UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

In-silico gene mining, characterisation and the phylogenetic relationships of bacteriocins from selected phyla

By

Ilana Lucille Krieg

Submitted for the partial fulfilment of the degree

Magister Scientiae Biochemistry

in the Faculty of Natural and Agricultural Sciences

University of Pretoria

Pretoria

2018

Submission Declaration

I, Ilana Lucille Krieg declare that the thesis/dissertation, which I hereby submit for the degree M.Sc. Biochemistry at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

SIGNATURE: 

DATE: November 2018

Plagiarism Declaration

The Division of Biochemistry places great emphasis upon integrity and ethical conduct in the preparation of all written work submitted for academic evaluation.

While academic staff teach you about referencing techniques and how to avoid plagiarism, you too have a responsibility in this regard. If you are at any stage uncertain as to what is required, you should speak to your lecturer before any written work is submitted.

You are guilty of plagiarism if you copy something from another author's work (eg a book, an article or a website) without acknowledging the source and pass it off as your own. In effect you are stealing something that belongs to someone else. This is not only the case when you copy work word-for-word (verbatim), but also when you submit someone else's work in a slightly altered form (paraphrase) or use a line of argument without acknowledging it. You are not allowed to use work previously produced by another student. You are also not allowed to let anybody copy your work with the intention of passing if off as his/her work.

Students who commit plagiarism will not be given any credit for plagiarised work. The matter may also be referred to the Disciplinary Committee (Students) for a ruling. Plagiarism is regarded as a serious contravention of the University's rules and can lead to expulsion from the University.

The declaration which follows must accompany all written work submitted while you are a student of the Department of Biochemistry. No written work will be accepted unless the declaration has been completed and attached.

Full names of student:	Ilana Lucille Krieg
Student number:	11057352
Topic of work:	*In-silico* Gene mining, characterisation and the phylogenetic relationships of Bacteriocins from selected Phyla

Declaration

1. I understand what plagiarism is and am aware of the University's policy in this regard.

2. I declare that this M.Sc. thesis (eg essay, report, project, assignment, dissertation, thesis, etc) is my own original work. Where other people's work has been used (either from a printed source, Internet or any other source), this has been properly acknowledged and referenced in accordance with departmental requirements.

3. I have not used work previously produced by another student or any other person to hand in as my own.

4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

SIGNATURE

Acknowledgements

Praise is dedicated to the divine trinity for the opportunity to partake in this study and His hand in all aspects of this project.

I would like to express my sincere gratitude to Dr Mervyn Beukes, for his guidance, valuable time and significant insights dedicated to the successful completion of this project.

I am deeply indebted to my friends and colleagues, Demetra Demetriou and Renaan Thompson for their countless collaboration opportunities and constructive suggestions for improvement of the contents of this document.

Further I extend my warm appreciation to my family, Adèle Krieg, Thelma Agenbach and Johan Koch for their unwavering support throughout the duration of my project.

"Whether you think you can or you think you can't, you're right."

Henry Ford

Abstract

Bacteriocins are post-translationally modified antimicrobial peptides of bacterial origin. Bacteriocins have been suggested for various applications including food preservation and alternates to antibiotic treatments. Over the last 2 decades discovery of these peptides evolved from expensive, time-consuming lab based screening methods to reasonably high throughput computer-based alternatives via *in silico* gene mining such as BAGEL. In this project 932 genomes were mined for the presence of bacteriocin gene clusters. Of the 932, a total of 11 Eukaryotic genomes were used as a negative control. Analysis was performed to identify the type of bacteriocins and distribution amongst following phyla: Actinobacteria, BV4, Firmicutes, Glidobacteria and Proteobacteria. Novel bacteriocins identified were characterised *in silico* with respect to their physiochemical properties, including molecular weight, pI, extinction coefficient, aliphatic index, hydropathy index and estimated half-life in mammalian cells, yeast and bacteria. Characterisation of the 3D structures was done by homology modelling. All bacteriocins of the Actinobacteria, Firmicutes and Proteobacteria were then subjected to phylogenetic analysis. In this study the *in silico* mining of 932 genomes identified 407 novel bacteriocins. The physiochemical characterisation and homology models provide predicted chemical and structural characteristics of the identified bacteriocins. Sequence alignments indicate conservation of genetic clusters within the phyla. Phylogenetic trees were inferred from these alignments to study evolutionary relationships between bacteriocins and between producer species. The phylogeny of these species interestingly supports horizontal as well as lateral gene transfer patterns. The findings from this project will assist in laboratory investigations to further explore the growing field of bacteriocins and their possible application.

Keywords

In silico, bioinformatics, bacteriocins, gene mining, homology modelling, physiochemical properties, sequence alignment, phylogeny

Table of Contents

List of Figures

List of Tables

List of Abbreviations

3D	Three dimensional
ABC	ATP Binding Casette
AOI	Area of Interest
ATP	Adenosine triphosphate
BAGEL	Bacteriocin genome-mining tool
BLASTP	Basic Local Alignment Search Tool for Proteins
CAMEO	Continuous Automated Model Evaluation
DDT	Dichlorodiphenyltrichloroethane
DHR10	Designed helical repeat protein
DNA	Deoxyribonucleic acid
EDTA	Ethylenediamine tetra-acetic acid
GMQE	Global model quality estimate
GRAS	Generally regarded as safe
GRAVY	Grand average of hydropathy
HIV	Human immuno deficiency virus
kDa	Kilo Dalton
Lan	Meso-lanthionine
LAP	Linear Azol(ine) peptides
MDRSA	Multi-drug resistant *Staphylococcus aureus*
MeLan	3-mehtyllanthionine
MIC	Minimal inhibitory concentration
MRSA	Methicillin-resistant *Staphylococcus aureus*
MW	Molecular weight
NCBI	National Center for Biotechnology Information
NMR	Nuclear magnetic resonance
NNI	Nearest-neighbour-interchange
ORF	Open reading frame
Pfam	Protein family database
pI	Isoelectric point
pKa	Logarithmic dissociation constant
RNA	Ribonucleic acid

rRNA	Ribosomal ribonucleic acid
SAM	S-adenosyl methionine
SCIFF	Six Cys in forty-five- System
TB	Tuberculosis
tRNA	Transfer ribonucleic acid
VRE	Vancomycin-resistant *Enterococci*

Chapter 1: Literature Review

1.1 Problem Statement

Infectious diseases like respiratory infections, diarrheal diseases and tuberculosis are identified among the top ten contributers to global mortality (Baylor College of Medicine, 2016). The rapidly expanding variety of infectious diseases, bacterial resistance, decreasing food safety and food-security scores, intensifies despite desperate efforts to resolve these crises (Tijani, 2016). Infectious diseases like HIV and TB are less manageable due to co-infection and drug resistance (Nugent, 2010). Developing countries carry a heavier burden of infectious diseases due to suboptimal health care systems. In the agricultural sector, drought accompanies severe food shortage which intensifies the urgency for increased food safety and food security measures (Tijani, 2016). Bacteriocins are natural anti-microbial peptides effective against several food- as well as human pathogens. Previously tested Lasso peptides are harmless to humans, making them attractive alternatives for treatment of malaria, tuberculosis (TB) and human immuno deficiency virus (HIV). There is very little resistance known to bacteriocins and disease treatment shows similar effectivity to current antibiotic treatment (Rupnik et al., 2009). Bacteriocins have been identified as attractive alternatives to traditional antibiotics to alleviate the current problems stated above.

1.2 Background

Bacteriocins are biologically active peptides ribosomally synthesized by all major bacterial and archaeal lineages (Klaenhammer, 1993). Bacteriocins facilitate bacterial fitness to the producing strain by attacking other bacteria in the local microenvironment (Gillor *et al.*, 2008).

1.2.1 Bacteriocin structure

Bacteriocins stem from a functionally conserved gene cluster as described in figure 1.1. The operon contains a regulatory gene, structural gene, various modification genes, transport gene and an immunity gene.

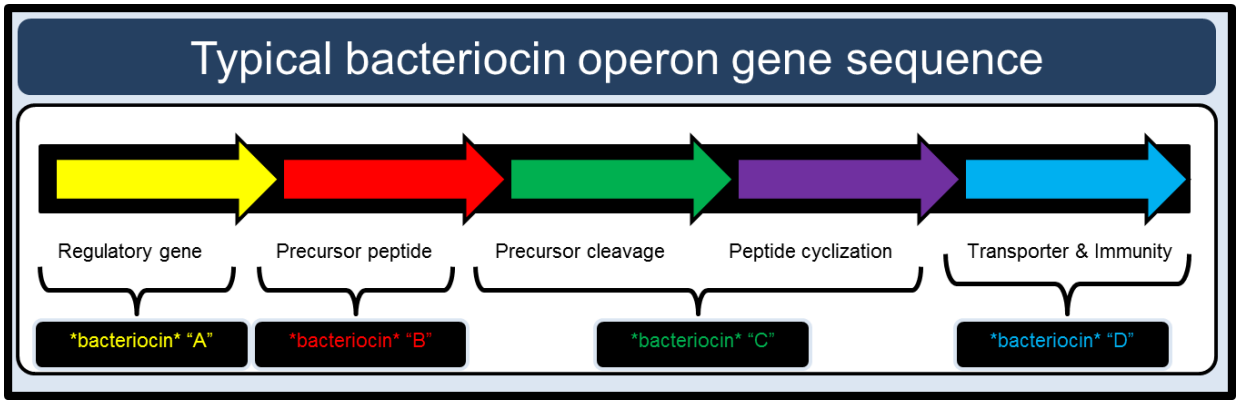

Figure 1.1: The genes encoding bacteriocins are generally sequentially conserved. The regulatory gene (yellow) regulates the operons transcription frequency. The precursor peptide gene (red) is followed by the modification enzyme genes (cleavage (green) and cyclization (purple)). The Lasso peptide transporter gene (blue) also serves as an immunity factor. Genes are specified as "bacteriocin name" and the corresponding "letter" E.g. Microcin J25: McjA, McjB, McjC, McjD (adapted from (Knappe *et al.*, 2008, van Heel *et al.*, 2013a))

Gene clusters encoding bacteriocins are conserved among species (Arnison *et al.*, 2013, van Heel *et al.*, 2013a). The regulatory gene is activated by a signal cascade. The precursor peptide is then modified after translation (Ditu *et al.*, 2014). The secreted mature bacteriocins lyse other bacteria in the micro -environment, providing a fitness advantage to the producing strain with the immunity gene (Beukes and Hastings, 2001).

1.2.2 Bacteriocin biosynthesis

Bacteriocins are ribosomally synthesized, modified and secreted from the cell to interact with the environment. A typical bacteriocin biosynthesis pathway is presented in figure 1.2.

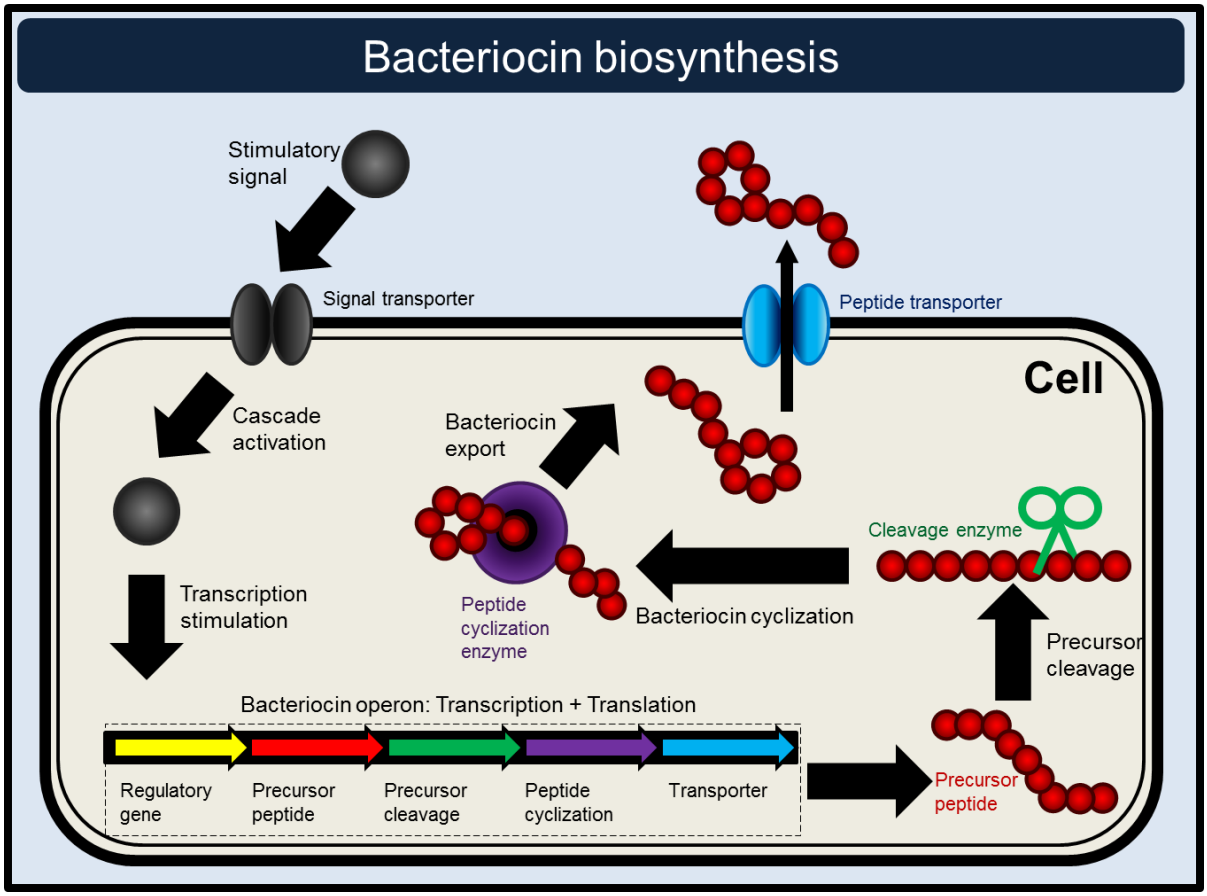

Figure 1.2: Bacteriocin biosynthesis pathway starts with an external signal to stimulate bacteriocin operon transcription via a signalling cascade. The transcript is translated and exported for modification like cleavage and cyclization. The mature bacteriocin peptide is then secreted from the cell via its respective transporter [adapted from (Ditu *et al.*, 2014)]

As an example, the Lasso peptide cluster is transcribed and translated within the bacterial cytoplasm. The linear precursor peptide synthesized by the ribosomes consists of a 5-45 residue leader sequence and 16-21 residues for the mature peptide. Microcin J25 (MccJ25) is the most studied and largest (58 amino acids) of the Lasso peptides. McjA is a regulatory gene for transcription of the Lasso peptide operon. McjB codes for a unique Adenosine triphosphate (ATP)-dependent cysteine protease responsible for leader sequence cleavage.

McjC codes for a lactam synthetase enzyme responsible for folding of the peptide in the signature Lasso structure. The lactam synthetase enzyme shares homology to asparagine synthetase B and is consequently predicted to form an isopeptide bond between the position 1 glycine and the side chain of the position 8 glutamic acid after being activated by an acyl- Adenosine monophosphate (AMP) intermediate. Enzymes coded for by McjB and McjC are interdependent for function (Arnison *et al.*, 2013). The ATP Binding Casette (ABC) transporter, coded for by McjD, secretes the mature Lasso peptide from cell cytoplasm to external environment and also serves as an immunity factor (Pan, 2012).

The secreted bacteriocin destroys susceptible cells in the environment by cell lysis. The producing strain is immune to the produced bacteriocin and can utilize all available nutrients unrivaled (Beukes and Hastings, 2001).

1.2.3 Bacteriocin functions

Bacteriocins have different structures and mode of actions which leads to different microbial potency (Gillor *et al.*, 2008). An ecological perspective on bacteriocin function suggested 3 main mechanisms of probiotic action. Bacteriocins can act as colonizing peptides that increase survival competitiveness by probiotic action (Hillman *et al.*, 1987, Casey *et al.*, 2007, Dawid *et al.*, 2007, Gillor *et al.*, 2009, Bhardwaj *et al.*, 2010), killing peptides that directly attack other pathogens (Cursino *et al.*, 2006, Dobson *et al.*, 2012), or signaling peptides that activate other bacteria or the host immune system to attack via the histidine protein receptor (Di Cagno *et al.*, 2010, Di Cagno *et al.*, 2011). Bacteriocins have great pharmacological potential. Bioactivities of bacteriocins include antimicrobial-(Zimmermann *et al.*, 2013), antiviral- (against human immunodeficiency virus (Frechet *et al.*, 1994)), inhibitory activity on enzymes and receptor agonists (like the glucagon receptor that plays a role in type 2 diabetes (Potterat *et al.*, 2004)).

1.2.4 Bacteriocin classes

Bacteriocin classification has been the subject of much debate (Jack *et al.*, 1995). The most used classification with 4 bacteriocin classes as proposed by Klaenhammer are described in figure 1.3 (Klaenhammer, 1993).

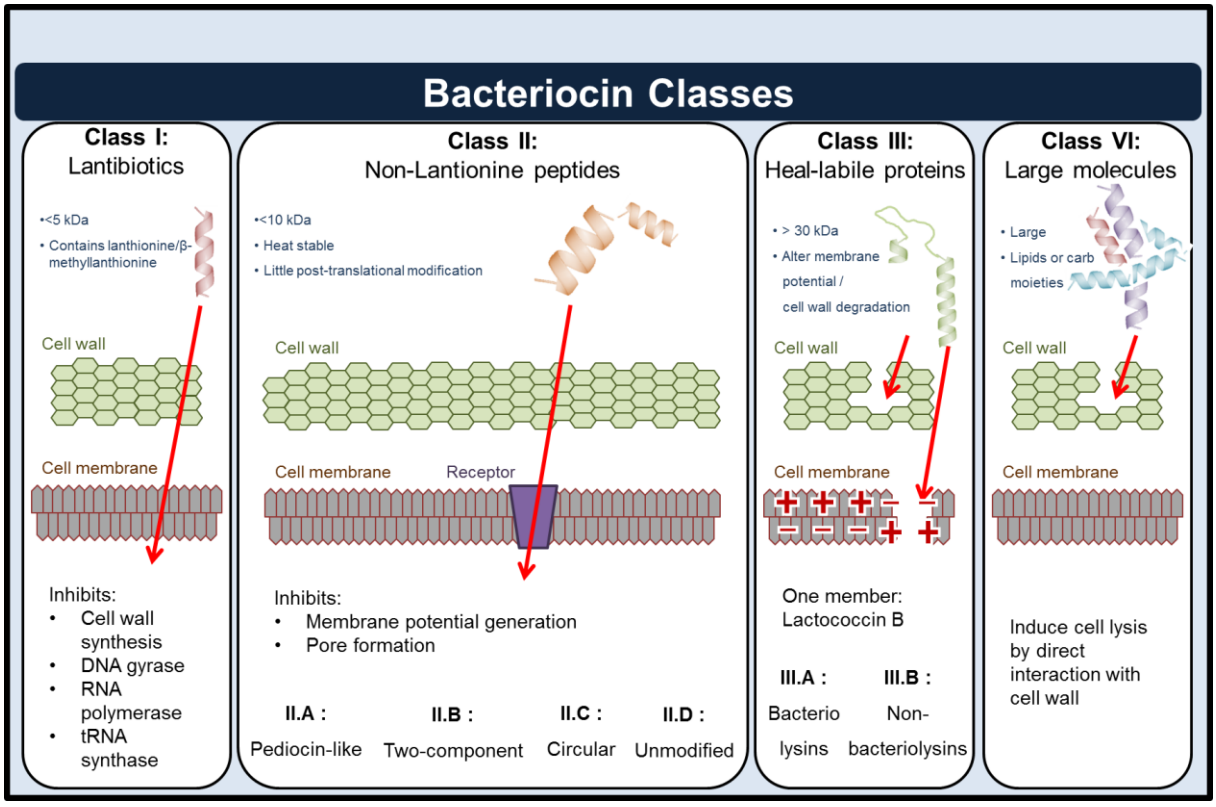

Figure 1.3: Structural classes of bacteriocins are distinguished. Class I Lanthibiotics are smaller than 5 kDa and contain modifications (either lanthionine or β-methlyllanthionine). Class II bacteriocins do not contain lanthionines, are smaller than 10 kDa and increase fitness pressure by interaction with other cell receptors. Class III bacteriocins are larger than 10 kDa, smaller than 30 kDa and cause cell lysis by either disrupting the membrane potential or direct cell wall degradation. Class VI contains the large molecules that directly interact with another cell wall and cause cell lysis (adapted from (Klaenhammer, 1993))

The Klaenhammer classification does not accommodate all bacteriocins discovered after 1993, thus allowing for improvement. Other classification models were proposed, but none yet universally agreed upon (Arnison *et al.*, 2013, Bastos Mdo *et al.*, 2015, Jack *et al.*, 1995).

1.3 Bacteriocin types

1.3.1 Bottromycins

Bottromycins contain unique macrocyclic amidines and a series of C-terminal decarboxilated thiazoles (follower peptide) instead of the N-terminal leader peptide. The Radical-S-Adenosyl Methionine (SAM) methyltransferases are responsible for C-methylation of the proline and phenylalanine residues and methylation of 2 valine residues on the precursor peptide. The gene cluster encodes, among others, cyclodehydration and thiazoline- forming enzymes for precursor peptide maturation (Arnison *et al.*, 2013). Producing strains obtain competitive fitness by secretion of Bottromycins. The putative Bottromycin-transporter grants immunity to the producing strain by facilitating Bottromycin export from the cell (Huo *et al.*, 2012). These bacteriocins inhibit translation of susceptible strains by binding to the A-site of the 50S ribosome (Arnison *et al.*, 2013).

Interest in the therapeutic potential of Bottromycins recently increased. Multi-drug resistant *Staphylococcus aureus* (MDRSA) and vancomycin-resistant enterococci (VRE) is inhibited by Bottromycin A2 and its derivatives, providing alternatives to these resistant infectious pathogens (Kobayashi *et al.*, 2010).

1.3.2 Cyanobactins

Cyanobactins consists of an N-C macrocyclic peptide. The precursor peptide (described as “partial bacteriocin name” “E” for example Patellamide precursor: PatE) is matured by cleavage and macrocyclisation by serine proteases “A” and “G”. Other common modifications include cyclohydratases, methyltransferases and pre-phenyltransferases (Arnison *et al.*, 2013).

Antimicrobial activity of Cyanobacteria against bacteria, viruses and cancer, as well binding to transition metals has been documented (Houssen and Jaspars, 2010). Attraction to Cyanobactins as putative cancer-therapy is due to the specificity to certain human cancel cell lines (Houssen and Jaspars, 2010).

1.3.3 Glyocins

The Glyocin class bacteriocins are characterised by their glycosylated post-translational modifications. The characterised members, Sublancin 168 and Glycocin F are unusually glycosylated at cysteine residues. Glyocins are not dependent on a leader peptide for activity. The mode of action, and biosynthesis pathway has not yet been fully characterised (Arnison *et al.*, 2013).

1.3.4 Head-to-tail cyclic peptides

The Head-to-tail cyclic peptides are macrocyclic- with a peptide bond between the N- and C-terminus residues. These large bacteriocins that are produced by gram-positive bacteria are resistant to proteolytic cleavage, temperature up to 100°C and pH changes (Arnison *et al.*, 2013). Post-translational modifications include bromination, disulphide bond formation, chlorination and amidation (Falanga *et al.*, 2017). Susceptible cells are lysed by ion efflux through pores formed in the bacterial cell membrane (Arnison *et al.*, 2013).

Subclasses of Head-to-tail cyclic peptides are distinguished based on physiochemical properties. Type I contains Head-to-tail cyclic bacteriocins cationic in nature with lower isoelectric point (pI). Members of this group include Circularin A, Lactocyclicin Q and Uberolysin. Type II Head-to-tail cyclic peptides are smaller than members of type I and have 3 N- to C-terminal crosslinks (Arnison *et al.*, 2013).

Head-to-tail cyclic bacteriocins are efficient antifungal agents (Lee and Kim, 2015), have broad spectrum antimicrobial activity and are applied in nanomedicine to enhance particle stability, improve nanoparticle effectivity and coat nanoparticles to kill biofilm-forming cells (Falanga *et al.*, 2017).

1.3.5 Lasso peptides

Lasso peptides are named according to their distinctive lasso structure. The penultimate threonine on the precursor peptide serves as recognition residue for modification enzymes (Maksimov *et al.*, 2012b). Association of the first N-terminal glycine or cysteine with the acidic aspartate or glutamate residue at position 8 or 9 on the chain forms a macrolactam ring. The C-terminal tail contains 2 amino acids with large side chains to physically prevent unfolding of the Lasso peptide (Duquesne *et al.*, 2007).

Precursor peptides of Proteobacteria have a conserved positively charged residue (histidine, lysine or arginine) to sterically trap the tail within the ring (Maksimov *et al.*, 2012b)

The different Lasso peptide classes and structural characteristics of each are shown in figure 1.4.

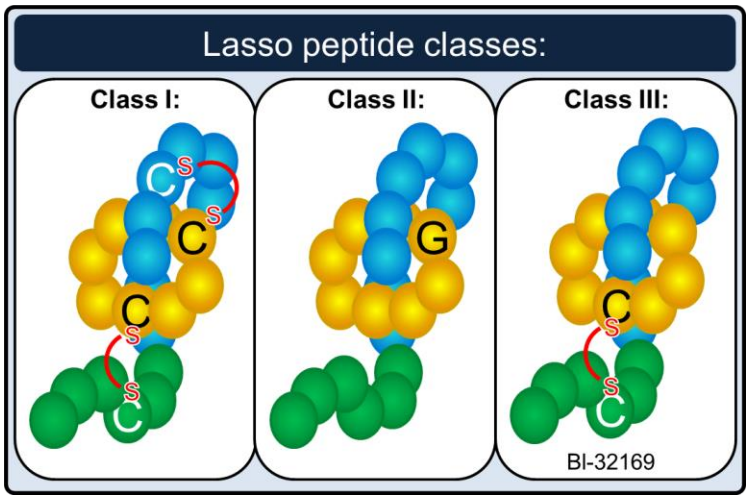

Figure 1.4: Classes I, II and III of Lasso peptides highlighted with their unique structural characteristics. Class I contains 4 cysteine residues, including 1 in position 1 of the loop, forming 2 disulphide bonds. Class II has a glycine in position 1 and Class III has 1 member, Bl-32169, which contains 2 cysteine residues forming 1 disulphide bond. Rings (yellow), loops (blue) and tails (green) are shown

Each respective class of Lasso peptides have different mechanisms of action. The class I Lasso peptide, Siamycin I, inhibit HIV infection *in vitro*, hypothetically via Glycoprotein 41 (a viral membrane protein) fusogenic activity inhibition (Ebadi, 2010). The class II Lasso peptide, MccJ25, inhibits transcription by binding and blocking the secondary channel of RNA Polymerase (Rebuffat, 2012). The sole member of class III Lasso peptide, Bl-32169, functions as a glucagon receptor antagonist which causes a decrease in blood sugar levels (Maksimov *et al.*, 2012a).

The unique fold of Lasso peptides gives extreme structural stability and a strong resistance against heat-, chemical and proteolytic degradation which make them attractive for drug design (Duquesne *et al.*, 2007). Previous research suggests that *Salmonella Newport* (Lopez *et al.*, 2007), Mersacidin resistant *Staphylococcus aureus* (MRSA) (Kruszewska *et al.*, 2004) and *Clostridium difficile* infections could be treated with respective bacteriocins with similar effectivity to conventional antibiotics, like Metronidazole and Vancomycin (Rupnik *et al.*, 2009). Lasso peptides like Siamicin I (MS-271) is active against HIV which broadens Lasso peptide drug applications to viral infections. Specific Lasso peptides show activity against human receptors and enzymes. The RES-701-1 Lasso peptide is an Endothelin receptor antagonist which is associated with a number of pathologies like cardiovascular disease, kidney disease and cancer. The BI-32169 Lasso peptide antagonizes the glucagon receptor to decrease blood glucose and is significant to the treatment of type I diabetes (Li *et al.*, 2014).

1.3.6 Linear azoline-containing peptides (LAPS)

Linear Azol(in)e-containing peptides (LAPs) contain oxazole and thiazole heterocycles, derived from cysteine, serine and threonine residues of precursor peptides. Biosynthesis of LAPs requires a precursor peptide “A” and a heterotrimeric enzyme complex consisting of a dehydrogenase “B”, cyclodehydratase “C” and “D”. The unmodified N-terminal leader peptide facilitates recognition by the “BCD” enzyme complex (Letzel *et al.*, 2014). After enzymatic maturation by post-translational modification, LAPs are secreted by an ABC transporter (Arnison *et al.*, 2013). Members of the LAP class typically exhibit antimicrobial activity against different *Streptomyces* and *Bacillus* species (Mills *et al.*, 2017).

1.3.7 Lanthibiotics

These bacteriocins contain meso-lanthionine (Lan) and 3-mehtyllanthionine residues (MeLan). Two steps after translation are required for Lanthibiotic maturation. Dehydration of serine and threonine residues on the precursor peptide is subsequently followed by addition of cysteine residues to form thioester crosslinks (Arnison *et al.*, 2013). Lanthibiotic subclasses are distinguished by the different Lan and MeLan installing enzymes and have different properties as described in table 1.1 (Arnison *et al.*, 2013).

Table 1.1: Lanthibiotics as classified by different modifying enzymes.

Lanthibiotic Class	Dehydration Enzyme	Cyclisation Enzyme
Lanthibiotic Class I	LanB dehydratase	LanC cyclase
Lanthibiotic Class II	Bifunctional synthetase: N-terminal dehydration	Bifunctional synthetase: LanC cyclase-like enzyme
Lanthibiotic Class III	Bifunctional synthetase: Central kinase	Bifunctional synthetase: Labionin-forming enzyme
Lanthibiotic Class IV	Bifunctional synthetase: Central kinase	Bifunctional synthetase: LanC cyclase-like enzyme

Class I Lanthibiotics have separate enzymes performing dehydration and cyclisation. Class II-IV Lanthibiotics have a bifunctional synthase performing modification steps in succession. The bifunctional synthases either dehydrate the precursor peptide at the N-domain (class II) or in the central domain (class III-IV). A carbon-carbon crosslink is formed for cyclisation of class III Lanthibiotics. Class IV Lanthibiotics are modified by a LanC-like cyclisation enzyme, similar to the enzymes for class I and II. Different post-translational modifications have been documented suggesting that leader peptides do not dictate the type of modification for maturation of precursor peptides to Lanthibiotics (Arnison *et al.*, 2013). Database genome mining suggested that Lanthibiotics are not restricted to Firmicutes and Actinobacteria, but synthesised in a variety of other phyla including Proteobacteria, Cyanobacteria, Bacteriodetes, etc. (Arnison *et al.*, 2013).

Lanthibiotics, like Nisin, has been extensively used to combat food pathogens for more than 40 years (Arnison *et al.*, 2013). In addition to antimicrobial activity, the morphogenetic Lanthibiotics SapB and SapT also function as bio surfactants allowing application within the biofuel waste disposal sector (Bezza *et al.*, 2015).

1.3.8 Linaridins

The first identified Linaridin, Cypemycin, was misclassified as a Lanthibiotic due to the similarities in structure. Linaridins may also contain a Lan or MeLan residue, but lack the appropriate dehydration enzyme genes within the bacteriocin coding operon (table 1.1 referred). In addition, the aminovinyl cysteine residues in Linaridins are composed of cysteines, as opposed to carbon residues in Lanthibiotics (Arnison *et al.*, 2013).

1.3.9 Microcins

The Microcin class members contain high serine and glycine content with low to now cysteine residues (Duquesne *et al.*, 2007). These hydrophobic peptides are highly resistant to heat, pH and protease degradation (Arnison *et al.*, 2013). Transporters for Microcins are predicted to be involved in either proteolytic cleavage for maturation of the precursor peptide, or self-immunity of the producing strain (Duquesne *et al.*, 2007). Before Microcins were structurally defined, all ribosomally synthesised antimicrobial peptides below 10 kDa from Enterobacteriaceae were named Microcins (Duquesne *et al.*, 2007). Recently however, Duquesne *et al.* proposed a more accurate classification system for Microcins as described in table 1.2.

The proposed classification system accommodates all 14 current Microcin members. Class I contains the small (≤3 kDa) Microcins that undergoes extensive modifications and class II the bigger Microcins (>3 kDa) with lesser modifications (Duquesne et al., 2007).

Table 1.2: Physiochemical properties and examples of members in Microcin classes.

Class		Size	Properties	Example Members
Class I		1-3 kDa	Extensive modification	MccB17, MccC7/C51, MccJ25
Class II	A	>3 kDa	Disulfide bond and siderophore modifications	MccV, MccL, Mcc24
	B	>3 kDa	No disulphide bonds, serine rich C-terminal region	MccE492, MccM, MccH47

The minimal inhibitory concentration (MIC) for antimicrobial activity is in the nano molar range. Unlike most bacteriocins, Microcins not only targets cell membranes, but also interact with intracellular targets at very high specificity (Duquesne *et al.*, 2007). The MccE492 Microcin inhibits mannose permease function, starving the susceptible cell of nutrients due to an inability to metabolise mannose and related hexoses. MccB17 inhibits DNA gyrase by irreversibly trapping the enzyme and blocking DNA replication. Other intracellular Microcin targets include aspartyl-tRNA synthetase and RNA polymerase. Their high specificity and small size make Microcins promising agents for probiotics, anti-tumural agents and antimicrobials for health and food preservation (Duquesne et al., 2007).

1.3.10 Sactipeptides

The cysteine sulphur linkage to alpha-carbon defines members of the Sactipeptide class. Nuclear Magnetic Resonance (NMR) spectroscopy studies unveiled the hairpin structure with exterior hydrophobic residues. An amide bond cyclises the N-terminal and cysteine sulphur bond stabilise the structure.

These hydrophobic bacteriocins lose their stability at basic pH to reveal an open hairpin structure (Arnison *et al.*, 2013). Broad spectrum antibiotics are known to disrupt homeostasis of normal gastrointestinal microbiota, leading to chronic

complications like inflammatory bowel disease(Rea *et al.*, 2011). *Clostridium difficil*, a gut pathogen, was exposed to Thuricin CD (Sactipeptide) treatment. Thuricin CD showed high activity and specificity to *C. difficil*, without disturbing normal gut flora (Rea *et al.*, 2011). Subtilocin A has spermicidal activity, is active against gram positive bacteria and can also lyse gram negative bacteria in the presence of high concentrations of ethylenediamine tetra-acetic acid (EDTA). The high specificity of Sactipeptides makes them attractive targets as antibiotics (Arnison *et al.*, 2013).

1.3.11 Thiopeptides

Six-membered nitrogenous rings and sulphur atoms characterise the Thiopeptide class of bacteriocins. The nitrogenous ring is recorded to be in either the piperidine, dehydropipieridine or pyridine oxidation states (Bagley *et al.*, 2005). Due to the structural complexity of these bacteriocins 6 enzymes are conserved for Thiopeptide production. The Thiopeptide precursor “A” is modified by dehydration enzymes “B” and “C”, cyclodehydratase “G” and dehydrogenase “E”. The metabolite “D” is predicted to assist in formation of 2 dehydroalanine residues. Peptide “F” is highly conserved, but has no homology to the enzymes coded for by other bacteriocin operons (Arnison *et al.*, 2013). Thiopeptides function by either binding near the GTPase-associated center disrupting translation, or by directly binding to and inactivating the 50S ribosomal subunit (Arnison *et al.*, 2013).

Thiopeptides show no activity against gram negative bacteria, but are potent inhibitors of gram positive protein synthesis, including many strains of MDRSA (Bagley *et al.*, 2005). The Thiopeptide member, Thiostreptin A, has activity against the malaria parasite interfering with proteasome function. (Arnison *et al.*, 2013). Thiostreptin A also shows anti-cancer activity by reducing function of the forkhead box M1 (FOXM1) transcription factor, active in tumours and cancer cell lines, leading to apoptosis (Hegde *et al.*, 2011, Pandit and Gartel, 2011)

1.4 Potential contribution

1.4.1 Infectious diseases

Infectious diseases refer to illness that is acquired directly or indirectly from an infected host or vector caused by infectious pathogens like bacteria, fungi, parasites and viruses or toxic products (Barreto *et al.*, 2006).

Table 1.3: Summary of the most prevalent infectious diseases globally.

Disease	Pathogen	Transmission	Global disease prevalence
Cholera	Bacteria	Oral-fecal route	4 500 000/year
HIV	Virus	Body fluid contact	36 700 000/year
Malaria	Parasite	Vector-mediated:*Anopheles* mosquito	214 000 000/year
TB	Bacteria	Airborne	10 400 000/year

(World health organisation, 2017)

Many efforts to combat infectious diseases failed, as the global disease prevalence per year as shown in table 1.3 is still high. Lasso peptides can decrease the prevalence of globally significant infectious diseases as they exhibit antimicrobial activity. Capistruin and MccJ25 are active against TB and the RP-71955 Lasso peptide has potent antiviral activity against HIV (Li *et al.*, 2014).

1.4.2 Bacterial resistance

Bacterial resistance is the ability of bacterial strains to tolerate specific antibiotics to which they previously showed susceptibility (Scientific Committees, 2017). An increasing number of multi drug resistant pathogens cause a decreased amount of common illnesses to be efficiently treated and cured. Although pathogenic drug resistance is a natural evolutionary process, it was enhanced by improper antibiotic use. Of the 71% of children that received antibiotics for upper respiratory infections, only 35% of cases were treated according to clinical guidelines. Children and elderly people are most vulnerable to infectious diseases due to weakened immune systems and carry the heaviest burden. Multi resistant pathogens cause economic strain; the cost to cure one person from drug-resistant TB, XDR-TB, is equal to the price to cure 200 patients with susceptible TB strains (Nugent, 2011).

Only half of antibiotics globally are obtained with a prescription. The improper use and distribution of antibiotics expands antibiotic resistance and increase the number of resistant pathogenic strains (Auta et al., 2019). Drug resistance is on a rapid increase like the Methicillin-resistant *Staphylococcus aureus* (MRSA) that increased from approximately 2% in 1974 to 50% in 2004 (Nugent, 2011). An increasing number of Malaria, pneumonia, cholera, TB and dysentery cases are caused by multidrug resistant strains which increase economic burden and prevent effective treatment (Nugent, 2011).

Lasso peptides are natural products effective against the most globally devastating infectious diseases namely malaria, TB and HIV. There is no resistance against the respective Lasso peptides and treatment shows similar effectivity to used antibiotics. The Lasso peptides active against the malaria, TB and HIV pathogens are harmless to humans, making them attractive alternatives to antibiotics.

1.4.3 Food security

Food security refers to the adequate amount of food to be supplied to feed a population. Sustained sources, confirmed supply, distribution and fair access are acknowledged to form part of the food security score (Fan *et al.*, 2017). Current challenges in South Africa regarding food security include inadequate emergency management systems, suboptimal fertile land utilization and little citizen knowledge of safe food handling (Department of social development, 2013).

Regional and sub-regional nutrition policies and programs were implemented between 2014 and 2015 (Tijani, 2016). Agricultural and economic resources are included in the food security dimensions which consist of availability, access and utilization of food. The international highest amount of food insecurity is however still present in sub-Saharan Africa. Climate changes like extreme weather events, loss of biodiversity and higher temperatures decrease food security.

Civil conflicts like farmer-worker disagreements, fuel strikes and crop or livestock plundering decrease food security by disrupting the food supply chain resulting in loss of assets, loss of income and an inability to deliver (Tijani, 2016).

The pesticides like herbicides for weeds, insecticides for insects and fungicides for fungi are used to increase crop yields and increase food security scores are toxic for human consumption. Acute dangers to pesticide exposure include nerve, skin and liver damage, nausea, fatigue and systemic poisoning. Chronic health effects of pesticide exposure include contribution to the risk of developing Autism spectrum disorders and cancers like leukemia, non-Hodgkins lymphoma and brain cancer. Multiple chemical sensitivity is a common condition associated with a body's inability to tolerate low exposure to chemicals and is intensified after pesticide exposure. Although banned in 1972, pesticides like dichlorodiphenyltrichloroethane (DDT) are still present causing endocrine system disruptions, hormone dysregulation and abnormal embryonic development which is associated with impaired brain development, behavioral disorders and incomplete sexual development (Broude *et al.*, 2015).

Ecosystem disruptions and human disease increases can be limited if an eco-friendly alternative like Lasso peptides could replace the use of pesticides. The antimicrobial activity of Lasso peptides can thus contribute to increase food security. The Lasso structure facilitates thermal stability resulting in optimal function regardless of the regional climate and weather.

1.4.4 Food safety

Food safety refers to the relative risk of illness when consuming available food. Chemical-and microbial contamination during growth and slaughter, processing, storage, shipping, and poor labelling decrease the food safety score (Wax, 2016). Foodborne illness can stem from poor food safety practices. Raw foods like meat, fish, eggs and milk need strict cold-chain transport procedures to prevent pathogen growth. *Salmonella* and *Listeria monocytogenes* are often found in raw and undercooked animal products like meat, fish and dairy products. *Shigella* spread through the oral-fecal route, contaminating food and water during preparation by improper hygiene practices. Pathogenic *Escherichia coli* is found in meat, dairy products and fresh produce like fruit juice. Presence of viruses, parasites and chemicals like pesticides also decreases food safety scores (National institute of diabetes and digestive and kidney diseases, 2014).

International spread of foodborne pathogens by tourists, immigrants and refugees increased the public health problem (Jones and van Niekerk, 2004).

Immune compromised individuals like young children, pregnant women and elderly people are at a heightened risk to have more severe symptoms. Pathogenic infection symptoms include vomiting, diarrhea, abdominal cramps and fever. These symptoms can lead to hemolytic uremic syndrome, dehydration and death if not treated. Chemical contaminated food can cause headaches, blurred vision, dizziness and paralysis. Symptoms can amplify chronic complications like irritable bowel syndrome, Guillain-Barrè syndrome, high blood pressure, diabetes and reactive arthritis (National institute of diabetes and digestive and kidney diseases, 2014). About 70% of infant diarrheal cases in South-African are due to contaminated food or water. Long term effects of dehydrated infants include immune system suppression, DNA damage that can lead to cancer and hormonal control abnormalities that can stomp growth and brain development (Puth, 2017). Socio-economic effects include loss of productivity and income due to loss of trade which can lead to loss of job opportunities and unemployment. This in turn relates to unbalanced diets, increased crime statistics and loss of tourism (Jones and van Niekerk, 2004).

Legislation stipulates that a certificate of acceptability is to be issued by local authority to allow a relevant party to store or handle food at the premises. The laws were implemented following the All Africa Games embarrassment in 1999, an international sports event where more than 600 school children participating in the event became sick from food served at the games (Jones and van Niekerk, 2004).

Various food control authorities including the national-and provincial departments of health are by law responsible for support, reinforcement and auditing of the associated institutions that handle or store food. District-and metro municipalities in terms of the Health Act 63 of 2003 are responsible for environmental health services like water quality monitoring, food control, waste management, chemical safety and disposal of the dead (Jones and van Niekerk, 2004). Consumers' food newspapers like Food Safety News and the New York have however published safety scandals as recent as 18 April 2017 (NewYorkTimes, 2017, NewsDesk, 2017). Due to the antimicrobial activity of Lasso peptides, fresh foods like meat, fruit and vegetables can be stored, transported and prepared with decreased risk of microbial

contamination. Lasso peptides are resistant to chemical and heat denaturation which keep the Lasso peptides stable during preservative processes.

The vast potential of bacteriocins as possible solutions to pressing issues across multiple research fields warrants further investigation of these peptides. This project aims to supplement literature, suggesting candidates that warrant further research for their contribution to solving pressing issues in medicinal and agricultural fields. This will be done via *in-silico* research.

In chapter 2 the project focuses on identifying bacteriocin gene clusters within species of various phyla by genome mining. The resultant statistics summarize the bacteriocin class distribution and estimate the amount of bacteriocin producing species amongst various phyla. The novel bacteriocins identified in chapter 2 are further studied in chapter 3 with regards to their physiochemical parameters and evolutionary relationships using bioinformatics software. The putative physiochemical properties, predicted 3 dimensional (3D) structures and inferred evolutionary relationships may assist laboratory experiments and further studies.

The results of this project, that are discussed in chapter 4, will contribute to the bacteriocin arsenal available and assist in solving food and pathogen related crises, encourage research in this field and ultimately improve quality of human life.

Chapter 2: Gene Mining and Bacteriocin Distribution Analysis

2.1 Research aim

This chapter aims to identify bacteriocins in a variety of species by genome mining and compare the bacteriocin types among selected phyla.

2.2 Research objectives

1) To select genomes of interest from Genbank
2) To identify bacteriocins of interest by gene mining using BAGEL3

2.3 Experimental design and methodology

Genomes obtained from an online database are mined through an online server. The methods and software used to obtain and analyse these genomes are summarised in figure 2.1 below.

2.3.1 Species genome sequences of interest obtained from Genbank and classified according to phylum

Genbank provided annotated bacteriocin genes (Benson *et al.*, 2005) which were manually selected to represent most of the bacterial phyla. Species of interest for further experimentation were selected. Genera-and species selection was based on availability of the relevant published resources in Genbank (Benson *et al.*, 2005). Selection was done to minimise further research time and costs. Gene mining was performed on the selected species.

2.3.2 Gene mining, characterisation and analysis

The process of gene mining entailed isolation of a significant gene from related genotypes (Dogra, 2017). Discovery analysis upstream and downstream of the gene was performed to determine sequence similarity.

Conserved sequences in associated genes assisted in further narrowing down of species to be analysed via the bacteriocin genome-mining tool (BAGEL3). BAGEL3 is an automated bacteriocin genome mining tool that uses a dual analysis method to identify putative bacteriocins and analyses input sequences independent of previously predicted open reading frames (ORF's). Direct and indirect analysis methods run in parallel. The direct method initiates a Glimmer ORF call, followed by a BLAST search against 3 databases and annotation by the protein family database (PFAM) software. The indirect mining strategy utilizes context based mining to initially identify ORF's with start and stop codons. The ORF's were then screened for protein domains, referred to as areas of interest (AOI's) to identify (Glimmer) and annotate (BLAST and PFAM) the smaller ORF's within (van Heel *et al.*, 2013b). DNA input analysis assisted in proper characterisation of the selected species. The data output generated clarified the significance of the properties studied. Conclusions about sequence conservation, gene associations and bacteriocin properties were drawn.

2.3.3 Comparative study

The frequency of different bacteriocin types in the genome mining data of the genera is summarised. The total amount of the types of bacteriocins are calculated and expressed as a percentage of the total amount of bacteriocins recorded within the phylum. The bacteriocin data summaries for the different phyla are then compared and contrasted.

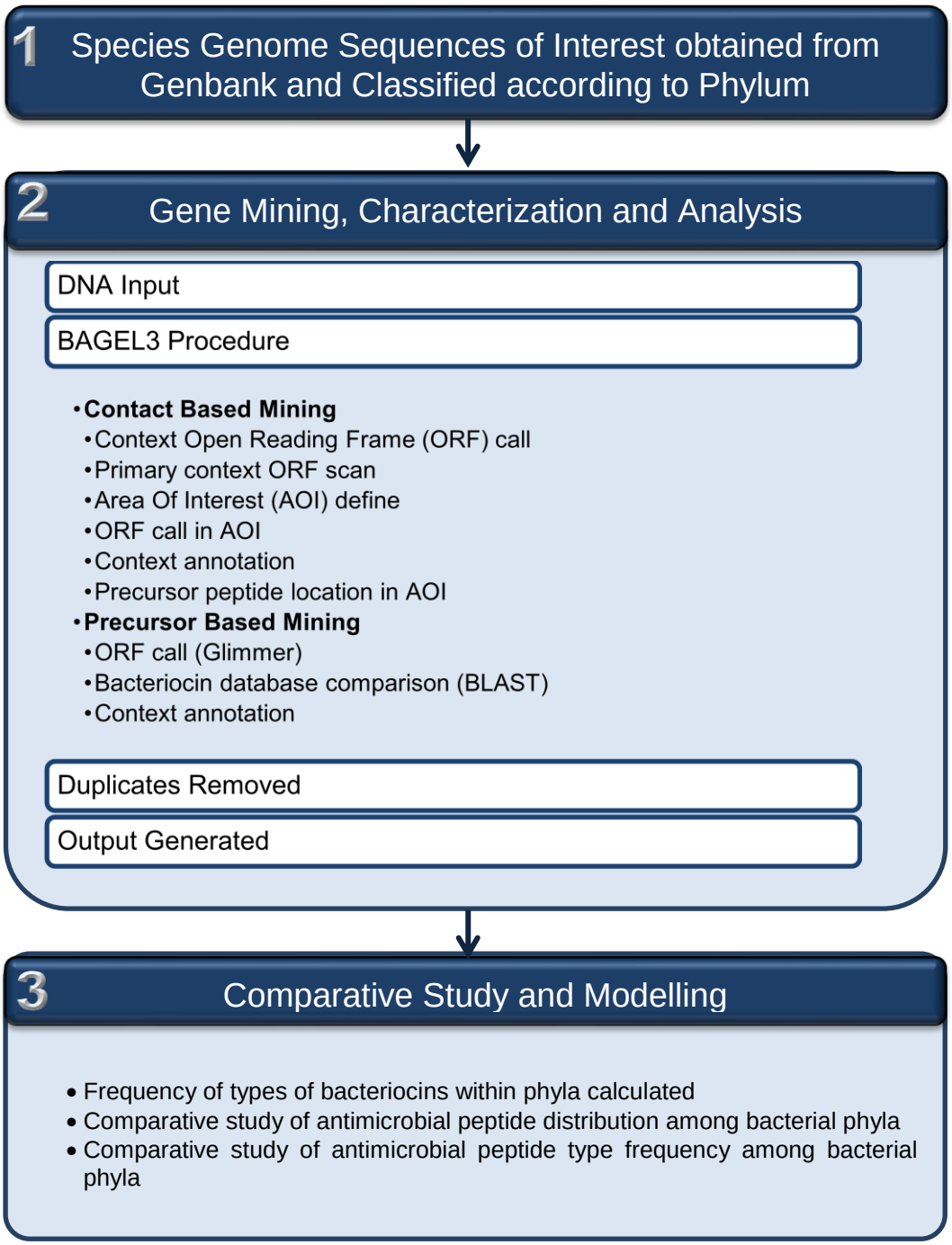

Figure 2.1: The diagram outlines the methods employed to accomplish the main objective; perform genomic mining for bacteriocins and analyse the frequency of different bacteriocin types among phyla

2.4 Results

A total of 932 genomes obtained from Genbank were analysed by the online genome mining tool, BAGEL3. Putative bacteriocin gene clusters that were identified were tabulated. The BAGEL3 folder containing resultant tables in excel format are available on Google drive at: https://drive.google.com/open?id=1PA0Mbrv3XYDXDi7IXDxSQ0hgXeCb1jYb

Examples of the bacteriocin gene cluster arrangement obtained from gene mining are represented in figure 2.2. The operons typically include a bacteriocin structural gene (green), immunity gene (red), transport/cleavage/modification gene (pink/blue) and undefined associated genes (grey).

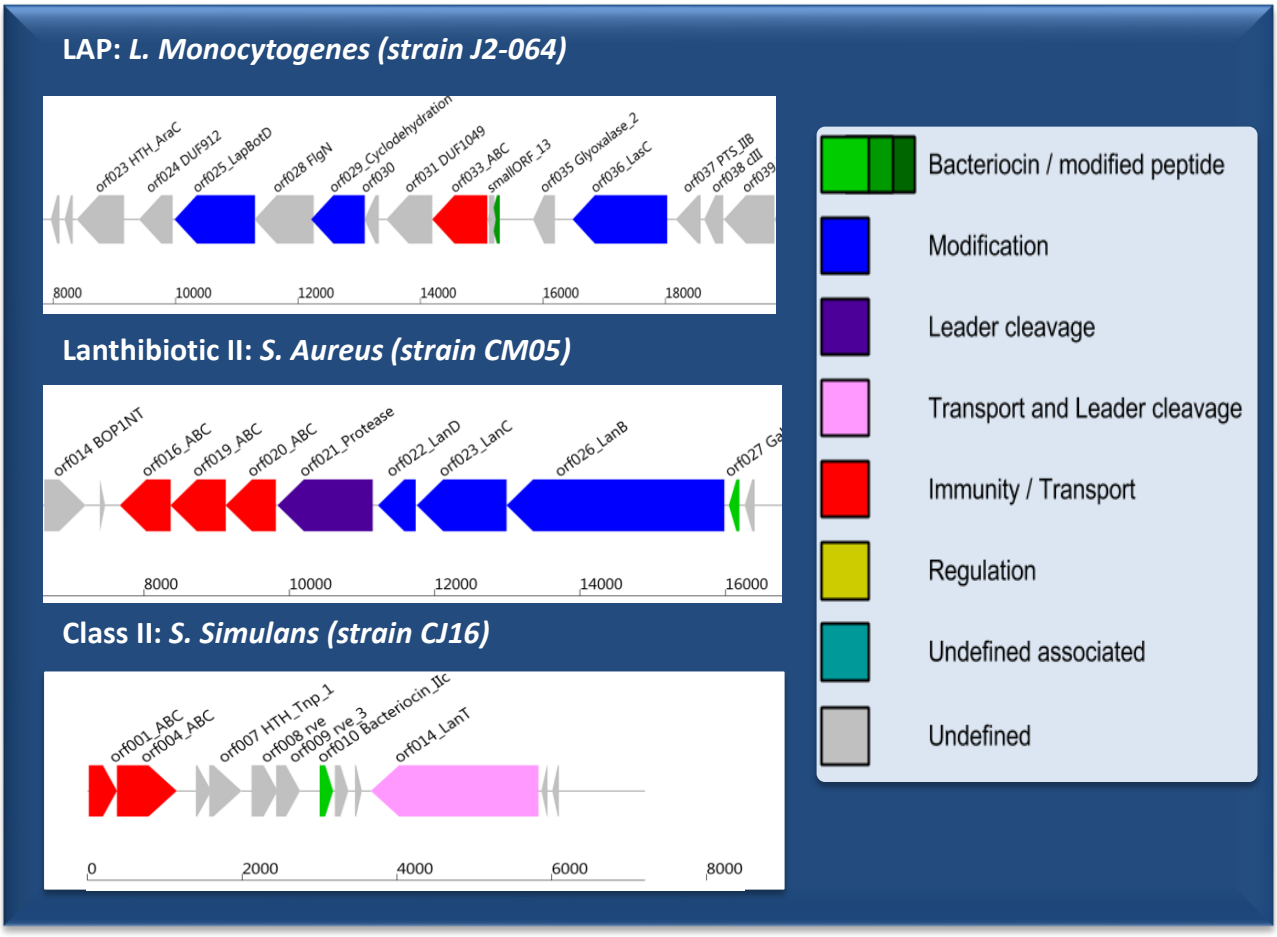

Figure 2.2: Conserved gene cluster organisation of 3 types of bacteriocins obtained from BAGEL3 genome mining

The gene cluster arrangement described in figure 2.2 is conserved among bacteriocin producers and among phyla. The amount of modification- and transport genes found, differed based on the type of bacteriocins. In the Actinobacteria phylum, 87 species were screened for the presence of bacteriocins. The types of bacteriocins found among these species are summarised in table 2.1. The main classes of bacteriocins present in the Actinobacteria phylum include Class III bacteriocins (18%), Head-to-tail cyclic peptides (12%) and class I Lanthibiotics (12%).

Table 2.1: Summary of bacteriocin class distribution among the Actinobacteria phylum. The amount of analysed species, producing species, bacteriocin class prevalence and the percentage that class members contribute to the total amount of bacteriocins analysed are included.

Total Actinobacteria Species	**87**	
Species with Bacteriocins	**48**	
Bacteriocin type	**Prevalence**	
Class III	21	18%
Head-to-tail cyclic	14	12%
Lanthibiotic I	14	12%
Sactipeptide	10	9%
Lanthibiotic III	10	9%
Lanthibiotic II	9	8%
Lasso peptide	9	8%
LAPs	8	7%
Thiopeptides	6	5%
Lanthibiotic IV	5	4%
Linaridin	5	4%
Bottromycin	2	2%
Class II	1	1%
Area of Interest	1	1%
Total Bacteriocins	**115**	

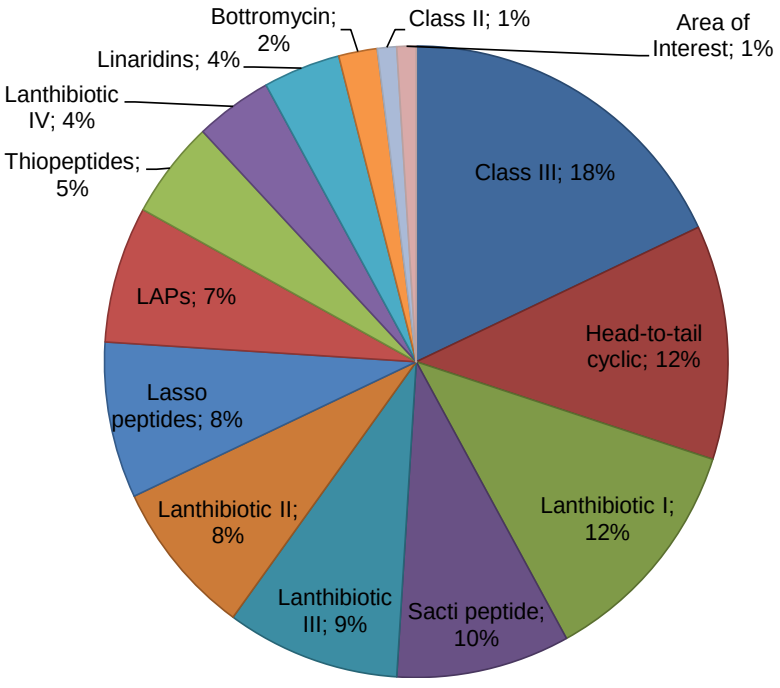

Figure 2.3: Pie chart depicting the percentage contribution of class members to the total amount of bacteriocins studied in the Actinobacteria phylum

A total of 115 bacteriocins were identified by genome mining within the Actinobacteria phylum. The 48 bacteriocin producing species yielded mostly Class III bacteriocins (18%), Head-to-tail cyclic peptides (12%) and class I Lanthibiotics (12%).

A summary of the bacteriocin types distributed among the species of the BV4 phylum are found in table 2.2. Most of the putative identified bacteriocins belong to Class III bacteriocins (31%), followed by class I Lanthibiotics (15%) and Head-to-tail cyclic peptides (13%).

Table 2.2: Summary of bacteriocin class distribution among the BV4 phylum. The amount of analysed species, producing species, bacteriocin class prevalence and the percentage that class members contribute to the total amount of bacteriocins analysed are included.

Total BV4 Species		**140**
Species with Bacteriocins		**50**
Bacteriocin type	**Prevalence**	
Class III	27	31%
Lanthibiotic I	13	15%
Head-to-tail cyclic	11	13%
Area of Interest	8	9%
Sactipeptide	7	8%
Class II	6	7%
Lasso peptide	4	5%
Lanthibiotic II	3	3%
LAP	3	3%
Glyocin	2	2%
Linaridin	2	2%
Bottromycin	1	1%
Cyanobactin	1	1%
Total Bacteriocins	**89**	

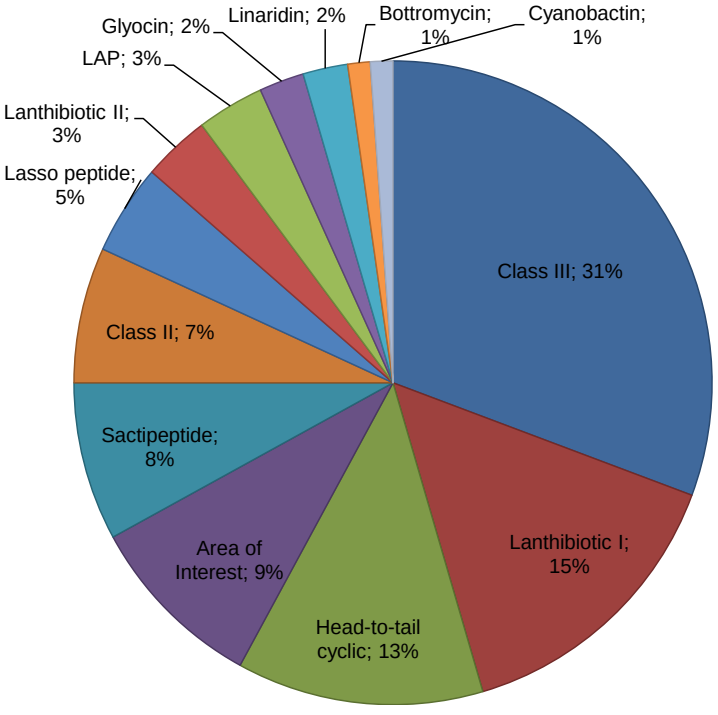

Figure 2.4: Pie chart depicting the percentage contribution of class members to the total amount of bacteriocins studied in the BV4 phylum

A total of 50 bacteriocin producing species were identified within the 140 BV4 species analysed. Class III bacteriocins are the most prevalent class of bacteriocins produced in the BV4 phylum at 31%. Other prevalent classes include class I Lanthibiotics (15%) and Head-to-tail cyclic peptides (12%).

Bacteriocins found within the phylum, Firmicutes were described in table 2.3. Based on BAGEL3 gene mining the main classes represented include Class II bacteriocins (19%) and Sactipeptides (19%).

Table 2.3: **Summary of bacteriocin distribution among the Firmicutes phylum. The amount of analysed species, producing species, bacteriocin class prevalence and the percentage that class members contribute to the total amount of bacteriocins analysed are included.**

Total Firmicutes Species	**157**	
Species with Bacteriocins	**87**	
Bacteriocin type	**Prevalence**	
Class II	38	19%
Sactipeptide	38	19%
Class III	29	15%
Head-to-tail cyclic	17	9%
Lanthibiotic II	16	8%
Lanthibiotic I	12	6%
Lasso peptide	10	5%
Area of Interest	8	4%
LAP	8	4%
Lanthibiotic III	5	3%
Lanthibiotic IV	5	3%
Bottromycin	3	2%
Thiopeptide	3	2%
Linaridin	2	1%
Unknown bacteriocin	2	1%
Total Bacteriocins	**196**	

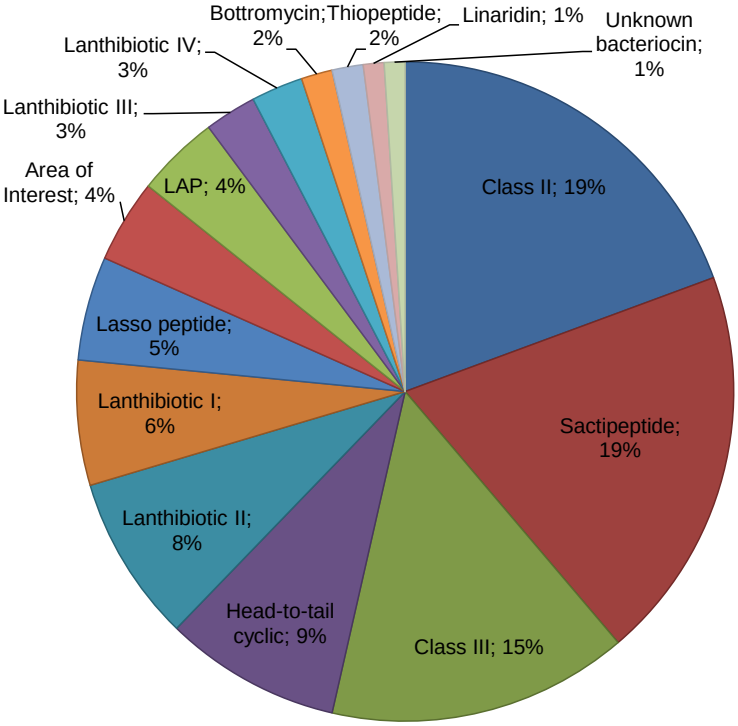

Figure 2.5: Pie chart depicting the percentage contribution of class members to the total amount of bacteriocins studied in the Firmicutes phylum

In the phylum, Firmicutes, Sactipeptides, Class II- and Class III bacteriocins are the most prevalent. More than 50% of the analysed genomes in this phylum potentially produce bacteriocins. The study found that on average 2 bacteriocins are produced for each analysed bacteriocin producing species within this phylum.

The study analysed 62 genomes from different species in the phylum Glidobacteria, summarised in table 2.4. Among the main classes of bacteriocins produced within this phylum are Head-to-tail cyclic bacteriocins (18%), Class III bacteriocins (14%), Sactipeptides (12%) and undefined areas of interest (10%).

Table 2.4: Summary of bacteriocin distribution among the Glidobacteria phylum. The amount of analysed species, producing species, bacteriocin class prevalence and the percentage that class members contribute to the total amount of bacteriocins analysed are included.

Total Glidobacteria Species	**62**	
Species with Bacteriocins	**42**	
Bacteriocin type	**Prevalence**	
Head-to-tail cyclic	14	18%
Class III	11	14%
Sactipeptide	9	12%
Area of Interest	8	10%
Cyanobactin	6	8%
Lanthibiotic II	6	8%
Lasso peptide	6	8%
Thiopeptide	6	8%
LAP	4	5%
Bottromycin	3	4%
Class II	2	3%
Microcin	1	1%
Unknown bacteriocin	1	1%
Total Bacteriocins	**77**	

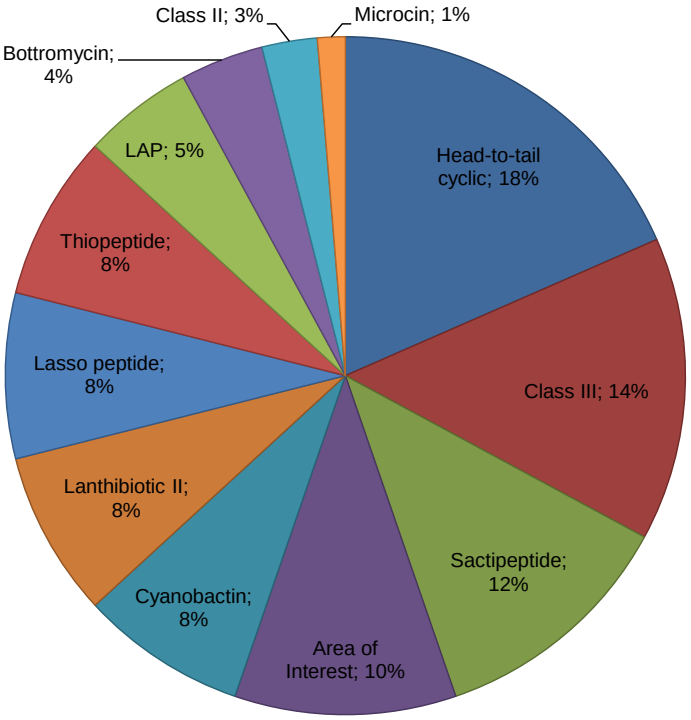

Figure 2.6: Pie chart depicting the percentage contribution of class members to the total amount of bacteriocins studied in the Glidobacteria phylum

Glidobacteria contained the highest ratio of bacteriocin producing species vs analysed species genomes. A 68% of the total Glidobacteria species produced bacteriocins (table 2.4) which include Head-to-tail cyclic peptides and Class III peptides.

The study revealed that within the phylum, Proteobacteria, most of the putative bacteriocins identified were Class III bacteriocins, as summarised in table 2.5.

Table 2.5: Summary of bacteriocin distribution among the Proteobacteria phylum. The amount of analysed species, producing species, bacteriocin class prevalence and the percentage that class members contribute to the total amount of bacteriocins analysed are included.

Total Proteobacteria Species	**373**	
Species with Bacteriocins	**170**	
Bacteriocin type	**Prevalence**	
Class III	105	38%
Microcin	40	14%
Bottromycin	39	14%
Lasso peptide	25	9%
Sactipeptide	23	8%
Lanthibiotic II	13	5%
LAP	11	4%
Head-to-tail cyclic	9	3%
Area of Interest	6	2%
Class II	4	1%
Cyanobactin	2	1%
Unknown bacteriocin	1	0%
Lanthibiotic IV	1	0%
Total Bacteriocins	**280**	

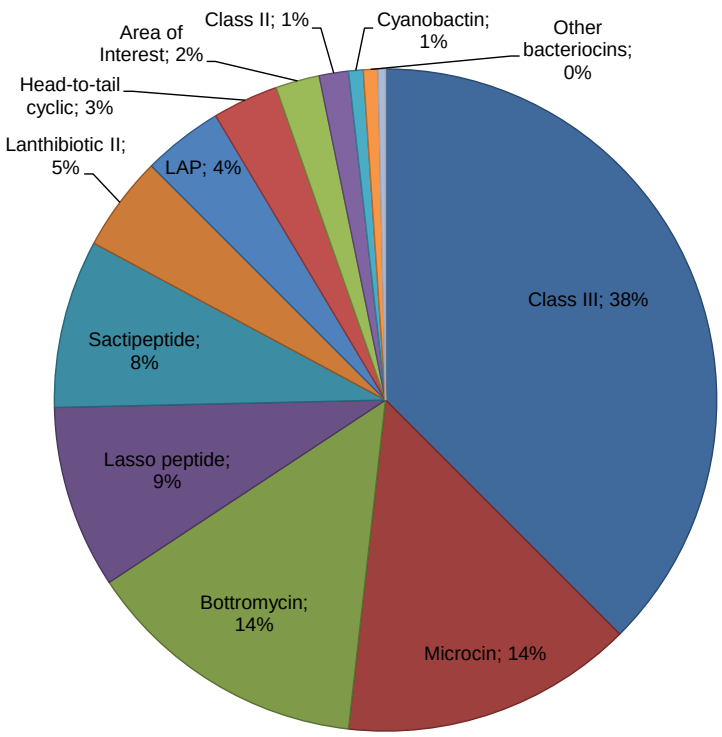

Figure 2.7: Pie chart depicting the percentage contribution of class members to the total amount of bacteriocins studied in the Proteobacteria phylum

The Proteobacteria phylum contained the highest amount of species analysed, consisting of 373 genomes. A total of 280 bacteriocins were identified of which 38% represent the Class III bacteriocins. The Microcins and Lasso peptides make up a large number when compared to the same classes in the other phyla.

For interest's sake, this study also analysed genomes represented form the archaea kingdom, summarised in table 2.6. The main classes present were identified as Sactipeptides (22%), Head-to-tail cyclic peptides (17%), Bottromycins (11%) and class II Lanthibiotics (11%).

Table 2.6: Summary of bacteriocin distribution among analysed archaea species. The amount of analysed species, producing species, bacteriocin class prevalence and the percentage that class members contribute to the total amount of bacteriocins analysed are included.

Total Archaea Species		**102**
Species with Bacteriocins		**41**
Sactipeptide	14	22%
Head-to-Tail Cyclic	11	17%
Bottromycin	7	11%
Lanthibiotic II	7	11%
Class II	6	9%
Class III	6	9%
Lasso Peptide	4	6%
Lanthibiotic I	3	5%
LAP	3	5%
Area of Interest	1	2%
Linaridin	1	2%
Microcin	1	2%
Total Bacteriocins	**64**	

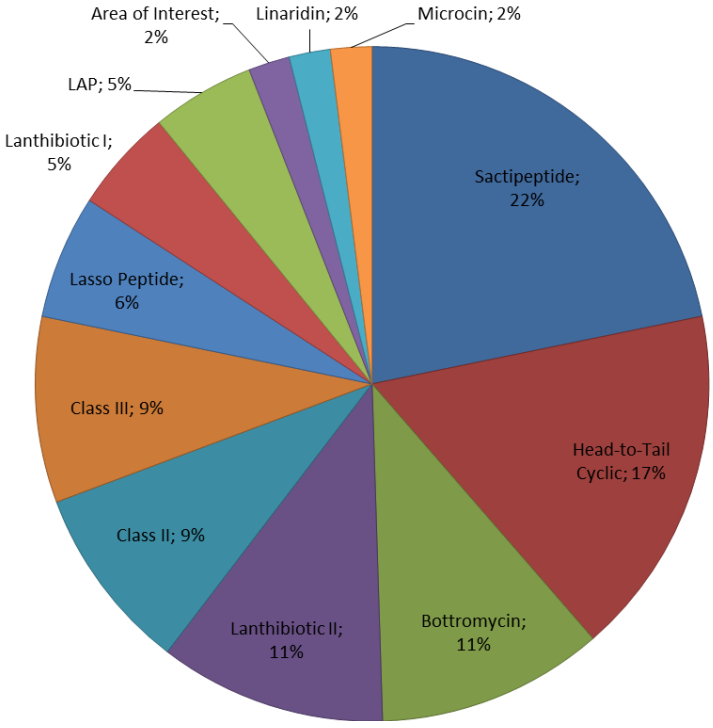

Figure 2.8: Pie chart depicting the percentage contribution of class members to the total amount of bacteriocins studied in the archaea kingdom

Less than half of the 102 analysed archaeal genomes are predicted to produce bacteriocins. Bacteriocin producing species typically only produced 1 bacteriocin. Sactipeptides (22%), Head-to-tail cyclic bacteriocins (17%) and Bottromycins (11%) are the most prevalent bacteriocins produced within this kingdom.

In an attempt to look at the type and distribution of bacteriocin producers across all the phyla analysed, table 2.7 was constructed. This revealed that at least 25% of identified bacteriocins belonged to Class III.

Table 2.7: **Summary of bacteriocin distribution among all species in this study. The amount of analysed species, producing species, bacteriocin class prevalence and the percentage that class members contribute to the total amount of bacteriocins analysed are included.**

Total Species		**921**
Species with Bacteriocins		**438**
Bacteriocin type	**Prevalence**	
Class III	199	24%
Sactipeptide	87	11%
Head-to-tail cyclic	72	9%
Class II	57	7%
Bottromycin	55	7%
Lasso peptide	54	7%
Lanthibiotic II	51	6%
LAP	48	6%
Lanthibiotic I	42	5%
Microcin	41	5%
Area of Interest	32	4%
Cyanobactin	20	2%
Lanthibiotic III	16	2%
Thiopeptide	15	2%
Lanthibiotic IV	12	1%
Linaridin	9	1%
Glyocin	5	1%
Unknown bacteriocin	5	1%
Total Bacteriocins	**820**	

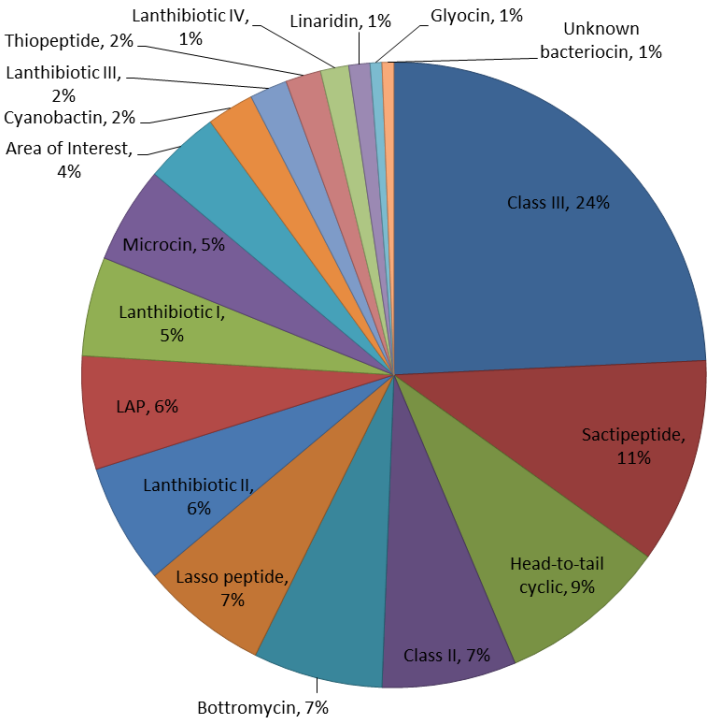

Figure 2.9: Pie chart depicting the percentage contribution of class members to the total amount of bacteriocins studied in all species of interest across all phyla

As can be seen from table 2.7 and figure 2.9, genome mining of 921 bacterial and archaeal genomes yielded 819 putative bacteriocins. From this, Class III bacteriocins (25%) and Sactipeptides (13%) were the most prevalent.

2.5 Discussion

The 921 genomes selected to represent all archaeal and bacterial phyla were obtained from the NCBI database (35 726 data entries). Bacteriocin mining was performed by the BAGEL3 software tool and the distribution of bacteriocins statistically determined by Microsoft Excel 2010. For this study, bacteriocin novelty or uncharacterised bacteriocin was defined as the absence of a literature reference in the BAGEL3 output file. A total of 819 bacteriocins were found in 431 species.

Several other gene mining studies conducted compared bacteriocin distribution among the same genera and class distribution in specific structures e.g. human gut studies, but a comparison of the distribution of bacteriocins within different archaeal- and bacterial phyla has not been studied before (Drissi *et al.*, 2015, Walsh *et al.*, 2015, Wang *et al.*, 2011). In Eukarya which was used as a negative control, no bacteriocins were identified. The very nature of bacteriocins is that they are produced by bacteria, making Eukarya the perfect control to test also the validity of the BAGEL3 software. Among the Actinobacteria species, it was identified that ~55% of species contain bacteriocin genes. This phylum, together with Firmicutes, also contains the only members identified in this study to produce class III Lanthibiotics. Class III bacteriocins (21%), Head-to-tail cyclic bacteriocins (14%) and class I Lanthibiotics (14%) contribute ~49% of bacteriocins present in this phylum (table 2.1). From total of 44 of the putative novel class I Lanthibiotics identified, as previously described in table 2.2, almost a third (31%) of bacteriocins synthesised by BV4 bacteria are of the Class III type. Head-to-tail cyclic bacteriocins (15%) and class I Lanthibiotics are also prevalent (13%). Of 101 putative Sactipeptides identified in this study, 38 are produced by bacteriocin producing Firmicutes members. Mostly Sactipeptides (19%) or Class II bacteriocins (19%) are present in the Firmicutes phylum as seen in table 2.3. All Glyocin bacteriocin producing species belong to the BV4 phylum. According to BAGEL3 results, Glidobacteria members have the highest percentage of species containing bacteriocin genes (~68%). mainly Head-to-tail cyclic bacteriocins (14%), Class III bacteriocins (11%) and Sactipeptides (9%), seen in table 2.4. The Proteobacteria had the highest amount of analysed species with Class III comprising ~40% of identified bacteriocins (table 2.5).

This is also the phylum that harbours the widest variety of bacteriocins, excluding Glyocins, class III Lanthibiotics, Linaridins and Thiopeptides.

In the Achaea, 41 species were identified to contain bacteriocin genes comprising of 64 bacteriocins, see table 2.6. The main classes of bacteriocins produced are Head-to-tail cyclic bacteriocins (17%) and Bottromycins (7%). Sactipeptides (22%) is the most prevalent type of bacteriocin in this phylum. A total of 31 undefined areas of interest and 2 unknown bacteriocins were identified by BAGEL3, which could prove intriguing. Other software mining tools and algorithms failed to classify these genes, therefore physical characterisation may yet yield novel classes of bacteriocins.

Natural products, as opposed to combinational synthesised pharmaceuticals, show more significant biological activity against multidrug-resistant pathogens and has thus became more attractive for further investigation (Bode and Muller, 2005). Genome mining software provided a valuable tool for discovering putative novel bacteriocins. Due to previously identified short-comings with BAGEL2 and similar software, developers allowed BAGEL3 to analyse data independent of predicted open reading frames and with a parallel direct and indirect procedure. This independent analysis technique not only broadens the inclusion of possible candidates, but also decreases manual evaluation of results (van Heel *et al.*, 2013b). The new version of the software, BAGEL4, was launched on the 1st February 2018 which may further refine the results reported on in this study.

Chapter 3: Physiochemical classification and phylogeny

3.1 Research aim

This section of the project aims to explore physiochemical properties and evolutionary relationships of the identified putative novel bacteriocins and the respective producing species.

3.2 Research objectives

1) To determine BAGEL3 bacteriocin identity (BLAST)
2) To obtain physiochemical properties of novel bacteriocins (Protparam)
3) To predict 3D protein model using homology modelling (SwissModel)
4) To align bacteriocin sequences (Clustal-Omega)
5) To infer phylogeny of the species with bacteriocins (MEGA X)

3.3 Experimental design and methodology

Genomes obtained from an online database are mined through an online server. The methods and software used to obtain and analyse these genomes are summarized in figure 3.1 below.

3.3.1 Identification of putative bacteriocins (BLAST)

The identity of putative bacteriocins, identified by BAGEL3 (van Heel *et al.*, 2013a), was obtained by the Basic Local Alignment Search Tool for Proteins (BLASTP) software (Altschul *et al.*, 1990, Madden *et al.*, 1996, States and Gish, 1994). The BLASTP setup phase entails capturing the user input (query) and search parameters where after the preliminary search and trace back steps are initiated. In the preliminary search, the software selects fixed length sequences (called 'words') to search for within the database (called 'subject sequences'). Each entry that contains the 'word' of interest is assigned a match-quality score (with and without gaps). These scores determine which entries are analysed in the final trace back phase for insertions and deletions, before the software generates the output (Madden, 2013).

3.3.2 Determination of physiochemical properties (Protparam)

Physiochemical properties of the novel putative bacteriocins were explored with Protparam software (Gasteiger *et al.*, 2005). The molecular weight of the input peptide is determined by adding up the respective isotopic masses without considering post-translational modifications. The isoelectric point (pI) is the pH where the peptide will have a net zero charge. The pI is calculated as the average of the acid dissociation constant on logarithmic scale (pKa) where the molecule has a net charge of +1 and -1 respectively. The software also reports on the charged amino acid composition. The extinction coefficient estimates the amount of light a protein absorbs during spectrophotometry analysis. It is determined by the product of molar extinction coefficients and the amounts of tyrosine, tryptophan and cysteine in the protein, respectively. Absorbance of the peptide is determined by dividing the extinction coefficient by the molecular weight. A more accurate coefficient can be determined experimentally, as the secondary- and tertiary structure is not taken into account by the software and it may influence this value.

In Vivo estimated half- life within humans, yeast and *E. coli* are predicted by investigating the N-terminal amino acid of the input sequence. The process of ubiquitin-mediated proteolytic degradation is highly influenced by the N-terminal amino acid. Predicted instability within an aqueous solution indicated by the instability index. An instability index below 40 classifies the peptide of interest as being stable. The aliphatic index indicates the relative volume occupied by the aliphatic side chains of residues. The grand average of hydropathy (GRAVY) is calculated by the average hydropathy of all the amino acids. GRAVY values predicts hydrophilicity (<0) or hydrophobicity (>0), indicating whether the nature of the protein is globular (hydrophilic) or membranous (hydrophobic) (Gasteiger *et al.*, 2005).

3.3.3 Predicted structure: homology modelling (SwissModel)

Templates similar to the input sequences mentioned above are obtained from BLAST and HHblit. Templates are ranked and top candidates compared to determine whether different conformational states or cover regions are present. Continuous Automated Model Evaluation (CAMEO) evaluates the generated multiple templates and output models built accordingly(Marco Biasini *et al.*, 2014).

3.3.4 Bacteriocin gene sequence alignment (Clustal Omega & Weblogo)

The Clustal Omega software (Sievers *et al.*, 2011) was used to align the various bacteriocin sequences within phyla. Pairwise alignment is performed by the k-tuple method, followed by sequence clustering using the mBed and k-means sequence clustering methods. The aforementioned clustering scores are converted to distance scores and the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) method used to construct a guide tree. Progressive alignment is performed by HHalign package which yields the multi sequence alignment output (Daugelaite *et al.*, 2013). The Weblogo software highlights similarities and differences of multiple sequence alignments and is used to visualise the sequence alignment result (Doerks *et al.*, 2002).

3.3.5 Phylogenetic tree modelling (MEGA X)

Phylogenetic trees were inferred from the bacteriocin gene sequences provided by BAGEL3 and the 16s ribosomal DNA sequences from producer organisms obtained from NCBI respectively.

Phylogenetic trees for 16s ribosomal ribonucleic acid (rRNA) nucleotide sequences and novel bacteriocin protein sequences were generated with 100 bootstrap repeats by MEGA X (Brunette *et al.*, 2015). A 16s rRNA maximum likelihood tree was built according to the Jukes-Cantor model following the nearest-neighbour-interchange (NNI) heuristic method. All codon positions were selected to form part of the data subset, patterns had uniform rates and 15 threads selected. The maximum likelihood tree for the bacteriocin sequences was inferred using the Poisson correction model (Brunette *et al.*, 2015).

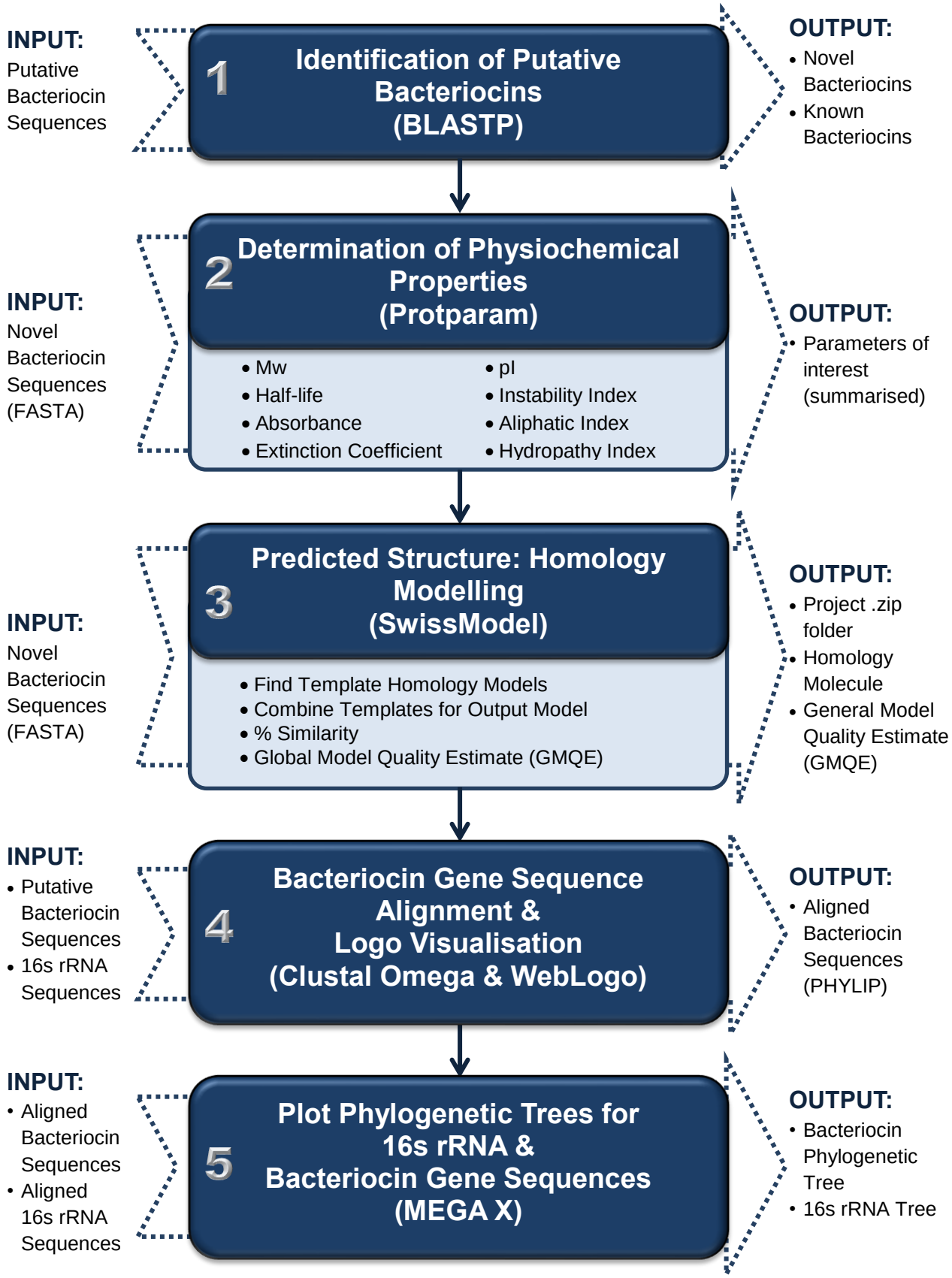

Figure 3.1: Schematic presentation of methods used in this study to predict bacteriocin identity and analyse the physiochemical properties and evolutionary relationships of the novel bacteriocins found

3.4 Results

3.4.1 Genome mining and physiochemical classification

Figure 3.2 displays the BAGEL3 output of an example strain. The operons can include bacteriocin structural genes (green), immunity (red), transport (red/pink), cleavage (pink/purple), modification gene (blue), undefined associated genes (tile) and undefined genes (grey).

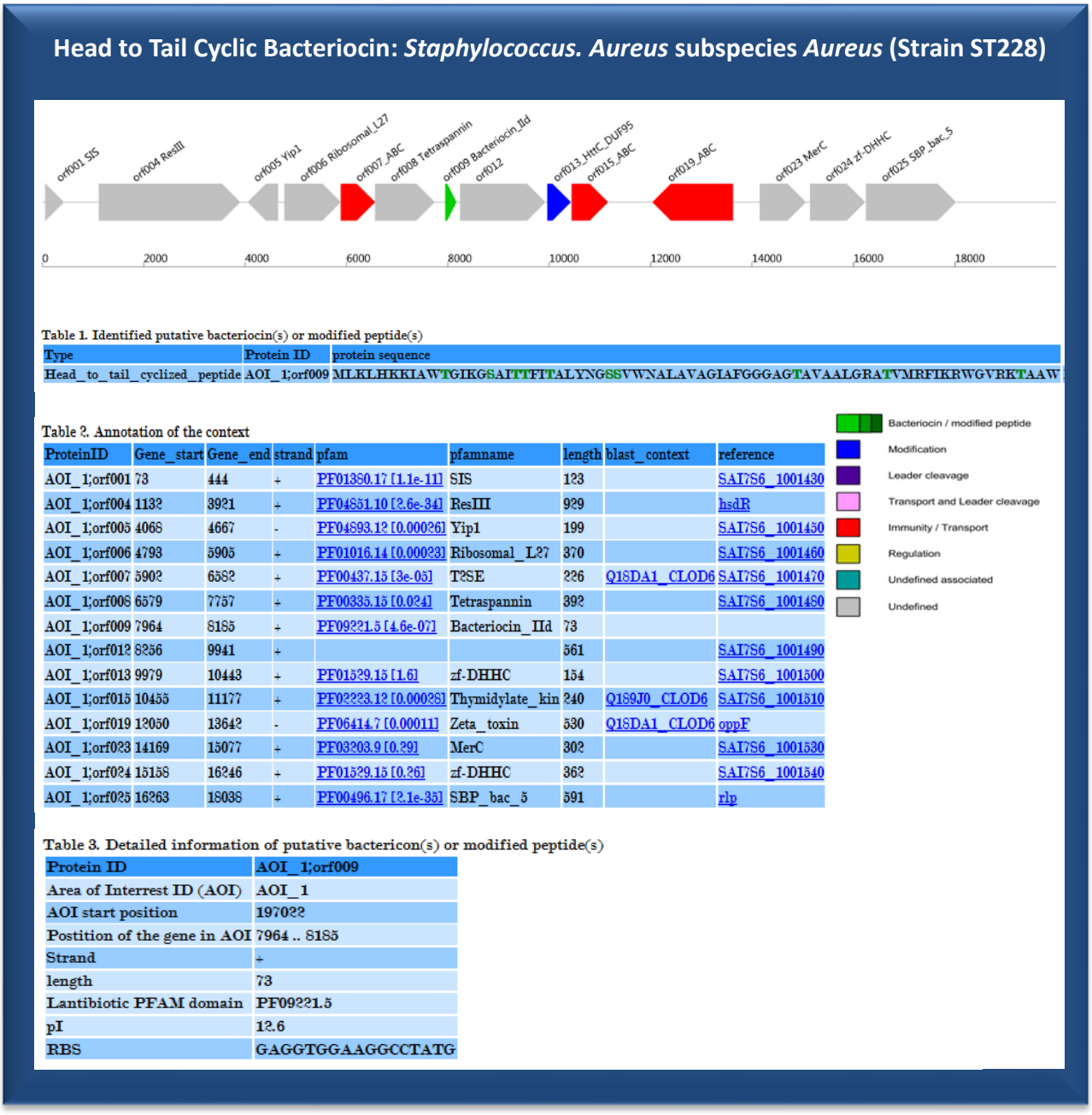

Figure 3.2: BAGEL3 output of *Staphylococcus aureus* strain ST228, illustrating the gene cluster arrangement within this operon after analysis. Included within this operon is a bacteriocin structural gene (green), immunity genes (red), a modification gene (blue) and undefined genes (grey)

BAGEL3 software locates putative bacteriocins gene elements from a DNA sequence query. The graphical output and accompanying tables show the bacteriocin type and gene sequence of the putative bacteriocin (figure 3.2). In figure 3.2 table 1 lists the type and the sequence of identified putative bacteriocins. Further, table 2 the open reading frames in the area of interest are annotated.

These annotations include the gene location within the area of interest, the protein family (pfam) entry description and associated literature. The last table, table 3, provides details on the identified bacteriocin including the bacteriocin sequence, length and putative isoelectric point.

Genome mining results of all analysed bacterial genomes and control genomes (Archaea and Eukaryotes) are available on Google drive at: https://drive.google.com/open?id=1qLWwcJMxkC9sFtNIXWGAb5OLjjD84Dnm
and
https://drive.google.com/open?id=1ZWb4GrjjocX_WbdNAdTHoxfG9W38zGXy
The BLASTP software employs local alignment search algorithms to determine the identity of input bacteriocin sequences. The graphical output describes the top rated identity results in decreasing fashion (Figure 3.3). A protein structure and function summary with appropriate references are tabulated for the top rated hit.

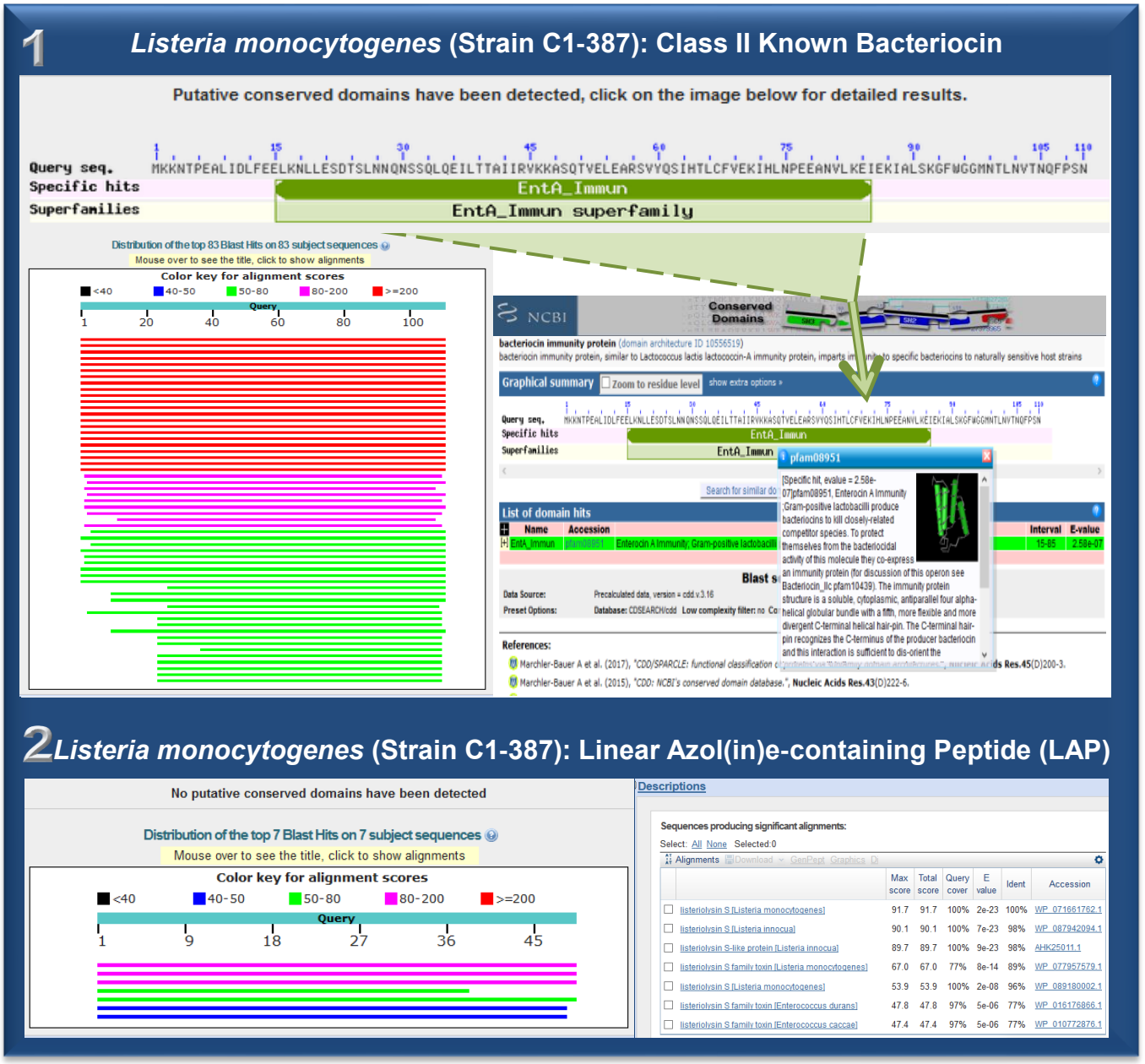

Figure 3.3: Bacteriocin identification through BLAST analysis of an example species as performed by the BLASTP algorithm. (1) The Class II bacteriocin detected by BAGEL3 from *Listeria monocytogenes* was identified as the ENtA-Immun protein with an alignment score of ≥200. (2) The *L. monocytotgenes* LAP yielded no BLASTP results and was thus identified as novel

Bacteriocin identified as novel as the above example in figure 3.3, allowed for selection of unknown sequences for further analyses. The physiochemical properties of the novel putative bacteriocins were analysed by Protparam software.

The novel putative bacteriocins, identified by genome mining, were named and their physiochemical properties calculated using the Protparam software. Physiochemical properties that were analysed include the peptide MW, pI, hydropathy index, aliphatic index and the predicted stability within yeast-, mammalian- and bacterial cells inferred from the stability index. Data is summarized in tables 3.1, 3.2 and 3.3, and contain the number of species that fits the criteria for the described categories.

Included in tables 3.1, 3.2 and 3.3 is a summary of the amount of Actinobacteria, Firmicutes and Proteobacteria species, their classification within each indicated category with regards to bacteriocin type and physiochemical properties.

Actinobacteria: The 114 mined bacteriocins in the Actinobacteria phylum, resulted in 75% being identified as novel (table 3.1). Most of the novel bacteriocins belonged either to the Head-to-tail cyclic class or class III Lanthipeptides. All the identified LAPs and Lasso peptides are novel. Most identified bacteriocins of the Actinobacteria phylum are unstable and hydrophilic in nature. pI's of the bacteriocins ranged from 2.92 to 12.35, with the majority being basic. The average molecular weight was also used as a criterion for comparison. The majority of identified bacteriocins were larger than the average MW. A Head-to-tail cyclic bacteriocin, a Lanthipeptide class IV and 2 Sactipeptides that were identified will prove challenging to analyse experimentally, since their *in vitro* or *in vivo* half-life is predicted to be below 5.5 hours and 3 minutes, respectively. Complete analysis of the Actinobacteria species used in this study is available on Google drive at: https://drive.google.com/open?id=1y6vHtOq9I1MDxjhKEp2YG2DXRHjuD6x6

Firmicutes: A total of 196 bacteriocins were identified with genome mining in the Firmicutes phylum (table 3.2). Class II bacteriocins contribute 20% of the 120 total novel identified bacteriocins. In contrast to the Actinobacteria, most of these peptides are stable and smaller than the average mass of the specific class of bacteriocin. Exceptionally, Sactipeptides are the only class that contain more unstable than stable peptides. Similar to the Actinobacteria, the majority of bacteriocins were hydrophilic and the pI's mostly basic. All identified bacteriocins of the Firmicutes phylum are predicted to be able to be studied *in vitro* (±30 hours half-life) and *in vivo* (10hours-20hours half-life in yeast and *E.coli* respectively).

Complete analysis of the Firmicutes species is available on Google drive at: https://drive.google.com/open?id=1ikVRjT03Y-DjWbqdheE0Q4Gqv7EhFvQL

Proteobacteria: Of the 373 analysed Proteobacteria species analysed, 170 species potentially produce bacteriocins. The bacteriocins mined resulted in 71% being novel with the majority belonging to Class III bacteriocins as seen in table 3.3. The identified Bottromycins, Cyanobactins, LAPs and Microcins was only associated with this phylum. Similar to the other phyla, the majority of bacteriocins tend to have a hydrophilic nature and the pI's are below 7. The class molecular weight average was higher than most members of this bacteriocin class (bacteriocin class members were small). Unlike the Actinobacteria and Firmicutes, the Proteobacteria phylum's bacteriocins have an equal distribution with regards to their stability. The instability associated with 10% of the novel bacteriocins prohibits *in vivo* and *in vitro* studies, but most can be analysed with spectrophotometry. Complete analysis of the Firmicutes species is available on Google drive at: https://drive.google.com/open?id=1gDS0uVl2yxtP26gJVg2Sb_OskZcZU8N1

Table 3.1 Summary of the amount of bacteriocin- producing Actinobacteria species grouped according to physiochemical properties.

Bacteriocin Class	Identification		Stability		Molecular Weight		Isoelectric Point		Hydrophobicity	
	Known	Novel	Stable	Unstable	Small	Large	Acidic	Basic	Yes	No
Bottromycin	0	2	0	3	2	1	2	1	0	3
Head-to-Tail Cyclic	1	13	8	5	9	4	3	10	7	6
Class II	0	1	0	0	0	0	0	1	0	0
Class III	20	1	0	1	0	0	0	1	0	1
Lanthipeptide Class I	2	12	1	9	5	5	8	2	1	9
Lanthipeptide Class II	1	8	2	8	6	6	5	7	3	9
Lanthipeptide Class III	6	4	2	10	8	4	9	3	5	7
Lanthipeptide Class IV	0	5	1	4	5	3	0	3	1	4
LAP	0	8	2	6	4	4	2	6	3	5
Lasso Peptide	2	7	2	5	4	3	6	1	1	6
Linaridin	0	5	1	4	4	1	5	0	3	2
Sactipeptide	2	8	1	7	4	4	2	6	2	6
Thiopeptide	0	6	0	5	3	2	1	4	0	5
	34	**80**	21	**67**	55	**37**	44	**45**	26	**64**

Table 3.2 Summary of the amount of bacteriocin- producing Firmicutes species grouped according to physiochemical properties.

Bacteriocin Class	Identification		Stability		Molecular Weight		Isoelectric Point		Hydrophobicity	
	Known	Novel	Stable	Unstable	Small	Large	Acidic	Basic	Yes	No
Bottromycin	0	2	0	2	1	1	0	2	1	1
Head-to-Tail Cyclic	3	17	12	5	15	2	4	13	13	4
Class I	0	1	1	0	0	1	0	1	1	0
Class II	16	23	21	2	11	12	12	11	10	13
Class III	20	10	8	2	6	4	6	4	2	8
Lanthipeptide I	2	7	6	1	6	5	7	4	3	8
Lanthipeptide II	6	8	5	3	5	3	5	3	2	6
Lanthipeptide IV	0	1	1	0	1	0	0	1	0	1
LAP	1	9	5	4	5	4	4	6	4	5
Lasso Peptide	3	6	5	1	3	3	4	2	2	4
Sactipeptide	10	34	14	20	21	13	16	18	18	16
Unknown Bacteriocin	0	2	2	0	1	1	0	2	2	0
	61	**120**	**80**	40	**75**	49	58	**67**	58	**66**

Table 3.3 Summary of the amount of bacteriocin- producing Proteobacteria species grouped according to physiochemical properties.

Bacteriocin Class	Identification		Stability		Molecular Weight		Isoelectric Point		Hydrophobicity	
	Known	Novel	Stable	Unstable	Small	Large	Acidic	Basic	Yes	No
Bottromycin	0	41	16	25	35	6	7	34	18	23
Cyanobactin	0	2	0	2	1	1	0	2	1	1
Head-to-Tail Cyclic	1	7	4	3	5	2	2	5	3	4
Class II	1	2	1	1	1	1	1	1	1	1
Class III	60	46	34	12	17	29	5	41	6	40
Lanthipeptide Class I	0	1	0	1	1	0	0	1	0	1
Lanthipeptide Class II	2	12	6	6	7	5	7	5	2	10
Lanthipeptide Class IV	0	1	1	0	1	0	0	1	0	1
LAP	0	9	2	7	7	2	4	5	4	5
Lasso Peptide	3	22	13	9	14	8	13	9	7	15
Micrcocin	0	40	17	23	24	16	12	28	18	22
Sacti Peptide	2	17	7	10	7	10	5	12	5	12
Unknown Bacteriocin	0	1	0	1	0	1	0	1	0	1
	69	**201**	**101**	100	**120**	81	56	**145**	65	**136**

3.4.2 Homology modelling

As part of the characterization of each novel putative bacteriocin, sequences were subjected to homology modelling to determine possible 3D structure (Figure 3.4).

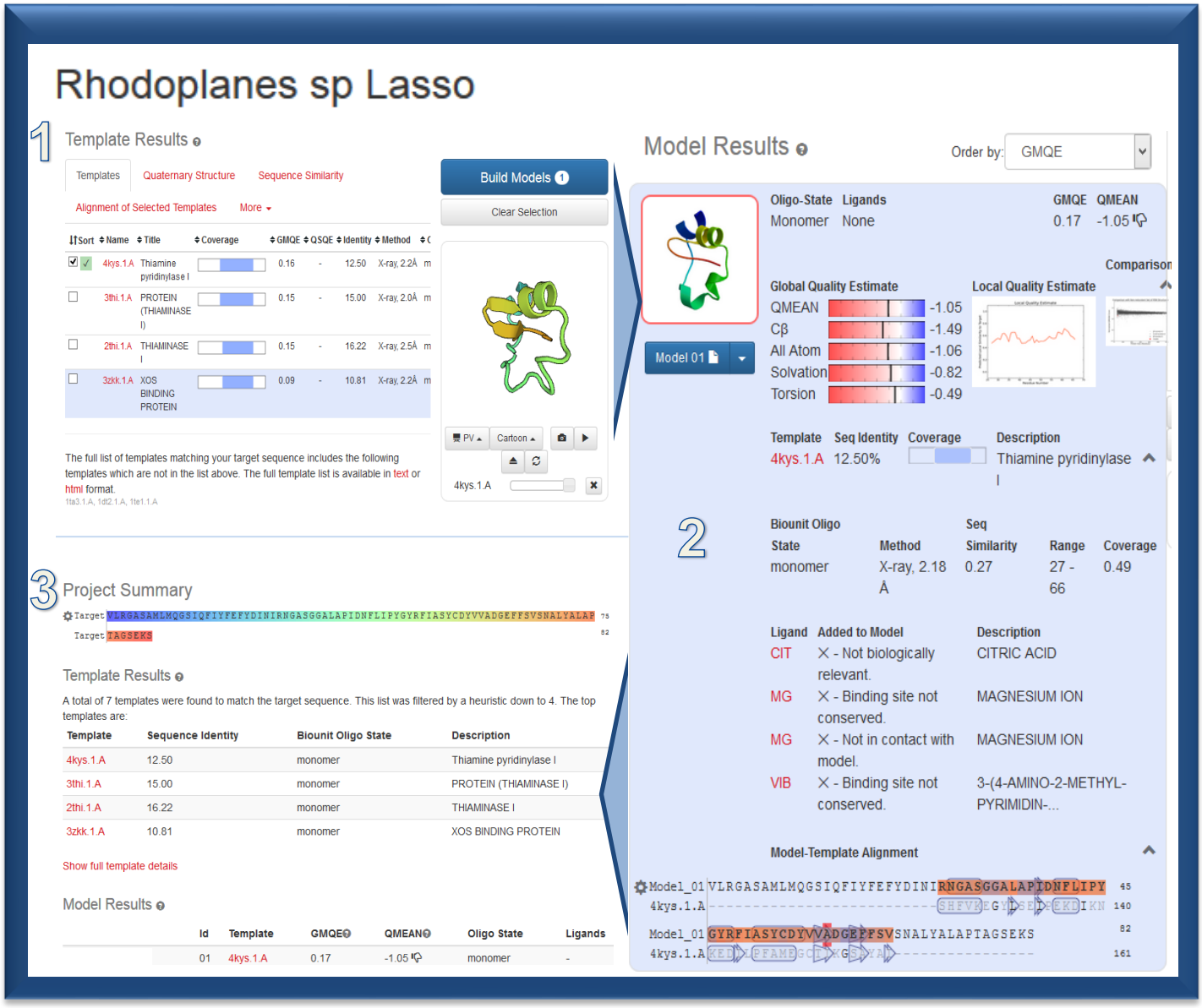

Figure 3.4: Typical homology modelling output when using the Swissmodel software. The figure shows the resultant templates, models and summary

The SwissModel software generates a list of templates of interest, a list of putative models and a project summary (Figure 3.4). Templates used for homology modelling of the query sequence are listed on the 'Template Results' page. Swissmodel estimates the quality of the predicted templates by assigning each template a global

model quality estimate (GMQE) score. The calculated global quality estimate components are listed for the putative models that were inferred from the listed templates. These elements are summarised in the 'Project Summary'.

Documents containing homology modelling results of all analysed phyla are compiled and available on Google drive. Modelling results for the novel putative bacteriocins produced by Actinobacteria species are available on Google drive at: https://drive.google.com/open?id=1nlGOT5-bVCckBA5C5_mKTkGZYSwA7yWt. Homology modelling results for bacteriocins produced by the Firmicutes are available on Google drive at https://drive.google.com/open?id=1_BWRyWQaJ3yHw5eNlUqHplVkAl5y_jQT. Predicted homology models for bacteriocins produced by the Proteobacteria are available at https://drive.google.com/open?id=1bPYudw6JPW1HnzBn62U-lgZVGnV_hphy.

Putative models were identified for 72.5% of the novel putative bacteriocins in the Actinobacteria phylum (table 3.4). GMQE scores (0-1) are considered favourable if the value is closer to 1. GMQE scores ranged between 0.09-0.82 while sequence coverage was below 50%, except for arobacin A, cellucin A and michigacin A which had higher sequence coverage.

The novel putative bacteriocins mined from Firmicutes were subjected to homology modelling (table 3.5). In the phylum 165 bacteriocins were analysed for which 75.8% homology models could be generated. Firmicutes bacteriocins had the best average GMQE score, 0.66, among all the phyla analysed. The average sequence coverage was 24.9%, with a maximum coverage of 80%.

In the Proteobacteria phylum, models were successfully generated for 75.8% of analysed species (table 3.6). The sequence coverage range of these models (8%-99.71%) was higher than that of the Actinobacteria (7.55%-65.71%) and Firmicutes (6.45%-80%). Although GMQE score averages were low 0.24, only 30 of the analysed 373 bacteriocins in this phylum do not show homology to the database models.

Table 3.4: Homology models of novel putative bacteriocins within the Actinobacteria phylum. Each model is accompanied by a global mean quality estimate (GMQE) and percentage similarity of the suggested model to the template model.

Actinobacteria Species	Bacteriocin		Homology Molecule Description	GMQE	%Similarity
A. phytoseiuli	Aphytocin	A	Regulator of ribonuclease activity A protein 1	0,27	13,89%
	Aphytocin	B	Cellulosome enzyme, Dockerin type I	0,29	34,48%
	Aphytocin	C	-	-	-
	Aphytocin	D	-	-	-
A. terrae	Aterraecin	A	Abscisic acid receptor PYL3	0,15	19,35%
	Aterraecin	B	Lipoprotein lprG	0,34	20,41%
			Interleukin-7 receptor subunit alpha	0,20	23,08%
			Eukaryotic translation initiation factor 4 gamma 2	0,17	25,00%
	Aterraecin	C	Protein (Chalone synthase)	0,21	13,16%
A. robiniae	Arobicin	A	Designed helical repeat protein	0,59	53,85%
	Arobicin	B	-	-	-
	Arobicin	C	-	-	-
	Arobicin	D	-	-	-
	Arobicin	E	Baculoviral IAP 4 and 7	0,21	25,93%
			Septum formation, penicillin binding protein 3, peptidoglycan synthetase	0,15	25,93%
	Arobicin	F	Mannan-binding lectin serine protease 1	0,37	15,38%
	Arobicin	G	-	-	-
Angustibacter sp	Angucin	A	25kDa structural protein VP25	0,49	40,00%
			Hypothetical protein TTHA1281	0,42	27,03%
			HOBA	0,36	26,83%
	Angucin	B	P1	0,42	37,50%
C. acidiphila	Cacidicin	A	-	-	-
	Cacidicin	B	65 kDa Yes-associated protein	0,26	22,22%
	Cacidicin	C	MSin3A-binding protein	0,15	25,00%
	Cacidicin	D	5'-amp-activated protein kinase catalytic subunit alpha-1	0,11	17,50%
	Cacidicin	E	-	-	-
C. acidiphila	Cacidicin	F	-	-	-
	Cacidicin	G	General negative regulator of transcription subunit 4	0,26	33,33%
	Cacidicin	H	-	-	-
	Cacidicin	I	Resistin-like beta	0,51	30,77%
C. cellulans	Cellucin	A	Uncharacterized protein SCO3027	0,66	50,94%
			Uncharacterized BCR	0,59	32,73%
			Uncharacterized protein XF2673	0,58	28,57%
C. michiganensis	Michigacin	A	Peptide antibiotic as-48	0,82	65,71%
	Michigacin	B	Putative cytoplasmic protein	0,32	28,57%
C. michiganensis subsp michigan	MichigacinM	A	Putative cytoplasmic protein	0,34	30,23%
			DNA-directed RNA polymerase II subunit RPB4	0,13	42,31%
	MichigacinM	B	MIC2-associated protein	0,23	29,73%
	MichigacinM	C	Geranylgeranyl diphosphate cyclase	0,17	32,00%
	MichigacinM	D	MIC2-associated protein	0,22	30,00%
C. aurantiacum	Auranticin	A	-	-	-
F. acidiphilum	Facidicin	A	YbaB	0,17	23,33%
F. thermotolerans	Fethercin	A	Bifunctional coenzyme PQQ synthesis protein C/D	0,44	17,50%
			Coenzyme PQQ synthesis protein D	0,42	9,64%
F. alni	Falnicin	A	-	-	-
G. africanus	Gafricin	A	TRUD	0,44	16,67%
G. harbinensis	Glyharcin	A	-	-	-
	Glyharcin	B	-	-	-
	Glyharcin	C	Biopolymer transport protein ExbB	0,20	16,67%
			Acidocin B	0,27	29,27%
	Glyharcin	D	NKR-5-3B	0,50	33,90%
	Glyharcin	E	NKR-5-3B	0,49	33,90%

Actinobacteria Species	Bacteriocin		Homology Molecule Description	GMQE	%Similarity
G. harbinensis	Glyharcin	F	NKR-5-3B	0,58	33,90%
	Glyharcin	G	-	-	-
H. album	Halaobucin	A	Probable tautomerase HP_0924	0,63	26,98%
	Halaobucin	B	PyrI4	0,10	36,00%
H. albus	Halobucin	A	Coenzyme PQQ synthesis protein D	0,54	19,72%
			Bifunctional coenzyme PQQ synthesis protein C/D	0,51	16,90%
			Flagellar motor switch protein FliG	0,31	7,55%
	Halobucin	B	-	-	-
H. massiliensis	Hemassicin	A	Growth hormone receptor	0,22	32,00%
I. variabilis	Isovaricin	A	ADP-ribosylation factor-binding protein GGA2	0,34	9,52%
J. endophyticus	Jatephycin	A	-	-	-
J. muralis	Jiamuracin	A	Burkholderia Lethal Factor 1 (BLF1)	0,47	38,46%
K. radiotolerans	Kiratocin	A	Acanthaporin	0,15	39,29%
K. flava	Kraflavacin	A	-	-	-
L. fradiae	Lefradiacin	A	Protein 2	0,50	40,00%
	Lefradiacin	B	Receptor tyrosine-protein kinase erbB-4	0,10	20,00%
			Phosphoprotein	0,09	21,51%
	Lefradiacin	C	MIX1	0,31	26,92%
	Lefradiacin	D	-	-	-
	Lefradiacin	E	-	-	-
	Lefradiacin	F	-	-	-
	Lefradiacin	G	-	-	-
	Lefradiacin	H	-	-	-
	Lefradiacin	I	Single-stranded DNA-binding protein WHY2, mitochondrial	0,24	31,25%
M. brevis	Mibrecin	A	Nodamura virus coat proteins	0,21	10,34%
N. veterana	Noveteracin	A	-	-	-
P. marina	Pamarinacin	A	-	-	-
P. minatonensis	Paminatocin	A	Epstein-Barr nuclear antigen 2	0,47	25,93%
P. rugosa	Prarugocin	A	Aldehyde oxidase medium subunit	0,20	16,00%
	Prarugocin	B	Helix-turn-helix domain protein	0,68	40,32%
P. lymphophilum	Propilymphocin	A	Myosin regulatory light chain	0,19	19,57%
			Myosin head	0,15	21,62%
R. defluvii	Rhodeflucin	A	MoeN5	0,20	32,26%
	Rhodeflucin	B	MoeN5,DNA-binding protein 7d	0,24	32,26%
R. albidiflava	Ralbidicin	A	Histone-binding protein RBBP4	0,31	16,28%
			COG3291: FOG: PKD repeat	0,20	21,05%
S. tropica	Satrocin	A	-	-	-
S. keddieii	Sakeddicin	A	Putative transcriptional regulator	0,73	39,58%
S. marina	Scimaricin	A	Signal recognition particle	0,28	17,50%
			Gingipain R2	0,19	20,00%
	Scimaricin	B	Hypothetical protein HD1797	0,17	21,88%
	Scimaricin	C	Primosomal protein 1	0,28	26,32%
	Scimaricin	D	E1 envelope glycoprotein	0,23	41,94%
	Scimaricin	E	DNA mismatch repair protein mutL	0,29	18,18%
T. album	Thalbucin	A	Secreted 45 kDa protein	0,33	13,33%
			Peptidoglycan endopeptidase RipA	0,28	15,59%
			Secreted 45 kDa protein	0,28	13,07%
			Alpha-actinin-2	0,27	10,10%
			Peptidoglycan endopeptidase RipA	0,26	11,11%
			Msmeg_4306	0,25	18,23%
			Protein regulator of cytokinesis 1	0,23	12,97%
			M23 peptidase domain protein	0,20	46,67%
T. tyrosinosolvens	Tsutysocin	A	-	-	-
Yonghaparkia sp	Yonghacin	A	-	-	-
Y. deserti	Yudesericin	A	-	-	-
	Yudesericin	B	-	-	-
	Yudesericin	C	Fusion glycoprotein F0	0,50	23,53%

Table 3.5: Homology models of novel putative bacteriocins within the Firmicutes phylum. Each model is accompanied by a global mean quality estimate (GMQE) and percentage similarity of the suggested model to the template model.

Firmicutes Species	Bacteriocin		Homology Molecule Description	GMQE	%Similarity
A. ethanolgignens	Acethanocin	A	Polymerase PB2	0,52	30,77%
A. acidiphilus	Alicacin	A	-	-	-
A. transvaalensis	Altracin	A	-	-	-
	Altracin	B	Aquaporin-1	0,36	29,63%
	Altracin	C	Caspase-6	0,46	18,00%
A. colihominis	Anaecolicin	A	M23 peptidase domain protein	0,57	28,63%
			Putative peptidase M23	0,41	32,00%
	Anaecolicin	B	Ataxin-7	0,56	34,38%
A. gonensis	Anogonicin	A	Putative alkaline phosphatase	0,14	26,92%
	Anogonicin	B	-	-	-
	Anogonicin	C	-	-	-
B. aerophilus	Bacaericin	A	NKR-5-3B	0,81	64,29%
	Bacaericin	B	NKR-5-3B	0,79	56,82%
	Bacaericin	C	AS-48 protein	0,62	80,00%
			Purine nucleoside phosphorylase	0,18	26,67%
			Probable glutamate/gamma-aminobutyrate antiporter	0,15	17,14%
	Bacaericin	D	Peptide antibiotic as-48	0,75	80,00%
B. megaterium	Bamegacin	A	-	-	-
B. seileri	Baseicin	A	-	-	-
	Baseicin	B	-	-	-
	Baseicin	C	Bacteriocin carnobacteriocin B2	0,17	20,00%
	Baseicin	D	Beta-hexosaminidase	0,55	21,05%
	Baseicin	E	-	-	-
B. brevis	Brevicin	A	Hemagglutinin	0,38	16,22%
	Brevicin	B	BPT4 gene 59 helicase assembly protein	0,30	20,00%
B. hungatei	Buthangacin	A	M23 peptidase domain protein	0,36	27,86%
			Zinc peptidase	0,36	28,90%
C. bescii	Calbecin	A	Bacteriocin AS-48	0,64	26,09%
			Probable butyrate kinase 2	0,36	22,50%
	Calbecin	B	Uncharacterized protein YKL091C	0,16	27,03%
C. saccharolyticus	Calsacin	A	Bacteriocin AS-48	0,69	28,57%
			Propionate kinase	0,26	22,22%
	Calsacin	B	Eco29kIR	0,32	35,29%
C. australicus	Calaucin	A	M23 peptidase domain protein	0,38	28,19%
			Designed protein OR327	0,11	15,17%
C. frackibacter	Cafrackicin	A	Diheme cytochrome c napb molecule: nitrate reductase	0,44	31,34%
			Transcription elongation factor s-ii, dst1	0,43	24,00%
	Cafrackicin	A	Cytochrome c family protein	0,27	28,26%
			Cytochrome c3	0,17	24,39%
C. hydrogenoformans	Cahydrocin	A	-	-	-
C. acetobutylicum	Clacetocin	A	Histone fold protein	0,39	18,75%
C. cadaveris	Clocadacin	A	Copper-sensitive operon repressor (CsoR)	0,76	39,08%
Sporomusa sp	Desmocin	A	Catalase HPII	0,16	23,53%
	Desmocin	B	Potassium channel toxin alpha-KTx 18.1	0,36	33,33%
D. meridiei	Desmericin	A	Possible Trp repressor	0,46	21,74%
			TrfB transcriptional repressor protein	0,52	6,45%
			Staphylococcal accessory regulator a	0,28	8,70%
D. carboxydivorans	Descarbocin	A	-	-	-
D. alkaliphilus	Dethalkacin	A	-	-	-
	Dethalkacin	B	FeS cluster assembly protein SufB	0,30	31,03%
D. pigrum	Dopigrucin	A	Hypothetical protein	0,19	16,67%
			DNA-binding protein SATB1	0,15	22,22%
	Dopigrucin	B	Hypothetical protein	0,20	15,15%
			DNA-binding protein SATB1	0,13	23,08%

Table 3.5 continued...

Firmicutes Species	Bacteriocin		Homology Molecule Description	GMQE	%Similarity
E. avium	Entavicin	A	P protein	0,40	10,42%
	Entavicin	B	P protein	0,40	10,42%
E. rectale	Eurectacin	A	-	-	-
F. arsenicus	Ficarsecin	A	-	-	-
G. cuniculi	Gecunicin	A	TrfB transcriptional repressor protein	0,24	13,33%
	Gecunicin	B	Acidocin B	0,52	41,38%
			Regulatory protein rop	0,13	20,00%
G. formicilis	Geformicin	A	ToxR-activated gene (TagE)	0,09	33,07%
			Ste24	0,08	17,46%
			Zinc peptidase	0,08	38,46%
	Geformicin	B	Capsid subunit VP1	0,28	29,73%
G. stearothermophilus	Geostearocin	A	Bacteriocin AS-48	69,00	32,86%
			Fructose-1,6-bisphosphate aldolase/phosphatase	0,19	24,14%
	Geostearocin	B	-	-	-
			M23 peptidase domain protein	0,49	32,33%
			Zinc peptidase	0,49	31,90%
			Cell wall-binding endopeptidase-related protein	0,28	21,51%
			Chitin elicitor-binding protein	0,23	19,86%
			Extracellular protein 6	0,22	24,29%
H. silvermanii	Hasilvecin	A	-	-	-
	Hasilvecin	B	-	-	-
H. halobius	Halobicin	A	Transcription elongation regulator 1	0,14	29,03%
			Toxofilin	0,13	21,43%
H. orenii	Halorecin	A	-	-	-
J. alimentarius	Jeoalicin	A	UPF0092 membrane protein yajC	0,28	16,67%
J. halophilus	Jeohalocin	A	Arginine/agmatine antiporter	0,25	7,55%
L. acetotolerans	Lacetocin	A	Peptide cyclase 1	0,16	18,80%
			Putative dipeptidyl aminopeptidase IV	0,03	15,63%
L. acetotolerans			Carboxy-cis,cis-muconate cyclase	0,03	9,68%
	Lacetocin	B	Nitrous-oxide reductase	0,10	11,24%
			R-SwaI protein	0,07	31,75%
L. garvieae	Lagaricin	A	K(+)-stimulated pyrophosphate-energized sodium pump	0,28	24,24%
	Lagaricin	B	Bacteriocin carnobacteriocin B2	0,41	17,86%
L. monocytogenes	Limonocin	A	Endoglucanase V-like protein	0,24	32,00%
M. mesophilum	Marimecin	A	Lactococcin 972	0,39	29,03%
	Marimecin	B	-	-	-
	Marimecin	C	Capsid protein VP1	0,43	17,65%
	Marimecin	D	Capsid protein VP1	0,43	17,65%
	Marimecin	E	Capsid protein VP1	0,43	17,65%
	Marimecin	F	Secretion chaperone, phage-display derived peptide	0,21	17,86%
O. oeni	Oenoecin	A	Bacteriocin AS-48	0,56	28,36%
	Oenoecin	B	NKR-5-3B	0,54	32,26%
			Bcl-2-like protein 2	0,44	15,25%
			Bcl-2 homologous antagonist/killer	0,43	21,05%
O. valericigenes	Oscivacin	A	Zinc finger protein 57	0,34	15,91%
			Uncharacterized HTH-type transcriptional regulator ygiT	0,44	13,21%
			NAD-dependent deacetylase	0,43	20,75%
			Probable lysine biosynthesis protein	0,39	24,49%
O. pfennigii	Oxopfennicin	A	P protein	0,32	18,37%
			Fusion glycoprotein F0	0,09	11,11%
P. polymyxa	Paepolycin	A	-	-	-
P. propionicus	Peprocin	A	-	-	-
	Peprocin	B	-	-	-
P. thermopropionicum	Pethermocin	A	Heparin lyase I	0,35	27,50%
P. niger	Penigecin	A	Ribonuclease MC	0,26	24,00%
P. vibrioides	Provicin	A	-	-	-
R. intestinalis	Rosintecin	A	Transcription regulatory protein MOTA	0,22	25,00%

Table 3.5 continued...

Firmicutes Species	Bacteriocin		Homology Molecule Description	GMQE	%Similarity
R. intestinalis	Rosintecin	B	-	-	-
	Rosintecin	C	Cytochrome C2	0,33	40,00%
S. peptonophila	Seipeptocin	A	Protein timeless homolog	0,19	12,12%
	Seipeptocin	B	Human protein tyrosine phosphatase receptor type J	0,41	13,16%
	Seipeptocin	C	-	-	-
	Seipeptocin	D	Putative uncharacterized protein	0,11	28,95%
S. ruminantium	Serumicin	A	Chaperone protein htpG	0,42	25,76%
			50S ribosomal protein L9, Elongation factor 4	0,29	18,37%
S. azabuensis	Shazacin	A	Surface glycoprotein	0,49	25,93%
S. kribbensis	Shikricin	A	GDP-mannose transporter 1	0,56	17,91%
	Shikricin	B	Restriction endonuclease ecorv	0,43	22,58%
	Shikricin	C	-	-	-
	Shikricin	D	-	-	-
S. nakayamae	Sponakacin	A	No mechanoreceptor potential C isoform L	0,26	13,51%
Sporomusa sp	Sporocin	A	-	-	-
S. psychrophila	Spopsychrocir	A		0,62	45,74%
			DNA-binding protein 7a	0,23	25,49%
S. capitis	Stacapicin	A	Subtilosin: Can't Create Monomer- Model Error	-	-
	Stacapicin	B	Subtilosin: Can't Create Monomer- Model Error	-	-
	Stacapicin	C	Subtilosin: Can't Create Monomer- Model Error	-	-
S. pseudintermedius	Stapseudicin	A	-	-	-
S. simulans	Stasimucin	A	-	-	-
	Stasimucin	B	Lactococcin 972	0,46	26,67%
S. warneri	Stawarcin	A	-	-	-
	Stawarcin	B	-	-	-
	Stawarcin	C	-	-	-
S. pyogenes	Strepyocin	A	Phycobilisome 32.1 kDa linker polypeptide, phycocyanin-associated, rod 1	0,20	26,67%
S. pyogenes	Strepyocin	B	Bacteriocin carnobacteriocin B2	0,34	17,07%
	Strepyocin	C	Bacteriocin carnobacteriocin B2	0,49	23,64%
			Pertussis toxin	0,33	14,00%
	Strepyocin	D	Bacteriocin carnobacteriocin B2	0,22	36,00%
S. schinkii	Syschicin	A	-	-	-
	Syschicin	B	Acetyl-CoA synthase	0,47	48,28%
	Syschicin	C	Telomere length regulator taz1	0,42	38,46%
T. halophilus	Tethalocin	A	N-acetylmuramoyl-L-alanine amidase domain-containing protein SAOUHSC_02979	0,40	43,48%
			Zoocin A endopeptidase	0,23	34,92%
T. phaeum	Thephaeucin	A	M23 peptidase domain protein	0,34	29,28%
			Zinc peptidase	0,33	25,59%
			Resuscitation-promoting factor RpfB	0,11	18,94%
			YGAU	0,09	19,30%
			Surface protein G	0,08	19,44%
			Cortical-lytic enzyme	0,07	17,71%
			Lmo1499 protein	0,06	10,42%
			Cell wall-binding endopeptidase-related protein	0,06	26,60%
	Thephaeucin	B	-	-	-
	Thephaeucin	C	-	-	-
	Thephaeucin	D	Beta-D-glucoside glucohydrolase	0,39	18,18%
T. toyohensis	Thetoyocin	A	Alr3090 protein	0,52	10,59%
T. aegyptius	Theagypcin	A	ORF134	0,31	8,51%
			V0 assembly protein 1	0,17	12,90%
T. thermosaccharolyticum	Thermosacin	A	-	-	-
T. dichotomicum	Therdichocin	A	KP6 killer toxin subunit beta	0,23	25,71%
V. necropolis	Vinecrocin	A	Ubiquitin-conjugating enzyme E2 D2		
V. soli	Visolicin	A	-	-	-

Table 3.6: Homology models of novel putative bacteriocins within the Proteobacteria phylum. Each model is accompanied by a global mean quality estimate (GMQE) and percentage similarity of the suggested model to the template model.

Proteobacteria Species	Bacteriocin		Homology Molecule Description	GMQE	%Similarity
A. xylosoxidans	Axylocin	A	M23 peptidase domain protein	0,53	30,73%
			Zinc peptidase	0,47	32,14%
			Conserved hypothetical secreted protein	0,47	29,91%
			LysM type receptor kinase	0,10	12,20%
A. suis	Actisuicin	A	DNA/RNA-binding protein Alba	0,36	31,25%
	Actisuicin	B	Glyceraldehyde 3-phosphate dehydrogenase (NADP+)	0,22	20,69%
A. bestiarum	Aebestiacin	A	-	-	-
A. popoffii	Aepocin	A	-	-	-
A. xiamenense	Axiamecin	A	-	-	-
A. donghaensis	Albidocin	A	DNA mismatch repair protein MutL	0,53	30,77%
A. dieselolei	Aldiesecin	A	-	-	-
A. dehalogenans	Adehalocin	A	M23 peptidase domain protein	0,56	30,97%
			ToxR-activated gene (TagE)	0,19	32,24%
			Lysostaphin	0,18	32,17%
			Lysostaphin	0,18	32,17%
A. heliothermus	Antheliocin	A	Protein KHNYN	0,44	24,14%
A. defluvii	Aquadeflucin	A	-	-	-
A. gephyra	Argephycin	A	Uncharacterized protein TM1086	0,35	38,46%
	Argephycin	B	-	-	-
	Argephycin	C	-	-	-
A. cryaerophilus	Arcryaerocin	A	Cytochrome c, iso-1	0,57	27,50%
			Cytochrome c552	0,56	30,00%
A. nasoniae	Arnasonicin	A	-	-	-
	Arnasonicin	B	Colicin IA	0,48	40,93%
			Colicin S4	0,27	30,65%
			Myosin-7	0,16	14,72%
A. nasoniae			Myosin-7	0,16	14,72%
			Anopheles Plasmodium-responsive Leucine-rich repeat protein 1	0,13	13,22%
			Myosin 2 heavy chain	0,13	12,61%
			Dynein heavy chain, cytoplasmic	0,08	13,54%
A. taihuensis	Astaihuencin	A	Oligosaccharyltransferase complex subunit OSTC	0,34	42,86%
	Astaihuencin	B	Ferredoxin-like protein	0,43	16,00%
			Translationally-controlled tumor protein	0,32	21,05%
A. coralicida	Aucoracin	A	Putative uncharacterized protein	0,23	36,36%
A. altamirensis	Auraltacin	A	-	-	-
A. caulinodans	Azocaulicin	A	Ribosomal protein, putative	0,31	31,91%
			Arf-GAP with Rho-GAP domain, ANK repeat and PH domain-containing protein 3	0,32	17,24%
			Protease nsP2	0,19	22,86%
A. chroococcum	Azochrocin	A	M23 peptidase domain protein	0,51	32,57%
			ToxR-activated gene (TagE)	0,18	40,61%
			Lysostaphin	0,16	29,37%
			Lysostaphin	0,16	29,37%
			Cell wall-binding endopeptidase-related protein	0,14	16,41%
	Azochrocin	B	Cell wall-binding endopeptidase-related protein	0,15	16,41%
			Lysostaphin	0,17	29,37%
			M23 peptidase domain protein	0,53	32,29%
			ToxR-activated gene (TagE)	0,19	40,61%
			Lysostaphin	0,17	29,37%
A. restrictus	Azorecin	A	Tumor necrosis factor receptor superfamily member 6	0,49	15,38%
			FAS	0,46	15,69%

Table 3.6 continued...

Proteobacteria Species	Bacteriocin		Homology Molecule Description	GMQE	%Similarity
Blastomonas sp	Blastocin	A	-	-	-
	Blastocin	B	M23 peptidase domain protein	0,45	30,61%
			Zinc peptidase	0,20	33,01%
			Extracellular Protein 6	0,08	13,45%
	Blastocin	C	Putative peptidase M23	0,50	34,02%
			Putative glycyl-glycine endopeptidase lytM	0,35	19,08%
Bosea sp	Boseacin	A	Zinc peptidase	0,39	22,76%
			M23 peptidase domain protein	0,38	19,18%
			Extracellular Protein 6	0,10	10,92%
Brenneria sp	Brennecin	A	-	-	-
	Brennecin	B	-	-	-
B. diminuta	Bredimicin	A	Packaging enzyme P4	0,32	28,57%
B. mallei	Bumalleicin	A	-	-	-
B. agrestis	Buttagrecin	A	Methylosome subunit pICln	0,33	8,00%
B.thuringiensis	Bathuricin	A	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1,RPN1	0,49	16,36%
			DNA gyrase subunit A	0,31	14,71%
	Bathuricin	B	Pesticidal crystal protein Cry6Aa	0,71	14,23%
			Colicin-E9	0,14	38,85%
			Apolipoprotein A-IV	0,08	15,25%
			Protein unc-13 homolog A	0,06	12,36%
			Putative tbpip family protein	0,06	8,51%
	Bathuricin	C	-	-	-
Chelatococcus sp	Chelacin	A	ATPase Inhibitor	0,38	24,00%
C. youngae	Citryocin	A	-	-	-
C. burnetii	Coburnecin	A	DNA-directed RNA polymerase subunit beta	0,35	28,00%
C. malonaticus	Cromalocin	A	-	-	-
C. basilensis	Cubasilencin	A	-	-	-
D. hydrothermale	Deshymecin	A	-	-	-
D. infernum	Desinfercin	A	-	-	-
D. aliphaticivorans	Desalicin	A	PDZ and LIM domain protein 5	0,17	20,00%
			Jumonji/ARID domain-containing protein 1D	0,16	16,67%
			DNA polymerase II large subunit	0,14	33,33%
	Desalicin	B	-	-	-
	Desalicin	C	Integration host factor subunit alpha	0,76	39,13%
			Integration host factor	0,75	35,48%
D. butyrativorans	Debutycin	A	Xylose isomerase domain protein TIM barrel	0,27	9,38%
D. sulfexigens	Desulfexin	A	-	-	-
	Desulfexin	B	-	-	-
D. multivorans	Demultivocin	A	-	-	-
D. retbaense	Deretbaecin	A	-	-	-
D. lacustre	Delacucin	A	Specialized acyl carrier protein	0,68	36,36%
D. conservatrix	Deconservaci	A	-	-	-
Desulfosarcina sp	Desulfosacin	A	Cyclic di-GMP phosphodiesterase YahA	0,32	22,73%
			TutF	0,28	33,33%
	Desulfosacin	B	Putative capsid protein	0,21	36,36%
			Phospholipase A1	0,17	14,29%
D. balticum	Debalticin	A	-	-	-
D. chrysanthemi	Dichrysacin	A	-	-	-
	Dichrysacin	B	-	-	-
Duganella sp	Duganecin	A	Membrane-bound lytic murein transglycosylase A	0,37	30,77%
	Duganecin	B	Conserved hypothetical secreted protein	0,15	34,06%
			Ste24p	0,15	13,23%
	Duganecin	C	E1 envelope glycoprotein	0,46	15,79%
			Serine/threonine-protein kinase MARK1	0,34	17,86%
	Duganecin	D	Tail fiber protein	0,54	44,44%
E. tarda	Etardacin	A	Penaeidin-3a	0,25	29,63%
E. corrodens	Eicorrocin	A	ATP-dependent CLP Protease Adapter Protein CLPS	0,78	38,81%

Table 3.6 continued...

Proteobacteria Species	Bacteriocin		Homology Molecule Description	GMQE	%Similarity
E. longus	Erylocin	A	Wiskott-Aldrich Syndrome Protein	0,09	12,00%
	Erylocin	B	-	-	-
E. fergusonii	Efergucin	A	-	-	-
	Efergucin	B	Colicin-E9	0,17	34,18%
			Uncharacterized protein	0,14	37,06%
			Haemolysin co-regulated protein	0,13	32,37%
			Colicin B	0,12	17,71%
			Colicin E7	0,10	16,88%
			Colicin E3	0,08	14,29%
E. americana	Ewamericin	A	-	-	-
	Ewamericin	B	Zinc peptidase	0,60	39,47%
			M23 peptidase domain protein	0,46	33,11%
G. anatis	Gallanacin	A	T-lymphoma invasion and metastasis-inducing protein 2	0,57	21,05%
G. hollisae	Grollicin	A	-	-	-
	Grollicin	B	Alpha-tubulin N-acetyltransferase	0,49	42,31%
H. alvei	Halveicin	A	Cholix toxin	0,45	32,26%
H. felis	Hefelicin	A	ToxR-activated gene (TagE)	0,51	73,66%
			Apocytochrome F	0,09	21,52%
			PilO	0,07	13,11%
			Centrosomal protein of 57 kDa	0,06	11,54%
I. sakaiensis	Ideosacin	A	-	-	-
	Ideosacin	B	M23 peptidase domain protein	0,55	32,77%
			ToxR-activated gene (TagE)	0,19	36,18%
	Ideosacin	C	-	-	-
	Ideosacin	D	Dermonecrotic toxin	0,25	25,71%
K. gyiorum	Kegyiocin	A	M23 peptidase domain protein	0,57	32,40%
			Conserved hypothetical secreted protein	0,49	31,88%
			Chitin elicitor-binding protein	0,09	10,53%
K. kingae	Kingaecin	A	Elongation factor G	0,39	15,38%
K. pneumoniae	Klepneumocin	A	-	-	-
K. ascorbata	Klascobacin	A	-	-	-
Limnobacter sp	Limnobacin	A	M23 peptidase domain protein	0,62	32,21%
			Conserved hypothetical secreted protein	0,48	36,36%
			ToxR-activated gene (TagE)	0,25	37,91%
L. enzymogenes	Lysenzymocin	A	Putative peptidase M23	0,57	35,08%
	Lysenzymocin	B	Uncharacterized protein Atu2773	0,32	24,39%
			Desmoplakin	0,16	15,38%
	Lysenzymocin	C	-	-	-
	Lysenzymocin	D	-	-	-
	Lysenzymocin	E	Endo-1,4-beta-xylanase A	0,35	27,27%
	Lysenzymocin	F	Subdomain Of Desmoplakin Carboxy-Terminal Domain (DPCT)	0,26	27,78%
L. gummosus	Lygummocin	A	Putative peptidase M23	0,56	34,96%
	Lygummocin	B	-	-	-
	Lygummocin	C	-	-	-
M. gryphiswaldense	Magryphicin	A	Aftrax	0,42	24,14%
	Magryphicin	B	M23 peptidase domain protein	0,55	33,52%
			Zinc peptidase	0,32	31,78%
			DNA-directed RNA polymerase beta' chain	0,09	20,83%
M. haemolytica	Manaemocin	A	Hypothetical Protein	0,50	29,03%
M. rhizophilum	Marhizocin	A	M23 peptidase domain protein	0,52	30,48%
			ToxR-activated gene (TagE)	0,17	36,60%
			Lysostaphin	0,16	33,10%
			Lysostaphin	0,16	33,10%
			Cell wall-binding endopeptidase-related protein	0,15	18,65%
M. loti	Meloticin	A	Probable tRNA threonylcarbamoyladenosine biosynthesis protein QRI7, mitochondrial	0,28	20,00%
	Meloticin	B	-	-	-
M. parvus	Meparvucin	A	-	-	-
M. stellata	Mestellacin	A	-	-	-

Table 3.6 continued...

Proteobacteria Species	Bacteriocin		Homology Molecule Description	GMQE	%Similarity
M. aerodenitrificans	Micraerocin	A	M23 peptidase domain protein	0,60	38,76%
			ToxR-activated gene (TagE)	0,19	33,12%
M. morganii	Morganiicin	A	-	-	-
	Morganiicin	B	Colicin-E9	0,41	32,18%
			Colicin E3	0,23	20,75%
			Colicin-E5	0,10	69,37%
			Colicin IA	0,08	18,11%
			Tropomyosin Beta Chain	0,06	21,25%
			Bridging integrator 2	0,06	13,92%
			Xrcc4-MYH7-1590-1657	0,03	12,68%
	Morganiicin	C	Colicin E3	0,53	39,58%
			Colicin E3	0,51	45,41%
			Plectin	0,08	19,53%
			Catenin Alpha-2	0,08	25,00%
			Nucleoporin Nup62	0,05	13,64%
			Rab guanine nucleotide exchange factor SEC2	0,05	17,65%
			Prefoldin subunit beta	0,04	19,05%
N. profundicola	Nauprocin	A	Peroxisomal Biogenesis Factor 19	0,56	28,57%
N. aquibiodomus	Nitraquibiocin	A	-	-	-
N. aestuarii	Nitraestuacin	A	M23 peptidase domain protein	0,58	32,20%
			Zinc Peptidase	0,48	33,13%
N. communis	Nicommunicir	A	M23 peptidase domain protein	0,69	36,06%
			ToxR-activated gene (TagE)	0,25	35,53%
O. indolifex	Ocindocin	A	DNA-Binding Protein H-NS	0,32	16,33%
	Ocindocin	B	N Utilization Substance Protein A	0,37	36,00%
	Ocindocin	C	-	-	-
O. alexandrii	Ocalexacin	A	-	-	-
O. granulosus	Oceagracin	A	-	-	-
O. antarctica	Olantacin	A	M23 peptidase domain protein	0,57	30,03%
			ToxR-activated gene (TagE)	0,19	31,58%
O. carboxidovorans	Olicarbocin	A	Cell Division Protein ZapD	0,33	31,03%
O. formigenes	Oxaformicin	A	Heat-labile enterotoxin B chain	0,39	36,00%
P. agglomerans	Pagglocin	A	-	-	-
P. betavasculorum	Pebetacin	A	2-Dehydro-3-Deoxyphosphooctonate Aldolase	0,28	12,12%
	Pebetacin	B	-	-	-
	Pebetacin	C	CST Complex Subunit STN1	0,48	17,24%
Pelomonas sp	Pelomonacin	A	Avirulence Protein	0,37	38,24%
P. zucineum	Phezucin	A	-	-	-
P. luminescens	Pholumicin	A	Colicin-E9	0,21	38,46%
			Uncharacterized protein	0,16	46,62%
			Colicin-E9	0,11	19,61%
			Putative uncharacterized protein	0,60	27,50%
	Pholumicin	B	Antitoxin DinJ	0,78	47,62%
	Pholumicin	C	-	-	-
P. oryzae	Pleorycin	A	Trypsin Inhibitor 3	0,12	32,00%
P. shigelloides	Pleshingecin	A	-	-	-
Polaromonas sp	Polarocin	A	M23 peptidase domain protein	0,58	30,53%
			Zinc peptidase	0,48	32,14%
			ToxR-activated gene (TagE)	0,21	29,03%
P. dicarboxylicus	Prodicarbocir	A	-	-	-
	Prodicarbocir	B	-	-	-
P. mirabilis	Promiracin	A	Colicin E3	0,81	67,39%
	Promiracin	B	Colicin E9	0,56	40,69%
			Colicin E3	0,18	34,48%
			Colicin IA	0,13	23,50%
P. sneebia	Prosneecin	A	-	-	-

Table 3.6 continued...

Proteobacteria Species	Bacteriocin		Homology Molecule Description	GMQE	%Similarity
P. violaceinigra	Pseudoviolac	A	Uncharacterized protein	0,45	17,24%
	Pseudoviolac	B	Ste24p	0,18	13,76%
			Conserved hypothetical secreted protein	0,15	34,07%
	Pseudoviolac	C	E1 envelope glycoprotein	0,46	15,79%
			Serine/threonine-protein kinase MARK1	0,34	17,86%
P. ferrugineus	Pseuferrucin	A	Modification Methylase Rsri	0,48	29,55%
	Pseuferrucin	B	-	-	-
P. ascidiaceicola	Pseudascin	A	Nitrite Reductase	0,28	50,00%
			Oplophorus-Luciferin 2-Monooxygenase Catalytic Subunit	0,23	36,36%
P. mexicana	Pseumecin	A	Putative Peptidase M23	0,55	34,98%
			Putative Glycyl-Glycine Endopeptidase Lytm	0,38	24,00%
R. planticola	Raoplanticin	A	-	-	-
	Raoplanticin	B	Major Prion Protein	0,12	15,38%
	Raoplanticin	C	-	-	-
R. fermentans	Rhofermecin	A	-	-	-
	Rhofermecin	B	M23 Peptidase Domain Protein	0,73	99,71%
			Cell Wall-Binding Endopeptidase-Related Protein	0,07	24,18%
R. ferrireducens	Rhoferricin	A	M23 Peptidase Domain Protein	0,55	31,37%
			Zinc Peptidase	0,29	34,62%
			ToxR-Activated Gene (TagE)	0,19	27,74%
	Rhoferricin	B	Protease Do-Like 2, Chloroplastic,Protease Do-Like 9	0,29	29,63%
Rhodoplanes sp	Rhodoplacin	A	Thiamine Pyridinylase I	0,17	12,50%
R. elongatum	Roselongacin	A	Ammonium Transporter Family Protein Rh50	0,34	15,63%
Roseibium sp	Roseicin	A	-	-	-
	Roseicin	B	-	-	-
R. halodurans	Roshalocin	A	Dipeptide Transport Protein DPPA	0,31	24,24%
R. nubinhibens	Ronubicin	A	NECAP1	0,19	33,33%
	Ronubicin	B	-	-	-
R. chamberiensis	Rouchambeci	A	-	-	-
R. cliftonensis	Rucliftocin	A	-	-	-
S. degradans	Sadegracin	A	Integrin Beta-2	0,27	22,22%
S. rosea	Saroseacin	A	Microtubule-Associated Protein, RP/EB Family, Member 1	0,20	20,69%
	Saroseacin	B	-	-	-
S. bongori	Sabongocin	A	Protein LSM14 Homolog A	0,42	18,52%
	Sabongocin	B	Type VI secretion system effector, Hcp1 family	0,08	21,05%
			Colicin D	0,09	11,49%
S.marcescens	Semarcin	A	-	-	-
	Semarcin	B	-	-	-
S. dysenteriae	Shidysecin	A	-	-	-
	Shidysecin	B	Hcp	0,15	26,37%
			Colicin D	0,09	10,99%
			Polyhedrin	0,05	16,36%
S. fusca	Shifuscacin	A	Putative Peptidase M23	0,56	35,96%
			Putative Glycyl-Glycine Endopeptidase LytM	0,39	22,41%
S. meliloti	Simelilocin	A	-	-	-
S. asaccharolytica	Sphasacin	A	Putative Peptidase M23	0,49	31,15%
	Sphasacin	B	-	-	-
	Sphasacin	C	Mandelate Racemase/Muconate Lactonizing Enzyme	0,38	18,03%
			Chorismate Mutase	0,30	13,73%
			Possible 2-Hydroxychromene-2-Carboxylate Isomerase	0,19	10,64%
S. macrogoltabida	Sphimacrocin	A	-	-	-
S. maltophilia	Stemaltocin	A	Putative Peptidase M23	0,56	36,07%
			Putative Glycyl-Glycine Endopeptidase Lytm	0,39	22,41%
	Stemaltocin	B	Profilin IA	0,35	17,86%
S. fumaroxidans	Syfumarocin	A	Carboxypeptidase Inhibitor	0,50	29,63%
T. robiniae	Tarobicin	A	Enoyl-[Acyl-Carrier-Protein] Reductase [NADH]	0,40	21,05%
T. norvegica	Thenovecin	A	Putative Uncharacterized Protein	0,52	37,04%

Table 3.6 continued...

Proteobacteria Species	Bacteriocin		Homology Molecule Description	GMQE	%Similarity
T. ingrica	Thingricacin	A	-	-	-
T. mobilis	Timobicin	A	Putative Excisionase	0,10	33,33%
	Timobicin	B	-	-	-
	Timobicin	C	Minor Core Protein Lambda 3	0,23	30,77%
	Timobicin	D	-	-	-
	Timobicin	E	Nucleocapsid Protein	0,17	14,29%
T. auensis	Tolauencin	A	Hemolysin II	0,40	20,00%
V. anguillarum	Vanguicin	A	Zinc Peptidase	0,74	51,43%
			M23 Peptidase Domain Protein	0,40	37,04%
			Apocytochrome F	0,08	14,81%
	Vanguicin	B	Glycosyl Hydrolase, Family 38	0,18	24,14%
X. autotrophicus	Xanautocin	A	Glycyl-Glycine Endopeptidase LytM	0,18	35,48%
X. bromi	Xabromicin	A	-	-	-
X. azovorans	Xenazovocin	A	M23 Peptidase Domain Protein	0,56	34,73%
			Zinc Peptidase	0,28	37,62%
			ToxR-Activated Gene (TagE)	0,19	35,29%
X. griffiniae	Xegrifficin	A	-	-	-
	Xegrifficin	B	Colicin-E9	0,61	45,71%
			Colicin E3	0,22	32,54%
	Xegrifficin	C	Colicin-E9	0,60	38,61%
			Colicin IA	0,17	23,18%
			Colicin E3	0,15	32,43%
			Cell Invasion Protein SipB	0,05	29,03%
			60S Ribosomal Protein L24E	0,03	20,45%
	Xegrifficin	D	-	-	-
	Xegrifficin	E	5-Carboxymethyl-2-Hydroxymuconate Delta-Isomerase	0,29	19,44%
X. hominickii	Xehomicin	A	Probable Exosome Complex Exonuclease 2	0,57	32,26%
	Xehomicin	B	Uncharacterized Protein	0,14	44,90%
			Colicin E7	0,08	15,23%
			Colicin D	0,04	21,74%
			Cell Wall Surface Anchor Family Protein	0,03	20,75%
	Xehomicin	C	PPC decarboxylase AtHAL3a	0,25	33,33%
Y. intermedia	Yerintecin	A	-	-	-
Y. regensburgei	Yoregencin	A	De Novo Protein M7	0,56	30,77%
	Yoregencin	B	-	-	-

3.4.2 Sequence alignment and phylogenetic tree analysis

Sequence alignments were performed with Clustal Omega for the bacteriocin sequences and the 16s rRNA of bacteriocin producing species within the Actinobacteria, Firmicutes and Proteobacteria phyla, respectively. These alignment results were visualised with Weblogo (figure 3.5).

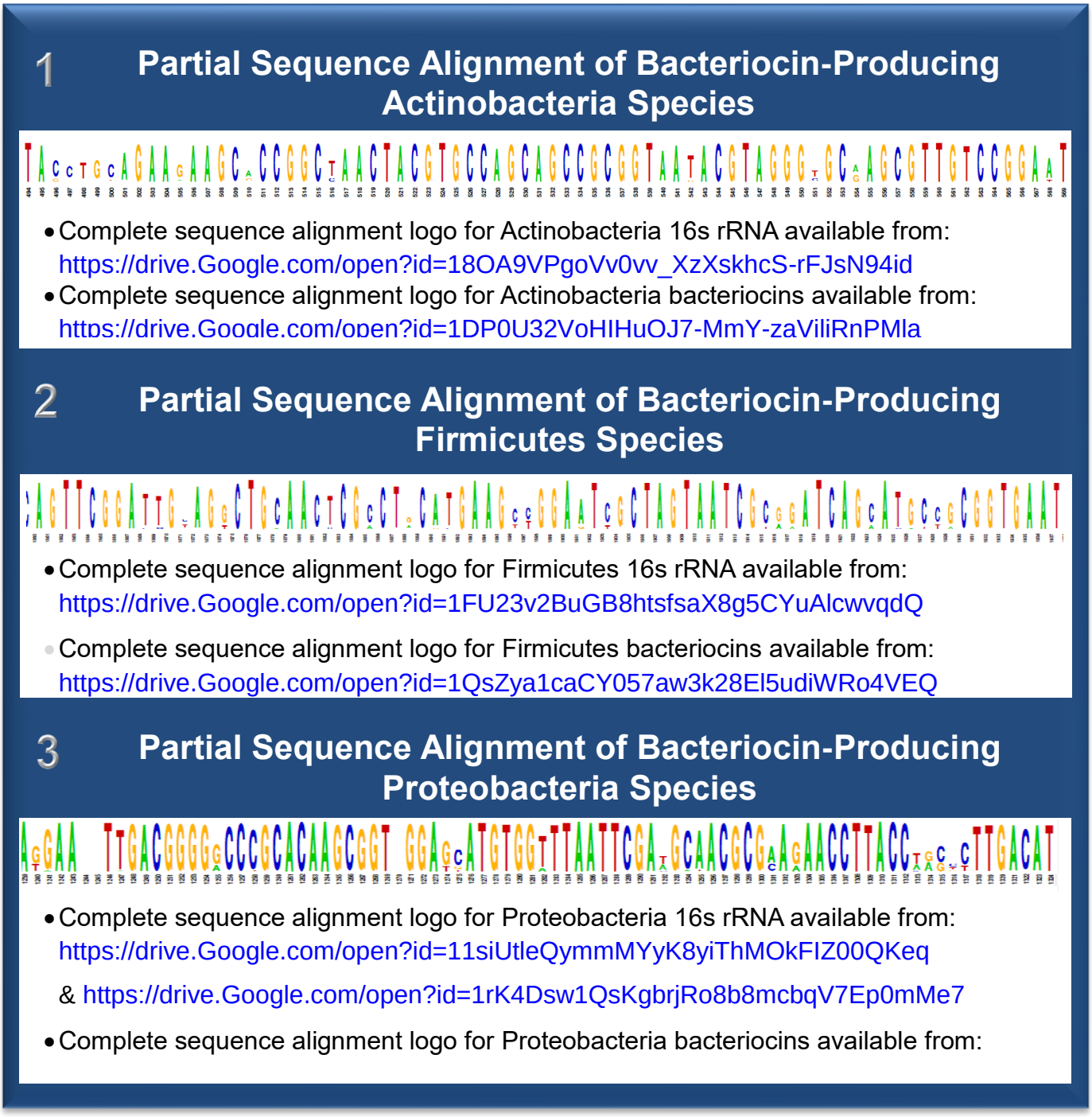

Figure 3.5: The 16s rRNA sequence alignment logo obtained from Clustal Omega and visualised with Weblogo for the (1) Actinobacteria, (2) Firmicutes and (3) Proteobacteria phyla

The aligned sequences indicate conservation of genetic clusters within the phyla. Phylogenetic trees were inferred from these alignments to study evolutionary relationships between bacteriocins and between producer species.

Phylogenetic trees were generated for comparison of different bacteriocin types and different species within a particular phylum, respectively. Reference trees for the different phyla were generated from the bacteriocin producer strains, using their 16s rRNA sequences. High resolution images (.pdf) are available for all trees on Google drive. Actinobacteria phylogenetic trees (figure 3.6) are available from: https://drive.google.com/open?id=1kn73PhcSUx3t4bWLzyOoizQ2pZVs89KZ. Firmicutes phylogenetic trees (figure 5.7) are available from: https://drive.google.com/open?id=1Cdb94P_dUTmaU787e5pSxDeRl4AQatBL and Proteobacteria phylogenetic trees (figure 5.8) are available from: https://drive.google.com/open?id=1KPVUeeJ8CK8Zi0Ar03EchOmTgJK1xzBN.

In addition to the phylogenetic relationships between producer strains, evolutionary relationships were also inferred for the different bacteriocin types within phyla. Linear forms of the reference and bacteriocin trees were contrasted for all analysed phyla and the high resolution document is available on Google drive at: https://drive.Google.com/open?id=15UaPZGEVd7hjja2nmcxHDGvJWCnwTC7Y

A reference tree for each phylum was compiled from the bacteriocin producing species 16s rRNA sequences. The bacteriocins tree of each phylum was compared and contrasted to their respective reference trees.

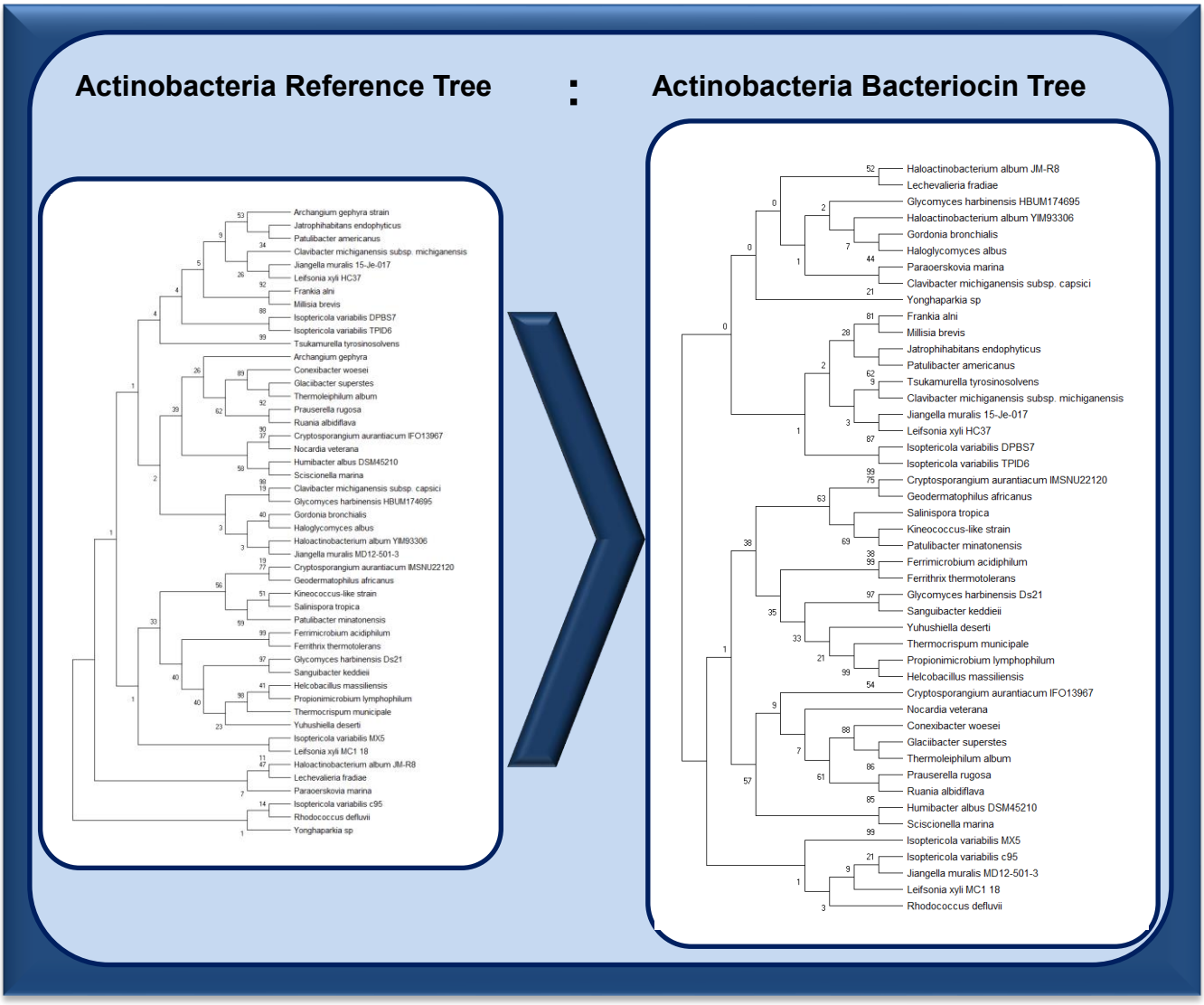

Figure 3.6: Comparison of the Actinobacteria reference tree of the 16s rRNA sequences of bacteriocin producer species with the Actinobacteria bacteriocin tree

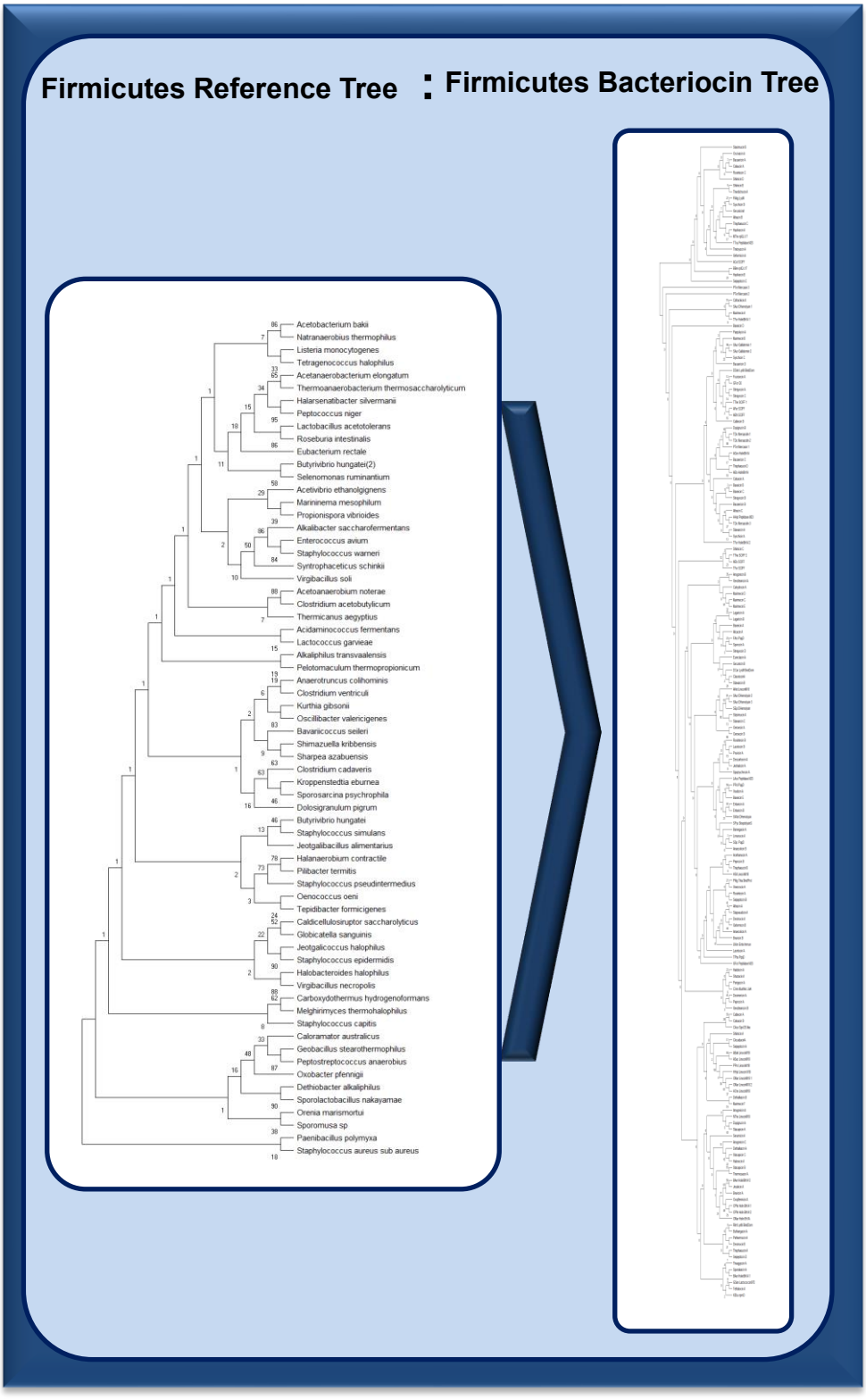

Figure 3.7: Comparison of the Firmicutes reference tree of the 16s rRNA sequences of bacteriocin producer species with the Firmicutes bacteriocin tree

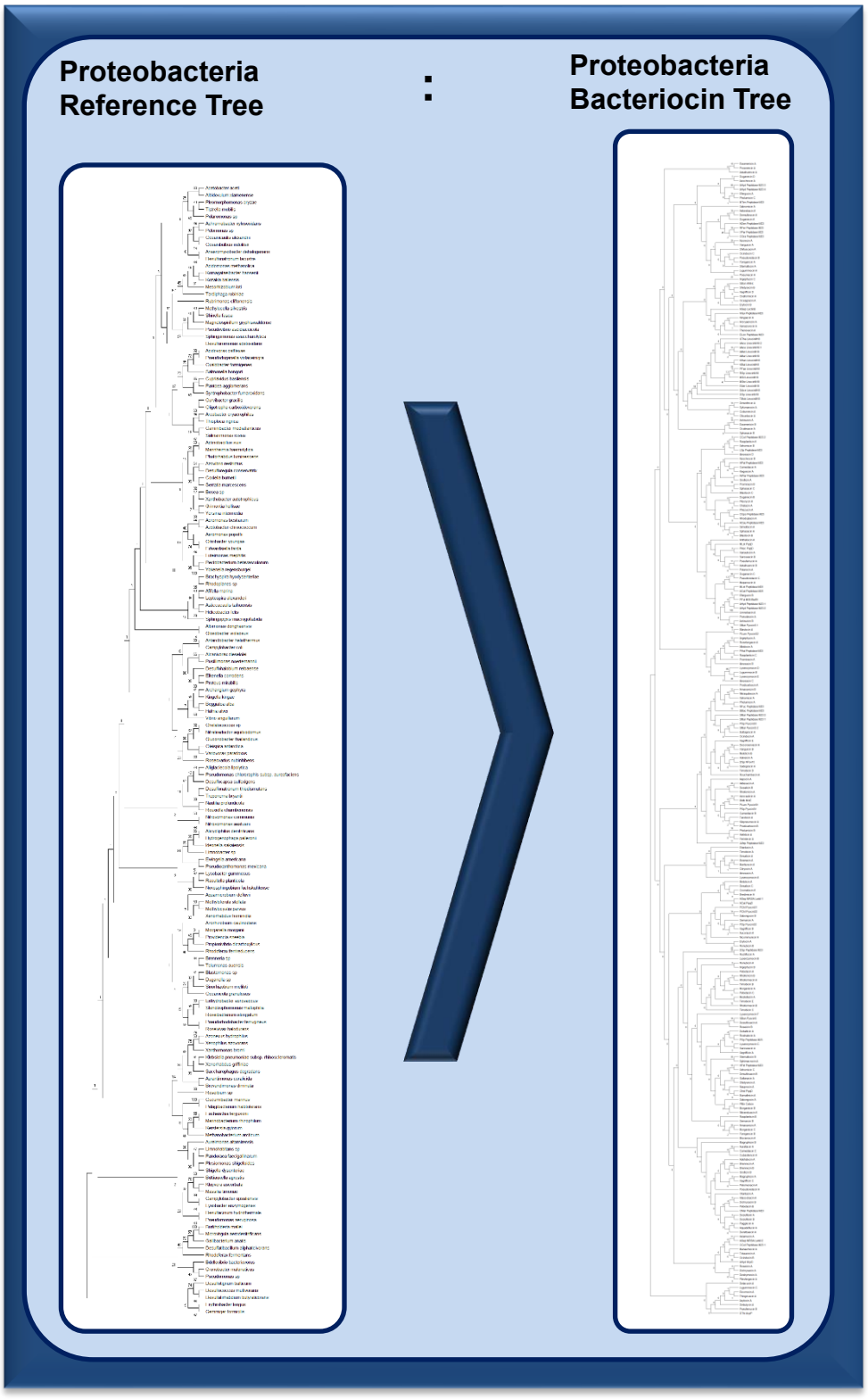

Figure 3.8: Comparison of the Proteobacteria reference tree of the 16s rRNA sequences of bacteriocin producer species with the Proteobacteria bacteriocin tree

Evolutionary analysis of members of the Actinobacteria, Firmicutes and Proteobacteria phyla was used to infer phylogenetic relationships among members within each phylum. In the Actinobacteria most of the class I Lanthipeptides, Class III bacteriocins and Head-to-tail cyclic peptides, respectively, clustered together. Other bacteriocin types were distributed widely on the tree with no obvious common ancestry. A Class III bacteriocin member, linocin M18, produced by different species, as expected cluster together. When compared to the reference tree, a possible common recent ancestor for these species was confirmed. Bacteriocins of *Clavibacter michiganensis* subspecies *michiganensis* unexpectedly cluster with *Haloactinobacterium album* and *Yuhushiella deserti* bacteriocins. *Y. deserti* was found to be unrelated to these species which suggests horizontal gene transfer as opposed to natural selection to form this phylogenetic pattern.

The Lasso peptides in the Firmicutes phylum are widely distributed among species within this phylum. This is clearly evident from the phylogenetic tree and suggests no visible inheritance pattern. The majority of mined bacteriocins in this phylum belonged to either class II bacteriocins or Sactipeptides. Interestingly Head-to-tail cyclic peptides and Sactipeptides often associated in clusters. Class II bacteriocins were mainly divided in 6 clusters with 2 or more members each. The majority of members of the Lanthipeptide classes could not be traced to a common ancestor.

Class III and Bottromycins from the Proteobacterial species, respectively clustered together as opposed to their producer species clustering together on the reference tree. In this phylum Bottromycins and Class III uniquely clustered together. Only 2 Cyanobactins were identified in the Proteobacterial phylum. Microcins identified were widely distributed on the bacteriocin tree and didn't correlate with producing species on the reference tree. This result could point to horizontal gene transfer events rather than genetic transfer from a common ancestor. The class II Lanthipeptides forms a single cluster, with a few outliers, none of which can be traced to recent common ancestry.

3.5 Discussion

A total of 617 genomes were selected for physiochemical and evolutionary analysis from the Actinobacteria, Firmicutes and Proteobacteria phyla. Genome mining located various putative bacteriocins from DNA sequences of the selected species within phyla. A local alignment search was performed to identify each putative bacteriocin.

Determining the identity of bacteriocins allowed for us to assign the novelty of the bacteriocins within the respective phyla. From 114 putative bacteriocins found in the Actinobacteria phylum, 87 were characterised as novel. The majority of identified bacteriocins (18%) belonged to Class III bacteriocins. Not surprisingly, most previously characterised Actinobacteria bacteriocins are Class III members, either linocin M18 or peptidase M23. Within the phylum, Firmicutes, 196 putative bacteriocins were identified of which ~61% were characterised as novel. Class II bacteriocins and Sactipeptides were the most prevalent of those identified as novel within this phylum. Class II bacteriocins are confirmed to be the largest class of bacteriocins within this phylum (Jakubovics *et al.*, 2014). The largest phylum analysed, 373 Proteobacteria species, yielded 280 putative bacteriocins located by genome mining were those from the Proteobacteria. A total of 201 of these bacteriocins have not yet been characterised. The majority of novel bacteriocins were represented by 105 Class III bacteriocins, 39 Bottromycins and 40 Microcins. Most novel bacteriocins within this phylum belonged to Class III bacteriocins. These bacteriocins are identified by genome mining solely on sequence homology since they are generally well conserved among species. The degree of scientific research focused on bacteriocin classes are well illustrated by the BAGEL database. Only 93 Class III members are listed, as opposed to 525 Class I and 851 Class II members (de Jong *et al.*, 2006). Unlike the Class I and Class II bacteriocins, Class III was only recently suggested to be employed in the industry possibly due to their sensitivity to proteolytic degradation (Parada *et al.*, 2007). An example of a Class III bacteriocin, megacin, is currently used in the industry to treat respiratory tract and urinary tract infections as well as skin and soft tissue infections (Abriouel *et al.*, 2011, MIMS Australia, 2015).

Physiochemical characterisation of the novel bacteriocins will elucidate favourable environmental conditions for bacteriocin production and function.

Physiochemical properties explored include molecular weight, isoelectric point, extinction coefficient, instability index, aliphatic index, hydropathy index, half-life *in vitro* in mammalian and *in vivo* in yeast and bacterial cells. All classes analysed of the Actinobacteria phylum are classified to be unstable. Molecular weight ranged widely among classes with most members having an isoelectric point falling within the acidic range. In contrast to the other classes, LAPs are predominantly hydrophilic rather than hydrophobic. This hydrophilic tendency is uncharacteristic as most bacteriocins lyse cells by pore formation or compromising membrane integrity. This feature requires either a hydrophobic or an amphipathic nature. Designated transporters into the target cell have not yet been identified for this group (de Lima and Filho, 2005). From 196 mined bacteriocins produced by members of the Firmicutes phyla, 120 bacteriocins are characterised as novel. Similar to Actinobacteria, they tend to be hydrophobic in nature. In contrast to Actinobacteria, the majority of Firmicutes bacteriocins are stable and have an isoelectric point within the basic range. The molecular weight of these bacteriocins is smaller than the average molecular weight for these types of bacteriocins. The majority of Sactipeptides were unstable, which is tin contrast to what was found in the other bacteriocin types.

Proteobacteria was the largest phylum analysed, in which 280 bacteriocins were mined with ~72% identified as novel. The novel bacteriocins identified were mainly representatives of the Class III bacteriocins, Microcins, Lasso peptides and Sactipeptides. The novel bacteriocins produced by Proteobacteria species are distributed evenly with respect to stability. Similar to the Firmicutes the majority of identified bacteriocins here have an isoelectric point that falls within the basic range and are hydrophilic. The 9 novel LAPs identified in this study seems intriguing. LAPs are previously reported to be made by species belonging to the Firmicutes and Spyrochaetes, but only now found to be produced by species of the Actinobacteria and Proteobacteria phyla (Letzel *et al.*, 2014). Literature suggests that Thiopeptides are only produced by species of the Actinobacteria and Firmicutes phyla, which is confirmed by this study as no Thiopeptides were identified within the Proteobacteria phylum (Arnison *et al.*, 2013).

The physiochemical properties of the novel bacteriocins mined from producing species of the Actinobacteria, Firmicutes and Proteobacteria phyla are reported in this study. These results can guide the full characterisation of these novel bacteriocins in the laboratory by providing a framework of known physiochemical properties. Elucidation of the tertiary structures can assist in providing a putative mode of action.

Novel bacteriocins were allocated putative 3D structures by homology modelling. Low bacteriocin sequence coverage, evident throughout the results, cannot accurately portray model quality as bacteriocin sequences may only partially cover the complete bacteriocin sequence of the putative models. This in turn will decrease the GMQE decreasing the threshold of interest to a GQME of 0.5. Since the mined bacteriocins did not exhibit sufficient DNA sequence similarity to other known sequences for identification, it is expected that these novel peptides will not show homology on protein level either. All novel bacteriocins of the Actinobacteria, Firmicutes and Proteobacteria phyla were subjected to homology modelling and a few intriguing results are discussed below.

The bacteriocin arobicin A, produced by the Actinobacteria species, *Actinospica robiniae*, was modelled with a high GMQE of 0.59, to a designed helical repeat protein DHR10. These types of proteins participate in molecular recognition and signalling (Brunette *et al.*, 2015). The naturally synthesised bacteriocin, arobicin A, may have similar functionality and mode of action as the designed proteins and can decrease the time and cost of production for applications similar to those of the designed DHR10 proteins.

The putative michigacin A sequence resembles that of the structural gene of the enterococcal peptide antibiotic AS-48 with a GQME score of 0.82. The bacteriocin lyses cells by pore formation (Martinez-Bueno M. *et al.*, 1994). The high aliphatic index (123.15) of the globular protein, michigacin A, may signify high thermal stability which sparks interest as a potential drug target (Ikai, 1980).

Prarugocin B was modelled to the aldehyde oxidase subunit of a flavoprotein. The putative novel bacteriocin might resemble the aldehyde oxidase subunit function by altering substrate binding within the inter membrane of the target cell. The predicted mode of action for prarugocin B is further supported by its physiochemical properties.

Its hydrophobicity allows for membrane insertion and the pI of 5.0 is similar to the periplasm pH of 5.5, ensuring optimal function under these conditions (Wilks and Slonczewski, 2007). A putative bacteriocin, sakeddicin A, might interfere with transcriptional regulation, inducing cell apoptosis, as opposed to membrane interactions. Yudesericin C is modelled with a GQME of 0.50 to fusion glycoprotein F0. The glycoprotein allows fusion of the viral capsid with the host cell membrane (Wong *et al.*, 2016). The hydrophobic nature and stability of yudesericin C supports the putative function, to facilitate entry of anti-microbial peptides into a target cell, similar to the viral fusion glycoprotein.

The 120 novel bacteriocins identified in the Firmicutes phylum by genome mining, were subjected to homology modelling. A total of 37 bacteriocins did not generate any hits in the SwissModel database. Putative proteins related to the well-studied Class III member, peptidase M23, were predicted in 5 novel producing species; *Actinospica transvaalensis*, *Butyrivibrio hungatei*, *Caloramator australicus*. *Caloramator silvermanii* and *Thermacetogenium phaeum*. Head-to-tail cyclic peptides were well represented in the Firmicutes phylum, of which 17 of these mined bacteriocins were identified as novel. The first member of this class described was bacteriocin AS-48 (also now known as enterocin AS-48) in 1986, only later followed by NRK-5-3-B and acidocin B. The AS-48 bacteriocin functions similar to enterocin, causing target cell lysis by pore formation. Novel producing strains include *Bacillus aerophilus*, *Caldicellulosiruptor bescii*, *Caldicellulosiruptor saccharolyticus*, *Geobacillus stearothermophilus* and *Oenococcus oeni*. Bacaericin A, Bacaericin B and Oenoecin B were modelled to NKR-5-3-B with high GQME scores of 0.81, 0.79 and 0.54, respectively. The NKR-5-3B circular bacteriocin elicits broad spectrum antimicrobial activity (Himeno *et al.*, 2015). Study results on open reading frames (ORFs) functions of the NKR-5-3-B locus included an ORF closely related the arsenical resistance 3 family, suggesting an additional role in the protection of the producing strain (Perez *et al.*, 2016). Homology modelling associated desmericin A with TrfB transcriptional repressor protein with a GQME of 0.52, but only 6.45% sequence coverage. The low percentage sequence coverage might be due to the small sequence query input only comprising a small part of the complete *TrfB* gene. The model of acidocin B closely related that of the novel bacteriocin, gecunicin B, suggesting a possible variant of acidocin B.

Variants of bacteriocins arise from environmental factors, gene duplication and subsequent diversification or random encounter (Arber, 2014).

In the phylum, Proteobacteria, 6% of the 201 novel bacteriocins yielded homology models. Several bacteriocins including axylocin A, aucoracin A, kegyiocin A manaemocin A, pholumicin A, pseudoviolacin A and thenovecin A supports the proposed existence of the respective hypothetical proteins with an average GMQE of 0.53. All hypothetical protein models are uncharacterised and have unknown function, supplying an interesting study for future research. The peptidase M23 (putative) proteins were suggested 28 times as a possible model matches for various novel bacteriocins. Bathuricin B was modelled with a GMQE of 0.71 to a pesticidal crystal protein produced by *Bacillus thuringiensis*, Cry6Aa, known to kill target cells of similar bacteria strains, coleopteran insects and nematode parasites by pore formation (Dementiev *et al.*, 2016). The model for integration host factor, responsible for stabilizing DNA during replication, transcription and recombination, was placed in parallel with desalicin C with a high GMQE of 0.76. Sequence coverage for this model amounted to 39.13%, possibly suggesting that desalicin C might be a mutated form on the integration host factor protein and can bind to the respective binding sites without enhancing DNA stability, thus halting DNA replication and transcription and causing cell apoptosis. Desalicin C might add to the growing literature which describes various bacteriocins acting on a genomic level rather than pore formation (Gavrish *et al.*, 2014, Kobayashi *et al.*, 2010, Mathavan *et al.*, 2014). Moganiicin C and promiracin A were modelled to colicin E3, a well-studied bacteriocin that degrades the 16s rRNA subunit of a target cell, halting protein synthesis and causing cell death (Zarivach *et al.*, 2002). Novel bacteriocins with suspected DNAse activity, promiracin B and rRNAse activity, xegrifficin B-C and xehomicin A, were also identified. Homology modelling determined pholumicin B similar to antitoxin DinJ with a high GMQE of 0.78. This result implores further research on connections between these seemingly unrelated proteins. TOP7, a protein synthesized *de novo* by computer simulations and confirmed with experimental data obtained fame for its uncharacteristic non-cooperative tertiary novel structure (Zhang and Chan, 2009). The TOP7 variant, M7, was designed to mimic the structure of TOP7, but incorporate typical cooperative folding as seen in naturally occurring proteins (Stordeur *et al.*, 2008). Yoregencin A was modelled to the M7 structure, not only demonstrating the precision of structure prediction technology, but also supplying

a natural host for future studies on the M7 protein's capabilities. In addition to the structures and inferred functions of the putative novel bacteriocins, the evolutionary study will shed more light on the origin of these antimicrobial peptides.

The phylogeny of all novel bacteriocins were scrutinised to potentially elucidate the origin of this wide variety of antimicrobial peptides. Types of bacteriocins were studied separately within all respective phyla. Within the phylum, Actinobacteria, 114 putative bacteriocins were identified. A single Class II peptide, Class III peptide and only 2 Bottromycins were identified within this phylum with no specific clustering between bacteriocins or species. These sample sizes doesn't permit comprehensive analysis and must be repeated with more samples for accurate results. The Head-to-tail cyclic peptides and Linaridins of the Actinobacteria species were generally unevenly distributed and were most likely obtained through horizontal gene transfer.

The only Head-to-tail cyclic bacteriocins clustered from a common ancestor species were isovaricin A and michigacin M as well as lefradiacin B, yonghacin A and glyharcin A-D respectively. In Class III bacteriocins, 2 clusters of linocin M18 peptides evolved separately. A large linocin M18 cluster evolved by synapomorphy and the small cluster by horizontal gene transfer to a common ancestor. *Jiangella muralis* that produces the linocin M18 outlier does not cluster with any of the associated producing species, suggesting that a horizontal gene transfer event may have occurred. Other class III bacteriocins show no specific pattern of inheritance and appear to be the result of horizontal gene transfer. The class I Lanthipeptides suggested common ancestry only for *Sciscionella marina* and *J muralis* and the respective bacteriocins. None of the other analysed class I Lanthipeptides cluster, implicating horizontal gene transfer occurred from neighbouring species. Interestingly, lefradiacins, produced by *Lechevalieria fradiae*, are highly dispersed and show no correspondence, other than the producing species. This possibly implies that the *L. fradiae* genome results from a wide variety of horizontal gene transfer events. The class II Lanthipeptides of *C. michiganensis* subspecies *michiganensis* forms a clear clade. Other bacteriocins of this class show no definitive pattern of inheritance. None of the analysed class III and class IV Lanthipeptides have a general evolutionary structure. These bacteriocins and their producing species do not cluster and are concluded to be the product of horizontal gene transfer.

LAPs produced by *C. michiganensis* stem from a common ancestor as expected, but no other LAPs followed suite. The Lasso peptides seem to have evolved from 3 common ancestors, which acquired their genes by horizontal gene transfer. All Sactipeptides, with the exception of mibrecin A and facidacin A, show no correlation on the species reference tree. The majority of Thiopeptides are widely distributed and arose by horizontal gene transfer. Only satrocin A and paminotocin A Thiopeptides cluster, suggesting common ancestry.

The 196 putative bacteriocins produced by Firmicutes species comprised of wide variety of bacteriocin classes. Like the Actinobacteria phylum, no inheritance pattern can be determined for the 2 Bottromycins and 1 class IV Lanthipeptide due to insufficient amounts data. Species producing Head-to-tail peptides obtained their genes predominantly by horizontal gene transfer. Only the bacteriocins produced by *O. oeni* suggest a recent common ancestor. The largest Firmicutes class analysed, Class II bacteriocins, contained 6 independent clusters. The 6 distinct clusters might suggest evolutionary events triggered horizontal gene transfer of these bacteriocins. Evolutionary events to cause an increase in the amount of times horizontal gene transfer takes place include viral or phage infection, structural flexibility and environmental mutagens (Arber, 2014). Stasimucin B first diverted from the initial ancestor where after the other Class II members followed. For the majority of cases the producing species of the clustered bacteriocins are not related, suggesting horizontal gene transfer events. Baseicin B and baseicin C seem to be the result of recent gene duplication. Bacterial gene duplication contributed significantly to the current functional diversity of bacteria (Serres *et al.*, 2009). Common ancestry is identified for oxopfennicin A and holing Bh1A produced by *Orenia marismortui*. Jeoalicin A, brevicin A and oxopfennicin A, clustered with holin Bh1A, but did not share similar physiochemical parameters, suggesting a genetically similar, but functionally different bacteriocin. A well-studied Class III member, linocin M18, strangely did not share common ancestry, implying that most of the producing species obtained this gene by horizontal gene transfer. Similar to linocin M18, the majority of the novel Class III bacteriocins arose from convergent evolution. The producing species of clustered Lanthipeptides, LAPs and Lasso peptides showed no relation on the reference tree.

Only *Paenibacillus polymyxa*, *S. Aureus* and *Geobacillus stearopthermophilus* as well as the mirimecins (C-E) produced by *Marininema mesophilum*, respectively, exhibited common ancestry.

Sactipeptides are abundant in the Firmicutes phylum. Common ancestry as well as horizontal gene transfer is suggested to result in the 2 clusters of SCIFF bacteriocins. Sactipeptides inexplicably often associates with Class III and Head-to-tail bacteriocins. The unknown bacteriocins identified by genome mining, bacaericin A and bacaericin B, are suggested to belong to the Head-to-tail class of bacteriocins by homology modelling. Physiochemical properties of the unknown bacteriocins compared to the average value of properties like the molecular weight, hydrophobicity and pI of Head-to-tail bacteriocins do not support this claim. Further research with regards to these unknown bacteriocins may prove interesting and supply valuable insight on their origin and function.

Proteobacteria with 373 species was the largest phylum analysed for physiochemical properties and evolutionary relationships. A total of 201 novel bacteriocins were identified in 170 Proteobacteria producing species. This result supports the literature which suggests that bacteriocin producing Proteobacteria species are less prevalent than those of Firmicutes (55% and 46% respectively). This study result however insists on a lower degree of production frequency difference of 12.64% compared to 49.30% in literature (Drissi *et al.*, 2015). Surprisingly very few bacteriocin producing species cluster as suggested by the associated bacteriocins tree clusters. Of the 45 Bottromycins identified in this study, 41 are produced by Proteobacteria. These often cluster with Class III bacteriocins, Microcins and Lasso peptides. In comparison to the other phyla, Proteobacteria yielded very few Head-to-tail peptides and Class II bacteriocins. Only 1 class II Lanthipeptide, 2 Class III peptides and 1 class IV Lanthipeptide was identified within this phylum. The amount of data does not permit accurate phylogenetic analysis. The class II Lanthipeptides have several novel bacteriocins identified, of which only lysenzymocin B, D and E can be traced to a common ancestor. Proteobacteria is the only phyla identified that contains novel Cyanobactins and Microcins, of which none of the producing species are evolutionary related. Members of the Sactipeptides, LAPs and Lasso peptide classes, respectively, do not share common ancestry either. This lack of common

ancestry tends to support the claim that the phenomenon of gene duplication and subsequent diversification is highly underestimated in bacteriocin evolution studies (Serres *et al.*, 2009). Class III bacteriocins contribute the majority of novel peptides in the Proteobacteria phylum.

Only members of 3 clusters share common ancestors respectively. Well studied Class III members studied include peptidase M23, pyocinS and linocin M18. All 15 linocin M18 peptides studied formed a single cluster, but species only correlate in groups of average 2 species on the reference tree. This result suggests various horizontal gene transfer and subsequent duplication events. The pyocinS producing species, *Photorhabdus luminescens*, *Serratia marcescens* and *Pseudomonas sp*, each contains 2 or more pyocinS genes. The genes for each species however do not associate on the phylogenetic tree, suggesting that gene duplication and subsequent considerable diversification caused this unexpected severance. Azochrocin A, a Class III member, joins the peptidase M23 cluster on the bacteriocin tree. Model results and physiochemical analysis confirms that azochrocin A might by a type of peptidase M23. Since the BLAST identification failed to draw the same conclusion, it stands to reason that azochrocin A might be a novel bacteriocin genetically related, but structurally and functionally different, to the Peptidase M23 family. Further investigation might not only clarify the situation regarding this bacteriocin, but also assist in elucidating the process of gene diversification and evolution of bacteriocins in general. Sadegracin A, the unknown bacteriocin identified by genome mining, does not show any physiochemical relationship to bacteriocins of clustering species, nor relevant models with a GQME above 0.50. Further analysis of this seemingly foreign bacteriocin, sadegracin A, can potentially supply another piece of the puzzle to the greatly unexplored world of antimicrobial peptides.

Chapter 4: Discussion and conclusion

The rapidly expanding variety of infectious diseases and increased pathogenic activity in the food and agricultural sectors intensifies despite desperate efforts to combat these crises. Bacteriocins are secreted peptides that kill related strains within the microenvironment to increase their fitness. Sakacin and nisin, food preservatives and megacin, a treatment for respiratory tract and urinary tract infections are a few examples currently employed in the industry (Abriouel *et al.*, 2011, López-Cuellar *et al.*, 2016, MIMS Australia, 2015). To assist in combating pathogen-related problems, this project focussed on quantitative analysis of the genomes of 921 bacteriocin producing species (35 656 Genbank entries) and yielded a total of 819 bacteriocins. Genome mining, physiochemical properties and evolutionary relationships were inferred for a total of 617 bacteriocins of which 407 bacteriocins are described for the first time in this study. Very little literature is available on quantitative bacteriocin studies and evolutionary relationship analyses. Hopefully this project will encourage further research in this intriguing field of study.

The genome mining performed comprised species of the Actinobacteria (87), BV4 (140), Firmicutes (157), Glidobacteria (62) and Proteobacteria (373) phyla. Bacteriocin class prevalence was analysed for each phyla respectively. The 932 species analysed yielded 819 bacteriocins. Class III bacteriocins contributed by far the majority, 25%, of the total mined bacteriocins, followed by 13% Sactipeptides and 10% Head-to-tail cyclic peptides. Class IV Lanthibiotics, Glyocins, Class I bacteriocins and unknown bacteriocins contributed less than 1% to the total of identified bacteriocins. Most bacteriocin classes are represented within all phyla, except for Glyocin identified only in the BV4 phylum, class I lanthibiotics, only identified within the Actinobacteria and BV4Glidobacteria phyla, and Microcins that were only identified within the Glidobacteria and Proteobacteria phyla. The diversity observed from these results led to a more narrow focus on mined bacteriocins within the selected Actinobacteria, Firmicutes and Proteobacteria phyla.

All respective classes represented in the Actinobacteria, Firmicutes and Proteobacteria phyla that were mined for, were subjected to a BLAST search to determine their identity and subsequent novelty. This study contributes 87 novel bacteriocins produced by Actinobacteria species, 120 Firmicutes novel bacteriocins and 201 novel bacteriocins from the Proteobacteria phylum to literature. Physiochemical properties were determined to assist in future research of these novel bacteriocins by laboratory procedures. Additional analyses consisting of homology modelling and evolutionary analysis were performed for all the novel bacteriocins. Homology modelling performed on 597 bacteriocin sequences suggested a diverse group of putative hits.

A total of 45 Bottromycins are identified among the Actinobacteria, Firmicutes and Proteobacteria phyla. Inexplicably the majority of Bottromycins, 41 found in the Proteobacteria phylum, tend to associate with Class III bacteriocins. None of the Bottromycins, or the species clustered on the bacteriocin tree share common ancestry. Homology modelling suggested the acidocin B model, a Class II bacteriocin, with a GMQE of 0.52 and 41.38% sequence coverage. Physiochemical properties could not clarify the probability of this suspicion as the Class II parameters are too diverse. As previously suspected, the class mostly represented within the Firmicutes phylum are class II bacteriocins (Jakubovics *et al.*, 2014). The majority of the 27 identified Class II bacteriocins belong to the Firmicutes phylum with only a single representative in Actinobacteria and 3 members in the Proteobacteria phylum. Firmicutes reveals 6 distinct clusters, suggesting triggering events that increase the frequency of horizontal gene transfer. Triggering events can include viral or phage infection, structural flexibility and environmental mutagens (Arber, 2014).

The 10 novel class III Lanthibiotics, 5 Linaridins and 6 Thiopeptides were only made by bacteriocin producing species in the Actinobacteria phylum. These classes, respectively, are widely dispersed in evolutionary time and clusters, with no apparent common ancestry, possibly suggesting a series of horizontal gene transfer events to yield this result. Thiopeptides are described in literature to be produced by bacteriocin producing members of the Actinobacteria and Firmicutes phyla (Arnison *et al.*, 2013). This study confirms the Thiopeptides presence in Actinobacteria producing species and the absence in members of the Proteobacteria phyla.

The bacteriocin producing Firmicutes species did not comply to provide Thiopeptides as expected from literature (Arnison *et al.*, 2013). All analysed phyla are represented by selected species, a list that is by no means exhaustive. Further research with different representative members of these phyla may prove valuable to understand not only their distribution among phyla, but also their seemingly unrelated origin.

The Class III bacteriocins are most abundant and most prevalent among all phyla, donating 57 novel members to literature. This class also dominates the other classes found in Actinobacteria and Proteobacteria respectively. Various new producing strains were identified for the well-studied members, peptidase M23 and linocin M18. linocin M18 have (a) prominent(s) in the Actinobacteria and Proteobacteria phyla, but do not exhibit any consistent evolutionary clusters in the Firmicutes phylum. This result may suggest that Firmicutes members were last of the studied phyla to obtain this gene by horizontal gene transfer from the other phyla, which shared an ancient common ancestor. PyocinS producing species may contain up to 2 genes of the same bacteriocin. Evolutionary analyses surprisingly yielded separate clusters for pyocinS genes of a single species. Upon further investigation, literature suggests that this lack of common ancestry may suggest an early gene duplication event. The phenomenon of gene duplication and subsequent diversification is highly underestimated in bacteriocin evolution studies due to the inability to draw parallels between the ancestral genes and the diversified gene sequences (Serres *et al.*, 2009). Lasso peptides and LAPs were represented within all analysed phyla. The identified Lasso peptides stemmed from 3 different ancestors in the Actinobacteria phylum, but showed no other pattern in the other phyla. LAPs were previously known to be made by species belonging to the Firmicutes and Spyrochaetes (Letzel *et al.*, 2014). This study discovered putative members of the Proteobacteria and Actinobacteria with the potential to produce LAPs, adding to the abundance of these peptides among phyla.

Novel Cyanobactins and Microcins were only identified within the Proteobacteria phylum producing species. It is however known that Cyanobactins are mainly produced by Cyanobacteria, which did not partake in this study, and Microcins also produced by Actinobacteria species (Rebuffat, 2012, Schmidt and Donia, 2009).

All 40 of the Microcins identified were novel with molecular weight ranging from 3 235.7 Da–11 596.25 Da. The majority of Microcins have basic pI's, have a hydrophilic nature, are unstable and have an estimated half-life of 30h, >20h and >10h in mammalian cells, yeast cells and *E. coli* cells respectively. Intriguingly, genome mining located 1 unknown bacteriocin in the Proteobacteria phylum and 2 unknown bacteriocins in the Firmicutes phylum. BLAST found no genomic hits to known bacteriocins. Homology modelling and physiochemical parameter determination was employed to further explore their identities. The unknown bacteriocin, sadegracin A, produced by the Proteobacteria species, *Saccharophagus degradans*, did not yield any suggested homology models, nor could its class be identified or speculated by physiochemical parameter comparison. These results might suggest this bacteriocin to be a new type of bacteriocin of a new class, definitely worthy of further study. The unknown bacteriocins in the Firmicutes phylum, bacaericin A and bacaericin B, are both produced by the *B. aerophilus* species. After failing to identify these on genomic level, homology modelling suggested these to form part of the Head-to-tail cyclic bacteriocin class. This however was contrasted by the Head-to-tail cyclic peptide parameters, encouraging further laboratory studies to elucidate their nature and function. Identification and characterisation of sadegracin A, bacaericin A and bacaericin B could prove an interesting future study, potentially able to provide a wealth of new information to known bacteriocin literature.

In conclusion, the rising amount of bacterial drug resistance pathogens and decreasing food safety and security scores stimulated intense research efforts to increase the quality of human life. Bacteriocins peaked research interest as possible solutions to these crises due their wide variety of applications and their GRAS (generally regarded as safe) status. Bacteriocins are also recruited for other applications including, but not limited to, cancer treatment, contraception, oral care, veterinary and agricultural applications. The majority of the 407 novel bacteriocins identified in this study can serve as promising putative candidates for applications in various sectors, paving the way for future research. The elucidated homology models, physiochemical parameters and evolutionary relationships inferred will assist future laboratory practise by providing a point of reference for the respective bacteriocins of interest.

The structural, chemical and evolutionary diversity of the identified novel bacteriocins supplies further studies with a pool of properties to choose from for their project's respective requirements.

References

Abriouel, H., Franz, C. M. a. P., Omar, N. B. & Gálvez, A. 2011. Diversity and applications of Bacillus bacteriocins. *Federation of European Microbiological Societies Microbiology Reviews,* 35**,** 201-232.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. 1990. Basic local alignment search tool. *Journal of Molecular Biology,* 215**,** 403-10.

Arber, W. 2014. Horizontal Gene Transfer among Bacteria and Its Role in Biological Evolution. *Life (Basel),* 4**,** 217-24.

Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., Camarero, J. A., Campopiano, D. J., Challis, G. L., Clardy, J., Cotter, P. D., Craik, D. J., Dawson, M., Dittmann, E., Donadio, S., Dorrestein, P. C., Entian, K.-D., Fischbach, M. A., Garavelli, J. S., Göransson, U., Gruber, C. W., Haft, D. H., Hemscheidt, T. K., Hertweck, C., Hill, C., Horswill, A. R., Jaspars, M., Kelly, W. L., Klinman, J. P., Kuipers, O. P., Link, A. J., Liu, W., Marahiel, M. A., Mitchell, D. A., Moll, G. N., Moore, B. S., Müller, R., Nair, S. K., Nes, I. F., Norris, G. E., Olivera, B. M., Onaka, H., Patchett, M. L., Piel, J., Reaney, M. J. T., Rebuffat, S., Ross, R. P., Sahl, H.-G., Schmidt, E. W., Selsted, M. E., Severinov, K., Shen, B., Sivonen, K., Smith, L., Stein, T., Süssmuth, R. D., Tagg, J. R., Tang, G.-L., Truman, A. W., Vederas, J. C., Walsh, C. T., Walton, J. D., Wenzel, S. C., Willey, J. M. & Van Der Donk, W. A. 2013. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Natural product reports,* 30**,** 108-160.

Auta, A., Hadi, M. A., Oga, E., Adewuyi, E. O., Abdu-Aguye, S. N., Adeloye, D., Strickland-Hodge, B. & Morgan, D. J. 2019. Global access to antibiotics without prescription in community pharmacies: A systematic review and meta-analysis. *Journal of Infection,* 78**,** 8-18.

Bagley, M. C., Dale, J. W., Merritt, E. A. & Xiong, X. 2005. Thiopeptide Antibiotics. *Chemical Reviews,* 105**,** 685-714.

Barreto, M. L., Teixeira, M. G. & Carmo, E. H. 2006. Infectious diseases epidemiology. *Journal of Epidemiology and Community Health,* 60**,** 192-195.

Bastos Mdo, C., Coelho, M. L. & Santos, O. C. 2015. Resistance to bacteriocins produced by Gram-positive bacteria. *Microbiology,* 161**,** 683-700.

Baylor College of Medicine. 2016. *Emerging Infections and Biodefense* [Online]. One Baylor Plaza, MS: BCM-385

Houston, TX 77030 Baylor College of Medicine. [Accessed 2019].

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Wheeler, D. L. 2005. GenBank. *Nucleic Acids Research,* 33**,** D34-D38.

Beukes, M. & Hastings, J. W. 2001. Self-Protection against Cell Wall Hydrolysis in Streptococcus milleri NMSCC 061 and Analysis of the Millericin B Operon. *Applied and Environmental Microbiology,* 67**,** 3888-3896.

Bezza, F. A., Beukes, M. & Chirwa, E. M. N. 2015. Application of biosurfactant produced by Ochrobactrum intermedium CN3 for enhancing petroleum sludge bioremediation. *Process Biochemistry,* 50**,** 1911-1922.

Bhardwaj, A., Gupta, H., Kapila, S., Kaur, G., Vij, S. & Malik, R. K. 2010. Safety assessment and evaluation of probiotic potential of bacteriocinogenic Enterococcus faecium KH 24 strain under in vitro and in vivo conditions. *International Journal of Food Microbiology,* 141**,** 156-164.

Bode, H. B. & Muller, R. 2005. The impact of bacterial genomics on natural product research. *Angew Chem Int Ed Engl,* 44**,** 6828-46.

Broude, S., Miller, C., Stokes, M., Jones, A., Kasper, S., Entin, L., Whiting-Jones, K. & Weilerstein, B. 2015. The problem with pesticides. *In:* Entin, L. (ed.) *Oranizing with residents to clean up and prevent pollution in New England since 1987.* USA.

Brunette, T. J., Parmeggiani, F., Huang, P.-S., Bhabha, G., Ekiert, D. C., Tsutakawa, S. E., Hura, G. L., Tainer, J. A. & Baker, D. 2015. Exploring the repeat protein universe through computational protein design. *Nature,* 528**,** 580-584.

Casey, P. G., Gardiner, G. E., Casey, G., Bradshaw, B., Lawlor, P. G., Lynch, P. B., Leonard, F. C., Stanton, C., Ross, R. P., Fitzgerald, G. F. & Hill, C. 2007. A five-strain probiotic combination reduces pathogen shedding and alleviates disease signs in pigs challenged with Salmonella enterica Serovar Typhimurium. *Applied and Environmental Microbiology,* 73**,** 1858-63.

Cursino, L., Smajs, D., Smarda, J., Nardi, R. M., Nicoli, J. R., Chartone-Souza, E. & Nascimento, A. M. 2006. Exoproducts of the Escherichia coli strain H22 inhibiting some enteric pathogens both in vitro and in vivo. *Journal of Applied Microbiology,* 100**,** 821-9.

Daugelaite, J., Driscoll, A. O. & Sleator, R. D. 2013. An Overview of Multiple Sequence Alignments and Cloud Computing in Bioinformatics. *ISRN Biomathematics,* 2013.

Dawid, S., Roche, A. M. & Weiser, J. N. 2007. The blp bacteriocins of Streptococcus pneumoniae mediate intraspecies competition both in vitro and in vivo. *Infection and Immunity,* 75**,** 443-51.

De Jong, A., Van Hijum, S. A., Bijlsma, J. J., Kok, J. & Kuipers, O. P. 2006. BAGEL: a web-based bacteriocin genome mining tool. *Nucleic Acids Res,* 34**,** W273-9.

De Lima, E. T. & Filho, R. L. A. 2005. Bacteriocins: Nomenclature, detection, mechanism of action and potential use in poultry production. *Food, Agriculture & Environment,* 3**,** 62-66.

Dementiev, A., Board, J., Sitaram, A., Hey, T., Kelker, M. S., Xu, X., Hu, Y., Vidal-Quist, C., Chikwana, V., Griffin, S., Mccaskill, D., Wang, N. X., Hung, S. C., Chan, M. K., Lee, M. M., Hughes, J., Wegener, A., Aroian, R. V., Narva, K. E. & Berry, C. 2016. The pesticidal Cry6Aa toxin from Bacillus thuringiensis is structurally similar to HlyE-family alpha pore-forming toxins. *BMC Biol,* 14**,** 71.

Department of Social Development 2013. National policy on food and nutrition security. *In:* Department of Agriculture, Forestry & Fisheries, (ed.). South Africa.

Di Cagno, R., De Angelis, M., Calasso, M. & Gobbetti, M. 2011. Proteomics of the bacterial cross-talk by quorum sensing. *Journal of Proteomics,* 74**,** 19-34.

Di Cagno, R., De Angelis, M., Calasso, M., Vincentini, O., Vernocchi, P., Ndagijimana, M., De Vincenzi, M., Dessì, M. R., Guerzoni, M. E. & Gobbetti, M. 2010. Quorum sensing in sourdough Lactobacillus plantarum DC400: Induction of plantaricin A (PlnA) under co-cultivation with other lactic acid bacteria and effect of PlnA on bacterial and Caco-2 cells. *PROTEOMICS,* 10**,** 2175-2190.

Ditu, L.-M., Chifiriuc, M. C., Pelinescu, D., Avram, I., Pircalabioru, G. & Mihaescu, G. 2014. Class I and II Bacteriocins: Structure, Biosynthesis and Drug Delivery Systems for the Improvement of their Antimicrobial Activity. *Current Proteomics* 11**,** 121-127.

Dobson, A., Cotter, P. D., Ross, R. P. & Hill, C. 2012. Bacteriocin Production: a Probiotic Trait? *Applied and Environmental Microbiology,* 78**,** 1-6.

Doerks, T., Copley, R. R., Schultz, J., Ponting, C. P. & Bork, P. 2002. Systematic identification of novel protein domain families associated with nuclear functions. *Genome Research,* 12**,** 47-56.

Dogra, V. 2017. Gene mining. USA: Scribd. Incorporated.

Drissi, F., Buffet, S., Raoult, D. & Merhej, V. 2015. Common occurrence of antibacterial agents in human intestinal microbiota. *Front Microbiol,* 6**,** 441.

Duquesne, S., Destoumieux-Garzon, D., Peduzzi, J. & Rebuffat, S. 2007. Microcins, gene-encoded antibacterial peptides from enterobacteria. *Natural Product Reports,* 24**,** 708-34.

Ebadi, M. 2010. *Pharmacodynamic Basis of Herbal Medicine*, CRC Press.

Falanga, A., Nigro, E., De Biasi, M. G., Daniele, A., Morelli, G., Galdiero, S. & Scudiero, O. 2017. Cyclic Peptides as Novel Therapeutic Microbicides: Engineering of Human Defensin Mimetics. *Molecules,* 22.

Fan, S., Ruel, M., Terero, M., Von Grebmer, K. & Haddad, L. 2017. Topic: Food security. *In:* International Food Policy Research Institute (ed.). USA: IFPRI.

Frechet, D., Guitton, J. D., Herman, F., Faucher, D., Helynck, G., Monegier Du Sorbier, B., Ridoux, J. P., James-Surcouf, E. & Vuilhorgne, M. 1994. Solution structure of RP 71955, a new 21 amino acid tricyclic peptide active against HIV-1 virus. *Biochemistry,* 33**,** 42-50.

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D. & Bairoch, A. 2005. *The proteomics protocols handbook*, Humana press.

Gavrish, E., Sit, C. S., Cao, S., Kandror, O., Spoering, A., Peoples, A., Ling, L., Fetterman, A., Hughes, D., Bissell, A., Torrey, H., Akopian, T., Mueller, A., Epstein, S., Goldberg, A., Clardy, J. & Lewis, K. 2014. Lassomycin, a ribosomally synthesized cyclic peptide, kills mycobacterium tuberculosis by targeting the ATP-dependent protease ClpC1P1P2. *Natural Chemical Biology,* 21**,** 509-518.

Gillor, O., Etzion, A. & Riley, M. A. 2008. The dual role of bacteriocins as anti- and probiotics. *Applied Microbiology and Biotechnology,* 81**,** 591-606.

Gillor, O., Giladi, I. & Riley, M. A. 2009. Persistence of colicinogenic Escherichia coli in the mouse gastrointestinal tract. *BMC Microbiology,* 9**,** 165.

Hegde, N. S., Sanders, D. A., Rodriguez, R. & Balasubramanian, S. 2011. The transcription factor FOXM1 is a cellular target of the natural product thiostrepton. *Nature Chemistry,* 3**,** 725-31.

Hillman, J. D., Dzuback, A. L. & Andrews, S. W. 1987. Colonization of the human oral cavity by a Streptococcus mutans mutant producing increased bacteriocin. *Journal of Dental Research,* 66**,** 1092-4.

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

Houssen, W. E. & Jaspars, M. 2010. Azole-based cyclic peptides from the sea squirt Lissoclinum patella: old scaffolds, new avenues. *Chembiochem,* 11**,** 1803-15.

Huo, L., Rachid, S., Stadler, M., Wenzel, S. C. & Muller, R. 2012. Synthetic biotechnology to study and engineer ribosomal bottromycin biosynthesis. *Chemistry & Biology,* 19**,** 1278-87.

Ikai, A. 1980. Thermostability and aliphatic index of globular proteins. *Journal of Biochemistry,* 88**,** 1895-8.

Jack, R. W., Tagg, J. R. & Ray, B. 1995. Bacteriocins of gram-positive bacteria. *Microbiological Reviews,* 59**,** 171-200.

Jakubovics, N. S., Yassin, S. A. & Rickard, A. H. 2014. Community interactions of oral streptococci. *Advances in applied microbiology,* 87**,** 43-110.

Jones, D. & Van Niekerk, N. 2004. Guidelines for environmental health practitioners on food safety control at special events. *In:* Department of Health Republic of South Africa (ed.). Africa.

Klaenhammer, T. R. 1993. Genetics of bacteriocins produced by lactic acid bacteria. *Federation of European Microbiological Societies Microbiology Reviews,,* 12**,** 39-85.

Knappe, T. A., Linne, U., Zirah, S., Rebuffat, S., Xie, X. & Marahiel, M. A. 2008. Isolation and structural characterization of capistruin, a lasso peptide predicted from the genome sequence of Burkholderia thailandensis E264. *Journal of the American Chemical Society,,* 130**,** 11446-54.

Kobayashi, Y., Ichioka, M., Hirose, T., Nagai, K., Matsumoto, A., Matsui, H., Hanaki, H., Masuma, R., Takahashi, Y., Omura, S. & Sunazuka, T. 2010. Bottromycin derivatives: efficient chemical modifications of the ester moiety and evaluation of anti-MRSA and anti-VRE activities. *Bioorganic & medicinal chemistry letters,* 20**,** 6116-20.

Kruszewska, D., Sahl, H. G., Bierbaum, G., Pag, U., Hynes, S. O. & Ljungh, A. 2004. Mersacidin eradicates methicillin-resistant Staphylococcus aureus (MRSA) in a mouse rhinitis model. *Journal of Antimicrobial Chemotherapy,* 54**,** 648-53.

Lee, D. W. & Kim, B. S. 2015. Antimicrobial Cyclic Peptides for Plant Disease Control. *The Plant Pathology Journal,* 31**,** 1-11.

Letzel, A.-C., Pidot, S. J. & Hertweck, C. 2014. Genome mining for ribosomally synthesized and post-translationally modified peptides (RiPPs) in anaerobic bacteria. *BMC Genomics,* 15**,** 983.

Li, Y., Zirah, S. & Rebuffat, S. 2014. *Lasso Peptides: Bacterial Strategies to Make and Maintain Bioactive Entangled Scaffolds*, Springer New York.

López-Cuellar, M. D. R., Rodríguez-Hernández, A.-I. & Chavarría-Hernández, N. 2016. LAB bacteriocin applications in the last decade. *Biotechnology & Biotechnological Equipment***,** 1-12.

Lopez, F. E., Vincent, P. A., Zenoff, A. M., Salomón, R. A. & Farías, R. N. 2007. Efficacy of microcin J25 in biomatrices and in a mouse model of Salmonella infection. *Journal of Antimicrobial Chemotherapy,* 59**,** 676-680.

Madden, T. 2013. The BLAST Sequence Analysis Tool. *The NCBI Handbook.* 2nd ed. Bethesda (MD): National Center for Biotechnology Information (US).

Madden, T. L., Tatusov, R. L. & Zhang, J. 1996. Applications of network BLAST server. *Methods in Enzymology,* 266**,** 131-41.

Maksimov, M. O., Pan, S. J. & James Link, A. 2012a. Lasso peptides: structure, function, biosynthesis, and engineering. *Natural product reports,* 29**,** 996-1006.

Maksimov, M. O., Pelczer, I. & Link, A. J. 2012b. Precursor-centric genome-mining approach for lasso peptide discovery. *Proceedings of the National Academy of Sciences of the United States of America,* 109**,** 15223-8.

Marco Biasini, Stefan Bienert, Waterhouse., A., Konstantin Arnold, Gabriel Studer, Tobias Schmidt, Florian Kiefer, Tiziano Gallo Cassarino, Martino Bertoni, Lorenza Bordoli & Schwede, T. 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research,* 42**,** W252-W258.

Martinez-Bueno M., Maqueda M., Galvez M., Samyn B., Van Beeumen B., Coyette J. & Valdivia, E. 1994. Determination of the Gene Sequence and the Molecular Structure of the Enterococcal Peptide Antibiotic AS-48. *Journal of Bacteriology,* 176**,** 6334-6339.

Mathavan, I., Zirah, S., Mehmood, S., Choudhury, H. G., Goulard, C., Li, Y., Robinson, C. V., Rebuffat, S. & Beis, K. 2014. Structural basis for hijacking siderophore receptors by antimicrobial lasso peptides. *Nature chemical biology,* 10**,** 340-2.

Mills, S., Ross, R. P. & Hill, C. 2017. Bacteriocins and bacteriophage; a narrow-minded approach to food and gut microbiology. *Federation of European Microbiological Societies Microbiology Reviews,,* 41**,** S129-S153.

Mims Australia 2015. Megacin. Australia.

National Institute of Diabetes and Digestive and Kidney Diseases 2014. Foodborne illnesses. *In:* Us Department of Health and Human Services Nih (ed.). USA.

Newsdesk 2017. Food safety news. Marler Clark.

Newyorktimes 2017. Food Safety. USA: The New York Times company.

Nugent, R. 2010. *The Race Against Drug Resistance*.

Nugent, R. 2011. The Race Against Drug Resistance. eSocialSciences.

Pan, S. J. 2012. *Biosynthesis and Engineering of Lasso Peptides.* Ph.D., Priceton university.

Pandit, B. & Gartel, A. L. 2011. Thiazole antibiotic thiostrepton synergize with bortezomib to induce apoptosis in cancer cells. *PLoS One,* 6**,** e17110.

Parada, J. L., Caron, C. R., Medeiros, A. B. P. & Soccol, C. R. 2007. Bacteriocins from Lactic Acid Bacteria: Purification, Properties and use as Biopreservatives. *Brazilian Archives of Biology and Technology,* 50**,** 521-542.

Perez, R. H., Ishibashi, N., Inoue, T., Himeno, K., Masuda, Y., Sawa, N., Zendo, T., Wilaipun, P., Leelawatcharamas, V., Nakayama, J. & Sonomoto, K. 2016. Functional Analysis of Genes Involved in the Biosynthesis of Enterocin NKR-5-3B, a Novel Circular Bacteriocin. *Journal of Bacteriology,* 198**,** 291-300.

Potterat, O., Wagner, K., Gemmecker, G., Mack, J., Puder, C., Vettermann, R. & Streicher, R. 2004. BI-32169, a bicyclic 19-peptide with strong glucagon receptor antagonist activity from Streptomyces sp. *Journal of natural products,* 67**,** 1528-31.

Puth, C. 2017. Long-term effects of dehydration. *In:* Consultants, V. W. (ed.). USA: WordPress.

Rea, M. C., Dobson, A., O'sullivan, O., Crispie, F., Fouhy, F., Cotter, P. D., Shanahan, F., Kiely, B., Hill, C. & Ross, R. P. 2011. Effect of broad- and narrow-spectrum antimicrobials on Clostridium difficile and microbial diversity in a model of the distal colon. *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 4639-4644.

Rebuffat, S. 2012. Microcins in action: amazing defence strategies of Enterobacteria. *Biochemical Society Transactions,* 40**,** 1456-1462.

Rupnik, M., Wilcox, M. H. & Gerding, D. N. 2009. Clostridium difficile infection: new developments in epidemiology and pathogenesis. *Nature Reviews Microbiology,* 7**,** 526-36.

Schmidt, E. W. & Donia, M. S. 2009. Chapter 23 Cyanobactin Ribosomally Synthesized Peptides—A Case of Deep Metagenome Mining. 458**,** 575-596.

Scientific Committees 2017. Bacterial resistance. *In:* Commission, E. (ed.) *Health and food safety glossary.* Europe.

Serres, M. H., Kerr, A. R., Mccormack, T. J. & Riley, M. 2009. Evolution by leaps: gene duplication in bacteria. *Biology Direct,* 4**,** 46.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., Mcwilliam, H., Remmert, M., Söding, J., Thompson, J. D. & Higgins, D. G. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology,* 7.

States, D. J. & Gish, W. 1994. Combined use of sequence similarity and codon bias for coding region identification. *Journal in computational biology,* 1**,** 39-50.

Stordeur, C., Dallüge, R., Birkenmeier, O., Wienk, H., Rudolph, R., Lange, C. & Lücke, C. 2008. The NMR solution structure of the artificial protein M7 matches the computationally designed model. *Proteins: Structure, Function, and Bioinformatics,* 72**,** 1104-1107.

Tijani, B. 2016. Africa regional overview of food security and nutrition. *In:* Fao (ed.). Africa: Food and Agriculture Organization,.

Van Heel, A. J., De Jong, A., Montalban-Lopez, M., Kok, J. & Kuipers, O. P. 2013a. BAGEL3: Automated identification of genes encoding bacteriocins and (non-) bactericidal posttranslationally modified peptides. *Nucleic Acids Research,* 41**,** W448-53.

Van Heel, A. J., De Jong, A., Montalbán-López, M., Kok, J. & Kuipers, O. P. 2013b. BAGEL3: automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. *Nucleic Acids Research,* 41**,** W448-W453.

Walsh, C., Guinane, C., Hill, C., Paul Ross, R., W O'toole, P. & Cotter, P. 2015. *In silico identification of bacteriocin gene clusters in the gastrointestinal tract, based on the Human Microbiome Project's reference genome database*.

Wang, H., Fewer, D. P. & Sivonen, K. 2011. Genome mining demonstrates the widespread occurrence of gene clusters encoding bacteriocins in cyanobacteria. *PLoS One,* 6**,** e22384.

Wax, R. D. E. 2016. UMMC- Food safety. *University of Maryland medical reference guide,* 2016.

Wilks, J. C. & Slonczewski, J. L. 2007. pH of the cytoplasm and periplasm of Escherichia coli: rapid measurement by green fluorescent protein fluorimetry. *Journal of Bacteriology,* 189**,** 5601-7.

Wong, J. J., Paterson, R. G., Lamb, R. A. & Jardetzky, T. S. 2016. Structure and stabilization of the Hendra virus F glycoprotein in its prefusion form. *Proceedings of the National Academy of Sciences of the United States of America,,* 113**,** 1056-61.

World Health Organisation. 2017. *Global health observatory* [Online]. Available: http://www.who.int/gho/epidemic_diseases/cholera/cases_text/en/ [Accessed 03/05/2017].

Zarivach, R., Ben-Zeev, E., Wu, N., Auerbach, T., Bashan, A., Jakes, K., Dickman, K., Kosmidis, A., Schluenzen, F., Yonath, A., Eisenstein, M. & Shoham, M. 2002. On the interaction of colicin E3 with the ribosome. *Biochimie,* 84**,** 447-54.

Zhang, Z. & Chan, H. S. 2009. Native topology of the designed protein Top7 is not conducive to cooperative folding. *Biophysical Journal,* 96**,** L25-7.

Zimmermann, M., Hegemann, Julian d., Xie, X. & Marahiel, Mohamed a. 2013. The Astexin-1 Lasso Peptides: Biosynthesis, Stability, and Structural Studies. *Chemistry & Biology,* 20**,** 558-569.