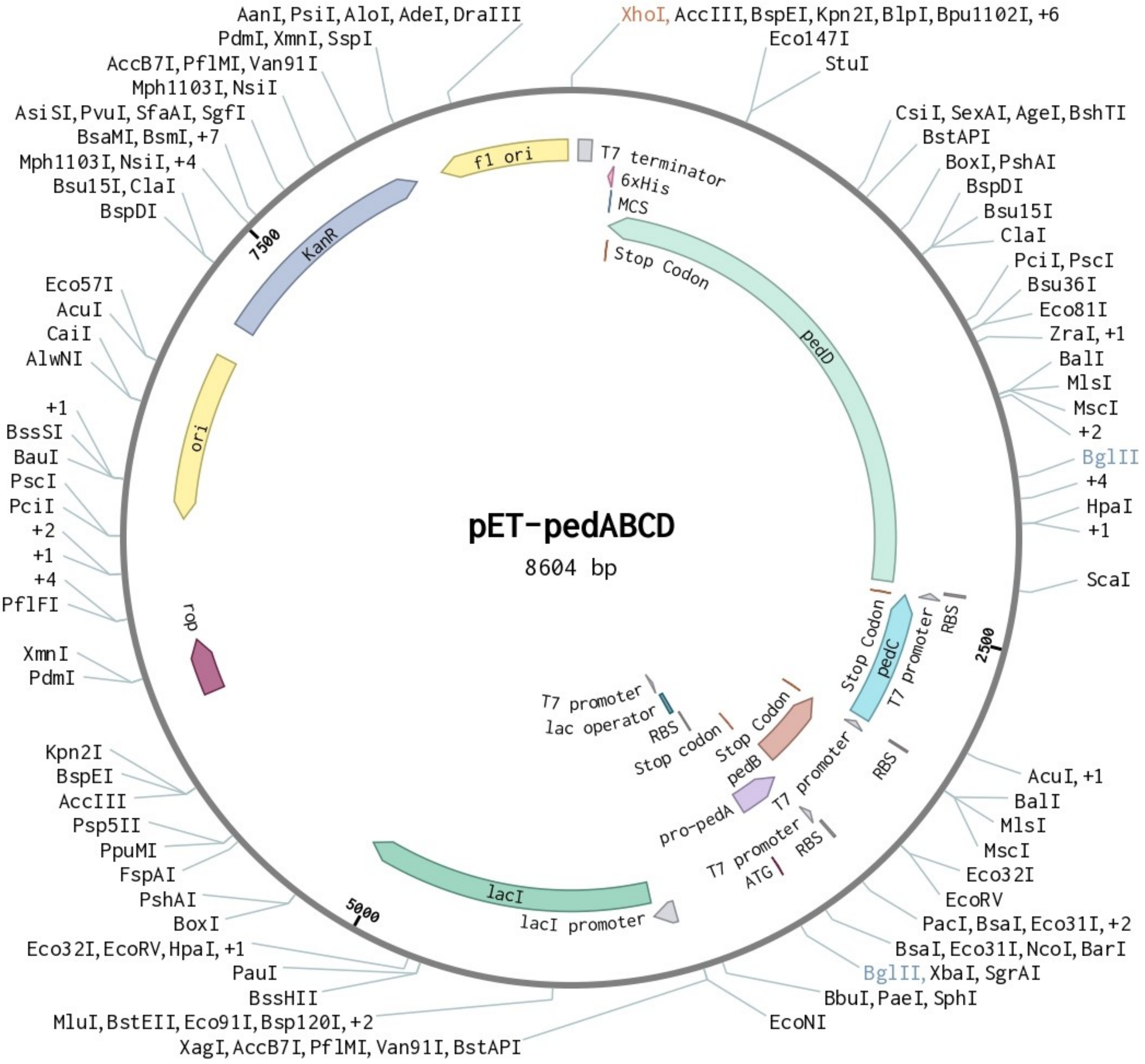


Figure 10 - The pET-pedABCD backbone. pET-pedABCD is a pET-28 a(+) based vector bearing four pediocin-related genes; *pro-pedA* (light purple) is the major one, responsible for the expression of pediocin-like precursor gene and plasmid modularity. *pedB* (light orange) encodes PedB the immunity protein against pediocin PA-1, which confers resistance for the

host. *pedC* (light blue) encodes PedC, the pediocin PA-1 biosynthesis protein that ensures correct disulfide bond formation. *pedD* (green) encodes PedD, the pediocin PA-1 transport/processing ATP-binding protein; an ABC transporter that functions during pediocin PA-1/AcH secretion in cells and also essential for pediocin production. lacI – lac repressor protein; KanR – Kanamycin resistance gene; f1 ori - phage f1 origin of replication; ori – origin of replication; rop – Regulatory protein Rop; RBS – Ribosome Binding Site; ATG – start codon.

The pediocin expression cassette (flanked by NcoI and XhoI) can be cloned in any other T7 based expression vector through digest and ligation reactions when NcoI and XhoI allow in-frame insertion (Figure 11). Also, XbaI digestion can be used to remove the control of the *lac T7* operon from the *pro-pedA* gene (upstream from operon lac). There are a plethora of bacteriocins heterologous expressed successfully in *E. coli* BL21(DE3) using pET derived and other vectors (MESA-PEREIRA et al., 2018b), but this is the first pET-28 based modular and customizable vector for the expression of pediocin-like bacteriocins.

It was seen that only the structural (*pedA*) and the transport (*pedD*) genes are exclusively required to produce a functional bacteriocin, but that the *pedC* gene conferred a higher yield of bacteriocin (MESA-PEREIRA et al., 2017). But previous works have expressed the entire ped operon, or just its precursor peptide

Table 9).

Various works used heterologous expression of bacteriocins as a tool to improve yield and expand the number of class IIa bacteriocins by rationally designed mutations; and the majority of them have chosen one of the available pET vectors and BL21 (DE3) *E. coli* strain. Recently, KIM et al., 2018 discovered a way to produce pediocin PA-1 through induction of the native host with phthalyl inulin nanoparticles; achieving 9-fold pediocin activity compared to control. Even so, only ~700 AU/mL of pediocin activity was achieved.

However, the production of pediocins to an industrial scale at affordable costs is prohibitive, since the cultivation of the original producing strain needs the fine-tuned control of several intrinsic factors to ensure high bacteriocin yield – 25,600 AU/mL (PAPAGIANNI; PAPAMICHAEL, 2014).

Table 9 – Heterologous expression of class IIa bacteriocins using *E. coli* BL21(DE3) strain and/or pET-based expression platform. [a] mature sequence (* -without leader); [b] IB, Inclusion bodies; SCF, Soluble Cellular Fraction (soluble fraction after cell pellet sonication); M, culture medium (cell-free supernatants), TE, Total cell extract. Adapted from MESA-PEREIRA et al., 2018a.

Bacteriocin	**Native host**	**_E. coli_ strain**	**Vector[a]**	**Location[b]**	**Culture conditions**	**References**
Pediocin PA-1 (Class IIa, two SS bonds)	_P. acidilactici_	BL21 (DE3)	pSuV1- _PedPA-1*_[c]	M	LB, 2 mM IPTG, ON, RT	INGHAM et al., 2005
	P. acidilactici PA003	BL21 (DE3)	pET32b- _pedA*_	IB	LB, 0.02 mM IPTG, 4 h, 37◦C	LIU; HAN; ZHOU, 2011
			pET20b- _pedA*_	SCF		
Carnobacteriocin B2 (Class IIa)	_C. maltaromaticum CP5_	BL21 (DE3)	pET32a-_CbnB2*_	TE	TB, 0.55 mM IPTG or 14.6 mM lactose 3 h, 37◦C	JASNIEWSKI et al., 2008
Carnobacteriocin BM1 (Class IIa)	_C. maltaromaticum CP5_	BL21 (DE3)	pET32a-Cbn_BM1*_ pET32a-CbnBM1M41V*	TE	TB, 0.55 mM IPTG or 14.6 mM lactose 3 h, 37◦C	JASNIEWSKI et al., 2008
Divercin AS7 (Class IIa, two SS bonds)	_C. divergens_ AS7	BL21 (DE3) pLys	pET28b-_AS7_	SCF	LB, 0.1 mM IPTG 24 h, 37˚C	OLEJNIK-SCHMIDT et al., 2014
Divercin V41	_C. divergens_ V41	Origami (DE3) pLysS	pET32b- _DvnV41*c_	SCF	TB, 1 mM IPTG 3 h, 37˚C	RICHARD et al., 2004
	C. divergens V41	Origami (DE3) pLysS/pCR03	pET32b- _DvnV41*_	SCF	TB or M9, 0-2 mM IPTG 3 h, 30˚C	YILDIRIM et al., 2007
E50-52 (Class IIa)	_Ent. faecium_ NRRL B-30746	BL21 (DE3)	pET SUMO-_rb50-52_	SCF	LB, 1.5 mM IPTG 5 h, 37˚C	(WANG et al., 2013)
Enterocin A (Class IIa)	_Ent. faecium_ ATB 197a	BL21 (DE3)	pET37b-_entA_	TE	LB, 0.1 mM IPTG, 1-4 h, 37◦C	KLOCKE et al., 2005
				SCF		
				M		
Enterocin B (Class II)	_Ent. faecium_ ATB 197a	BL21 (DE3)	pET37b-_entB_	TE	LB, 0.1 mM IPTG 1–4 h, 37˚C	KLOCKE et al., 2005
Enterocin CRL35 (Class IIa, one SS bond)	_Ent. mundtii_ CRL35	BL21 (DE3) pLysS	pET22b-_munA_	M	LB, 0.5 mM IPTG 1 h, 37˚C	MASÍAS et al., 2014
Enterocin P (Class IIa)	_Ent. faecium_	BL21 (DE3)	pTYB1-_EntP*_[c]	M	LB, 2 mM IPTG, ON, RT	INGHAM et al., 2005
	Ent. faecium P13	Tuner (DE3) pLacI	pETBlue-1-_entP_	M	M9, 0-1 mM IPTG, 3 h, 37◦C	GUTIÉRREZ et al., 2005
			pETBlue-1-_entP-entiP_	SCF		
Sakacin P (class IIa)	_Lb. sakei_	BL21 (DE3)	pET28a–sakP*	IB	LB, 0.8 mM IPTG, 3 h, 20 or 37◦C	CHEN et al., 2012
Piscicolin 126 (Class IIa, one SS bond)	_C. piscicola_	BL21 (DE3)	pSuV1- _pisA*_[c]	M	LB, 2 mM IPTG, ON, RT	GIBBS; DAVIDSON; HILLIER, 2004

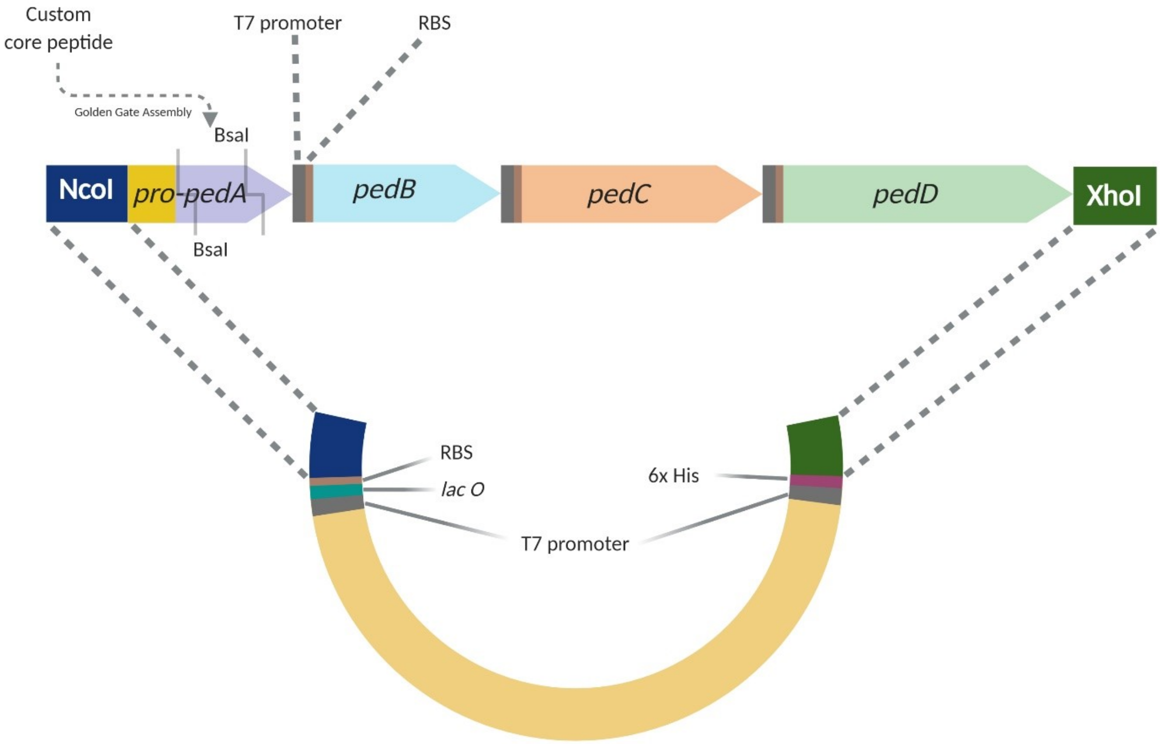


Figure 11 – Versatility of the expression cassette. To allow exchange between plasmids, the pediocin-like expression cassette is flanked by the NcoI (blue box) and XhoI (green box) restriction sites. *pro-pedA* gene (yellow and light purple arrow) holds several features to ensure the insertion of a custom core peptide right after the leader peptide codifying sequence (yellow box). Each arrowhead represents a stop codon. RBS – Ribosome Binding Site; *lac* O – lac operator; 6x His – poly-histidine tag. Created with Biorender.com.

There are many features in various pET-based heterologous expression platforms (Table 10) that can be exploited for class IIa bacteriocin production. In this work, pET-28 a(+) (bold line in Table 10) was chosen to be the backbone plasmid for the pediocin PA-1 expression cassette, due to all its features.

Table 10 - pET vectors features for heterologous protein expression. Amp, ampicillin; Kan, kanamycin; EK, enterokinase; Thr, thrombin. Adapted from MESA-PEREIRA et al., 2018a.

Vector	**Size (bp)**	**Promoter**	**Selection**	**Tags and fusion partners**	**Protease cleavage sites**	**Origin**	**Supplier**
pETBlue-1	3.476	T7*lac*	Amp	C-His	None	pUC	Novagen
pETcoco-2	12.417	T7lac	Amp	N-His	EK	Mini-F/RK2	Novagen
		pBAD		S			
				C-HSV tag			
pETDuet-1	5.420	T7	Amp	N-His	None	pBR322	Novagen
				S			
pET SUMO	5.643	T7*lac*	Kan	N-His	SUMO protease	pBR322	Invitrogen
				N-SUMO			
pET-14b	4.671	T7	Amp	N-His	Thr	pBR322	Novagen
pET-15b	5.708	T7	Amp	N-His	Thr	pBR322	Novagen
pET-20b (+)	3.716	T7	Amp	Signal sequence	None	pBR322	Novagen
				C-His			
pET-21c	5.441	T7*lac*	Amp	C-His	None	pBR322	Novagen
pET-22b (+)	5.493	T7*lac*	Amp	Signal sequence	None	pBR322	Novagen
				C-His			
pET-28a,b	**5.369**	**T7*lac***	**Kan**	**N-His**	**Thr**	**pBR322**	**Novagen**
				C- His			
pET-29a	5.371	T7*lac*	Kan	C-His	Thr	pBR322	Novagen
				Stag			
pET-32a,b	5.900	T7*lac*	Amp	N-Trx	Thr	pBR322	Novagen
				Internal His	EK		
				C-His			
pET-37b (+)	–	T7*lac*	Kan	Signal sequence	Thr	pBR322	Novagen
				N-CBD$_{cenA}$ C-His	Xa		
				S			

Bacteriocin expression modularity is conferred by the features presented in the *pro-pedA* gene (Figure 12). This gene was designed to allow the insertion of any core peptide using Golden Gate Assembly (ENGLER et al., 2009), through digestion with the BsaI restriction enzyme.

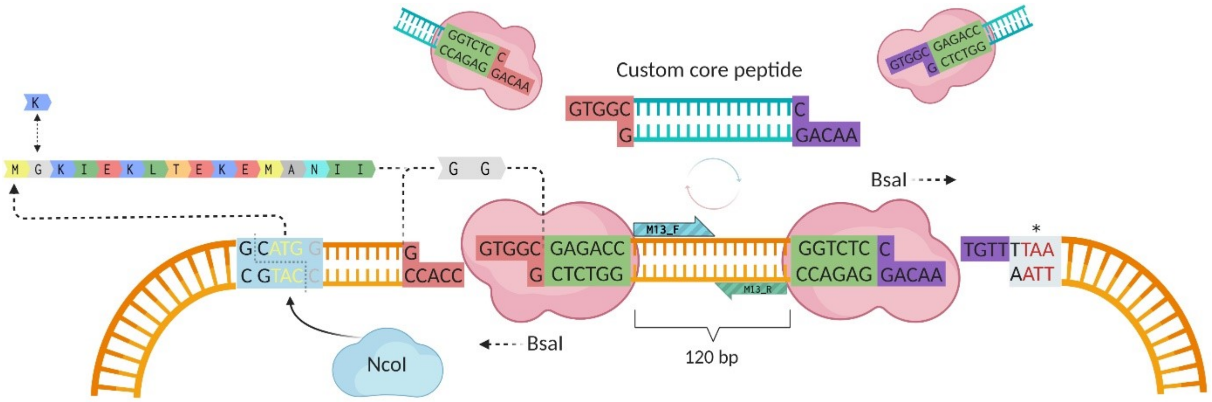
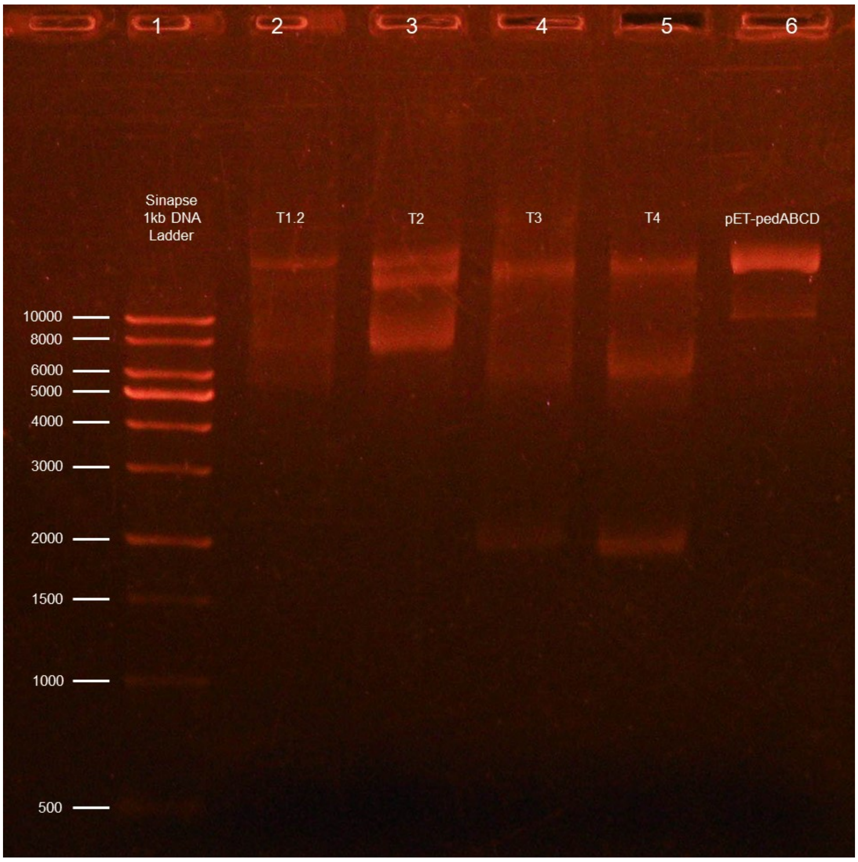


Figure 12 – Modularity and rational design of *pro-pedA*. The *pro-pedA* gene has a NcoI restriction site (light blue box) that contains the start codon (yellow ATG). The leader peptide (LP) of pediocin PA-1 (translated peptide) is universal to any pediocin inserted. To maintain the NcoI restriction site, a K2G mutation was created in the LP. BsaI recognition sites (green boxes) were invert placed to allow the scarless insertion (Golden Gate Assembly) of custom pediocins at the BsaI cutting sites (red and purple cohesive ends), which confers its modularity. The double glycine motif (red cohesive ends) is essential to the maturation of pediocins since the PedD protein cleaves the peptide right after it. M13 primers (half headed blue and green arrows) binding sites were placed between BsaI recognition sites to verify the ligation reaction using PCR. In which a 120 bp fragment arises when the control plasmid is amplified. The end of the precursor gene is marked by its stop codon (* / red TAA). Created with Biorender.com.

Our approach used the engineered (codon optimization, reformulation of *pedA* gene for modularity, avoidance of hairpins and XhoI, NcoI and BsaI restriction sites and insertion of NcoI and XhoI restriction sites at the ends) biosynthesis machinery of pediocin PA-1/AcH to allow the production and secretion of virtually any core peptide, both class IIa and class IId (due to its lack of post-translational modifications). It is expected that high bacteriocin activities could be achieved, in a low-cost, fast, and efficient manner. To enhance the overall scalability, it is suggested to change the inducer molecule from IPTG to lactose, which could decrease expression costs and increase the viability of the process.

5.5. Selection of Transformants and Plasmid DNA extraction

Clones T1.2 and T2 were similar in size to the positive control, possibly indicating the pET-pedABCD vector presence.



	1	2	3	4	5	6
Sample	Sinapse 1kb DNA ladder	T1.2 PLASMID	T2 PLASMID	T3 PLASMID	T4 PLASMID	pET-pedABCD
Ammount	3 uL	10uL	10uL	10uL	10uL	1 uL

Figure 13 - Plasmid Integrity Analysis. pET-pedABCD is the positive control at lane 6, with 8604 bp in length. To confirm plasmid integrity, all purified samples were submitted to 1% agarose gel. As seen in lanes 2 and 3, the transformant T1.2 and T2 seem to match the exact size of the positive control plasmid. Quantification of Plasmid DNA

Plasmid DNA quantification measures were done in technical and experimental triplicate to ensure minimal square deviation. Plasmid DNA from transformants was diluted by a factor of two, while the positive control was diluted by a factor of eight; both with the plate blank (deionized sterile water), and the real concentration of plasmid DNA was determined.

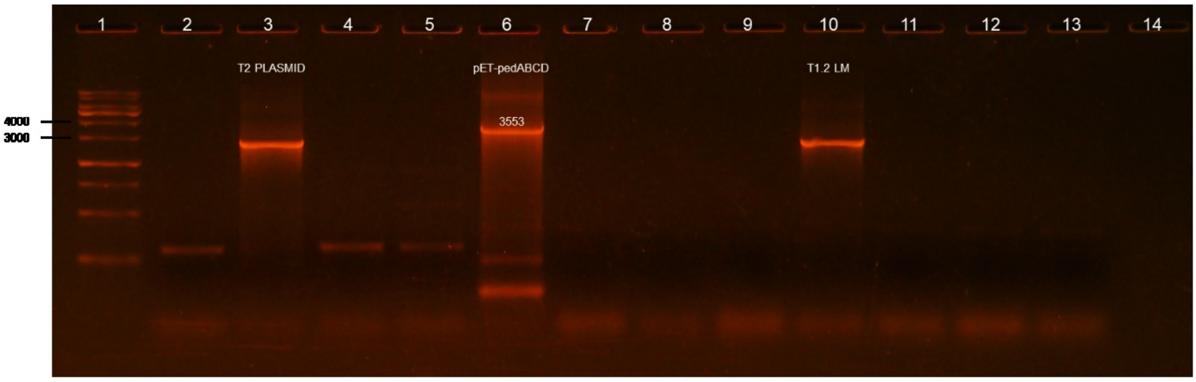
The T1.2 transformant was the cleanest yet weakest band, however, all plasmid DNAs concentrations were satisfactory on average.

Table 11 - Plasmid purified DNA quantification and quality control analysis.T1.2 to T4 plasmid DNAs were diluted by a factor of two, while pET-pedABCD by a factor of eight.

Sample	$A^{260-280}$	$A^{260-230}$	[ ] ng/µL - diluted	[ ] ng/µL - real	Amount (µL)	[ ] ng/µL - Qubit 3.0
T1.2	2.343	1.225	56.983	113.97	45	18.2
T2	2.2065	1.4135	101.148	202.3	45	22.2
T3	2.108	1.9595	70.733	141.47	45	22.4
T4	2.204	2.109	111.333	222.67	45	28.2
pET-pedABCD	1.5695	2.3215	23.183	185.464	45	41.8

5.6. Expression Cassette Confirmation

Out of four transformants, two carried the expression cassette; indicating the presence of the pET-pedABCD plasmid.

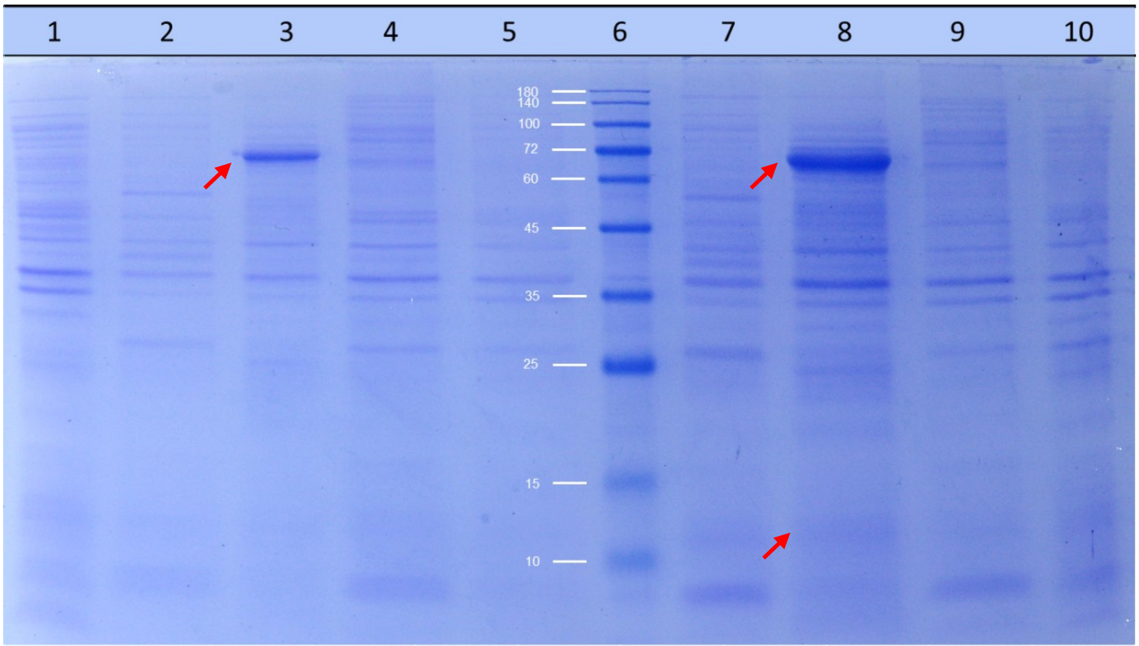


	1	2	3	4	5	6	7	8	9	10	11	12	13
Sample	Sinapse 1kb DNA ladder	T1.1 PLASMID	T2 PLASMID	T3 PLASMID	T4 PLASMID	pET-pedABCD	AE LM	C-	T1.1 LM	T1.2 LM	T2 LM	T3 LM	T4 LM
Ammount	3 uL	10uL	10uL	10uL	10uL	1 uL	10uL	10uL	10uL	10uL	10uL	10uL	10uL

Figure 14 - Pediocin-like expression cassette confirmation by PCR. The 3553 bp band was expected in transformants carrying the pediocin expression cassette; T2 and T1.2 were positive. Lanes 2 to 5 represent PCR with purified plasmid DNAs as a template, lane 6 the control plasmid. AE LM = Arctic Express Liquid Medium (negative control 1); C- (PCR without template DNA). Lanes 9 to 13 represents colony PCR of the transformants.

5.7. Protein Expression Analysis

Analyzing the protein expression in the 12,5% SDS-PAGE of the insoluble fraction, the enrichment of the band at approximately 72 kDa in lanes 3 and 8 (T1.2 before and after induction, respectively), could be explained by the expression of the PedD protein; an ABC transporter with membrane localization. Its expected size is 81,65 kDa; but electrophoretic shifts may occur and it is considered common for membrane proteins (RATH et al., 2009); preliminarily indicating the expression of pET-pedABCD proteins. As *pedD* is constitutively expressed, an increment in PedD content is expected due to the higher time of protein expression inside the cell (18h TPI) compared to the control (0h TPI).



	1	2	3	4	5	6	7	8	9	10
Sample	C-	T1.1	T1.2	T2	T3	M	T1.1	T1.2	T2	T3
TPI	18H	0H	0H	0H	0H		18H	18H	18H	18H

Figure 15 - 12,5% SDS-PAGE of the insoluble fraction of transformed *E. coli*. TPI - Time Post-Induction. The Blue Classic Prestained Protein Marker from Jena Bioscience Cat. No. PS-107 was used in lane 6 for protein size comparison. At lanes 3 and 8 probably is the PedD protein with 81,65 kDa and at ~13 kDa, possibly the PedB protein, which is also expected to be anchored to the membrane.

The resolution of the total cell extract and soluble cellular fraction SDS-PAGE's were not good enough to identify proteins expressed by transformed *E coli* clones, thus, this data was not displayed.

Subsequently, SDS-PAGE of total cell extract, soluble cellular fraction, and the insoluble fraction will be redone (same amount of protein each) to certainly verify the expression of PedB, PedC, and PedD proteins. The concentration of IPTG will be reduced to the minimum possible (0,5 mM; 0,25 mM; 0,1 mM) to avoid toxicity and maintain the yield of bacteriocin expression. Once IPTG concentration is adjusted, the protein expression will be scaled and the yield of bacteriocin production will be calculated. Another great benefit of engineered bacteriocins is the possibility to remove protease cutting sites to maintain bacteriocin integrity.

Four well know pediocin-like bacteriocins were engineered and already ordered (pediocin PA-1, enterocin-M, leucocin-A, and sakacin-A), and will be inserted in pET-pedABCD to verify if, even with distinct biosynthesis gene clusters, the engineered pediocin PA-1 biosynthesis machinery will be successful to produce active bacteriocins.

New pediocin-like bacteriocins discovered from dairy food metagenome mining will be engineered and ordered to verify the antimicrobial activity by a bacteriocin activity test. It is expected that the *E. coli* strain used will express high quantities of pediocin-like bacteriocins in the extracellular medium of soluble cell fraction, lowering the costs and raising activity compared to the producer strain. Moreover, the main goal is to achieve a functional expression of several pediocin-like bacteriocins using only one expression vector – pET-pedABCD.

6. CONCLUSION

The existence of a large data set of dairy food and bovine rumen metagenomes, rich in the genetic information of class IIa bacteriocins, foment the pursuit for new bacteriocins and the understanding of mechanisms of antimicrobial resistance in these complex microbial environments. In this work, were discovered new amino acid patterns with greater precision and sensibility than others previously used to identify the pediocin-like family of bacteriocins.

The bets pattern - memi-9 - was used to guide the Memi algorithm in the mining of dairy food and bovine ruminal metagenomes. Which yielded many new candidates of class IIa bacteriocins and 52 uniquely mapped pediocins with e-values ranging from 8e-04 to 2e-20. It demonstrates the richness of sequences from class IIa bacteriocins of the dairy food microbiome.

Furthermore, a new heterologous expression vector, named as pET-pedABCD, for class IIa and class IId bacteriocin protein expression was created and preliminary analysis suggests that PedD protein was expressed, possibly allowing pediocin export and maturation.

Our perspectives are to express 4 well-known bacteriocins to discover the vector's ability to express and secrete mature bacteriocins; perform the conventional well-diffusion method to determinate the bacteriocin activity of each expressed peptide, and evaluate the spectrum of action of newly discovered class IIa bacteriocins from the mining of dairy food metagenomes.

7. REFERENCES

AGILENT TECHNOLOGIES. ArcticExpress Competent Cells and ArcticExpress (DE3) Competent Cells Instruction Manual Catalog #230191. 2015.

AHMAD, V. et al. Antimicrobial potential of bacteriocins: in therapy, agriculture and food preservation. **International Journal of Antimicrobial Agents**, v. 49, n. 1, p. 1–11, 2017.

ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 403–410, out. 1990.

AZEVEDO, A. C. et al. Distribution and Genetic Diversity of Bacteriocin Gene Clusters in Rumen Microbial Genomes. **Applied and Environmental Microbiology**, v. 81, n. 20, p. 7290–7304, 2015.

BALANDIN, S. V; SHEREMETEVA, E. V; OVCHINNIKOVA, T. V. Pediocin Like Antimicrobial Peptides of Bacteria. v. 84, n. 5, p. 464–478, 2019.

BATEMAN, A. UniProt: A worldwide hub of protein knowledge. **Nucleic Acids Research**, v. 47, n. D1, p. D506–D515, 8 jan. 2019.

BÉDARD, F. et al. Synthesis, antimicrobial activity and conformational analysis of the class IIa bacteriocin pediocin PA-1 and analogs thereof /631/45/611 /631/326/22 /639/638/549/978 /639/638/309/2419 /639/638/403/977 /140/131 /82/6 /82/16 article. **Scientific Reports**, v. 8, n. 1, p. 1–13, 2018.

BÉDARD, F.; BIRON, E. Recent progress in the chemical synthesis of class II and S-glycosylated bacteriocins. **Frontiers in Microbiology**, v. 9, n. MAY, p. 1–14, 2018.

BHADRA, P. et al. AmPEP: Sequence-based prediction of antimicrobial peptides using distribution patterns of amino acid properties and random forest. **Scientific Reports**, v. 8, n. 1, p. 1697, 26 dez. 2018.

BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics**, v. 30, n. 15, p. 2114–2120, 1 ago. 2014.

BROWN, N. P.; LEROY, C.; SANDER, C. MView: a web-compatible database search or multiple alignment viewer. **Bioinformatics**, v. 14, n. 4, p. 380–381, 1 maio 1998.

CHEN, H. et al. Cloning and heterologous expression of a bacteriocin sakacin P from Lactobacillus sakei in Escherichia coli. **Applied Microbiology and Biotechnology**, v. 94, n. 4, p. 1061–1068, maio 2012.

COLLINS, F. W. J. et al. eincarnation of bacteriocins from the Lactobacillus Pangenomic Graveyard. **Frontiers in Microbiology**, v. 9, n. JUL, p. 1–9, 2018.

COTTER, P. D.; HILL, C.; ROSS, R. P. Bacterial lantibiotics: strategies to improve therapeutic potential. **Current protein & peptide science**, v. 6, n. 1, p. 61–75, fev. 2005a.

COTTER, P. D.; HILL, C.; ROSS, R. P. Bacteriocins: developing innate immunity for food. **Nature reviews. Microbiology**, v. 3, n. 10, p. 777–88, out. 2005b.

COTTER, P. D.; ROSS, R. P.; HILL, C. Bacteriocins-a viable alternative to antibiotics? **Nature Reviews Microbiology**, v. 11, n. 2, p. 95–105, 2013.

CROOKS, G. E. WebLogo: A Sequence Logo Generator. **Genome Research**, v. 14, n. 6, p. 1188–1190, 12 maio 2004.

DESRIAC, F. et al. Bacteriocin as weapons in the marine animal-associated bacteria warfare: Inventory and potential applications as an aquaculture probiotic. **Marine Drugs**, v. 8, n. 4, p. 1153–1177, 2010.

DIEP, D. B. et al. Common mechanisms of target cell recognition and immunity for class II bacteriocins. **Proceedings of the National Academy of Sciences of the United States of America**, v. 104, n. 7, p. 2384–2389, 2007.

DRAPER, L. A. et al. Lantibiotic Resistance. v. 79, n. 2, p. 171–191, 2015.

DRIDER, D. et al. The Continuing Story of Class IIa Bacteriocins. **Microbiology and Molecular Biology Reviews**, v. 70, n. 2, p. 564–582, 2006.

EDGAR, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**, v. 32, n. 5, p. 1792–1797, 2004.

ENGLER, C. et al. Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. **PLoS ONE**, v. 4, n. 5, p. e5553, 14 maio 2009.

ENGLER, C.; KANDZIA, R.; MARILLONNET, S. A one pot, one step, precision cloning method with high throughput capability. **PLoS ONE**, v. 3, n. 11, 2008.

FIMLAND, G. et al. Pediocin-like antimicrobial peptides (class IIa bacteriocins) and their immunity proteins: Biosynthesis, structure, and mode of action. **Journal of Peptide Science**, v. 11, n. 11, p. 688–696, 2005.

FUTURE, S. T. H. E.; DRUG-RESISTANT, F. NO TIME TO WAIT : SECURING THE FUTURE FROM DRUG-RESISTANT INFECTIONS. n. April, 2019.

GIBBS, G. M.; DAVIDSON, B. E.; HILLIER, A. J. Novel expression system for large-scale production and purification of recombinant class IIa bacteriocins and its application to piscicolin 126. **Applied and Environmental Microbiology**, v. 70, n. 6, p. 3292–3297, jun. 2004.

GUTIÉRREZ, J. et al. Cloning, production and functional expression of enterocin P, a sec-dependent bacteriocin produced by Enterococcus faecium P13, in Escherichia coli. **International Journal of Food Microbiology**, v. 103, n. 3, p. 239–250, 15 set. 2005.

HUGHES, D. Exploiting genomics, genetics and chemistry to combat antibiotic resistance. **Nature Reviews Genetics**, v. 4, n. 6, p. 432–441, 2003.

INGHAM, A. et al. The bacteriocin piscicolin 126 retains antilisterial activity in vivo. **Journal of Antimicrobial Chemotherapy**, v. 51, n. 6, p. 1365–1371, 2003.

INGHAM, A. B. et al. A versatile system for the expression of nonmodified bacteriocins in Escherichia coli. **Journal of Applied Microbiology**, v. 98, n. 3, p. 676–683, 2005.

JAMI, E. et al. Exploring the bovine rumen bacterial community from birth to adulthood. **ISME Journal**, v. 7, n. 6, p. 1069–1079, 2013.

JASNIEWSKI, J. et al. Optimization of the production and purification processes of carnobacteriocins Cbn BM1 and Cbn B2 from Carnobacterium maltaromaticum CP5 by heterologous expression in Escherichia coli. **Journal of Microbiological Methods**, v. 73, n. 1, p. 41–48, abr. 2008.

JOHNSEN, L.; FIMLAND, G.; NISSEN-MEYER, J. The C-terminal domain of pediocin-like antimicrobial peptides (class IIa bacteriocins) is involved in specific recognition of the C-terminal part of cognate immunity proteins and in determining the antimicrobial spectrum. **Journal of Biological Chemistry**, v. 280, n. 10, p. 9243–9250, 2005.

KIM, W. S. et al. A new way of producing pediocin in Pediococcus acidilactici through intracellular stimulation by internalized inulin nanoparticles. **Scientific Reports**, v. 8, n. 1, p. 1–14, 2018.

KLOCKE, M. et al. Heterologous expression of enterocin A, a bacteriocin from Enterococcus faecium, fused to a cellulose-binding domain in Escherichia coli results in a functional protein with inhibitory activity against Listeria. **Applied Microbiology and Biotechnology**, v. 67, n. 4, p. 532–538, jun. 2005.

KUMAR, B. et al. Cloning and expression of bacteriocins of Pediococcus spp.: A review. **Archives of Clinical Microbiology**, v. 2, n. 3, p. 1–18, 2011.

KUMARIYA, R. et al. Bacteriocins: Classification, synthesis, mechanism of action and resistance development in food spoilage causing bacteria. **Microbial Pathogenesis**, v. 128, n. January, p. 171–177, 2019.

LEVY, S. B.; BONNIE, M. Antibacterial resistance worldwide: Causes, challenges and responses. **Nature Medicine**, v. 10, n. 12S, p. S122–S129, 2004.

LIBBY, E. et al. Syntrophy emerges spontaneously in complex metabolic systems. **PLOS Computational Biology**, v. 15, n. 7, p. e1007169, 24 jul. 2019.

LIU, S. N.; HAN, Y.; ZHOU, Z. J. Fusion expression of pedA gene to obtain biologically active pediocin PA-1 in Escherichia coli. **Journal of Zhejiang University: Science B**, v. 12, n. 1, p. 65–71, jan. 2011.

LOZUPONE, C. A. et al. Diversity, stability and resilience of the human gut microbiota. **Nature**, v. 489, n. 7415, p. 220–230, set. 2012.

MADEIRA, F. et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. **Nucleic Acids Research**, v. 47, n. W1, p. W636–W641, 2 jul. 2019.

MAJEED, H. et al. Competitive interactions in Escherichia coli populations: The role of bacteriocins. **ISME Journal**, v. 5, n. 1, p. 71–81, 2011.

MANTOVANI, H. C.; RUSSELL, J. B. Bovicin HC5, a lantibiotic produced by Streptococcus bovis HC5, catalyzes the efflux of intracellular potassium but not ATP. **Antimicrobial Agents and Chemotherapy**, v. 52, n. 6, p. 2247–2249, 2008.

MASÍAS, E. et al. Co-expression and characterization of enterocin CRL35 and its mutant in Escherichia coli Rosetta. **Peptidomics**, v. 1, n. 1, 21 maio 2014.

MEDEMA, M. H. et al. Minimum Information about a Biosynthetic Gene cluster. **Nature Chemical Biology**, v. 11, n. 9, p. 625–631, 2015.

MESA-PEREIRA, B. et al. Controlled functional expression of the bacteriocins pediocin PA-1 and bactofencin A in Escherichia coli. **Scientific Reports**, v. 7, n. 1, p. 1–11, 2017.

MESA-PEREIRA, B. et al. Heterologous expression of biopreservative bacteriocins with a view to low cost production. **Frontiers in Microbiology**, v. 9, n. JUL, p. 1–15, 2018a.

MESA-PEREIRA, B. et al. Heterologous Expression of Biopreservative Bacteriocins With a View to Low Cost Production. **Frontiers in Microbiology**, v. 9, 26 jul. 2018b.

MILLER, K. W. et al. Isolation and characterization of pediocin AcH chimeric protein mutants with altered bactericidal activity. **Applied and environmental microbiology**, v. 64, n. 6, p. 1997–2005, jun. 1998.

MITCHELL, A. L. et al. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. **Nucleic Acids Research**, v. 47, n. D1, p. D351–D360, 8 jan. 2019.

NILSEN, T.; NES, I. F.; HOLO, H. Enterolysin A, a Cell Wall-Degrading Bacteriocin from Enterococcus faecalis LMG 2333. **Applied and Environmental Microbiology**, v. 69, n. 5, p. 2975–2984, maio 2003.

NISSEN-MEYER, J. et al. Structure-Function Relationships of the Non-Lanthionine-Containing Peptide (class II) Bacteriocins Produced by Gram-Positive Bacteria. **Current Pharmaceutical Biotechnology**, v. 10, n. 1, p. 19–37, 2009.

O’TOOLE, P. W.; COONEY, J. C. Probiotic Bacteria Influence the Composition and Function of the Intestinal Microbiota. **Interdisciplinary Perspectives on Infectious Diseases**, v. 2008, p. 1–9, 2008.

OLEJNIK-SCHMIDT, A. K. et al. Expression of bacteriocin divercin AS7 in Escherichia coli and its functional analysis. **Annals of Microbiology**, v. 64, n. 3, p. 1197–1202, 2014.

OPPEGÅRD, C. et al. The pediocin PA-1 accessory protein ensures correct disulfide bond formation in the antimicrobial peptide pediocin PA-1. **Biochemistry**, v. 54, n. 19, p. 2967–2974, 2015.

PAPAGIANNI, M.; PAPAMICHAEL, E. M. Production of pediocin SM-1 by Pediococcus pentosaceus Mees 1934 in fed-batch fermentation: Effects of sucrose concentration in a complex medium and process modeling. **Process Biochemistry**, v. 49, n. 12, p. 2044–2048, 1 dez. 2014.

PASOLLI, E. et al. Large-scale genome-wide analysis links lactic acid bacteria from food with the gut microbiome. **Nature Communications**, v. 11, n. 1, p. 1–12, 2020.

PORTO, M. C. W. et al. Pediococcus spp.: An important genus of lactic acid bacteria and pediocin producers. **Biotechnology Advances**, v. 35, n. 3, p. 361–374, 2017.

RATH, A. et al. Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. **Proceedings of the National Academy of Sciences**, v. 106, n. 6, p. 1760–1765, 10 fev. 2009.

RICHARD, C. et al. Heterologous expression and purification of active divercin V41, a class IIa bacteriocin encoded by a synthetic gene in Escherichia coli. **Journal of Bacteriology**, v. 186, n. 13, p. 4276–4284, jul. 2004.

RILEY, M. A.; WERTZ, J. E. Bacteriocins: Evolution, Ecology, and Application. **Annual Review of Microbiology**, v. 56, n. 1, p. 117–137, 2002.

RÍOS COLOMBO, N. S. et al. Pediocin-like bacteriocins: new perspectives on mechanism of action and immunity. **Current Genetics**, v. 64, n. 2, p. 345–351, 2018.

RODRÍGUEZ, J. M.; MARTÍNEZ, M. I.; KOK, J. Pediocin PA-1, a wide-spectrum bacteriocin from lactic acid bacteria. **Critical Reviews in Food Science and Nutrition**, v. 42, n. 2, p. 91–121, 2002.

SHARMA, R. et al. “Unculturable” bacterial diversity: An untapped resource. **Current Science**, v. 89, n. 1, p. 72–77, 2005.

STEWART, R. D. et al. Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. **Nature Biotechnology**, v. 37, n. 8, p. 953–961, 2019.

SULLIVAN, O. O. et al. **Encyclopedia of Metagenomics**. New York, NY: Springer New York, 2013.

SUSHIDA, H. et al. Evaluation of leader peptides that affect the secretory ability of a multiple bacteriocin transporter, EnkT. **Journal of Bioscience and Bioengineering**, v. 126, n. 1, p. 23–29, 2018.

UMU, Ï¿½ZGÏ¿½N C. O. et al. The potential of class II bacteriocins to modify gut microbiota to improve host health. **PLoS ONE**, v. 11, n. 10, p. 1–22, 2016.

UMU, Ö. C. O. et al. Potential applications of gut microbiota to control human physiology. **Antonie van Leeuwenhoek**, v. 104, n. 5, p. 609–618, nov. 2013.

VAN HEEL, A. J. et al. BAGEL4: A user-friendly web server to thoroughly mine RiPPs and bacteriocins. **Nucleic Acids Research**, v. 46, n. W1, p. W278–W281, 2018.

VILLARANTE, K. I. et al. Purification, characterization and in vitro cytotoxicity of the bacteriocin from Pediococcus acidilactici K2a2-3 against human colon adenocarcinoma (HT29) and human cervical carcinoma (HeLa) cells. **World Journal of Microbiology and Biotechnology**, v. 27, n. 4, p. 975–980, 25 abr. 2011.

WAGHU, F. H. et al. CAMP R3 : a database on sequences, structures and signatures of antimicrobial peptides: Table 1. **Nucleic Acids Research**, v. 44, n. D1, p. D1094–D1097, 4 jan. 2016.

WAGHU, F. H.; BARAI, R. S.; IDICULA-THOMAS, S. Leveraging family-specific signatures for AMP discovery and high-throughput annotation. **Nature Publishing Group**, p. 1–7, 2016.

WANG, Q. et al. Production of bacteriocin E50-52 by small ubiquitin-related modifier fusion in Escherichia coli. **Polish journal of microbiology**, v. 62, n. 4, p. 345–50, 2013.

YILDIRIM, S. et al. Production of recombinant bacteriocin divercin V41 by high cell density Escherichia coli batch and fed-batch cultures. **Applied Microbiology and Biotechnology**, v. 77, n. 3, p. 525–531, nov. 2007.

ZHAO, X. et al. LAMP: A Database Linking Antimicrobial Peptides. **PLoS ONE**, v. 8, n. 6, p. e66557, 18 jun. 2013.

ZHONG, Z. et al. Challenges and advances in genome mining of ribosomally synthesized and post-translationally modified peptides (RiPPs). **Synthetic and Systems Biotechnology**, v. 5, n. 3, p. 155–172, 2020.

ZOU, J. et al. Strategies for screening, purification and characterization of bacteriocins. **International Journal of Biological Macromolecules**, v. 117, p. 781–789, 2018.

ZOUHIR, A. et al. A new structure-based classification of gram-positive bacteriocins. **Protein Journal**, v. 29, n. 6, p. 432–439, 2010.