Biosynthetic gene cluster identification in plasmids and characterization of plasmids from animal-associated microbiota

A Thesis Submitted to the
College of Graduate and Postdoctoral Studies
In Partial Fulfillment of the Requirements
For the Degree of Master of Science
In the Department of Veterinary Microbiology
University of Saskatchewan
Saskatoon

By

RAÍZA DE ALMEIDA MESQUITA

Permission to use

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other uses of materials in this thesis in whole or part should be addressed to:

Head of the Department of Veterinary Microbiology
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5B4 Canada

OR

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9

Abstract

Individual bacteria in complex microbial communities can acquire and accumulate new traits. These traits are reflective of their environment, being niche-specific. A major player in trait sharing is horizontal gene transfer (HGT). Plasmids, extrachromosomal DNA molecules, have a role in HGT and can change the host's phenotype. Considering the transformative role of plasmids in bacterial lifestyle, we investigated the prevalence, distribution and products of biosynthetic gene clusters (BGCs) present in plasmids. Sequences available on the National Center for Biotechnology Information (NCBI) database (n=101 416) were run through two bioinformatic pipelines for BGC detection that apply different approaches, deepBGC and antiSMASH (antibiotics and secondary metabolites analysis shell). The highest percentage of plasmids with BGCs was detected in Actinobacteria but, apart from Chlamidiae and Tenericutes, all phyla had BGCs in their plasmids, with predictions varying according to the software used. The BGCs identified comprised a range of classes, indicating that plasmid-encoded BGCs could be leveraged for the discovery of new molecules. In order to apply that concept to real-life examples, plasmids were isolated from animal-associated microbial communities and characterized. Plasmids from *Escherichia coli* isolated from wild birds (n=36) were screened for phenotypes of interest in human and animal health. Seven isolates displayed plasmid-encoded antibiotic resistance. Taxonomic identification of the hosts of plasmids isolated from bovid-associated microbiomes (n=38) was determined via 16S rRNA gene, and placed the majority of the isolated in the phylum Firmicutes, apart from a single *Klebsiella pneumoniae* isolate. Twelve plasmids were sequenced. Three plasmids from different hosts (pRAM-12, pRAM-19-2 and pRAM-30-2) shared 100% nucleotide sequence and a gene cluster for the bacteriocin cloacin. Two of those hosts shared not one, but two plasmids, pRAM-19-1 and pRAM-30-1, despite being in different phyla. This highlights the intimacy of gene sharing and the importance of HGT. pRAM-28 and pRAM-21 shared a plasmid that harbors the BGC for the bacteriocin aureocin A70, the only four peptide bacteriocin known to date. Additional analysis revealed two putative novel lanthipeptide gene clusters in pRAM-2. These results suggest that the plasmidome is a neglected source of secondary metabolites with the potential for molecule discovery. Furthermore, it can be leveraged to study genetic exchange in a community and how plasmid-encoded featured can mediate interactions in a microbiome.

Acknowledgements

My deepest and most sincere gratitude to my brilliant supervisor, Dr. Janet Hill, for being an admirable leader and example of kindness, always an inspiration. My committee members, Dr. Vikram Misra, Dr. Joe Rubin and Dr. Harold Bull, for the continuous support through this journey; Linda Nemeth and Lana Abrey, for always being so welcoming and always making me feel at home, and the Western College of Veterinary Medicine and Saskatchewan Health Research Foundation for funding this project.

A special and sincere thank you and my eternal love and debt to the brilliant people that supported me throughout my life, from near and far. My parents, brother and grandma, and my dear sisters from other mothers, Laura Coutinho e Letícia da Costa e Silva, there are no words to express how much you inspire me! The amazing Renê da Luz, Guilherme Hassemer and André Batista, for always keeping a smile in my face. I am eternally thankful to Carolina Malgarin, Daniel Watanabe and Thaísa Meira Sandini, for showing me that home away from home is still a home.

The University of Saskatchewan allowed me to meet exquisite people that I will forever cherish: Stephanie Saundh, Michelle Sniatynski, Kayla Buhler, Chris Zink, Alex Foley-Eby and Chris Aiello, thank you for being thoughtful and encouraging. Jenny Liang, Kathyana Deeyagahage, Radwa Asar, Dr. Ruzzini, Richa Jain, Aimen Khan, as well as all the colleagues in Hill Lab, your insights and inputs will always be appreciated.

To everyone that crossed my path and, in any way, shape, or form, made my days a little brighter, thank you.

Table of contents

List of Tables

List of Figures

List of Abbreviations

HGT: horizontal gene transfer

BGC(s): biosynthetic gene cluster(s)

RiPP(s): ribosomally synthesized and post-translationally modified peptide(s)

PK(s): polyketide(s)

PKS: polyketide synthase

AT: acyltransferase

KS: ketosynthase

KR: ketoreductase

TE: thioesterase

ER: enoylreductase

DH: dehydratase

ACP: acyl carrier protein

CoA: coenzyme A

NRP(s): non ribosomal peptide(s)

NRPS: non-ribosomal peptide synthetases

A: adenylation domain

PCP: peptidyl carrier protein

C: condensation domain

antiSMASH: antibiotic and secondary metabolite analysis shell

pHMM(s): profile Hidden Markov Model(s)

Pfam: protein family

NCBI: National Center for Biotechnology Information

BiLSTM: bidirectional long short-term memory

RNN: recurrent neural network

ESBL: extended-spectrum beta-lactamase

LB: lysogeny broth

TRACA: transposon-aided capture

CAS: chrome azurol S

1 Introduction

Although we have studied individual pathogens since the nineteenth century, only recently have we come to appreciate the role microbial collectives play in health and disease. These microbiomes, which are defined by members and habitat, play roles in the prevention, causation, or aggravation of human and animal disease. A common feature of microbiomes is functional redundancy among their members[1,2]. This redundancy is created by two different paths: recruitment of bacteria that share traits or trait sharing between bacteria. Specific recruitment implies a long, well-established system that has co-evolved. Trait-sharing among bacteria can happen on a much shorter timescale.

Despite the redundancy encoded within any microbiome, an individual bacterial genome is dynamic[3–5]. A bacterial genome generally consists of a single chromosome and can include one or more plasmids. The conserved chromosomal genes (or core genome) allow us to relate individuals to each other and a common ancestor. The flexible genome, the genes that are not conserved in a species, is known for passing on specialized biological functions. These are called accessory genes and can facilitate rapid adaptation. Therefore, metabolic functions and habitat-specific interactions can be understood by studying the flexible genome[3].

Accessory genes can be acquired through horizontal gene transfer (HGT), a process that allows the acquisition of new genetic traits and can contribute greatly to bacterial adaptation and functional innovation[3,6]. These can carry big implications for both an individual and a microbial community since bacteria can transition from a commensal to a pathogenic lifestyle via HGT[4,5,7]. Gene transfer can have a significant effect on a bacterium phenotype and the structure of the microbial community[4,5]. A primary example of accessory genes that are often exchanged among bacteria is plasmids. These extrachromosomal, self-replicating DNA units can be associated with the adaptation to environmental pressures and the emergence of new traits that a bacterium can leverage to thrive in challenging settings.

Current plasmidomic studies are often the result of nucleotide sequence-based metagenomics in which the total DNA content of a microbial community is sequenced[8,9]. The annotation of genes and gene clusters harbored in plasmids are frequently the result of the study of specific genes or biosynthetic gene clusters (BGCs) on a case-by-case basis. Therefore, plasmid-encoded traits are not uncovered by the prioritization of this

genetic space, but by a plasmid-by-plasmid basis. Furthermore, metagenomics studies analyse plasmids without knowledge about their host bacteria. This thesis uses a systematic approach that prioritizes the plasmidome of culturable bacteria as a rich source for the discovery of gene clusters with products that can be leveraged for both human and animal health. Focusing on the plasmidome and its phenotypes can allow us to shed a light on virulence mechanisms related to disease emergence, bacterial interactions, and antimicrobial resistance. Improving our understanding of how virulence factors operate, how resistance is mounted, and how bacteria compete with each other can, ultimately, be used as basis for development of antimicrobial treatment strategies and new therapies.

2 Literature review

2.1 Plasmid-encoded phenotypes of interest for human and animal medicine

Plasmids are self-replicating extrachromosomal DNA molecules that can be shared by bacteria through several mechanisms[10]. Conjugation, transformation, and transduction are the three canonical mediators of DNA transfer in bacteria[4,5]. Conjugation occurs when a donor bacterium makes direct cell contact with a recipient bacterium, using a conjugative pilus. Transformation is the natural process of DNA uptake from the environment, such as from surrounding bacteria that have been lysed, while transduction occurs when a bacteriophage acquires a fragment of bacterial genetic information and infects another bacterium. Recently, other mechanisms of DNA transfer were identified, such as nanotubes[11], which are structurally different from conjugative pili, and extracellular vesicles (exosomes)[12].

Gene transfer can have a significant effect on a bacterium phenotype and the structure of the microbial community[4,5]. A primary example of accessory genes that are often exchanged among bacteria is plasmids. These extrachromosomal, self-replicating DNA can be associated with the adaptation to environmental pressures and the emergence of new traits that a bacterium can leverage to thrive in challenging settings.

The acquisition or loss of a plasmid can drastically alter a bacterial phenotype. It has been correlated to the development of virulent phenotypes attributed to small molecules[13–16]. This extrachromosomal DNA can also encode for protein toxins, antibiotic resistance, secretion systems, and iron-scavenging molecules named siderophores[17,18]. Plasmid-encoded pathogenicity has been shown for *Vibrio crassostreae*, a benign colonizer of oysters that becomes pathogenic when carrying the plasmid pGV1512. Interestingly, though the pathogenic phenotype has emerged, none of the genes encode for known virulence factors[15]. The fish pathogen *Vibrio anguillarum* has two phenotypes related to plasmid acquisition: the production of a siderophore named anguibactin, and a virulence system that encodes for a potent enterotoxin, both encoded by the plasmid pJM1[13,14]. The bacteria *Staphylococcus aureus* can produce an exotoxin that causes blisters in humans and animals when carrying an ETB plasmid. Moreover, the emergence of ETB plasmids containing multiple antibiotic resistance genes, which is a potential problem for human and animal health, has also been reported[19]. Despite the bias toward the study of pathogens, plasmids can also help non-pathogenic bacteria overcome limited resources and niche occupancy, as is the case of plasmid-encoded small

molecules such as 9-methoxyrebeccamycin, an analog of the antitumoral agent rebeccamycin encoded on plasmid pBCI2-2[20–24].

2.2 Classes of natural products

2.2.1 Ribosomally synthesized and post-translationally modified peptides

One of the classes of small molecules that has attracted attention from both the academy and industry due to the rapid discovery of new molecules, structural diversity and functional variability is the Ribosomally synthesized and post-translationally modified peptides (RiPPs)[25–28]. Bioactivity from these small molecules ranges from targeting DNA gyrase[29] and RNA polymerase[30] to cell membranes[31,32]. Despite great diversity of structures, their biosynthesis and minimal gene cluster composition has common features, which allow researchers to classify them as RiPPs.

A precursor peptide that includes an N-terminal leader and a C-terminal core peptide and modifying enzymes constitute the minimal components of a RiPP BGC[28]. Different post-translational modification enzymes will install different moieties, which will result in the various classes of RiPPs[26,28]. The biosynthesis starts with the synthesis of a precursor peptide. Most RiPP precursors have a leader peptide attached to the N-terminal of the core peptide. An exception are the bottromycins, where the leader region is at the C-terminus and was termed follower peptide[33–35]. The leader sequence is recognized by the post-translational modification enzymes, as well as by the export system, while the core region is modified to become the mature, active RiPP[25–28] (Figure *2.1*).

Recently, new sequencing techniques coupled with genome mining approaches allowed researchers to link known product classes with biosynthetic gene clusters, and the exploration of novel BGCs resulted in new RiPP classes. These new classes, as well as defining characteristics of all RiPPs families, covering research up to June 2020 were reviewed in detail by Montalbán-López et al. (2020)[36]. Seventeen new classes were recently described, and Table 2.1 presents a summary of all currently known classes of RiPPs, their defining features, as well as class representatives of plasmid-encoded peptides, if known.

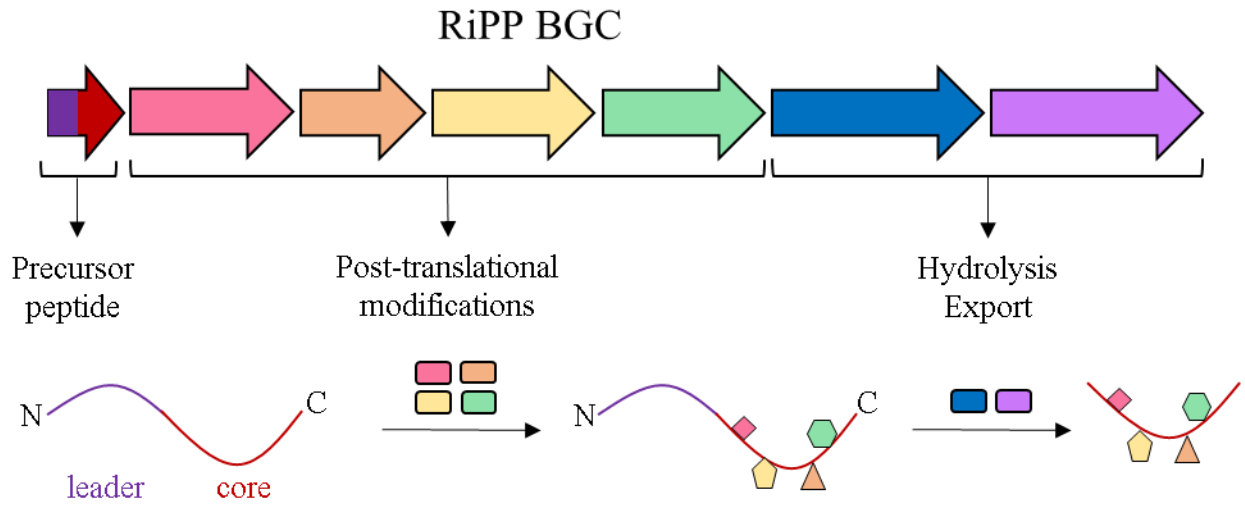

Figure 2.1: Schematic overview of the biosynthesis of RiPPs. The precursor peptide is composed of a leader and a core region. The core peptide will be transformed in the mature peptide. The post-translational modifications enzymes vary depending on the family of RiPPs, and are guided by the recognition of the leader peptide and recognition sequences. After tailoring reaction(s), the leader peptide is cleaved from the core, and the mature peptide is exported from the cell.

Table 2.1: Known RiPP classes, their defining features, class representative and, if applicable, plasmid-encoded example (Adapted from Montalbán-López et al. (2020)).

Class	Example	Defining feature	Plasmid-encoded
Lanthipeptides	Nisin	(methyl)lanthionine	epidermin[37], staphylococcin C55[38], salivaricins A2 and B[39], pep5[40]
Linaridins	Cypemycin	Dehydrobutyrine (Dhb), no lanthionines	
Proteusins	Polytheonamides	Nitrile hydratase leader peptide	
Linear azol(in)e-containing peptides (LAPs)	Streptolysin S	Azol(in)es	Microcin B17[29,41–43]
Cyanobactins	Pattelamide	N-terminal protease	
Thiopeptides	Thiostrepton	[4 + 2] cycloaddition of two dehydroalanines (Dha)	lactocillin[44], MP1[45]
Bottromycins	Bottromycin A2	Macrolactamidine	
Microcins	Microcin C	Low molecular mass peptides, produced by *Enterobacteriaceae*	Microcins C (both C7[46] and C51[47]), microcin D93[48], microcin PDI[49]
Lassopeptides	Microcin J25	Macrolactam with threaded C-terminal tail	Microcin J25[30,50–52]; *citrocin** [53]
Graspetides	Microviridin A	Macrolactones and/or macrolactams	
Sactipeptides	Subtilosin A	Sactionine crosslink	Bacthuricin F4[54,55]
Bacterial head-to-tail cyclized peptides	Enterocin AS-48	N-to-C cyclization	Acidocin B[56]
Glycocins	Sublancin 168	S, O-glycosylation of Ser/Cys	ASM1[57]
Autoinducing peptides	AIP-1	Cyclic ester/thioester	
ComX	ComX168	Indole cyclization and prenylation	
Methanobactins	Methanobactin	Oxazolones	
Thioamitidies	Thioviridamide	Backbone thioamide	
Dikaritins	Ustiloxin	Tyr-Xxx ether crosslink	
Guanidinotides	Pheganomycin	a-Guanidino acid	
Mycofactocin	Mycofactocin	Val–Tyr crosslink	
Streptides	Streptide	Trp–Lys crosslink	
Borosins	Omphalotin	Amide backbone N-methylation, N-to-C cyclization	
Crocagins	Crocagin	Indole-backbone cyclization	
Epipeptides	YydF	D-Amino acids	
Lyciumins	Lyciumin A	Pyroglutamate, Trp–Gly crosslink	
Lipolanthines	Microvionin	C-terminal labionin/avionin, N-terminal FAS/ PKS segment	
Spliceotides	PlpA	b-Amino acids	
Ranthipeptides	Freyrasin	Sulfur-to-non-Cα thioether crosslink	
Cyclotides	Kalata B1	N-to-C cyclization, disulfide(s)	
Pearlins	Thiaglutamate	aa-tRNA derived	
Atropitides	Tryptorubin	Aromatic amino acids crosslinked resulting in a non-canonical atropisomer	
Cittilins	Cittilin A	Biaryl and aryl–oxygen–aryl ether crosslinks	
Orbitides	Cyclolinopeptide A	N-to-C cyclization; no disulfides	
Pantocins	Pantocin A	Glu–Glu crosslink	
Rotapeptides	TQQ	Oxygen-to-a-carbon crosslink	
Sulfatyrotides	RaxX	Tyrosine sulfation	
Pyrroloquinoline quinones	PQQ	Glu–Tyr crosslink	
Amatoxins/phallotoxins	Phalloidin	N-to-C cyclization, Cys–Trp crosslink	

* Suggestive evidence that the gene cluster is plasmid-encoded.

2.2.2 Polyketides

Polyketides (PKs) comprise a class of natural products found in bacteria, fungi, plants and animals, that presents distinct structures and widely diverse clinical applications[58–60]. This class includes the antibiotics erythromycin[61] and tetracycline[62], the antifungal amphotericin[63], as well as the antiparasitic ivermectin[64]. The mycolactone produced by *Mycobacterium ulcerans* is a cytotoxic macrolide that is plasmid encoded[18], and so is mycolactone F, a unique toxin produced by the fish pathogen *Mycobacterium marinum*[65]. Two macrolide antibiotics, lankacidin and lankamycin, are encoded on a plasmid[66]. Curiously, the same plasmid carries two additional BGCs, for the production of a cryptic type II polyketide and carotenoids, making two-thirds of the plasmid responsible for secondary metabolism genes.

The complex biosynthesis of polyketides involves the multifunctional enzymes polyketide synthases (PKSs)[60,67]. A simplified version of mycolactone synthesis is shown in Figure 2.2. These multi-domain enzymes harbor acyltransferase (AT), ketosynthase (KS), and thioestherase (TE), as well as optional domains, such as ketoreductase (KR), enoylreductase (ER) and dehydratase (DH). Based on the structural architecture and enzymatic mechanism, PKSs have been divided into three types[60,68,69]. Type I PKSs are multienzyme complexes with modules fused covalently. Each individual module has several domains (AT, KS, TE, KR, etc.), in order for the catalyzing reactions to assemble the final polyketide. Type II PKSs are monofunctional enzymes, each responsible for a specific reaction in the polyketide assembly line. This type of PKS is mainly found in bacteria and generates aromatic compounds. Type III PKSs are usually found in plants, although three PKSs identified in mycobacterium genome also belong to type III[60]. These PKSs are simple homodimers, and function independently of the acyl carrier protein (ACP) domain.

The biosynthesis of polyketides has been divided into *cis*-AT PKS and *trans*-AT PKS[70]. In the *cis*-AT biosynthesis, the starter unit acyl-Coenzyme A (CoA) is loaded on the ACP, catalyzed by the AT domain[59]. The elongation of the carbon chain occurs catalyzed by the KS domain. Different structures can be added by other additional domains, such as KR, DH and ER. The TE domain then terminates the elongation process by hydrolysis or cyclization of the polyketide chain from the ACP domain. The *trans*-AT biosynthesis involves PKSs that lack the AT domains[70]. The activity of these domains in each elongation step is provided by proteins encoded in the BGC.

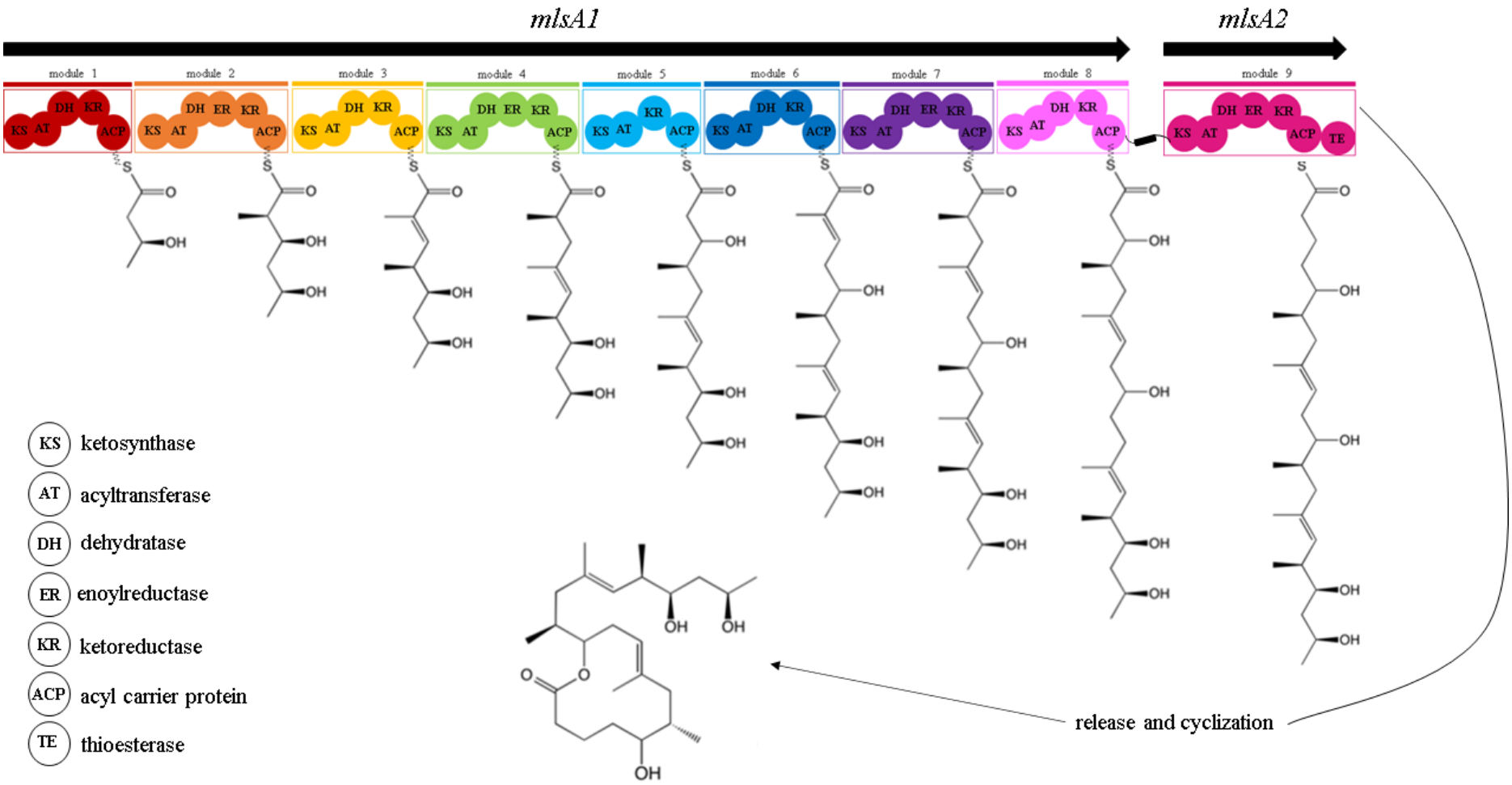

Figure 2.2: Simplified biosynthesis of the polyketide mycolactone, product of a BGC harbored by the pMUM001 plasmid. MLSA1 and MLSA2 are modular PKSs encoded by the genes *mlsA1* and *mlsA2*. The starter unit acyl-CoA is loaded on the ACP, catalyzed by the AT domain (loading module not depicted). The carbon chain is elongated by the KS domain. Further elongation and modifications occur in each module, until the elongation is terminated by the TE domain via hydrolyzation and/or cyclization of the completed molecule from the ACP domain.

2.2.3 Non-ribosomal peptides

With diverse structure range, and a vast sphere of activity, non-ribosomal peptides (NRPs) can be differentiated from ribosomally synthesized peptides through structural features and their biosynthesis[71,72]. NRPs have structures that are usually macrocyclic or branched macrocyclic, with dimers and trimers of identical elements. Non-proteinogenic amino acids can be included, such as ornithine or (di)hydroxyphenyl-glycine. These peptides can also have fatty acids incorporated, and acetate and propionate units can be inserted. N-methylations, N-formylations and glycosylations can also be present. Anguibactin, a siderophore produced by *Vibrio anguillarum*, is one of the known plasmid-encoded NRPs[73]. Recently, a pediocin-like peptide was identified, with broad spectrum activity against the pathogen *Listeria monocytogenes*[74]. This peptide also has its BGC harbored in a plasmid.

The biosynthesis of NRPs is dependent on the megaenzymes non-ribosomal peptide synthetases (NRPSs)[71]. These enzymes are able to process hundreds of monomers and have a modular organization, where each module (section) of the NRPS is responsible for the addition of one amino acid to the final peptide[75]. The minimal components for the peptide elongation step are an adenylation (A) domain, a peptidyl carrier protein (PCP) domain, and a condensation (C) domain[71,75,76]. The chain initiation involves an A domain selecting the substrate and activating it as an aminoacyl-adenylate. The PCP domain is a transport unit, which allows the movement of the activated amino acids and elongation intermediates between catalytic centers. It carries the acyl-intermediates on the –SH group of its cofactor 4'-phosphopantetheine (acyl-S-PCP intermediate). The two modules are condensed by the C domain, which catalyses a peptide bond between the adjacent modules. The termination of the peptide chain is done by the release of the peptide, by a thioesterase (TE) domain, by hydrolysis or cyclization (Figure 2.3).

Hybrid BGCs of polyketides and non-ribosomal peptides can be found in nature, with different biological activities such as the antibiotic leinamycin[77], the antitumor drug bleomycin[78], and the siderophore yersiniabactin[79]. The combination of PKs and NRPs is typically synthesized by PKS and NRPS modules in certain order in the assembly line[80]. A second, different mode of biosynthesis can be found in fungi and was reviewed by Fisch (2013). These hybrids BGCs can also be found in plasmids, as is the case of the myxobacterium toxin sandarazol[81].

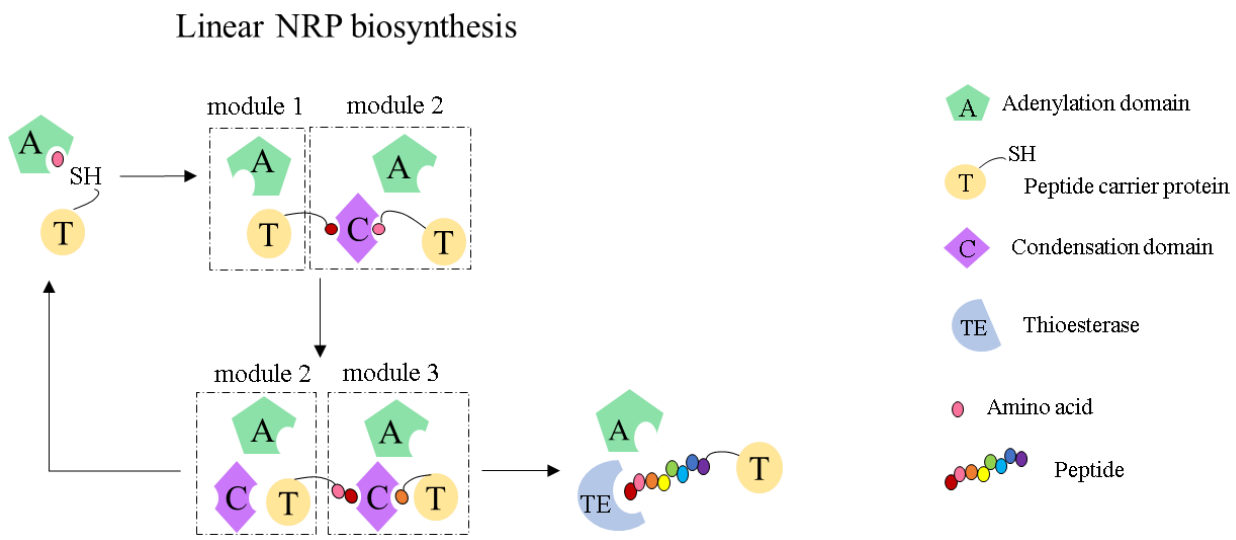

Figure 2.3: Schematics of a linear non ribosomal peptide biosynthesis. Each module is responsible for the addition of a single residue. Using an assembly-line-like structure, the substrate is adenylated (activated). The thiol group of the pantetheine cofactor of the peptide carrier protein is used as a shuttle between the modules (catalytic domains). The amide bond formed between the substrates is catalyzed by the condensation domain. A thioesterase recognizes the mature peptide and cleaves it from the NRPS machinery, often macrocyclizing it during the release.

2.2.4 Terpenes

Terpenes are a structurally diverse class of secondary metabolites and, despite being the largest class of secondary metabolites (over 80,000 known terpenes[82]), only a small fraction of the terpenoid metabolites have been connected to bacteria[82–84]. Recent bioinformatic screening studies have shown that terpene synthases are widely distributed in prokaryotes[84,85]. A highlight of this is the linear megaplasmid pSCL4, isolated from *Streptomyces clavuligerus* ATCC 27064, which harbors twelve putative BGCs with one or more terpene synthases or cyclases[86].

The terpene biosynthesis starts with the creation of a linear polyene with branching methyl groups. This is done by joining multiple five-carbon units (isoprenes). This linear methyl-branched polyene is held in a defined conformation by a terpene cyclase, initiating a series of rearrangements and cyclizations. This hydrocarbon skeleton can be modified to a wide variety of conformations, and sugars, fatty acids and amino acids can be added to the structure[83].

2.3 Bioinformatic approaches for the discovery of BGCs

Computational tools have been used to identify BGCs in nucleotide sequences since the beginning of bacterial genome sequencing[87–89]. It started with simple comparison techniques and, currently, a spectrum of software tools is available. Years of bacterial genome analysis revealed that specialized metabolites are often synthesized by metabolic pathways encoded by genes assembled next to each other. Therefore, leveraging gene clustering in prokaryotes, coupled with the understanding of the biosynthetic logic of different classes of natural products is used in several bioinformatic methods to detect secondary metabolites.

Currently, a range of tools are available to detect bacterial BGCs (Table 2.2). Fungal secondary metabolites can be detected by antiSMASH[90], SMURF[91] and TOUCAN[92]. Bioinformatic tools have been successfully applied to genome mining, resulting in the discovery of various molecules with a range of bioactivities, belonging to different families of natural products. Such is the case of the lasso peptides burhizin-23[93], mycetohabin-16[93], mycetohabin-15[93], and specialicin[94], the terpenoid antibiotic family tiancilactone[95], the non ribosomal peptides brevicidine[96], laterocidine[96], and paenibacterin B[96], the alkaloids argimycins P[97], and a newly reported class V of lanthipeptides[98].

Table 2.2: Currently available bioinformatic tools for detection of bacterial secondary metabolite gene clusters.

Bioinformatic tool	Key features	Reference
antiSMASH	Identifies different types of BGCs, with in depth analysis for dome classes; has a manually curated database	99
deepBGC	Applies a deep learning strategy coupled with random forest classifiers to predict compound classes and potential chemical activity	100
PRISM	Detects 22 secondary metabolites types; allows for structure prediction	101
ClusterFinder	Uses a probabilistic algorithm to identify BGCs of known and unknown classes	102
EvoMining	Incorporates evolutionary principles with phylogenomics to genome mining	103
C-Hunter	Identifies BGCs based on shared gene ontology information	104
NP.searcher	Screens for natural product BGCs; outputs NRPs' and PKs' chemical structures	105
ClustScan	Detects modular enzymes such as PKS, NRPS and hybrid PKS/NRPS enzymes	106
CLUSEAN	Integrates BLAST and HMMer to identify motifs and domains in NRPSs and PKSs	107
NRPSpredictor	Applies machine learning to predict substrate specificity of NRPSs	108
BAGEL	Uses core peptide database and HMMs to identify RiPPs and bacteriocins	109
DeepRiPP	Incorporates genomic and metabolomic data	110
SeMPI	Specialized screening of available databases to predict PK and NRP scaffolds	111
NeuRiPP	Trained neural network on precursor peptides (PP) datasets, allowing for identification of known PP as well as likely PP sequences	112
RiPPMiner	Uses a machine learner classifier coupled with a curated database of >500 characterized RiPPs	113
RODEO	Combines HMMs and machine learning to predict precursor peptides, although initially focused solely on lasso peptides	114
PKMiner	Classifies domains to predict BGCs with type II PKSs and aromatic polyketides based on aromatase and cyclase domains	115
RiPPer	Identifies precursor peptides independent of the family of RiPPs	116
RRE-Finder	Detects RiPPs based on the RiPP recognition element, which binds to the precursor peptide	117
decRiPPter	Integrative algorithm that allows the discovery of new classes of RiPPs	118

This thesis focuses on two programs used to detect and identify a range of secondary metabolite gene clusters: antiSMASH and deepBGC. antiSMASH (antibiotic and secondary metabolite analysis shell) was developed in 2011[119] and is currently in its sixth version[99]. It comprises a software pipeline that can be used either in the web-server form (http://antismash.secondarymetabolites.org/) or as a stand-alone on a personal computer. The software uses the machine learning algorithm Prodigal or the interpolated Markov modeller Glimmer to detect open reading frames (ORFs) in the raw input sequences. The detection of the biosynthetic gene clusters is done by applying profile Hidden Markov Models (pHMMs) through the HMMer tool. All protein-encoding genes are analyzed with pHMMs based on alignments of signature proteins and protein domains to a library that provides models of signature genes and scaffolds for a range of secondary metabolites. False positive pHMMs are used to avoid misclassification of homologous structures, such as fatty acid synthases. Rules that define what needs to exist in order to constitute a biosynthetic gene cluster are manually curated and validated. The current version contains rules for 71 different BGCs. Another set of pHMMs is used to detect NRPS/PKS domains and predicts substrate specificity, stereochemistry and structure of the molecule. A comparison tool using the annotated database is then applied to attempt a functional understanding of the BGC. In order to predict unknown BGCs that could be missed by the antiSMASH detection module, a framework for automated detection of BGCs is used. Predicted Pfam domains are fed to an HMM. This allows for detection of BGCs in a more generalized way. The results of this pipeline are visualized in an interactive XHTML page. Gene clusters that were identified are shown in different colors, based on the classification. Furthermore, antiSMASH has a database of BGCs (https://antismash-db.secondarymetabolites.org/) detected in nucleotide sequences available from the National Center for Biotechnology Information (NCBI) GenBank[120]. An evolutionary context can be drawn from the comparison of the queried BGC with all known gene clusters, resulting in a better understanding of the secondary metabolite and assumption of gene functions based on the sequence homology. It also permits the user to browse by phylogeny or metabolite type, and provides statistics about the natural products in the database.

deepBGC has been available since 2019[100]. It also uses Prodigal to predict ORFs. However, it exploits the Pfam database using HMMer to predict protein domains. Because protein families represent functional elements in the gene clusters, they are useful for BGC identification. The Pfam domains are converted into numeric vector representations that take superfamily similarities into account. This is fed to a bidirectional long short-term memory (BiLSTM) neural network, composed of three layers. The input layer, which is comprised of

Pfam domains in their genomic order, in the form of sequential numerical vectors. The BiLSTM layer, composed of a network of forward and backward LSTM layers. These are basic memory cells with 128-dimension hidden state vector. This acts as the neural network memory, holding the information on the data the network processed before. The output from all the LSTM cells is refined by a single output layer with a sigmoid function: this provides a single value for each Pfam. This score for the Pfam domain represents the BGC classification score. The algorithm allows the user to set the minimum score Pfams need to achieve to be considered for a putative BGC, the default being 0.5. Consecutive genes are assembled into putative BGCs, and compound classes and biological activity are predicted using random forest classifiers.

Despite the variety of bioinformatic tools available to detect BGCs, the use of software to predict secondary metabolites gene clusters has its limitations[121]. With the constant update on the current knowledge of biosynthetic pathways, new families can be missed or misclassified. Prediction of compound structures based on genetic knowledge is often used to dereplicate natural products and focus on new molecules. However, tailoring enzymes cannot be predicted as precisely as core biosynthetic enzymes, which in turn results in inaccurate structure prediction. Additionally, bioactivity cannot always be inferred, challenging activity-based prioritization. Translating bioinformatic results to novel natural products is also a potential challenge, since silent and/or low-expressing BGCs require synthetic biology tools to be developed and accessible, in order to study their natural products.

Objectives

Taking in consideration that plasmids are key players in HGT, can contribute to phenotypes of interest for human and animal health and, historically, the plasmidome is not a prioritized genetic space, this work aimed to:

1. Determine the prevalence, taxonomic distribution, and type of product encoded by BGCs present in publicly available plasmid sequences using two different bioinformatic approaches.

2. To purify and characterize plasmids present in animal-related microbiomes.

3 A bioinformatic analysis confirms the plasmidome as a fruitful source of biosynthetic gene clusters of interest for human and animal health

Raíza de Almeida Mesquita, Antonio Ruzzini, Janet Hill

Raíza de Almeida Mesquita: Writing – Original Draft, Investigation, Project Administration, Formal Analysis, Validation, Visualization. **Antonio Ruzzini:** Resources, Methodology, Conceptualization. **Janet Hill:** Conceptualization, Visualization, Writing – Review & Editing.

3.1 Abstract

The role played by microbial communities in prevention, causation and aggravation of health states has only recently become appreciated. The members of any microbiome have functional redundancy. This redundancy occurs in two ways: recruitment of bacteria that share traits or trait sharing between bacteria. While the first is a co-evolved system, the second can happen on a shorter timescale and results in rapid evolution through the acquisition and accumulation of new traits. These traits can reflect a bacterium's environment, since they are niche-specific. Horizontal gene transfer (HGT) plays a major role in trait redundancy, since traits are encoded by genes. Plasmids, self-replicating extrachromosomal DNA molecules, are major participants in HGT. The acquisition or loss of a plasmid can drastically alter an individual's phenotype. Plasmid-encoded phenotypes include antibiotic resistance, virulence factors and bioactive small molecules. To better understand the prevalence, taxonomic distribution and products of BGCs harbored in plasmids, the sequences of complete plasmids available on the National Center for Biotechnology Information (NCBI) database were analysed with the software deepBGC and antiSMASH (antibiotics and secondary metabolites analysis shell) to predict the presence of secondary metabolites biosynthetic gene clusters (BGCs). Actinobacteria was the phylum with the highest percentage of plasmids with BGCs, followed by Cyanobacteria and Proteobacteria. At least one BGC was identified in 8.48 to 25.5% of the plasmids, varying according to software used. Averages of 1.1 and 2.64 BGC/plasmid were observed with antiSMASH and deepBGC, respectively. BGCs were detected across all phyla, suggesting valuable opportunity to explore less studied phyla for the discovery of new molecules.

3.2 Introduction

Bacterial metabolites (secondary or specialized) comprise a preeminent source of bioactive compounds. These molecules can be classified based on an array of chemical structures or biological activities[87–89]. Secondary metabolites are biosynthesized by metabolic pathways encoded by adjacent genes. These biosynthetic gene clusters (BGCs) encode the necessary enzymes, regulatory proteins, immunity proteins and transporters for the biosynthesis and export of the specialized metabolite. With the development of computational toolkits, these characteristics allow for computational identification of BGCs in DNA sequences.

Among the current available approaches to detect BGCs, deepBGC[100] is the newest software available to identify BGCs of different product classes. It uses the Prodigal algorithm to predict genes in the raw input sequence. Each of the genes detected are assigned to a protein family (Pfam) domain using HMMer. The software transforms each of the Pfam domains in a numeric vector, which is input to a bidirectional long short-term memory recurrent neural network (BiLSTM RNN). The BiLSTM layer of the software analyzes each Pfam domain in genomic order. The vector has binary flags that indicate where the domain is found in the protein (beginning or end). The memory cell processes the input layer and all previously seen Pfam, while the backward layer does the vector analysis in reverse order. The output from both memory cells is converted and results in a BGC score for the Pfam domain. Based on the classification scores, consecutive candidate genes are assembled to putative BGCs. Random forest classifiers are used to predict compound class and biological activity, which are then output to the user.

antiSMASH (antibiotic and secondary metabolite analysis shell) has been a popular free computational toolkit since it was established in 2011[99], with over 750 000 jobs processed in the web server[120]. Similar to deepBGC, antiSMASH uses Prodigal to detect open reading frames (ORFs) in the raw input sequence. A set of profile hidden Markov models (pHMMs) related to BGCs is applied in the input data. A set of manually curated and validated rules for different BGCs is used in the pHMMs. These specialized libraries allow the software to detect and catalog the various subclasses of the secondary metabolites. An algorithm to identify regions rich in Pfam domains runs in parallel. Finally, a filter for the cut-offs using the known minimal core components of each BGC class is applied. antiSMASH has a database of BGCs detected in nucleotide sequences available on GenBank. The database is used to compare the identified gene cluster to all known gene clusters, resulting in an evolutionary context that

provides more understanding of the role of the specialized metabolite. This allows assumption of gene functions based on sequence similarity, which is then output to the user.

To better understand the prevalence, taxonomic distribution and products of BGCs harbored in plasmids, the sequences of complete plasmids available in the National Center for Biotechnology Information (NCBI) nucleotide database were analysed with deepBGC and antiSMASH to predict the presence of secondary metabolite BGCs.

3.3 Materials & Methods

3.3.1 Input sequences

Complete sequences of plasmids available in the National Center for Biotechnology Information (NCBI) nucleotide database were downloaded in FASTA format, which is accepted by both software. The query was performed in November 2020, using "plasmid" and "complete sequence" as key words. The database allows for dividing the results by either organism classification (e.g., *Escherichia coli*, *Staphylococcus aureus*) or by phylum. The second option was chosen and the sequences were downloaded divided by phylum of plasmid host.

3.3.2 BGC identification and classification

The analysis was performed in the High-Performance Computing (HPC) Centre at the University of Saskatchewan. The software used were deepBGC version 0.1.26, and antiSMASH version 6.0 beta, with the default settings. This version of the deepBGC software provides a .json file output, that can be uploaded in the antiSMASH website, along with the sequence file. The version of antiSMASH that allows the upload of files from other software was only available by the website access, not in the standalone mode, at the time of this study.

By uploading the .json file from the deepBGC output in the antiSMASH website, the results of both programs can be seen side by side. However, the maximum size of input to the website is 150 MB. The sequence files that were over this limit were split and uploaded individually. Outputs were assessed individually and summarized by phyla.

3.3.3 Data analysis

In order to facilitate comparison of BGCs identification and classifications, the output of the both software had to be standardized. The nomenclature utilized for comparison of results was the one used by deepBGC, and it is shown in Table 3.1. deepBGC has an additional category, "no confident class", with no equivalent in the antiSMASH output. The data from this class were analyzed separately.

Table 3.1: Biosynthetic gene cluster product classification by the software used in this study. In the first column, the nomenclature used by deepBGC and applied to the classification results from both software. In the second column, the nomenclature used by antiSMASH

deepBGC	**antiSMASH**
Polyketides	Type I PKS Type II PKS Type III PKS Atypical PKS
RiPPs	Bacteriocins Bottromycin Cyanobactins Glycocin Head-to-tail cyclised peptide Lanthipeptides Lasso peptide Linaridin Linear azol(in)e-containing peptides Microviridin Proteusin Sactipeptide TfuA-related RiPP Thiopeptide
NRPs	NRPS Atypical NRPS Thioamide-containing NRPs
Terpenes	Terpenes
Alkaloids	Alkaloids
Others	Acyl amino acids Aminoglycosides Aryl polyenes Beta lactams Beta lactones Butyrolactones Ectoines Furan Homoserine lactone Indoles Ladderane lipids Melanins NAGGN Nucleosides Phenazine Phosphonate Resorcinol Siderophores Tropodithietic acid

3.4 Results

A total of 101 416 plasmid sequences were retrieved from the Genbank search: Fusobacteria (89); Chlamydiae (295); Tenericutes (296); Cytophaga, Fusobacterium, and Bacteroides (CFB) (566); Cyanobacteria (1377); Spirochaetes (5529); Actinobacteria (2186); Firmicutes (18170); and, Proteobacteria (72908).

3.4.1 deepBGC

The software deepBGC detected an average of 2.64 biosynthetic gene clusters per plasmid, with a total of 86139 BGCs identified. According to this software, 25.5% of the total set of plasmids harbor BGCs (Table 3.2). The distribution of BGC product classes detected by deepBGC in the total plasmid data set analyzed is shown in Figure 3.1.

The compound classes are predicted by random forest classifiers. One feature of this software is the "No confident class" category, for BGCs that cannot be positively assigned and, therefore, represent a plausible source for the discovery of novel gene cluster products and classes. The range of BGCs that could not be confidently classified by the software was between 64 (Firmicutes) and 94% (Spirochetes) of the detected gene clusters across all phyla analyzed (Figure 3.2).

Although virulence factors and antibiotic resistant phenotypes have been linked to the presence of plasmids in Chlamydiae, no BGCs were identified by deepBGC in plasmids from the Chlamydiae phylum. The other phyla had at least four biosynthetic gene clusters identified in their set of plasmids (Figure 3.3). The only detected product class that was common across all phyla was ribosomally synthesized and post-translationally modified peptides (RiPPs). With the exception of Cyanobacteria and Actinobacteria, all the other phyla had the largest portion of their gene cluster products identified as RiPPs, with all the detected gene clusters on Tenericutes plasmids being classified as such. In Cyanobacteria and Actinobacteria plasmids, the predominant product class was polyketides.

The proportion of non-ribosomal peptides (NRPs) products among the plasmids of different phyla ranged from 4.3 (CFB) to 27.9% (Cyanobacteria). NRPs gene clusters were not identified in Fusobacteria, Tenericutes or Spirochaetes plasmids. Terpene gene clusters comprised 50% of the BGCs detected in Fusobacteria and were not detected in plasmids from Tenericutes, CFB or Spirochetaes hosts. In the remaining phyla, terpene gene clusters

accounted for 4.3% (Firmicutes) to 8.9% (Cyanobacteria) of classified BGCs. Actinobacteria was the only phylum that had alkaloid gene clusters (0.14%) identified on its plasmids.

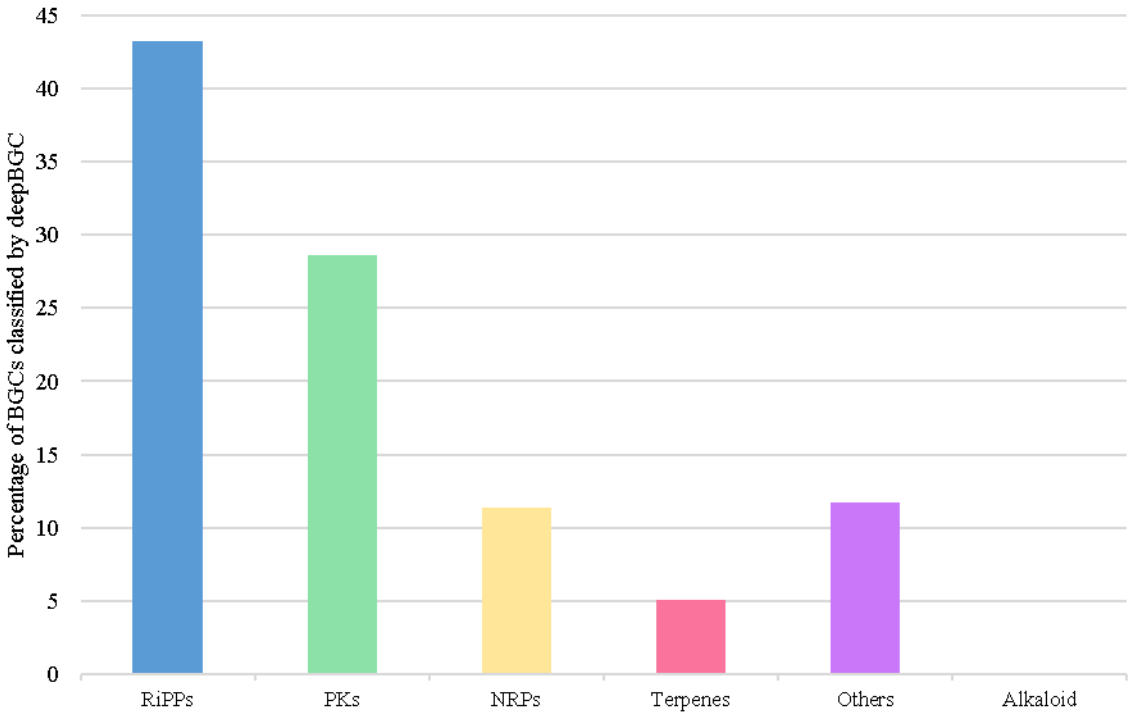

Figure 3.1: Percentage of gene cluster product classes detected on the total plasmid data set by deepBGC.

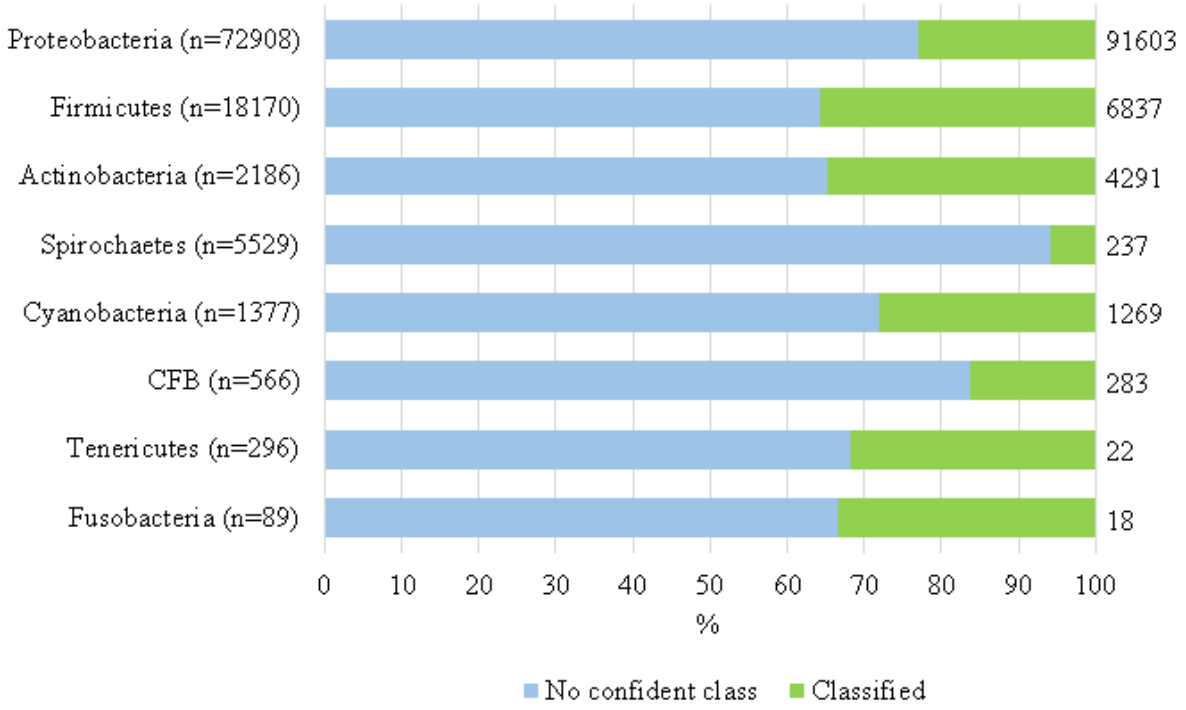

Figure 3.2: Proportion of plasmid biosynthetic gene clusters detected by the software deepBGC that could not be confidently categorized. Number in parentheses beside the phylum name is the number of plasmid sequences analysed. Number on the right-hand end of the graphic is the total number of BGCs for each phylum.

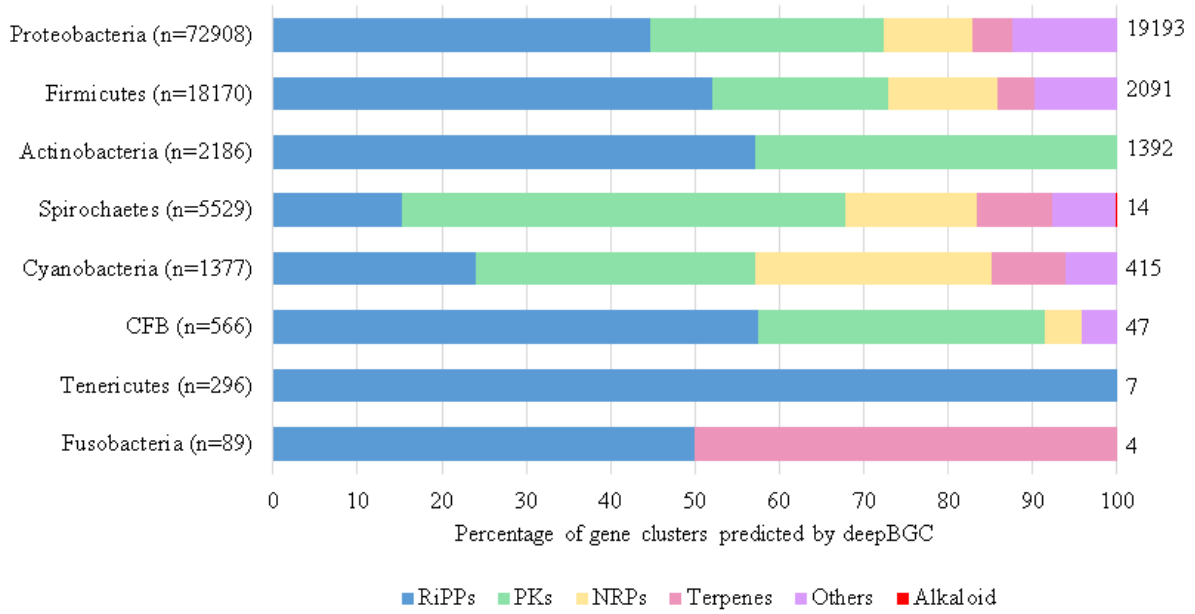

Figure 3.3: Biosynthetic gene cluster product classification by deepBGC and their distribution across phyla. Number in parentheses beside the phylum name is the number of plasmids analysed. Number on the right-hand of the graphic is the total number of BGCs that were classified by the software.

3.4.2 antiSMASH

A total of 12592 BGCs were identified by antiSMASH, across 8.48% of the plasmid sequences. An average of 1.1 BGC/plasmid was detected (Table 3.2). antiSMASH uses a set of manually curated pHMMs for different BGCs in its pipeline. The application of specialized libraries and a constantly updated database of BGCs permit the identification of various subclasses of the secondary metabolites, giving a better resolution of the results. However, to facilitate comparison with deepBGC results, these additional subclasses were combined according to Table *3.1*. Classes of the products detected in BGCs by antiSMASH and their distribution is shown in Figure 3.4.

Similar to the results of the deepBGC, no BGCs were identified in the Chlamydiae phylum. antiSMASH also did not detect any BGCs on plasmids from Tenericutes. The other phyla had at least two gene clusters identified on their set of plasmids (Figure 3.5). Two phyla had 100% of the gene clusters identified on its plasmids belonging to the same class, Fusobacteria (NRPs) and Spirochetes (RiPPs). The proportion of BGCs classified as RiPPs across phyla varied from 0 (Fusobacteria) to 74.1% (Firmicutes), being the largest parcel of BGCs detected in CFB (38%) and Firmicutes. In the other phyla, its representation ranged from 20 (Actinobacteria) to 30.6% (Proteobacteria). Apart from Fusobacteria, Cyanobacteria plasmids had the highest proportion of the detected gene clusters products classified as NRPs (42%). This classification otherwise varied from 9.5 (CFB) to 21.4% (Actinobacteria). Terpene gene clusters were the least common across all phyla, ranging from 1.2% (Firmicutes) to 13.9% (Actinobacteria).

The classification “Others” was the second most common, being responsible for the largest part of the BGCs identified in Actinobacteria and Proteobacteria. A detailed distribution of each of the classes assigned as “Others” and the frequency of which each was observed can be seen in Figure 3.6. Homoserine lactones and siderophores were the most observed products, accounting for 41 and 34.9% of all “Others” classification, respectively.

Siderophores were the only subclass common to all the five phyla that had BGCs classified as “Others” (Figure 3.7). Proteobacteria plasmids presented exclusive classifications of BGCs, and are responsible for 100% of the tropodithietic acid, N-acetylglutaminylglutamine amide (NAGGN), phenazine, phosphonate, homoserine lactone, and acyl amino acids gene clusters. Actinobacteria was the only phylum in which melanin products were observed. Known antibiotic molecules (β lactams and aminoglycosides) BGCs were detected on plasmids belonging to Actinobacteria and Proteobacteria.

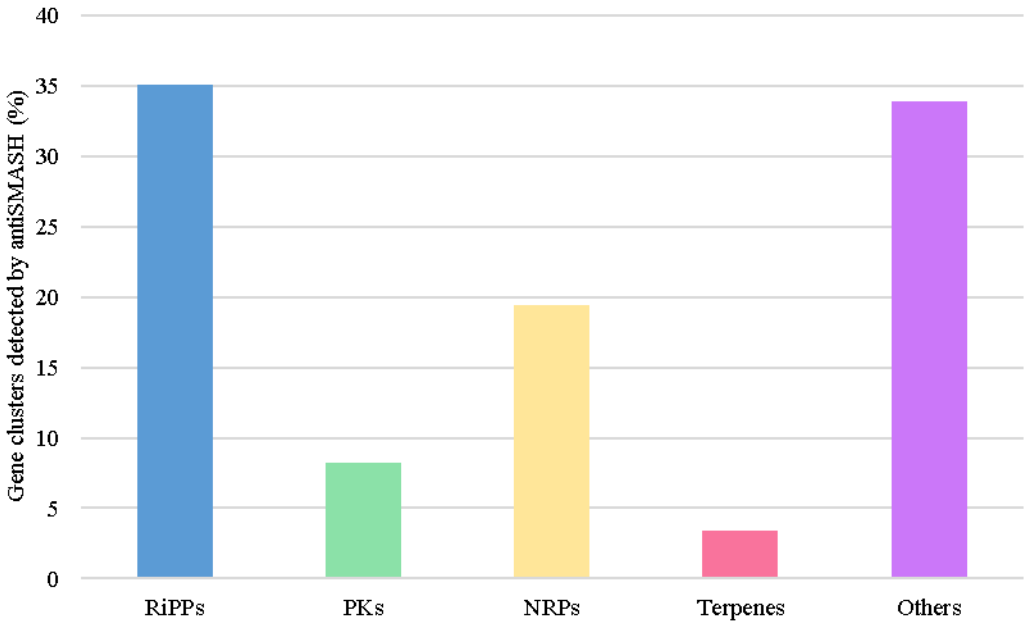

Figure 3.4: Distribution of gene cluster product classes detected on plasmids from different phyla by antiSMASH.

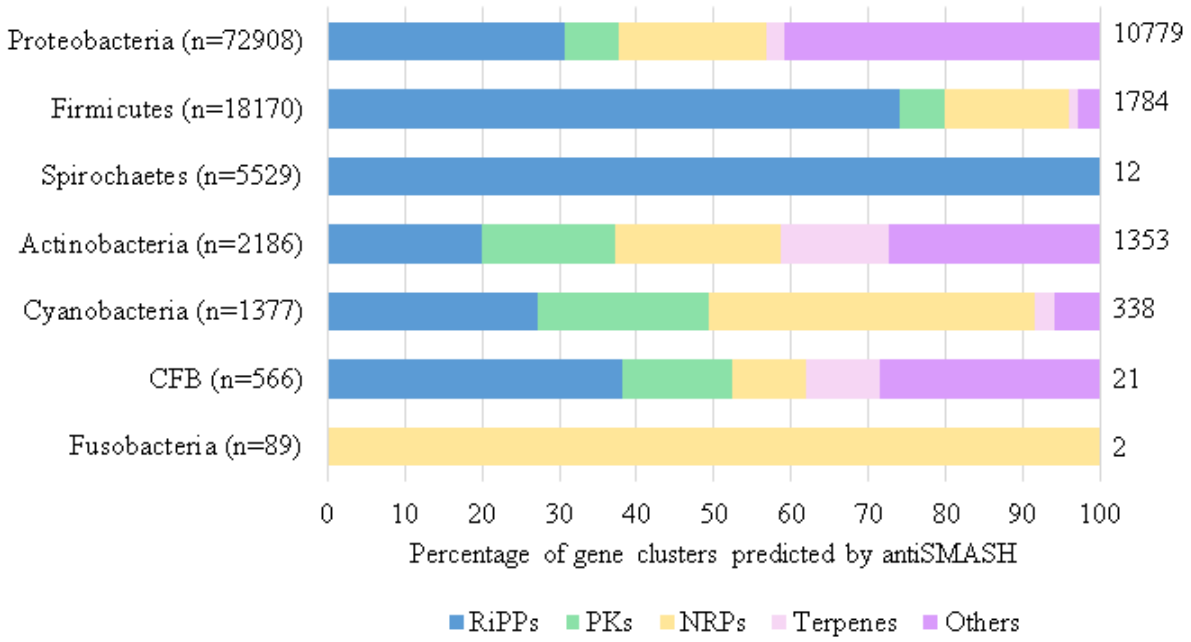

Figure 3.5: Distribution of the biosynthetic gene clusters products classes across phyla, identified by antiSMASH. Number on the right-hand of the graphic is the total number of BGCs that were classified by the software.

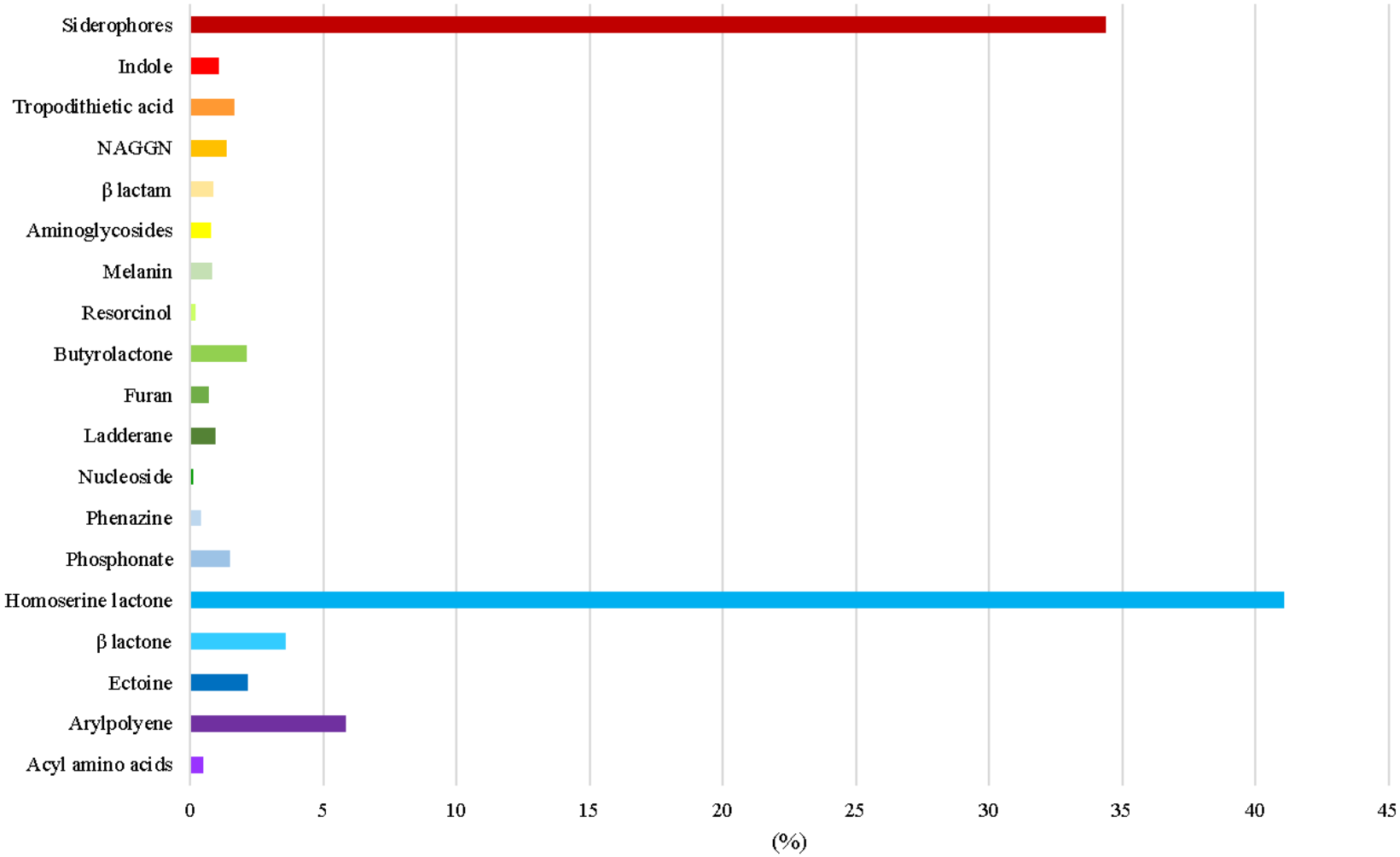

Figure 3.6: Distribution of BGCs classified by antiSMASH as “Other” (n=4852, all phyla combined).

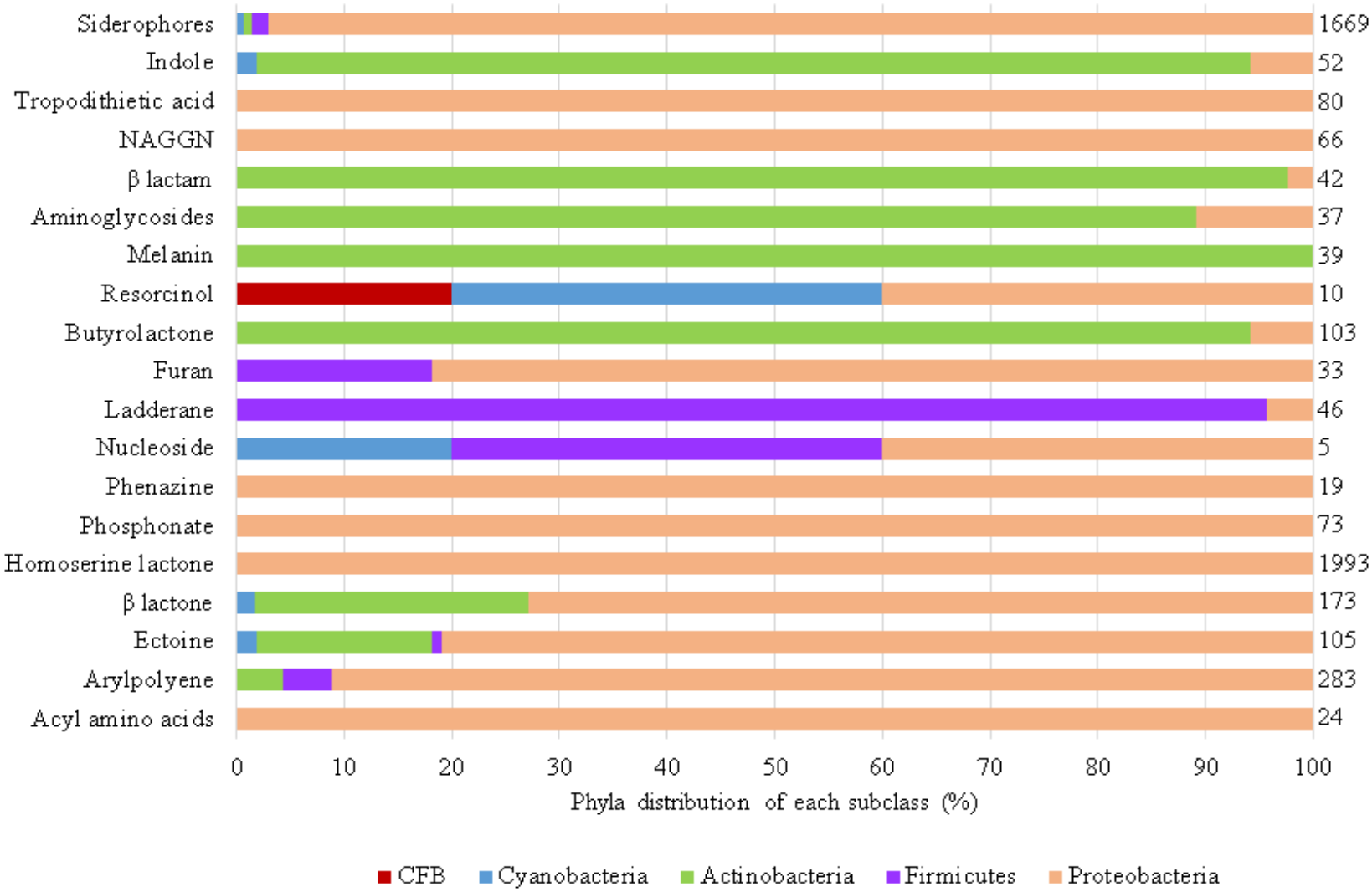

Figure 3.7: Distribution of BGCs products classifications grouped as "Other" across phyla. Number in the right-hand of the graphic is the total number of BGCs of each subclass.

3.4.3 Comparison of BGCs detected by deepBGC and antiSMASH

An overall summary of the BGCs detected, as well as BGC/plasmid and the percentage of plasmids that harbored BGCs are shown in Table 3.2. deepBGC detected over six times the number of BGCs detected by antiSMASH, in addition to twice the average number of BGC/plasmid. It is worth highlighting that, in the cases of Proteobacteria, Spirochaetes, Actinobacteria, CFB and Fusobacteria, the number of BGCs identified by deepBGC was an order of magnitude higher than the detected by antiSMASH. A total of 4732 (4.7%) plasmid sequences had BGCs detected by both programs. However, in most cases, the BGCs detected did not overlap, despite being in the same plasmid sequence.

A comparison of the numbers of classes of natural products detected by each software can be seen in Figure 3.8. RiPPs, PKs and Terpenes were identified by deepBGC 2, 5.65 and 2.45 times more, respectively. antiSMASH detected 150 more BGCs belonging to NRPs than deepBGC. While antiSMASH identified more BGCs belonging to "Others", it is worth noting that the parameters for this classification in the deepBGC software are not clear.

The distribution of the BGCs classes identified by each software, divided by phylum, is shown in Figure *3.9* and Table 3.3. While predictions by the two programs did not agree fully in any phylum, some disagreements were more perceptible. The difference in numbers can be observed mostly in the classes of RiPPs and PKs in all phyla. Fusobacteria (Figure 3.9G) showed no agreement in classifications, despite having two plasmid sequences where gene clusters were detected by both programs. All the gene clusters detected by antiSMASH were classified as NRPs, while deepBGC classified 50% as RiPPs and 50% as Terpenes.

Table 3.2: Number of BGCs identified, BGC/plasmid and percentage of plasmids that had BGCs according to each software used. The last column shows the number of plasmid sequences that had BGCs detected by both software.

	deepBGC			**antiSMASH**			**Sequences that overlap**
	BGC identified	BGC/plasmid	% plasmids with BGCs	BGC identified	BGC/plasmid	% plasmids with BGCs	
Proteobacteria (n=72908)	76047	6.5	29.9	9862	2.1	9.0	3922
Firmicutes (n=18170)	5391	2.3	13	1673	1.2	7.6	434
Spirochaetes (n=5529)	227	1.2	3.4	12	1	0.2	6
Actinobacteria (n=2186)	3185	3.3	44	779	1.9	18.9	258
Cyanobacteria (n=1377)	989	2.2	33	239	1.3	13.4	102
CFB (n=566)	260	2.4	19.4	25	1.4	3.2	8
Tenericutes (n=296)	22	1.4	5.4	0	0	0	0
Chlamydiae (n=295)	0	0	0	0	0	0	0
Fusobacteria (n=89)	18	4.5	4.5	2	1	2.25	2
TOTAL	**86139**	**2.64**	**25.5**	**12592**	**1.1**	**8.48**	**4732**

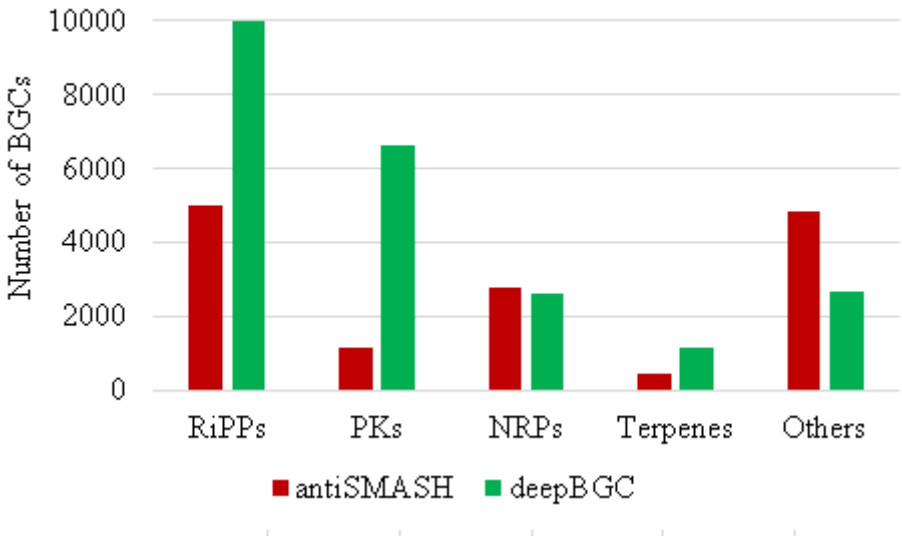

Figure 3.8: Comparison of the total number of BGCs of each natural product class in the plasmid sequences detected by each software tested.

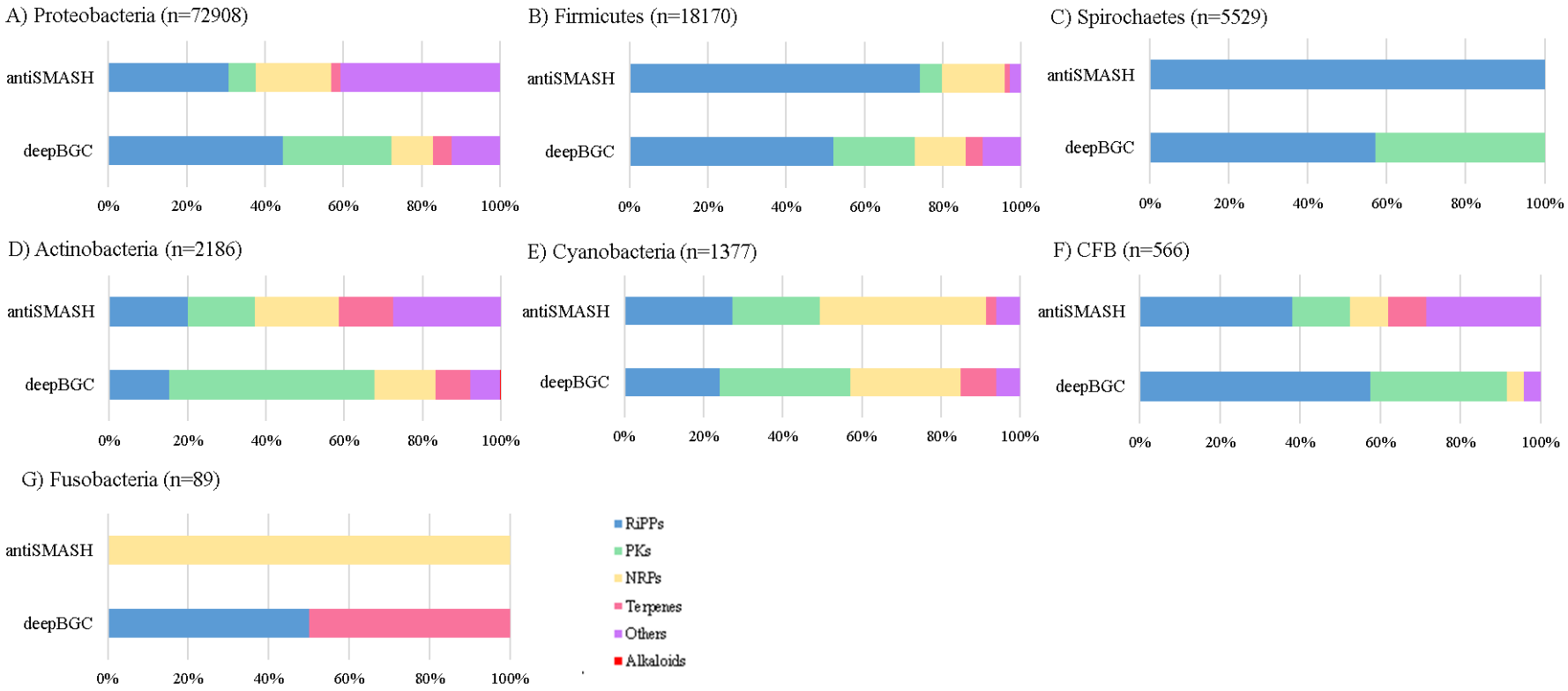

Figure 3.9: Comparison of the percentage of each class of BGC identified by the software, by individual phylum.

Table 3.3: Percentage of gene clusters classified in the different natural product families, according to each software.

	Proteobacteria		**Firmicutes**		**Spirochaetes**		**Actinobacteria**		**Cyanobacteria**		**CFB**		**Fusobacteria**	
	deepBGC	antiSMASH	deepBGC	antiSMASH	deepBGC	antiSMASH	deepBGC	antiSMASH	deepBGC	antiSMASH	deepBGC	antiSMASH	deepBGC	antiSMASH
RiPPs	44.7	30.7	52.1	74.2	57.1	100	15.2	20	24.1	27.2	57.4	38.1	50	0
PKs	27.6	7.1	20.7	5.7	42.8	0	52.6	17.1	33	22.2	34	14.3	0	0
NRPs	10.5	19.1	12.9	16.1	0	0	15.6	21.4	27.9	42	4.2	9.5	0	100
Terpenes	4.8	2.4	4.4	1.2	0	0	8.9	13.9	8.9	2.7	0	9.5	50	0
Others	12.3	40.8	9.8	2.9	0	0	7.5	27.4	6	5.9	4.2	28.6	0	0
Alkaloids	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0

3.5 Discussion

Sequences of complete plasmids available in the NCBI nucleotide database were analysed with deepBGC and antiSMASH to detect the presence of secondary metabolite BGCs. This information can be used to better understand the prevalence, taxonomic distribution and products of BGCs harbored in plasmids. Ultimately, these results can be used to streamline and optimize the discovery of new molecules. As much as bacteria are known for being a rich source of bioactive metabolites, apart from Actinobacteria and Cyanobacteria, their potential has been mostly underexplored[122]. In this aspect, this work appears to be the first to shine a light on secondary metabolites gene clusters on Proteobacteria, Firmicutes, Spirochaetes, Tenericutes, CFB, and Fusobacteria as a phylum, not focusing in specific genus and/or species.

Marine Proteobacteria have recently been in the spotlight of genome mining for the discovery of new bioactive molecules[122–125]. Despite being an abundant phylum in aquatic environments (50 to 80% of aquatic bacteria), few compounds have been described from Proteobacteria when compared to other phyla[122]. Among NRPs produced by marine Proteobacteria, indigoine and thiomarinols and have antibacterial activity; and myxothiazols can act as an antifungal.

Actinobacteria is one of the most ecological diverse phyla, and that diversity can be observed in the variety of secondary metabolites produced[126,127]. This phylum is responsible for two thirds of the antibiotic scaffolds used in clinic today, as well as other bioactive compounds such as immunosuppressants, herbicides and antivirals, among otheres[126]. Multiple efforts focusing in marine Actinobacteria have been successful in reporting new polyketides, phenazines, isoprenoids and terpenes[128–130]. Resources were also used to mine genomes for ribosomally synthesised and post-translationally modified peptides[116,131,132]. Poorinmohamma et al. (2019) identified at least one RiPP BGC in 25.5% of the tested genomes (n=629), totaling 477 BGCs in 185 strains[131]. Remarkably, all the subclasses of RiPPs known at the time were detected. The proportion of BGCs identified as RiPP in 2,186 Actinobacteria plasmids in our current study ranged from 6.7% (deepBGC) to 3.8% (antiSMASH). This discrepancy can be explained by the total number of sequences used in each study, as well as the fact this study was focused on the plasmidome. This phylum appears to be a rich source of lanthipeptides (subclass of RiPPs)[131,133–136], and 1163 lanthipeptide BGCs were reported in 830 actinobacterial genomes[137].

Insect-associated actinobacteria also have been investigated, and case studies have shown the production of alkaloids, phenylpyrrolines, lanthipeptides and polyketides can have influence in maintaining the health of the insect and its nest[138,139]. Endophytic actinobacteria associated with medicinal plants displayed a prevalence of genes encoding polyketide synthases I and/or II in 33 of the 52 strains (63.46%)[140]. Marine actinobacteria isolated from Deep Sea core sediments (n=123) harbored genes for PKSs type I and/or II (39.13 and 56.52%, respectively) and for NRPSs (69.57%)[141]. While all of these numbers come from very specific bacterial sources and limited number of sequences, the biggest discrepancy when compared to the results obtained in this study is regarding NRPs. The detection of NRPs in this study ranged from 15.59% (deepBGC) to 21.43% (antiSMASH). However, the fact that the genetic space focused in this study is the plasmidome should be taken in consideration.

Actinobacteria was the only phylum in which detection of alkaloid BGCs was observed (Table 3.3). Despite the low incidence (0.14% of BGCs classified by deepBGC), multiple alkaloids produced by actinobacteria have been reported. Again, the focus on marine Actinobacteria seems to be successful. The approach resulted in the discovery and characterization of anandins A and B[142], actinobenzoquinoline and actinophenanthrolines A-C[143], and unnamed alkaloids produced by *Nocardiopsis* sp. NCS1[144]. Soil actinobacteria are also successful producers of alkaloids, such as nonocarbolines A-E[145], pyridine-2,5-diacetamide[146], N-acetyltyramine and N-acetyltryptamine[147].

Cyanobacteria are recognized for producing a range of secondary metabolites, mostly NRPs and PKs, or hybrids NRP-PKs[148,149]. Multiple genome mining experiments were conducted in this particular phylum, successfully showing that it is a prolific source for the discovery of molecules[148–153]. Wang et al. (2011) reported a total of 145 BGCs detected in 43 genomes, mostly classified as RiPPs, as precursor peptides were identified[148]. RiPPs BGCs were also detected by Laikoski et al. (2012), in the form of cyanobactin pathways, in 24.6% of the genomes analyzed[150]. The range of RiPPs in Cyanobacteria detected in this study varied between 7.9% (deepBGC) and 3.7% (antiSMASH). Larsen et al. (2021) reported the presence of type III PKSs in 17% of the 517 cyanobacteria genomes analyzed[152]. The rates of PKs observed in our study are lower (between 10.9 and 3.1%). However, these studies do not differentiate the gene clusters harbored in plasmids from the ones in the chromosome, or make it clear whether or not plasmids were even analyzed, so comparisons of prevalence are difficult.

A single study (unpublished) on the importance of plasmids in the production of natural products in Cyanobacteria reported at least one BGC in 17% of the total plasmids (n=424)[154]. BGCs were identified using antiSMASH. The proportion of Cyanobacteria plasmids with BGCs in the present study ranged from 13.36 (antiSMASH) to 33% (deepBGC). This difference among results can be explained by the fact that the number of sequences analyzed in our current study (n= 1377) is over three times that examined by Popin et al (2020). Regarding the classification, the presence of hybrid NRPs/PKs was reported to account for more than half of the gene clusters, followed by NRPs and RiPPs, with Terpenes being the least observed product. While the classification of hybrid gene clusters was not used in this study, NRPs and PKs were the most detected class of BGCs, and the frequency of RiPPs and Terpenes is in agreement with what was identified by Popin et al. (2020).

The use of two programs that apply different methods for BGC detection and classification resulted in disagreements in identification of BGCs. Machine learning approaches, such as deepBGC, have a bigger potential to detect novel BGCs and, therefore, completely new molecules[155]. However, a higher rate of false positives has been noted when compared to rule-based approaches, as is the case of antiSMASH. Hrab et al. (2021) analyzed the complete genome sequence of *Streptomyces cyagenus* S136 using an array of bioinfomatic tools, including antiSMASH and deepBGC[156]. The number of BGCs detected were 102 (deepBGC) and 33 (antiSMASH). Manual comparison of the output of all the software used revealed 41 BGC. This "control" is only achievable in a small sample size study, as was the case (n=1). The authors highlight that 12.5% of the BGCs detected were considered novel, but did not explain how the manual analysis agreed with the results shown by either software used. This study observed a range of differences between the detection and identification of BGCs by antiSMASH and deepBGC. As mentioned previously, deepBGC has the capacity of detecting novel BGCs, and that is supported by the number of gene clusters that could not be confidently categorized, which was over 60% across all phyla (Figure 3.2). deepBGC detected, in most cases, an order of magnitude more BGCs than antiSMASH (Table 3.2). This pattern was also observed by Yamani (2021), where the number of gene clusters detected by antiSMASH were two orders of magnitude smaller than the total BGCs detected by deepBGC[157].

3.6 Conclusions

Even with the advantage of next generation sequencing (NGS), researchers continue to focus on the same phyla (Actinobacteria and Cyanobacteria), since it has been proven time and again that these are rich suppliers of bioactive molecules. However, this work points to bacterial plasmids across many phyla being a prolific source of biosynthetic gene clusters with potential bioactivity. This knowledge can be leveraged to focus on specific phyla of bacteria depending on the class of molecules that the research is focused on. The results of this work suggest that, in order to study plasmid-encoded RiPPs, a researcher should focus on Proteobacteria, Firmicutes, Spirochaetes and CFB isolates, whereas for polyketides, Actinobacteria plasmids would be more fruitful. Cyanobacteria and Fusobacteria plasmids would harbor more NRPs (Table *3.3*). However, if the goal is to screen plasmids for any kind of secondary metabolite gene cluster, Actinobacteria showed the highest rates of plasmids harboring BGCs (Table *3.2*).

When adding the uncategorized BGCs and the higher orders of magnitude of detection by deepBGC, even with the chance of false positives, it seems like a fruitful bioinformatic approach for mining novel natural products. The fact that it also assigns activity of the BGC product can be leveraged to BGC-prioritization for research. However, antiSMASH shows a higher resolution of the classes of the BGCs detected, allowing the user to optimize methods for product isolation when going from the bioinformatic tool to the lab bench. Furthermore, it displays a percentage similarity of the predicted product with the products available in the antiSMASH database, which can be useful to avoid rediscovery of molecules.

Transition statement

In Chapter 3, we showed by using bioinformatic tools, that the plasmidome is a fruitful source of molecules of interest for human and animal health. We then aimed to apply these bioinformatic tools in an investigation of the plasmid-encoded BGCs of bacteria isolated from animal-associated microbiomes.

4 Screening, purification and characterization of plasmids from animal microbiomes

Raíza de Almeida Mesquita, Jenny Liang, Yenuki Rajapaksha, Michelle Sniatynski, Joe Rubin, Antonio Ruzzini, Janet Hill

Raíza de Almeida Mesquita: Writing – Original Draft, Investigation, Project Administration, Formal Analysis, Validation, Visualization. **Jenny Liang:** Investigation. **Yenuki Rajapaksha:** Investigation, Formal Analysis. **Michelle Sniatynski:** Investigation, Methodology, Validation. **Joe Rubin:** Supervision, Resources, Funding Acquisition, Conceptualization. **Antonio Ruzzini:** Supervision, Resources, Methodology, Funding Acquisition, Conceptualization. **Janet Hill:** Conceptualization, Visualization, Writing – Review & Editing.

4.1 Abstract

In complex microbial communities, individuals can acquire and accumulate new traits that allow them to adjust to different conditions in their environment by sharing genes. Sharing plasmids, which are self-replicating extrachromosomal DNA molecules, allows significant exchange of genetic material and can contribute greatly to bacterial evolution. In light of the potential transformative role of plasmids in bacterial evolution and lifestyle, we purified and characterized plasmids from organisms isolated from companion animals, wildlife, and livestock – sources that humans encounter daily. Isolates from animal-associated microbial communities were screened for the presence of plasmids: cat-associated microbial community (n=50), *Escherichia coli* isolated from wild birds (n=65) and isolates from bovid-associated microbiomes (n=250). At least one plasmid was detected in 47/50, 36/65 and 38/250 of these isolate collections, respectively. Phenotypic screening was done for the plasmids recovered from wild bird isolates, and plasmid-encoded antibiotic resistance was detected in seven isolates. The taxonomic identity of plasmid-harboring strains from bovid-associated microbiomes was determined using the 16S rRNA gene and showed that most of the identified plasmid harboring isolates were Gram-positive. Apart from *K. pneumoniae*, all isolates belong to the phylum Firmicutes. Twelve plasmids from seven host isolates were chosen to be sequenced and three plasmids (pRAM-12, pRAM-19-2 and pRAM-30-2) shared 100% nucleotide sequence identity. Curiously, two of the parent strains shared more than one plasmid in common, despite being in different phyla: pRAM-19-1 and pRAM-30-1 were also identical. Host RAM-19 was identified as *K. pneumoniae*, a Proteobacteria; and RAM-30 as *B. licheniformis*, a Firmicute. pRAM-28 from *S. aureus* contained genes encoding the bacteriocin aureocin A70; and pRAM-21 has 100% nucleotide identity to pRAM-28. Additional analysis of sequences of plasmids from the bovid isolate collection resulted in the detection of three other bacteriocins: cloacin and two putative gene clusters for lanthipeptides. The results of this work suggest that the plasmidome is an important source of potential unknown secondary metabolites that are used by bacteria to compete with each other within and between microbiomes. Genetic exchange and the apparent plasmid-sharing highlights the intimacy of interactions within a community.

4.2 Introduction

It is well known that the interface between humans, domestic animals, wildlife and the environment influences the health status of all living beings involved. This interface has been entangled in the emergence of infectious diseases and new and re-emerging zoonoses[158,159]. Changes to the food production system and closer contact with companion animals can explain the increase of animal-borne zoonoses, since there is increased opportunity for direct transmission with and without vectors[160]. However, by approaching animals as a host-microbe ecosystem, we can develop new insights into the maintenance of human health, especially by recognizing that the interaction between bacteria and other organisms are central for the health status of both the individual and the environment[161]. Although the role of individual bacteria in infectious disease has been studied since the nineteenth century, research on microbiomes and their roles in prevention, cause or aggravation of diseases only began in the recent decades[162].

Microbiomes are defined by their bacterial membership and environmental niche. A common feature of microbiomes is functional redundancy among their members[1,2]. There are two major paths to this redundancy: recruitment of bacteria that share traits or trait sharing between bacteria. Specific recruitment implies a long, well-established system that has co-evolved. Trait-sharing among bacteria can happen on a much shorter timescale. The latter is facilitated by interactions between individuals within a microbiome via horizontal gene transfer (HGT), which allows for the acquisition and accumulation of new adaptive traits that reflect their environment. Among the known mechanisms of HGT, plasmid exchange is one that allows for rapid transformative effects from the acquisition of large gene collectives. A plasmid can drastically alter an individual's phenotype. Their acquisition has been correlated to the development of antibiotic resistance[24,163,164], virulent phenotypes[13,15,165,166], and iron-scavenging molecules[17,18].

Nearly all plasmidomic studies are simply nucleotide sequence-based metagenomics[8,9]. This results in knowledge of plasmids regardless of host bacteria. In contrast, the functional annotation of plasmid-encoded genes has typically been done by studying the genes or gene clusters on a case-by-case basis rather than through the prioritization of this genetic space. More often than not, these functional studies are not focused on plasmid-encoded traits and this genomic context is a chance finding, not the emphasis of the study. As a result, plasmid-encoded traits have been revealed on a

plasmid-by-plasmid basis rather than attempting a systematic approach that prioritizes plasmid isolation and functional annotation through phenotypic screens prior to nucleotide sequencing. Therefore, the objective of this project was to prioritize this genetic space. We purified and characterized plasmids from known bacterial isolates isolated from animal-related microbiomes. We focused on plasmids from microbial communities that have the greatest potential to influence our own microbiomes, those with overlapping environment and that are associated with companion animals, wildlife and livestock. From our samples from dairy and beef cattle in Saskatoon, Saskatchewan, we re-discovered a four-membered bacteriocin system that was originally described from *Staphylococcus aureus* isolated from milk in Brazil; two putative gene clusters for lanthipeptides; as well as a cloacin gene cluster. Our results show that a prioritization of the plasmidome along with a systematic evaluation of plasmids from related microbiomes is a productive approach to small molecule discovery.

4.3 Materials and Methods

4.3.1 Bacterial isolation and growth

Escherichia coli isolated from wild birds

A collection of *Escherichia coli* isolated from wild birds was donated by the Rubin Lab. Sixty-five crows (*Corvus brachyrhynchos*) that were taken to the Veterinary Medical Center at the University of Saskatchewan had swabs from their cloacas. Swabs were plated on chromogenic extended-spectrum beta-lactamases (ESBL) agar and incubated at 37 ºC overnight. This agar allows for the selective isolation of ESBL-producing isolates, with high sensitivity and specificity, based on color difference[167,168]. *E. coli* presents itself in a dark pink to reddish color. These colonies were re-streaked on Columbia Blood Agar (5% sheep's blood) at 37 °C overnight. For plasmid screening and purification, the isolates were cultivated in 50 mL lysogeny broth (LB) overnight, at 200 rpm and 37 ºC.

Cat-associated microbiota

Fifty isolates obtained from cat feces were received from Dr. Ruzzini. The media used for bacterial isolation were R2A and Brain Heart Infusion (full-strength or 1⁄10). For isolation experiments, samples were plated and incubated at 30 °C. After growing the isolates in liquid broth, monocultures were verified by re-streaking single colonies, using the same medium that was used for isolation. For plasmid screening and isolation experiments, the isolates were grown in 50 mL of the same medium used for isolation, overnight, at 200 rpm and 30 ºC.

Bovid-associated microbiomes

Bacteria were isolated from five distinct bovid-associated microbiomes: dirty and clean bedding from a dairy barn, milk and teat canal swabs of dairy cattle, and bovine feces (beef and dairy cattle). The milk and mammary swabs were collected from both healthy and mastitic cattle. This work was designed and conducted in accordance with the Canadian Council for Animal Care and approved by the University Animal Care Committee at the University of Saskatchewan (Protocol Nº AUP20080015). Three

distinct media were used for bacterial isolation: R2A and 1/10 Brain Heart Infusion supplemented with cyclohexamide (200 μg/mL), or Columbia Blood Agar (5% sheep's blood). For isolation experiments, serial dilutions of samples were plated and incubated at room temperature, 30 ºC and 37 ºC. Single colonies were picked from isolation plates to generate monocultures. Morphological diversity was used to prioritize colony selection for further propagation. Monocultures were verified by re-streaking single colonies that were picked and grown in liquid broth, typically the same medium that was used for isolation but without the addition of the antifungal agent (cyclohexamide).

For plasmid screening and purification experiments, the isolates were cultivated in 50 mL LB overnight, at 200 rpm and 37 ºC. For more fastidious organisms, tryptic soy broth supplemented with additional yeast extract (3 g/L) was used for bacterial propagation.

4.3.2 Plasmid screening and purification

All the isolates were grown in 50 mL cultures. After harvesting the cells by centrifugation, the pellet was subjected to an alkaline lysis[169,170]. Briefly, the pellet was resuspended in 4 mL of solution I (10 mM Tris-HCl, 1mM EDTA pH 8.0) containing 10 μg/mL of RNase A. After full resuspension, 6 mL of solution II (0.2 M NaOH, 1% sodium dodecyl sulfate) was carefully added and mixed by inversion. 8 mL of solution III (3 M potassium acetate) was used to neutralize the mixture. The addition of solution III was followed by careful inversion and an incubation on ice for 15 minutes. After another centrifugation at 20627 × g, 4 ºC for 10 minutes, the supernatant was mixed with one volume of cold isopropanol and kept at –20 ºC for two hours. Following the incubation time, the DNA was pelleted by centrifugation at 20627 × g, 4 ºC for 30 minutes. The supernatant was discarded and the pellet washed with ethanol twice. After centrifuging again, for the removal of the ethanol, the pellet was air-dried. The dry DNA pellet was resuspended in 50-100 μL of TE buffer (10 mM Tris-HCl, 1mM EDTA pH 8.0). 250 ng of the obtained DNA were loaded on a 0.8% agarose gel with ethidium bromide for analysis. Isolates that harbored plasmids were grown at a larger scale (200 mL) and plasmids were purified using a commercial kit (PureLinkTM, HiPure Plasmid Midiprep Kit, Invitrogen®) for better yield and purity.

4.3.3 Phenotypic screening plasmids from wild bird *E. coli* isolates

A plasmid-encoded phenotypic screening pipeline (Figure 4.1) was developed and applied to the *E. coli* isolate collection from wild birds. First, chemically competent NEB® 5α competent *E. coli* (New England Biolabs, Whitby, ON) was transformed with pools of purified plasmids. Pools of five plasmids were created, using 50 ng of each plasmid. After pooling, the mixture was dried by speed vacuum and resuspended in 10 µL of TE buffer. A 50 µL aliquot of chemically competent *E. coli* cells was thawed on ice and 3 µL of the plasmid pool was added into the vial and mixed gently by tapping. The mixture was incubated on ice for 30 minutes and heat-shocked in a 42 °C water bath for 30 seconds. Another ice incubation for five minutes was done before adding 950 µL of room-temperature LB. This was followed by an incubation at 37 °C for 1 hour at 200 rpm in a shaking incubator. After the incubation period, 100 µL of cells were spread on LB agar plates supplement with kanamycin (50 µg/mL), ampicillin (100 µg/mL), chloramphenicol (40 µg/mL), tetracyline (25 µg/mL), and sulfamethoxazole (20 µg/mL). The plates were incubated at 37 °C overnight before being accessed for growth. Plates that presented growth had the plasmids from the pools transformed individually, to associate plasmids with the observed phenotype.

Plasmids that did not have a natural antibiotic resistance gene had *in vitro* transposon insertions of an antimicrobial resistance cassette, in order to construct plasmids that could be propagated in *E. coli*[163,171–173]. By using a transposase, we were able to insert a kanamycin (kan^R) resistance marker randomly into the plasmid DNA. The reaction was set up with 1 µL of EZ-Tn5 10X Reaction Buffer, 0.2 µg of target DNA and the molar equivalent of EZ-Tn5 Transposon, 1 µL of EZ-Tn5 Transposase (1 U) and sterile water to a reaction volume of 10 µL. The mixture was incubated at 37 °C for two hours. The addition of 1 µL of EZ-Tn5 10X Stop Solution was followed by incubation at 70 °C to stop the reaction.

Chemically competent NEB® 5α competent *E. coli* was transformed using 5 µL of the insertion reaction. The transformation was done as previously described. After the incubation period, 100 µL of cells were spread on LB agar plates with kanamycin (50 µg/mL). The plates were incubated at 37 °C overnight. To construct the transposon-aided capture (TRACA) library, 192 colonies were randomly picked from the plates and used to inoculate two 96-well plates. Each well contained 200 µL of LB plus kanamycin. The

plates were incubated at 37 °C overnight with shaking at 200 rpm. This library was used as a starting point for the phenotypic screening assays.

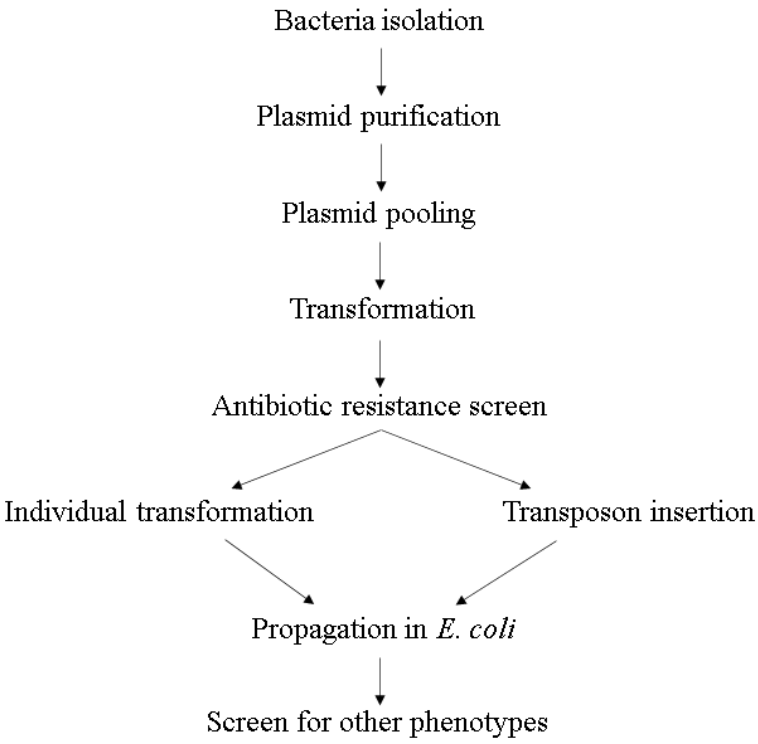

Figure 4.1: Fluxogram of the steps used in wild bird *E. coli* plasmid-associated phenotype screening experiment.

In order to identify the production of a broad range of pathogenic effectors, we used an assay inspired by a naturally occurring interaction in the soil: the predator-prey relationship between amoeba and bacteria. The amoeba *Dictyostelium discoideum* requires bacteria as a food source. Consumption of edible bacteria (bacteria that do not produce any virulence factors) leads to the development of macroscopic fruiting bodies, while inedible bacteria (isolates expressing virulence factors) prevent amoeba1 growth. The co-cultivation of amoeba and bacteria has been done to identify pathogens[174] and as a starting point for the discovery of small molecules[175,176], since amoeba toxicity is a proxy for toxicity to other eukaryotes. The TRACA library was used as a food source. After growing the bacterial library for 6 hours, 5 μL of each isolate were spotted on SM/5 agar (per liter: 2 g glucose, 2 g bacto peptone, 0.2 g yeast extract, 0.2 g $MgSO_4*7H_2O$, 1.9 g KH_2PO_4, 1.0 g K_2HPO_4, and 15 g agar). A suspension containing an axenic culture with 7 x 10^4 cells per mL of *D. discoideum* cells was spotted (5 μL) on top of the dry bacterial food source spots. The plates were incubated at room temperature for 48 hours and the presence or absence of amoeba growth on each spot was assessed visually.

Iron is an essential element for the growth of bacteria and is often correlated to pathogenesis. Siderophores are small molecules that scavenge iron from the host or the environment [177]. Siderophore production was monitored by spotting 10 μL of each isolate from an overnight culture of TRACA library on a nutrient agar containing chrome azurol S (CAS), iron (III) and hexadecyltrimethylammonium bromide[178]. These components form a complex that is disrupted by siderophore activity, resulting in colour change from blue to orange. Plates were incubated at 37 °C overnight before being assessed for colour change.

To detect plasmid-encoded antimicrobial production, we used an assay that has been successfully implemented to identify biologically active small molecules from cosmid libraries [179–181]. The *E. coli* clones from the TRACA library were spotted on LB agar plates containing kanamycin and incubated at 37 °C overnight. A thin layer of soft agar (0.4%) containing an intruder bacterium is overlaid. *E. coli* harboring the plasmid pET-41 was used as an intruder, since this plasmid confers kanamycin resistance to its host. The clones that have the ability of producing antimicrobial molecules were identified after incubation at 37 °C, by the inability of the intruder strain to grow nearby, creating a zone of inhibition that is easily recognizable from the lawn of the intruder strain growth.

4.3.3 Isolate identification

The taxonomic identities of isolates that harbored plasmids were determined by 16S rRNA gene sequencing. The PCR was carried out using universal primers 8F (5'-AGA GTT TGA TCC TGG CTC AG-3') and Eub1492R (5'-ACG GCT ACC TTG TTA CGA CTT-3'). Each PCR reaction (50 μL) contained 1X Dream Taq Buffer (proprietary buffer that contains KCl, $(NH_4)_2SO_4$ and $MgCl_2$), 0.2 mM dNTP, 0.3 μM of each primer, approximately 500 ng of template DNA, and 5 U of Dream Taq (ThermoFisher, Waltham, MA). The thermocycler parameters were: 95 °C for 5 minutes, followed by 30 cycles of denaturation at 95 °C for 30 seconds, annealing at 50 °C for 30 seconds, extension at 72 °C for 90 seconds, and final extension for 5 minutes.

PCR products were visualized on a 0.8% agarose gel, and purified using QIAquick PCR Purification Kit (Qiagen, Mississauga, ON). The Sanger sequencing of the PCR amplicons was done by Macrogen, Korea. An 800 bp region of each gene sequence was used as a blastn query of the NCBI 16S Microbial database.

4.3.4 Plasmid sequencing, de novo assembly and annotation

Plasmid DNA libraries were prepared using Nextera DNA Flex Library Prep kit (Illumina) according to the manufacturer's instructions and sequenced using an Illumina iSeq-100 at the University of Saskatchewan Next-Generation Sequencing Facility. Paired end reads were assembled into contigs using Geneious software[182], applying default settings. Contigs with one order of magnitude higher coverage than the rest (suggestive of plasmid derived contigs) were annotated through Rapid Annotation using Subsystem Technology (RAST). Manual analysis of the annotations, repeated sequences and protein-coding genes were used generate circular plasmid DNA sequences. The plasmid sequences were analyzed with antiSMASH and deepBGC as well, to identify potential biosynthetic gene clusters (BGC).

4.4 Results

4.4.1 Plasmid screening

A total of 365 bacterial isolates from animal microbiomes were screened for plasmids. At least one plasmid was identified in 121 isolates. The cat related collection had 94% (47/50) of the isolates harboring at least one plasmid. The bovid-associated microbiome collection plasmid screen resulted in 15.2% (38/250) of isolates, while 55.3% (36/68) of the *E. coli* collection from wild birds harbored plasmids.

Plasmids were purified from each isolate, and twelve plasmids from bovid-associated isolates were deliberately chosen for sequencing. These were prioritized based on bacterial source, bacterial species, and differences between their apparent size when visualized using ethidium bromide-stained agarose gels. We explicitly included plasmids from two prominent pathogens in our study. One from *Staphylococcus aureus* was included, since this pathogen is a priority for Canadian Dairy industry stakeholders[183]. A second from *Klebsiella pneumoniae* was selected, since it was the most prominent infectious agent in the barn during the collection period of spring and summer of 2019. Another main factor when choosing the plasmids to be sequenced was the fact that they appeared smaller than the expected average plasmid size.

4.4.2 Phenotypic screen

Plasmids from 36 *E. coli* isolates recovered from *C. brachyrhynchos* were pooled and used to transform chemically competent *E. coli*. The transformants were plated on LB plates containing different antibiotics, in order to screen for antibiotic resistance. Pools that contained resistant transformants had their plasmids transformed individually to associate phenotype with plasmid.

Plasmid-encoded antibiotic resistance was observed in seven cases. In four instances, plasmids mediated resistance to two antibiotics, ampicillin and tetracyline. Resistance to ampicillin only was observed in one case, while resistance in tetracyline exclusively was detected in two circumstances. In all seven cases, it allowed for growth in sulfamethoxazole (Table 4.1).

A library composed of 192 TRACA transformants was screened for the phenotypes of interest (antibiotic resistance, virulence factors, and small molecule

production). None of the transformants presented these phenotypes, indicating that the plasmids did not harbor biosynthetic gene clusters of interest for this project.

Table 4.1: Antibiotic resistance conferred by plasmids purified from seven *E. coli* isolates from wild birds (*C. brachyrhynchos*).

Source isolate	Ampicillin	Tetracycline	Sulfamethoxazole
106A		+	+
107A	+		+
112A	+	+	+
113A	+	+	+
114A	+	+	+
115A	+	+	+
135A		+	+

4.4.3 Host isolate identification

The taxonomic identity of plasmid-harboring strains from cat feces and bovid-associated microbiomes was determined using 16S rRNA gene sequences. Taxonomy classification was deemed successful when over 98% identity was detected. The 16S rRNA gene identification showed that the vast majority of the identified plasmid harboring isolates from bovid-associated microbiomes are Gram-positive (Table 4.2). Apart from *Klebsiella pneumoniae*, all the known isolates belong to the phylum Firmicutes (97.1% of isolates), with the genus *Bacillus* composing the majority of isolates. This trend was also observed in the cat-associated microbiota, where 15/24 (62.5%) of the identified isolates belonged to the Firmicutes phylum. However, the majority of these isolates belonged to the genus *Staphylococcus*. The remaining 37.5% were part of Proteobacteria and, more specifically, belonged to the class Gammaproteobacteria.

Table 4.2: Taxonomic ID of plasmid-harboring isolates from feline- and bovid- associated microbiomes.

Source	Isolate ID
Cat	*Salmonella enterica* (n=1)
	Escherichia fergusonii (n=1)
	Enterococcus faecalis (n=4)
	Staphylococcus epidermidis (n=10)
	Streptococcus canis (n=1)
	Pseudomonas viridiflora (n=1)
	Escherichia coli (n=6)
Dairy cattle	*Staphylococcos sciuri* (n=1)
	Bacillus licheniformis (n=13)
	Bacillus rhizospherae (n=1)
	Bacillus pumilus (n=6)
	Bacillus pervagus (n=2)
	Bacillus subtilis (n=1)
	Lysinibacillus fusiformis (n=1)
	*Bacillus circulans (*n=1)
	Staphylococcus auricularis (n=2)
	Lysinibacillus pakistanensis (n=1)
	Staphylococcus equorum (n=2)
	Klebsiella pneumoniae (n=1)
	Staphylococcus aureus (n=1)
Beef cattle	*Staphylococcus aureus* (n=2)

4.4.4 Plasmid sequencing, de novo assembly and annotation

Sequencing of 13 plasmids from the bovid microbiome isolates was done using paired-end Illumina technology. The average number of reads/sample was 371,248. After assembling the reads into contigs, the mean coverage was used to determine which contigs were possibly plasmids. Contaminating genomic DNA was present in all samples although as expected, coverage was much lower for these contigs. Contigs with a mean coverage that was at least one order of magnitude higher than that of the majority of sequences in the dataset were selected for bioinformatic annotation using RAST. Manual inspection of the contigs revealed repetitive sequences, which was used define their circular nature. The annotated plasmids are described in Table 4.3.

Table 4.3: Plasmids assembled in this study.

Host	Plasmid	Size (bp)	Average coverage
Bacillus licheniformis (RAM-2)	pRAM-2	27,419	1,063 x
Bacillus pumilus (RAM-4)	pRAM-4	7,143	871 x
Bacillus pumilus (RAM-9)	pRAM-9	4,299	66 x
Bacillus licheniformis (RAM-12)	pRAM-12	5,723	144 x
Staphylococcus auricularis (RAM-15)	pRAM-15	7,497	815 x
Klebsiella pneumoniae (RAM-19)	pRAM-19-1	4,399	244 x
	pRAM-19-2	5,723	151 x
	pRAM-19-3	7,819	127 x
Staphylococcus auricularis (RAM-21)	pRAM-21	8,245	95 x
Staphylococcus aureus (RAM-28)	pRAM-28	8,052	310 x
Bacillus licheniformis (RAM-30)	pRAM-30-1	4,399	313 x
	pRAM-30-2	5,723	140 x
	pRAM-30-3	16,336	634 x

Plasmid pRAM-2 (Figure 4.2A) was analyzed with antiSMASH and deepBGC. Both identified a RiPP biosynthetic gene cluster although the results from the two programs differed. While antiSMASH classified all of the genes as belonging to a single cluster, deepBGC detected only the second gene cluster. Manual analysis of the predicted protein sequences showed that the pRAM-2 plasmid has two putative class II lanthipeptide BGC. The first gene cluster (Figure 4.2B) is composed of (i) a gene that encodes for the 71 amino acid lanthipeptide precursor; (ii) lanM, an enzyme that contains regions responsible for Ser and Thr dehydration in the N-terminal domains, whereas the C-terminal region catalyses the lanthionine bridge formation[26]; (iii) a serine peptidase coding sequence; and (iv) a predicted ABC transporter that expels the mature peptide out of the cell. The precursor peptide shares 100% sequence similarity in its last 54 amino acids to a lanthipeptide produced by *Mammaliicoccus sciuri* that belongs to the LchA2/BrtA2 family (Accession number WP_199194657.1).

The second putative lanthipeptide gene cluster (Figure 4.2C) contains (i) a gene responsible for the 63 amino acid lanthipeptide precursor; (ii) lanM, a modification enzyme; (iii) putative ABC transporter; and (iv) ABC transporter permease CDS, both part of the efflux system to transport the active peptide to the surrounding environment. The lanthipeptide produced shares 98.4% nucleotide similarity with a lanthipeptide from the plantaricin C family, curiously also produced by *M. sciuri* (Accession number WP_107602421.1). When aligned with the plantaricin C lanthibiotic (Accession number WP_064511516.1) produced by *Lactobacillus plantarum*, only 32.8% nucleotide similarity was observed. Manual annotation of the rest of plasmid pRAM-2 resulted in identification of other coding sequences that are not involved in specialized metabolite production (Table 4.4).

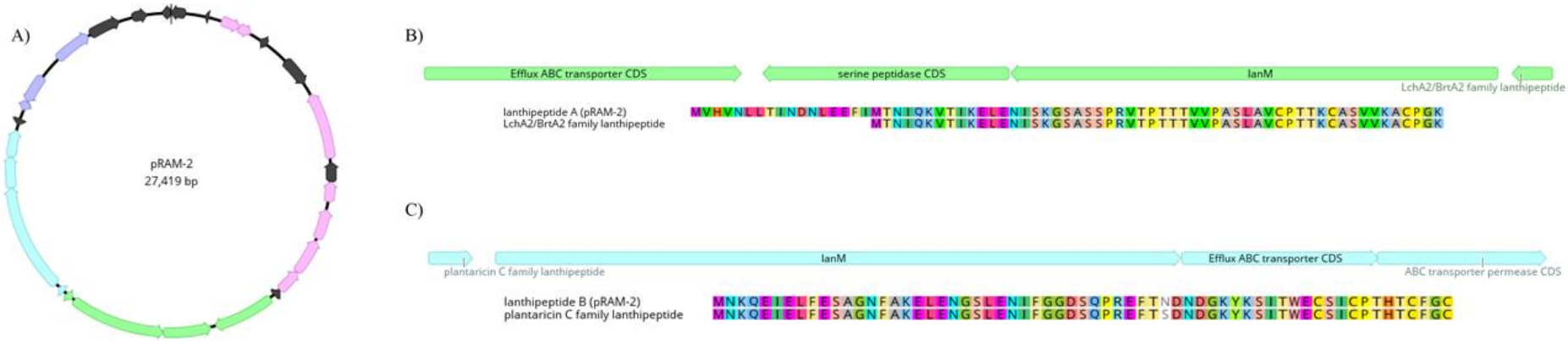

Figure 4.2: Plasmid pRAM-2 (A) harbors two putative lanthipeptides gene clusters (shown in green and blue). Genes responsible for partition and replication are shown in purple. Displayed in pink are other known CDS. Genes with unknown functions are dark grey. The first BGC (B) is responsible for the production of a lanthipeptide that shares 100% nucleotide sequence in the region that overlaps with a previously identified LchA2/BrtA2 family lanthipeptide. Genes responsible for the modification enzyme LanM, serine peptidase and transport protein are also part of this BGC. The second lanthipeptide gene cluster (C) has as a product 98.4% similarity with a plantaricin C family lanthipeptide of M. sciuri, a LanM post-translational modification enzyme and genes responsible for the transport protein.

Table 4.4: Plasmid pRAM-2 coding sequences that are not involved in secondary metabolite production, regulation or transport.

CDS		**Putative function**
From	To	
1405	1920	Transcription factor regulator
1917	2276	Adenylyltransferase
6471	4630	ATP-dependent endonuclease
7650	7093	Resolvase/integrase
8707	7874	Methyltransferase
9774	8779	Histidine kinase
10445	9780	Transcription factor regulator
22575	22324	Plasmid replication associated protein
23367	22576	Partition protein A
24025	25050	Replication initiation protein A

Curiously, two of the parent strains share more than one plasmid in common, despite these hosts being in different phyla: RAM-19 was identified as *K. pneumoniae*, a Proteobacteria; and RAM-30 as *B. licheniformis*, a Firmicute. Nonetheless, pRAM-19-1 (Figure 4.3A) and pRAM-30-1 have 100% consensus identity (Figure 4.3B). However, no biosynthetic gene cluster was identified in pRAM-19-1. Apart from the genes involved in mobilization and replication, gene functions were unknown.

Plasmids pRAM-12, pRAM-19-2 and pRAM-30-2 were also found to be identical to each other (Figure 4.3C and D). These plasmids all harbor a gene cluster responsible for the production, immunity and export of cloacin (Figure 4.3E), identified manually. The cloacin peptide encoded by these plasmids shares 100% amino acid identity with the peptide produced by several species of Enterobacteriaceae (Accession number WP_101972327.1) (Figure 4.3F). This gene cluster was first described in 1969[184], observed in *Enterobacter cloacae* harboring the plasmid CloDF13 and its composed by (i) cloacin peptide; (ii) immunity protein[185]; and (iii) cloacin release. Interestingly, this BGC was identified in diverse Proteobacteria, and similar gene clusters (with over 92% identity) were detected in Firmicutes, when compared to the non redundant protein database.

Plasmids pRAM-4, pRAM-9, pRAM-15, pRAM-19-3 and pRAM-30-3 (Figure 4.4) had few genes with known functions. Unknown genes represented from 42.9 (pRAM-19-1) to 100% (pRAM-9) of the genes on the plasmids. Identified coding sequences and their predicted functions, as well as the percentage of unknown genes, are shown in Table 4.5.

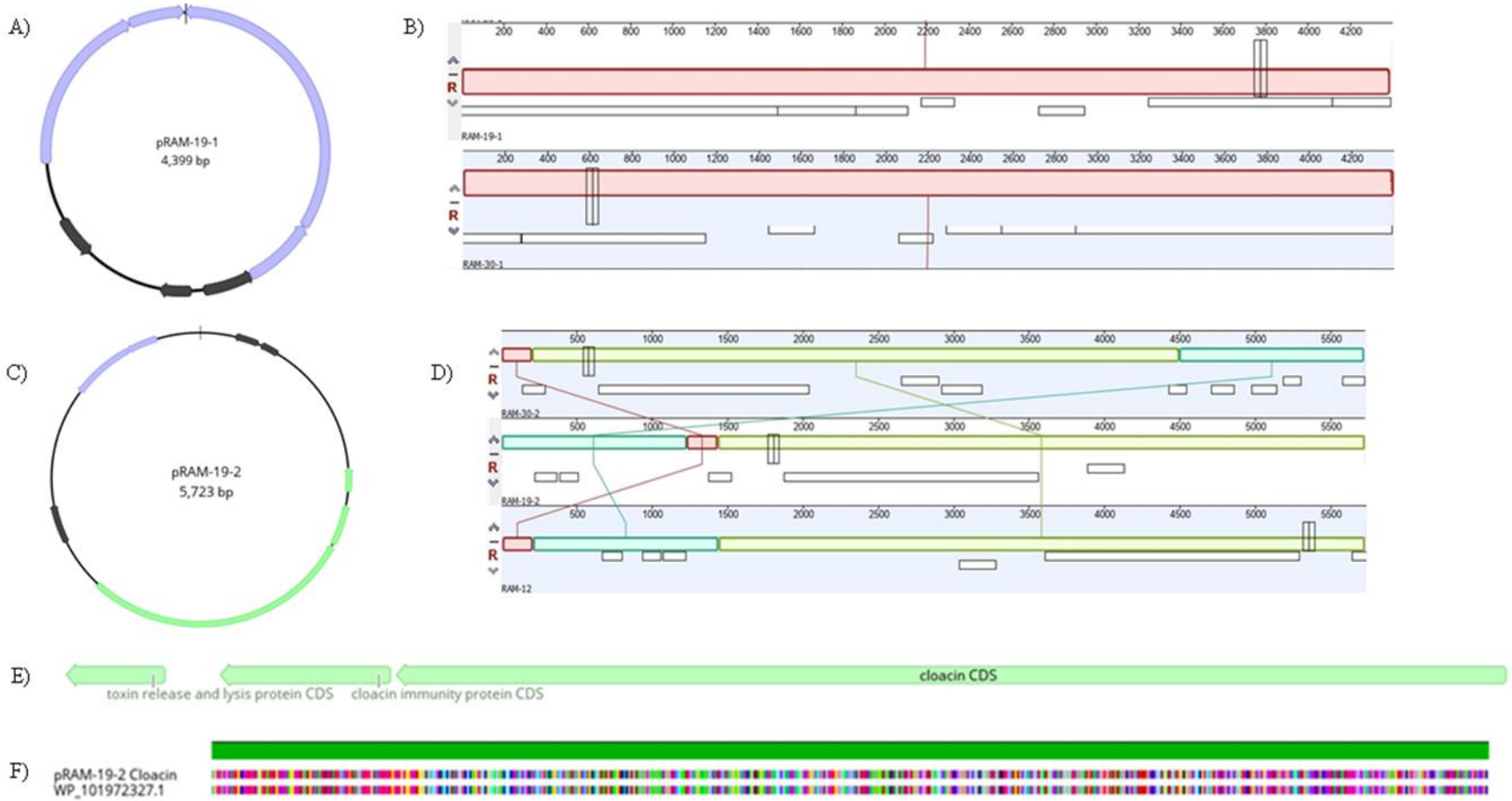

Figure 4.3: Plasmid pRAM-19-1 (A) shares 100% nucleotide sequence with the plasmid pRAM-30-1(B), despite hosts belonging to different phyla. Plasmid pRAM-19-2 (C) shares 100% nucleotide identity to plasmids pRAM-12 and pRAM-30-2 (D). These plasmids harbor a biosynthetic gene for the production, immunity and export of cloacin (E). The cloacin produced has 100% identity to the cloacin produced by members of the Enterobacteriaceae (F). Genes in purple are responsible for mobilization and replication, and genes in grey are unknown.

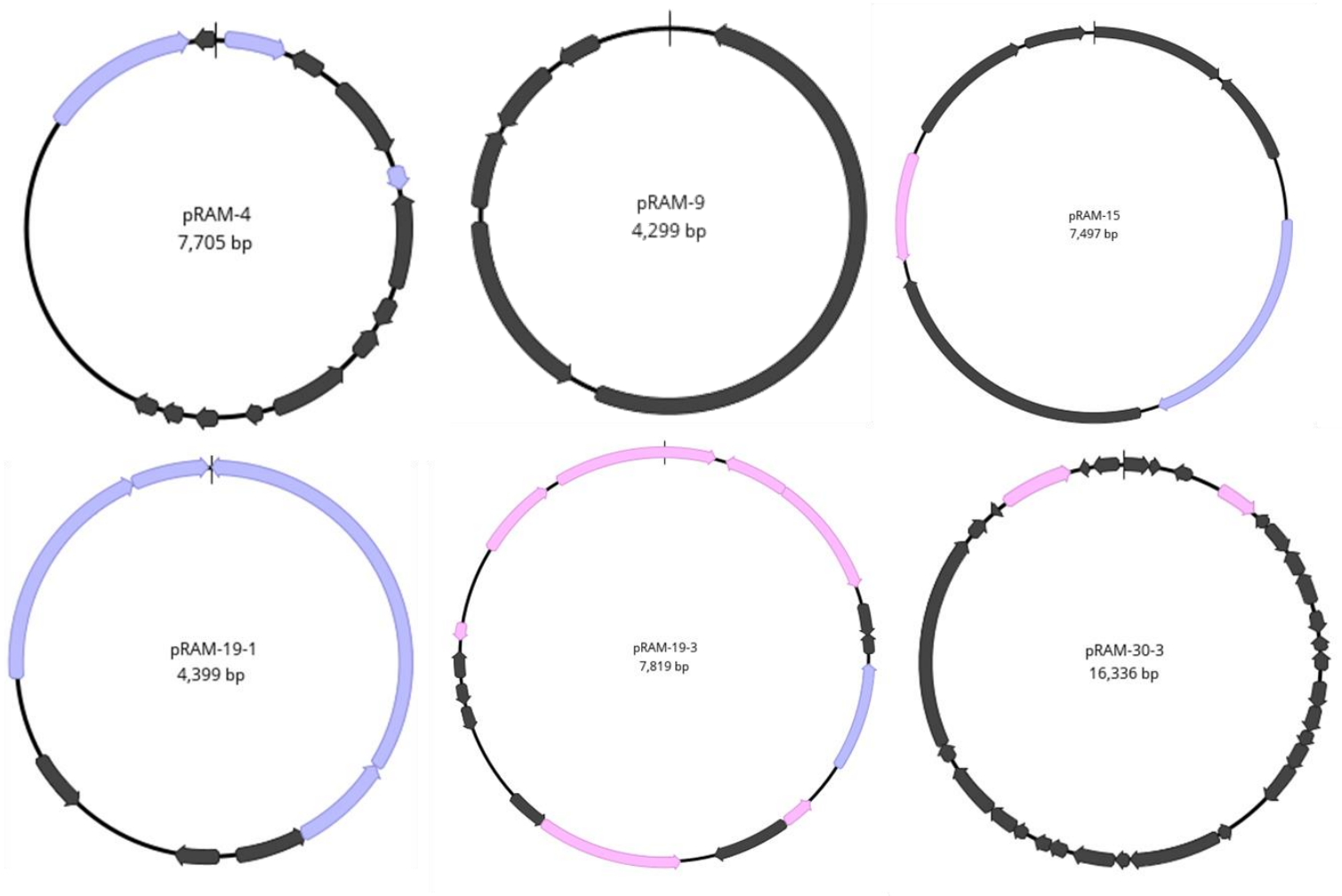

Figure 4.4: Plasmids reassembled in this study that did not harbor known BGCs. Genes in purple are responsible for mobilization and replication. Pink colored genes are genes that were identified but are not involved in known gene clusters. Genes in grey are unknown.

Table 4.5: Assembled plasmids that did not harbor a known BGC, and their known genes.

Plasmid	CDS		Putative function	% of genes with products of unknown identity
	From	To		
pRAM-4	67	468	Plasmid replication	68.8
	1516	1665	Plasmid replication	
	4910	6028	Aspartate phosphatase	
	6018	6134	Rap protein	
	6509	7534	Plasmid replication	
pRAM-9	-	-	-	100
pRAM-15	1847	3337	Replication initiation protein A	71.4
	6073	5399	IS6 family transposase	
pRAM-19-1	1497	1	Relaxase	42.9
	1862	1494	Mobilization protein C	
	3246	4112	Replication initiation protein	
	4116	4391	Plasmid replication	
pRAM-19-3	777	385	IS3 family transposase	46.7
	768	1520	IS3 family transposase	
	2662	1997	Replication initiation protein	
	3121	2927	IS1 family transposase	
	4700	3813	Integrase	
	6078	5968	Small membrane protein	
	6556	7059	Transposase	
	7146	319	Acyltransferase	
pRAM-30-3	1318	1863	Integrase	87.5
	3302	2886	Lipase	
	8935	9162	Transcription regulator	
	14688	15653	Integrase	

The plasmid isolated from *S. auricularis* (pRAM-21) was identical to the plasmid isolated from *S. aureus*, pRAM-28 (Figure 4.5A). Both of these plasmids share 89% nucleotide sequence identity with plasmid pRJ6, a plasmid from *S. aureus* that is notorious for harboring the only known four-member bacteriocin gene cluster[186]. Production of these peptides, collectively referred to as aureocin A70, is the hallmark feature of pRJ6 and now pRAM-21 and pRAM-28. However, this BGC was only detected through manual annotation.

The A70 BGC (Figure 4.5B) is composed of (i) *aurR*, a regulator[187]; (ii) *aurI*, that encodes for a protein that gives the host strain immunity[188]; (iii) the *aurABCD* operon, composed of four genes that code for individual peptides that form the bacteriocin; and (iv) *aurT*, an ABC transporter responsible for the efflux of the bacteriocin outside of the cell[186]. The nucleotide sequence similarity shared among the presumed operon *aurABCD* from pRJ6 and the operon from the plasmids isolated in this study is 99.8%. A non-synonymous mutation within the *aurD* coding sequence is the only difference among them (Figure 4.5C). The mobilization genes found on both of the plasmids were substantially more divergent, sharing 78.9% nucleotide sequence identity with their orthologous sequences on pRJ6.

A bioinformatic search for aureocin A70 outside of South America was performed using the Basic Local Alignment Search Tool (BLAST)[189]. Sequences similar to the aureocin A70 gene cluster were identified, including eight matches between 99 and 99.94% identity at the nucleotide level. Three were isolated in Brazil, two from bovid-microbiome (Accession numbers AF241888.2 and MK796167.1), and one from a human case of meconium aspiration syndrome (Accession number CP021143.1). One isolate from a human in Germany was also a match, but no further information was available (Accession number CP047834.1). The other four were all isolated in the United States, three in a Methicillin-Resistant *Staphylococcus aureus* (MRSA) outbreak in a New York city/state hospital (Accession numbers CP030522.1; CP030402.1; CP030460.1), and one is present in the Food and Drug Administration's Center for Disease Control Antimicrobial Resistant Isolate Bank (Accession number CP029651.1).

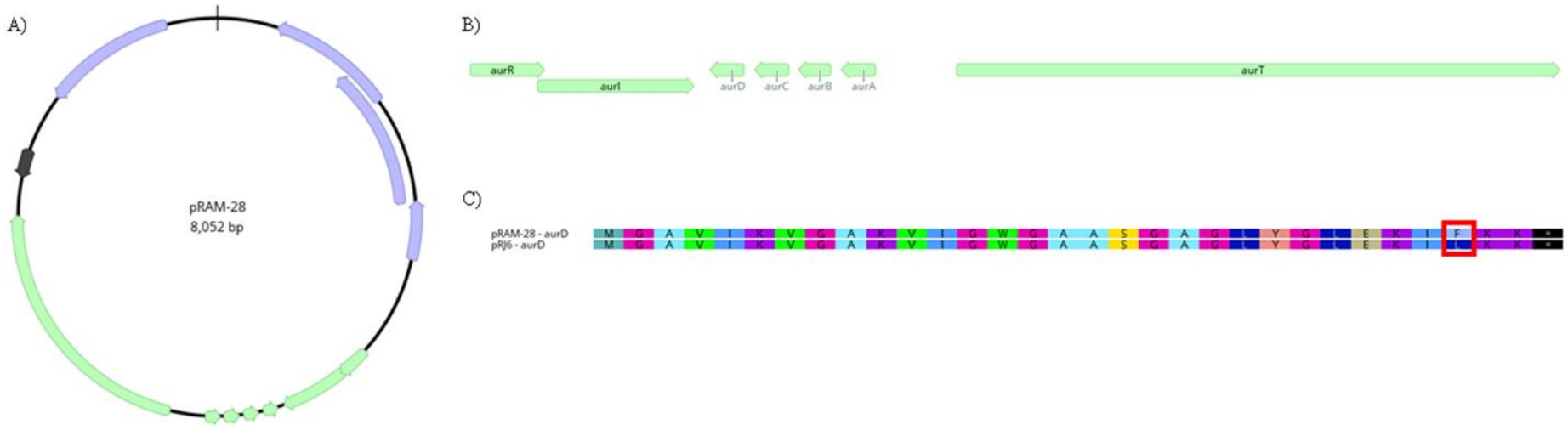

Figure 4.5: Plasmid pRAM-21 and pRAM-28 (A) share 100% nucleotide identity. These plasmids carry a BGC that is responsible for the regulation, production, immunity and transport of aureocin A70 (B), previously detected on plasmid pRJ6. The only difference between the gene cluster detected in pRAM-28 and pRJ-6 is a non-synonymous mutation L29F on the aurD gene (C). Purple genes are responsible for mobilization and replication.

Table 4.6: Coding sequences with known function from plasmid pRAM-28 that are not involved in secondary metabolite production.

CDS		**Putative function**
From	To	
1216	401	Mobilization protein B
1925	918	Relaxase
2290	1907	Mobilization protein C
7727	6870	Replication initiation protein B

4.5 Discussion

The acquisition or loss of a plasmid can drastically alter a bacterium's phenotype. However, plasmids are rarely the target of microbiome research. In this study, we focused on the plasmidome of microbiomes that often overlap and can influence our own microbiome. By using a combination of phenotypic and genomic methods, phenotypes of interest for both human and animal health, in the form of antibiotic resistant genes, were detected. Secondary metabolites gene clusters were identified, and evidence of plasmid sharing among phyla was observed.

Plasmids isolated from *E. coli* (n=36) recovered from wild crows (*C. brachyrhynchos*) that were brought to the Veterinary Medical Center at the University of Saskatchewan were screened for phenotypes of interest. Although no siderophore or antimicrobial activities were observed, resistance to three classes of antibiotics: β-lactams, tetracyclines and sulfonamides, was encoded by 19.4% (n=7) of the plasmids (Table 4.1). All seven of these plasmids harbored antibiotic resistance genes for sulfonamides. Although chromosomal variants of the *folP* gene result in sulfa resistance, the majority of clinical resistance is plasmid-borne[190]. Currently, four genes (*sul1-4*) are known to result in sulfonamide resistance, although only three (*sul1-3*) of them have been identified in clinical settings. Over eighty-five years of use of sulfa drugs, as well as the large presence on plasmids reported results in no surprise that all the plasmids that showed antibiotic resistance phenotype carried one of the *sul* genes.

Five and six of the plasmids from wild bird *E. coli* also resulted in ampicillin and tetracycline resistance of host bacteria, respectively, with four plasmids conferring resistance to all three antibiotics. Plasmid-mediated resistance to ampicillin arises from the production of β-lactamases, enzymes that hydrolyze the β-lactam ring, rendering the antibiotic inactive[164,191,192]. Over 2000 unique β-lactamases have been reported in different settings[191,193]. Both the fact that β-lactamases coding sequences are largely encoded in plasmids and mobile elements across a myriad of Enterobacteriaceae, as well as their high adaptability to expand their activity spectrum as new modified antibiotics are introduced, are decisive for the success of the spread of these genes and a threat to the effectiveness of β-lactams[194,195]. Tetracycline resistance was first reported in 1953[196,197] and became largely associated with the presence of plasmids and other mobile genetic elements[196,198,199]. Three different methods of tetracycline resistance have been identified: efflux of tetracycline by energy dependent membrane associated proteins, ribosome

protection proteins and enzymatic alteration of the drug. Tetracycline efflux from the cell is the most common mode of resistance acquired by bacteria[196,199–202], with 26 genes responsible for transmembrane efflux pumps identified. Of those, at least 14 are plasmid-encoded[202]. Future sequencing work can be done to determine which genes are at play in the resistance phenotype observed in wild bird *E. coli*, as well as to contribute with the knowledge of which genes are circulating in the environment.

Four of the plasmids purified and screened in this section of the work encoded for resistance to three antibiotics (Table 4.1). Multidrug resistance plasmids have been identified in Enterobacteriaceae isolated from avian sources in different settings. When studying *E. coli* isolated from wild and domestic birds in Bangladesh, Hasam et al. (2012) observed that resistance to tetracycline, trimethoprim/sulfamethoxazole and ampicillin had the highest incidence among isolates[203]. Both Ahmed et al. (2013)[204] and Enany et al. (2019)[205] analysed *E. coli* isolates from septicemic broilers and environmental and avian sources, from Egyptian farms, respectively. All (100%) of the tested isolates in both studies were resistant to ampicillin and tetracycline, among other antibiotics tested. The first study observed resistance in all isolated to the combo trimethoprim/sulfamethoxazole, while the second study reported resistance to sulfamethoxazole in 80.92% of the isolates. Similar trends of resistance were also observed in avian pathogenic *E. coli* responsible for colibacillosis in poultry farms in Pakistan[206,207] and in an outbreak of multidrug resistant pathogenic *E. coli* in canaries, in Brazil[208].

Although the screening process did not result in sequencing to associate genes with the resistance phenotype observed, this approach allows us to maintain the connection of host and plasmid, which does not occur in metagenomic approaches. This link is of relevance because it provides information relevant in two aspects: first, knowing what genes are circulating in determined non-pathogenic, commensal microbiota allows us to infer what pool of traits is available to be picked up by pathogens interacting with that community. Secondly, when determined to have antibiotic resistance genes in pathogenic bacteria, it provides us with a much more urgent scenario that required careful considerations. The same can be said if a virulent phenotype were detected in a previously non-pathogenic bacterium.

The high incidence of plasmid-mediated antibiotic resistance in *E. coli* isolates obtained from birds, both observed in this study (19.4%) and in the literature suggests that the use of antibiotics is selecting the phenotype of resistance further from human and

veterinary clinical use. It also indicates that wild birds could be used for surveillance of spread of resistance, as suggested by Parker et al. (2016)[209].

Since the phenotypic screen conducted in the first part of this study demonstrated the utility of the culture-based approach used for the identification plasmids that can be screened systematically for interesting phenotypes, plasmids were purified from feline and bovid microbiomes and we investigated the distribution of these plasmids across phyla, as well as the presence of biosynthetic gene clusters.

Domestic cat fecal isolates carrying plasmids belonged to the phyla Firmicutes (62.5%) and Proteobacteria (37.5%). This is in accordance with previous descriptions of feline fecal microbiomes, although it is important to highlight that only plasmid-harboring isolates were taxonomically identified in this study. Other studies of the feline fecal microbiota identified Firmicutes in a range of 13 to 92% of the microbiota, followed by Proteobacteria (6-14%) and Actinobacteria (7%)[210–213]. The differences among the results of these studies is not only influenced by method and feline individual, but also by age[214], diet[215], obesity and whether or not the animal is neutered[216].

Regarding the plasmid-harboring isolates from bovid-associated microbiomes, the most abundant phyla is Firmicutes (97.1%). Herbivore gut microbiota has, reportedly, an abundance of Firmicutes bacteria[217–219], which can also be observed in the rumen microbiota[219]. The high incidence of Firmicutes plasmids purified from Bacilli was also observed by Shintani et al (2015), when analyzing plasmid sequences available on NCBI database[220].

Twelve plasmids from bovid-associated isolates were selected for sequencing, and the plasmids reassembled are shown in Figures 4.2-5. Software designed for the detection of BGCs (deepBGC and antiSMASH) was only successful in identifying gene clusters in the case of pRAM-2, and manual annotation was necessary to resolve differences in their results. Manual annotation was used to assign BGCs in the other plasmids. This resulted in the identification of a cloacin gene cluster in three plasmids from different hosts, as well as the detection of aureocin A70 gene cluster in two other plasmids. The cloacin gene cluster appears to be widespread in bacteria.

Plasmids pRAM-21 and pRAM-28 harbor a BGC responsible for the production of aureocin A70. This bacteriocin system was initially described in *Staphylococcus* isolates from milk in Brazil[186]. Further studies by the same group suggested a broader geographical range of this biosynthetic gene cluster in South America, more specifically in cows suffering from subclinical mastitis from Argentinian herds[221]. Bacteriocins

identical or similar to aureocin A70 were also present in 34 strains involved in bovine mastitis in southeast Brazil[222]. When screening staphylococci isolates from milk of healthy cows in Brazil for bacteriocin production, Brito et al. (2011) determined that 58 of 111 isolates provided PCR products when primers for the aureocin A70 gene cluster were used[223]. This suggests that the presence of bacteriocin is not exclusive to clinical cases of mastitis. This widespread presence in healthy animals may suggest that aureocin A70 is not directly involved in pathogenesis. It is interesting to note that the *aurABCD* operon of pRAM-28 is the most divergent yet differs by a single nucleotide mutation, leading to an L29F substitution in AurD. In fact, no single gene varies by more than a handful of substitutions. The remarkable conservation is evidence for dissemination via HGT, and the phylogeny of hosts suggests that the transfer is most likely mediated by conjugation between *Staphylococcus*.

By focusing on the plasmidome, we were able to observe evidence of HGT in the dairy barn from which samples were collected. Two identical set of plasmids (pRAM-19-1 and pRAM-19-2, as well as pRAM-30-1 and pRAM-30-2) were recovered from hosts from distinct bacterial phyla. In one case, a Gram-negative pathogen, *K. pneumoniae*, that was determined to be the causing agent of mastitis in the animal it was recovered from. It harbored two plasmids that were identical to plasmids found in *B. licheniformis* (a common Gram-positive member of the rumen microbiome[224]) isolated from a healthy host intramammary swab sample.

Horizontal gene transfer has been reported in a myriad of environments, such as in the gut microbiome[163,225,226], plant surface[227], food waste composting[228], cheese rind[229], and hospital inhalable particulate matter[230]. However, this is the first time, to our knowledge, where evidence of plasmid transfer has been identified among two different cattle hosts sharing a barn. While gene exchange within phyla is widely reported[231–235], HGT across phyla is reportedly more rare. It has, nonetheless, been reported between Spirochaetes and Firmicutes[236], Proteobacteria and Firmicutes, and Proteobacteria and Actinobacteria[237]. Recently, it has been observed between *Mycobacterium* (Tenericutes) and Actinobacteria[238], as well as between *Mucispirillum schaedleri* (Deferribacteres) and Proteobacteria[239].

A bioinformatic analysis done by Caro-Quintero & Konstantinidis (2014) evaluated quantitatively inter-phylum HGT, taking in consideration the environment and ecological conditions[237]. Their pipeline estimated the HGT across taxa while minimizing the effect of taxonomic representation. Not surprisingly, this study pointed out that shared

ecology, oxygen tolerance and other physiological parameters influence the frequency of gene exchange among bacteria. Interestingly, the removal of the phylum Firmicutes in a sample set reduced the horizontal gene transfer by 97%, indicating that this phylum is an important and promiscuous part in HGT. Furthermore, their results suggest that lateral transfer between distantly related organisms can be favored when both organisms have overlapping ecology. This study agrees with these findings, since Firmicutes were part of all the sets of identical plasmids detected and the bacterial hosts were isolated from similar ecological niches.

Evans et al. (2020) suggest that highly identical plasmids are evidence of recent transfer, since the mutation rates of plasmids are similar to the chromosome and may not have the necessary time to diverge and adapt to its bacterial hosts[231]. The fact that the cloacin peptide encoded by pRAM-19-2, pRAM-12 and pRAM-30-2 was 100% identical to the peptide produced by diverse members of Enterobacteriaceae, while being only 93.4 to 95.5% identical to the cloacin peptide previously detected in the Firmicutes phylum could suggest the direction of genetic exchange in this case. While we cannot determine with certainty the direction of the genetic exchange, this result highlights the fact that pathogens are interacting with the host's natural flora: RAM-19 was identified as the causative agent of mastitis in the animal it was isolated from, while RAM-12 and RAM-30 are common members of the cattle microbiome.

4.6 Conclusions

By focusing on a specific genetic context – plasmids – we were able to detect four different BGCs encoding small molecules, including the only known four-peptide leaderless bacteriocin. This approach also showed great potential for identifying genes responsible for antibiotic resistance and identifying uncharacterized genes for future study. More so, it allows the link between host and plasmid to be used in decision making after gene identification. Overall, this work suggests that the plasmidome is an important source of potential small molecules that are used by bacteria to compete with each other within and between microbiomes. By concentrating on plasmids purified from cultured host bacteria, we were also able to provide evidence for genetic exchange via plasmids, illustrating the intimacy of interactions within animal-associated microbial communities.

5 General Discussion

5.1 Summary and limitations of this work

This work aimed to shine a light on a frequently neglected genetic space: the plasmidome. Prioritization of the plasmidome can allow for the discovery of new biosynthetic gene clusters, as shown in this work. Using computational tools, we were able to show that plasmids have great potential as sources of secondary metabolites BGCs. In a practical approach, plasmids were isolated from animal microbiomes. Sequencing of the plasmids allowed the identification of four bioactive gene clusters in six plasmids.

Chapter 3 intended to analyze publicly available plasmid sequences to determine the type of product, taxonomic distribution and prevalence of biosynthetic gene cluster. Using the software deepBGC and antiSMASH, we determined that the different classes of BGCs are present in most phyla. Since the two programs use distinct approaches for detection and identification, the number of BGCs as well as their product classes varied. Despite the discrepancy in the results, we were able to show that plasmids are a productive source for secondary metabolites with the potential to be used in medicine and veterinary sciences. However, the computational work done in Chapter 3 has its limitations, regardless of the individual restraints of each software. The biggest of which is the lack of a "control group". Ideally, a manual investigation would allow us to infer which software was more accurate in the predictions. However, due to the sample size (n=101 415), that was not practical. The sample size also restrained sequence dereplication.

Chapter 4 aimed to use plasmid screening techniques to characterize plasmids present in animal-related microbiomes. Plasmids isolated from different sources were screened for different purposes. One of the limitations of the study of plasmid-mediated antibiotic resistance developed in this work is that the genes responsible remain unknown, since no sequencing or gene identification via PCR was done. As for the cat microbiota and bovid-related microbiome members, only bacteria that harbored plasmids were identified, which does not represent the whole bacteria community found in these spaces. Furthermore, only culturable bacteria were screened. This represents only a fraction of the microbial species present on Earth[240]. Plasmids were isolated from 38 bovid microbiome hosts. However, the method used to purify these plasmids is biased towards high molecular weight plasmids. Only a fraction of the plasmids was sequenced (12/38),

which is also a limitation of this study. Regarding plasmid annotation, manual intervention resulted in gene clusters that were not detected by either software used. This shows not only a constraint of this work, but also that, even thought we have consolidated bioinformatic tools available, there's still room for development and growth of the computational tools accessible to researchers.

The focus on plasmids permitted the detection of four BGCs encoding small molecules in six plasmids. The plasmidome approach also displayed potential in the identification of uncharacterized genes and gene clusters for future studies, as well as identifying plasmids connected to antibiotic resistance. Evidence of genetic exchange via plasmids, showing the close interactions of animal-associated microbiomes, was also provided. This work suggests the plasmidome as not only a source of small molecules, but as means for the deeper understanding of microbial interactions, antibiotic resistance and virulence mechanisms.

5.2 Future prospects

The need for new molecules is not new and is not going to go away anytime soon. As novel molecules get discovered, *in vitro* tests often quickly show that resistance is just around the corner. While that is not restricted to antibiotics, as this pattern has been seen in antiparasitic, antifungal and also in insecticide products, bacteria seem to be always a step ahead of us. Recent technological advances, such as NGS, and deeper understanding of microbial communities and the role played by their secondary metabolites, allowed the development of computational approaches to mine genomes for biosynthetic gene clusters with potential new bioactive products.

As new bioinformatic tools and software are developed, more accurate BGCs predictions will be possible. However, the need to attempt to keep the bioinformatic approaches up-to-date with the recent literature and novel BGC classes discovered will always be a challenge. Said challenge is also extended to the ability of expression of cryptic and silent BGCs *in vitro*. As this work has shown the potential of plasmids for molecule discovery, it is expected that more research prioritizing this previously neglected space will surface. With that, the knowledge of how plasmids can mediate host colonization and microbial interactions will surely expand.

6 References

1. Doolittle, W. F. & Booth, A. It's the song, not the singer: an exploration of holobiosis and evolutionary theory. *Biol. Philos.* **32**, 5–24 (2017).
2. Turnbaugh, P. J. *et al.* The human microbiome project: exploring the microbial part of ourselves in a changing world. *Nature* **449**, 804–810 (2007).
3. Beiko, R. G., Harlow, T. J. & Ragan, M. A. Highways of gene sharing in prokaryotes. *Proc. Natl. Acad. Sci.* **102**, 14332–14337 (2005).
4. Hall, J. P. J., Brockhurst, M. A. & Harrison, E. Sampling the mobile gene pool: Innovation via horizontal gene transfer in bacteria. *Philosophical Transactions of the Royal Society B: Biological Sciences* vol. 372 (2017).
5. Emamalipour, M. *et al.* Horizontal Gene Transfer: From Evolutionary Flexibility to Disease Progression. *Frontiers in Cell and Developmental Biology* vol. 8 (2020).
6. Jain, R., Rivera, M. C., Moore, J. E. & Lake, J. A. Horizontal gene transfer accelerates genome innovation and evolution. *Mol. Biol. Evol.* **20**, 1598–1602 (2003).
7. Ochman, H. & Moran, N. A. Genes Lost and Genes Found: Evolution of Bacterial Pathogenesis. **1096**, 1096–1099 (2001).
8. Culligan, E. P., Sleator, R. D., Marchesi, J. R. & Hill, C. Metagenomics and novel gene discovery: Promise and potential for novel therapeutics. *Virulence* vol. 5 399 (2014).
9. Bengtsson-Palme, J. The diversity of uncharacterized antibiotic resistance genes can be predicted from known gene variants-but not always. *Microbiome* **6**, 1–12 (2018).
10. Smillie, C., Garcillan-Barcia, M. P., Francia, M. V., Rocha, E. P. C. & de la Cruz, F. Mobility of Plasmids. *Microbiol. Mol. Biol. Rev.* **74**, 434–452 (2010).
11. Dubey, G. P. & Ben-Yehuda, S. Intercellular nanotubes mediate bacterial communication. *Cell* **144**, 590–600 (2011).
12. Fulsundar, S. *et al.* Gene transfer potential of outer membrane vesicles of Acinetobacter baylyi and effects of stress on vesiculation. *Appl. Environ. Microbiol.* **80**, 3469–3483 (2014).
13. Crosa, J. H., Schiewe, M. H. & Falkow, S. Evidence for Plasmid Contribution to the Virulence of the Fish Pathogen Vibrio anguillarum. **18**, 509–513 (1977).
14. Actis, L. A. *et al.* Characterization of Anguibactin , a Novel Siderophore from Vibrio anguillarum 775 (pJM1). **167**, 57–65 (1986).
15. Bruto, M. *et al.* Vibrio crassostreae, a benign oyster colonizer turned into a pathogen after plasmid acquisition. *ISME J.* **11**, 1043–1052 (2017).
16. Lee, C. *et al.* The opportunistic marine pathogen Vibrio parahaemolyticus becomes virulent by acquiring a plasmid that expresses a deadly toxin. **112**, 1–6 (2015).
17. Baquero, F., Bouanchaud, D. & Fernandez, C. Microcin Plasmids : a Group of

Extrachromosomal Elements Coding for Low-Molecular-Weight Antibiotics in Escherichia coli. **135**, 342–347 (1978).

18. Stinear, T. P. *et al.* Giant plasmid-encoded polyketide synthases produce the macrolide toxin of Mycobacterium ulcerans. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 1345–1349 (2004).

19. Hisatsune, J. *et al.* Emergence of Staphylococcus aureus Carrying Multiple Drug Resistance Genes on a Plasmid Encoding Exfoliative Toxin B. **57**, 6131–6140 (2013).

20. Sit, C. S. *et al.* Variable genetic architectures produce virtually identical molecules in bacterial symbionts of fungus-growing ants. **112**, (2015).

21. Van Arnam, E. B., Ruzzini, A. C., Sit, C. S., Currie, C. R. & Clardy, J. A Rebeccamycin Analog Provides Plasmid-Encoded Niche Defense. *J. Am. Chem. Soc.* **137**, 14272–14274 (2015).

22. Ruzzini, A. C. & Clardy, J. Minireview Gene Flow and Molecular Innovation in Bacteria Lessons from Chemical Ecology. *Curr. Biol.* **26**, R859–R864 (2016).

23. Arnam, E. B. Van *et al.* Selvamicin , an atypical antifungal polyene from two alternative genomic contexts. **113**, 12940–12945 (2016).

24. Bennett, P. M. Plasmid encoded antibiotic resistance: Acquisition and transfer of antibiotic resistance genes in bacteria. in *British Journal of Pharmacology* vol. 153 S347 (Wiley-Blackwell, 2008).

25. Arnison, P. G. *et al.* NÃO. *Natural Product Reports* vol. 30 108–160 (2013).

26. Hudson, G. A. & Mitchell, D. A. RiPP antibiotics: biosynthesis and engineering potential. *Current Opinion in Microbiology* vol. 45 61–69 (2018).

27. Travin, D. Y., Bikmetov, D. & Severinov, K. Translation-Targeting RiPPs and Where to Find Them. *Frontiers in Genetics* vol. 11 (2020).

28. Zhong, Z., He, B., Li, J. & Li, Y. X. Challenges and advances in genome mining of ribosomally synthesized and post-translationally modified peptides (RiPPs). *Synthetic and Systems Biotechnology* vol. 5 155–172 (2020).

29. Heddle, J. G. *et al.* The antibiotic microcin B17 is a DNA gyrase poison: Characterisation of the mode of inhibition. *J. Mol. Biol.* **307**, 1223–1234 (2001).

30. Delgado, M. A., Rintoul, M. R., Farías, R. N. & Salomón, R. A. Escherichia coli RNA polymerase is the target of the cyclopeptide antibiotic microcin J25. *J. Bacteriol.* **183**, 4543–4550 (2001).

31. McAuliffe, O., Ross, R. P. & Hill, C. Lantibiotics: structure, biosynthesis and mode of action. *FEMS Microbiol. Rev.* **25**, 285–308 (2001).

32. Bierbaum, G. & Sahl, H.-G. Lantibiotics: Mode of Action, Biosynthesis and Bioengineering. *Curr. Pharm. Biotechnol.* **10**, 2–18 (2009).

33. Gomez-Escribano, J. P., Song, L., Bibb, M. J. & Challis, G. L. Posttranslational β-methylation and macrolactamidination in the biosynthesis of the bottromycin complex of ribosomal peptide antibiotics. *Chem. Sci.* **3**, 3522–3525 (2012).

34. Huo, L., Rachid, S., Stadler, M., Wenzel, S. C. & Müller, R. Synthetic biotechnology to study and engineer ribosomal bottromycin biosynthesis. *Chem. Biol.* **19**, 1278–1287 (2012).

35. Hou, Y. *et al.* Structure and biosynthesis of the antibiotic bottromycin D. *Org. Lett.* **14**, 5050–5053 (2012).

36. Montalbán-López, M. *et al.* New developments in RiPP discovery, enzymology and engineering. *Natural Product Reports* vol. 38 130–239 (2021).

37. N, S. *et al.* Analysis of genes involved in the biosynthesis of lantibiotic epidermin. *Eur. J. Biochem.* **204**, 57–68 (1992).

38. MA, N., HG, S. & JR, T. Identification of genes encoding two-component lantibiotic production in Staphylococcus aureus C55 and other phage group II S. aureus strains and demonstration of an association with the exfoliative toxin B gene. *Infect. Immun.* **67**, 4268–4271 (1999).

39. Hyink, O. *et al.* Salivaricin A2 and the novel lantibiotic salivaricin B are encoded at adjacent loci on a 190-kilobase transmissible megaplasmid in the oral probiotic strain Streptococcus salivanus K12. *Appl. Environ. Microbiol.* **73**, 1107–1113 (2007).

40. C, K. *et al.* Pep5, a new lantibiotic: structural gene isolation and prepeptide sequence. *Arch. Microbiol.* **152**, 16–19 (1989).

41. Yorgey, P., Davagnino, J. & Kolter, R. The maturation pathway of microcin B17, a peptide inhibitor of DNA gyrase. *Mol. Microbiol.* **9**, 897–905 (1993).

42. Severinov, K., Semenova, E. & Kazakov, T. Class I Microcins: Their Structures, Activities, and Mechanisms of Resistance. in *Prokaryotic Antimicrobial Peptides* 289–308 (Springer New York, 2011). doi:10.1007/978-1-4419-7692-5_15.

43. Collin, F. & Maxwell, A. The Microbial Toxin Microcin B17: Prospects for the Development of New Antibacterial Agents. *Journal of Molecular Biology* vol. 431 3400–3426 (2019).

44. Donia, M. S. *et al.* A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* **158**, 1402 (2014).

45. Bennallack, P. R., Burt, S. R., Heder, M. J., Robison, R. A. & Griffitts, J. S. Characterization of a novel plasmid-borne thiopeptide gene cluster in Staphylococcus epidermidis strain 115. *J. Bacteriol.* **196**, 4344–4350 (2014).

46. MA, N., L, D.-G., JL, S. M. & F, M. Cloning and mapping of the genetic determinants for microcin C7 production and immunity. *J. Bacteriol.* **168**, 1384–1391 (1986).

47. NE, K., EI, B., AZ, M., DA, Z. & IA, K. Cloning and mapping of the genetic determinants for microcin C51 production and immunity. *Mol. Gen. Genet.* **241**, 700–706 (1993).

48. JL, M. & JC, P.-D. Isolation, characterization, and mode of action on Escherichia coli strains of microcin D93. *Antimicrob. Agents Chemother.* **29**, 456–460 (1986).

49. Eberhart, L. J. *et al.* Characterization of a novel microcin that kills enterohemorrhagic escherichia coli O157:H7 and O26. *Appl. Environ. Microbiol.*

78, 6592–6599 (2012).

50. Bellomio, A., Vincent, P. A., De Arcuri, B. F., Farias, R. N. & Morero, R. D. Microcin J25 has dual and independent mechanisms of action in Escherichia coli: RNA polymerase inhibition and increased superoxide production. *J. Bacteriol.* **189**, 4180–4186 (2007).

51. Vincent, P. & Morero, R. The Structure and Biological Aspects of Peptide Antibiotic Microcin J25. *Curr. Med. Chem.* **16**, 538–549 (2009).

52. Yan, K. P. *et al.* Dissecting the Maturation Steps of the Lasso Peptide Microcin J25 in vitro. *ChemBioChem* **13**, 1046–1052 (2012).

53. Cheung-Lee, W. L., Parry, M. E., Cartagena, A. J., Darst, S. A. & Link, A. J. Discovery and structure of the antimicrobial lasso peptide citrocin. *J. Biol. Chem.* **294**, 6822–6830 (2019).

54. I, B. F., Z, F., F, K., S, T. & S, J. Isolation of the Bacillus thuringiensis plasmid carrying Bacthuricin F4 coding genes and evidence of its conjugative transfer. *J. Infect. Dev. Ctries.* **8**, 727–732 (2014).

55. Sit, C. S., Van Belkum, M. J., McKay, R. T., Worobo, R. W. & Vederas, J. C. The 3D solution structure of thurincin H, a bacteriocin with four sulfur to α-carbon crosslinks. *Angew. Chemie - Int. Ed.* **50**, 8718–8721 (2011).

56. Van Der Vossen, J. M. B. M. *et al.* Production of acidocin B, a bacteriocin of Lactobacillus acidophilus M46 is a plasmid-encoded trait: Plasmid curing, genetic marking by in vivo plasmid integration, and gene transfer. *FEMS Microbiol. Lett.* **116**, 333–340 (1994).

57. P, M. *et al.* Bacteriocin ASM1 is an O/S-diglycosylated, plasmid-encoded homologue of glycocin F. *FEBS Lett.* **594**, 1196–1206 (2020).

58. Risdian, C., Mozef, T. & Wink, J. Biosynthesis of Polyketides in Streptomyces. *Microorg. 2019, Vol. 7, Page 124* **7**, 124 (2019).

59. Wang, J. *et al.* Biosynthesis of aromatic polyketides in microorganisms using type II polyketide synthases. *Microb. Cell Factories 2020 191* **19**, 1–11 (2020).

60. Gokulan, K., Khare, S. & Cerniglia, C. Metabolic Pathways: Production of Secondary Metabolites of Bacteria. *Encycl. Food Microbiol. Second Ed.* 561–569 (2014) doi:10.1016/B978-0-12-384730-0.00203-2.

61. Donadio, S. & Katz, L. Organization of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin formation in Saccharopolyspora erythraea. *Gene* **111**, 51–60 (1992).

62. Pickens, L. B. & Tang, Y. Decoding and engineering tetracycline biosynthesis. *Metab. Eng.* **11**, 69–75 (2009).

63. Caffrey, P., Lynch, S., Flood, E., Finnan, S. & Oliynyk, M. Amphotericin biosynthesis in Streptomyces nodosus: deductions from analysis of polyketide synthase and late genes. *Chem. Biol.* **8**, 713–723 (2001).

64. MacNeil, D. J. *et al.* Complex organization of the Streptomyces avermitilis genes encoding the avermectin polyketide synthase. *Gene* **115**, 119–125 (1992).

65. Ranger, B. S. *et al.* Globally distributed mycobacterial fish pathogens produce a novel plasmid-encoded toxic macrolide, mycolactone F. *Infect. Immun.* **74**, 6037–6045 (2006).

66. Arakawa, K. Genetic and biochemical analysis of the antibiotic biosynthetic gene clusters on the Streptomyces linear plasmid. *Biosci. Biotechnol. Biochem.* **78**, 183–189 (2014).

67. M, K. & M, G. Engineering strategies for rational polyketide synthase design. *Nat. Prod. Rep.* **35**, 1070–1081 (2018).

68. Nivina, A., Yuet, K. P., Hsu, J. & Khosla, C. Evolution and Diversity of Assembly-Line Polyketide Synthases. *Chem. Rev.* **119**, 12524–12547 (2019).

69. Das, A. & Khosla, C. Biosynthesis of Aromatic Polyketides in Bacteria. (2008) doi:10.1021/ar8002249.

70. Helfrich, E. J. N. & Piel, J. Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* **33**, 231–316 (2016).

71. Dirk Schwarzer, Robert Finking & A. Marahiel, M. Nonribosomal peptides : from genes to products. *Nat. Prod. Rep.* **20**, 275–287 (2003).

72. Walsh, C. T. Insights into the chemical logic and enzymatic machinery of NRPS assembly lines. *Nat. Prod. Rep.* **33**, 127–135 (2016).

73. Lorenzo, M. Di *et al.* A Nonribosomal Peptide Synthetase with a Novel Domain Organization Is Essential for Siderophore Biosynthesis in Vibrio anguillarum. *J. Bacteriol.* **186**, 7327 (2004).

74. Kalmokoff, M. L., Banerjee, S. K., Cyr, T., Hefford, M. A. & Gleeson, T. Identification of a New Plasmid-Encoded sec-Dependent Bacteriocin Produced by Listeria innocua 743. *Appl. Environ. Microbiol.* **67**, 4041–4047 (2001).

75. Marahiel, M. A. A structural model for multimodular NRPS assembly lines. *Nat. Prod. Rep.* **33**, 136–140 (2016).

76. Finking, R. & Marahiel, M. A. Biosynthesis of nonribosomal peptides. *Annu. Rev. Microbiol.* **58**, 453–488 (2004).

77. Tang, G. L., Cheng, Y. Q. & Shen, B. Leinamycin Biosynthesis Revealing Unprecedented Architectural Complexity for a Hybrid Polyketide Synthase and Nonribosomal Peptide Synthetase. *Chem. Biol.* **11**, 33–45 (2004).

78. Du, L., Sánchez, C., Chen, M., Edwards, D. J. & Shen, B. The biosynthetic gene cluster for the antitumor drug bleomycin from Streptomyces verticillus ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase. *Chem. Biol.* **7**, 623–642 (2000).

79. Pelludat, C., Rakin, A., Jacobi, C. A., Schubert, S. & Heesemann, J. The yersiniabactin biosynthetic gene cluster of Yersinia enterocolitica: Organization and siderophore-dependent regulation. *J. Bacteriol.* **180**, 538–546 (1998).

80. Maria Fisch, K. Biosynthesis of natural products by microbial iterative hybrid PKS–NRPS. *RSC Adv.* **3**, 18228–18247 (2013).

81. Panter, F., Bader, C. D. & Müller, R. The Sandarazols are Cryptic and Structurally

Unique Plasmid-Encoded Toxins from a Rare Myxobacterium**. *Angew. Chemie Int. Ed.* **60**, 8081–8088 (2021).

82. Dickschat, J. S. Terpenes. *Beilstein J. Org. Chem. 15292* **15**, 2966–2967 (2019).

83. Helfrich, E. J. N., Lin, G.-M., Voigt, C. A. & Clardy, J. Bacterial terpene biosynthesis: challenges and opportunities for pathway engineering. *Beilstein J. Org. Chem. 15283* **15**, 2889–2906 (2019).

84. Yamada, Y. *et al.* Terpene synthases are widely distributed in bacteria. *Proc. Natl. Acad. Sci.* **112**, 857–862 (2015).

85. Reddy, G. K. *et al.* Exploring novel bacterial terpene synthases. *PLoS One* **15**, (2020).

86. Medema, M. H. *et al.* The Sequence of a 1.8-Mb Bacterial Linear Plasmid Reveals a Rich Evolutionary Reservoir of Secondary Metabolic Pathways. *Genome Biol. Evol.* **2**, 212 (2010).

87. Fedorova, N. D., Moktali, V. & Medema, M. H. Bioinformatics approaches and software for detection of secondary metabolic gene clusters. *Methods in Molecular Biology* vol. 944 23–45 (2012).

88. Medema, M. H. & Fischbach, M. A. Computational approaches to natural product discovery. *Nature Chemical Biology* vol. 11 639–648 (2015).

89. Chavali, A. K. & Rhee, S. Y. Bioinformatics tools for the identification of gene clusters that biosynthesize specialized metabolites. *Briefings in bioinformatics* vol. 19 1022–1034 (2018).

90. Blin, K. *et al.* AntiSMASH 5.0: Updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **47**, W81–W87 (2019).

91. N, K. *et al.* SMURF: Genomic mapping of fungal secondary metabolite clusters. *Fungal Genet. Biol.* **47**, 736–741 (2010).

92. Almeida, H., Palys, S., Tsang, A. & Diallo, A. B. TOUCAN: a framework for fungal biosynthetic gene cluster discovery. *NAR Genomics Bioinforma.* **2**, (2020).

93. Bratovanov, E. V. *et al.* Genome Mining and Heterologous Expression Reveal Two Distinct Families of Lasso Peptides Highly Conserved in Endofungal Bacteria. *ACS Chem. Biol.* **15**, 1169–1176 (2019).

94. Kaweewan, I., Hemmi, H., Komaki, H., Harada, S. & Kodani, S. Isolation and structure determination of a new lasso peptide specilicin based on genome mining. *Bioorg. Med. Chem.* **26**, 6050–6055 (2018).

95. Dong, D. L.-B., Rudolf, D. J. D., Deng, D. M.-R., Yan, D. X. & Shen, P. D. Ben. Discovery of the Tiancilactone Antibiotics by Genome Mining of Atypical Bacterial Type II Diterpene Synthases. *Chembiochem* **19**, 1727–1733 (2018).

96. Li, Y.-X., Zhong, Z., Zhang, W.-P. & Qian, P.-Y. Discovery of cationic nonribosomal peptides as Gram-negative antibiotics through global genome mining. *Nat. Commun. 2018 91* **9**, 1–9 (2018).

97. Ye, S. *et al.* Identification by Genome Mining of a Type I Polyketide Gene Cluster from Streptomyces argillaceus Involved in the Biosynthesis of Pyridine and

Piperidine Alkaloids Argimycins P. *Front. Microbiol.* **0**, 194 (2017).

98. Xu, M. *et al.* Functional Genome Mining Reveals a Class V Lanthipeptide Containing a d-Amino Acid Introduced by an F420H2-Dependent Reductase. *Angew. Chemie Int. Ed.* **59**, 18029–18035 (2020).

99. Blin, K. *et al.* antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* **49**, W29–W35 (2021).

100. Hannigan, G. D. *et al.* A deep learning genome-mining strategy for biosynthetic gene cluster prediction. *Nucleic Acids Res.* **47**, e110 (2019).

101. Skinnider, M. A., Merwin, N. J., Johnston, C. W. & Magarvey, N. A. PRISM 3: expanded prediction of natural product chemical structures from microbial genomes. *Nucleic Acids Res.* **45**, W49 (2017).

102. Cimermancic, P. *et al.* Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158**, 412 (2014).

103. P, C.-M. *et al.* Phylogenomic Analysis of Natural Products Biosynthetic Gene Clusters Allows Discovery of Arseno-Organic Metabolites in Model Streptomycetes. *Genome Biol. Evol.* **8**, 1906–1916 (2016).

104. G, Y., SH, S. & MR, T. Identifying clusters of functionally related genes in genomes. *Bioinformatics* **23**, 1053–1060 (2007).

105. Li, M. H., Ung, P. M., Zajkowski, J., Garneau-Tsodikova, S. & Sherman, D. H. Automated genome mining for natural products. *BMC Bioinformatics* **10**, 185 (2009).

106. A, S. *et al.* ClustScan: an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures. *Nucleic Acids Res.* **36**, 6882–6892 (2008).

107. T, W. *et al.* CLUSEAN: a computer-based framework for the automated analysis of bacterial secondary metabolite biosynthetic gene clusters. *J. Biotechnol.* **140**, 13–17 (2009).

108. Röttig, M. *et al.* NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* **39**, W362 (2011).

109. AJ, van H. *et al.* BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Res.* **46**, W278–W281 (2018).

110. Merwin, N. J. *et al.* DeepRiPP integrates multiomics data to automate discovery of novel ribosomally synthesized natural products. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 371 (2020).

111. Zierep, P. F., Ceci, A. T., Dobrusin, I., Rockwell-Kollmann, S. C. & Günther, S. SeMPI 2.0—A Web Server for PKS and NRPS Predictions Combined with Metabolite Screening in Natural Product Databases. *Metabolites* **11**, 1–27 (2021).

112. ELC, de L. S. NeuRiPP: Neural network identification of RiPP precursor peptides. *Sci. Rep.* **9**, (2019).

113. P, A., S, K., M, G., N, S. & D, M. RiPPMiner: a bioinformatics resource for deciphering chemical structures of RiPPs based on prediction of cleavage and

cross-links. *Nucleic Acids Res.* **45**, W80–W88 (2017).

114. Tietz, J. I. *et al.* A new genome-mining tool redefines the lasso peptide biosynthetic landscape. *Nat. Chem. Biol.* **13**, 470 (2017).

115. J, K. & GS, Y. PKMiner: a database for exploring type II polyketide synthases. *BMC Microbiol.* **12**, (2012).

116. Santos-Aberturas, J. *et al.* Uncovering the unexplored diversity of thioamidated ribosomal peptides in Actinobacteria using the RiPPER genome mining tool. *Nucleic Acids Res.* **47**, 4624–4637 (2019).

117. Kloosterman, A. M., Shelton, K. E., van Wezel, G. P., Medema, M. H. & Mitchell, D. A. RRE-Finder: a Genome-Mining Tool for Class-Independent RiPP Discovery. *mSystems* **5**, (2020).

118. Kloosterman, A. M. *et al.* Integration of machine learning and pan-genomics expands the biosynthetic landscape of RiPP natural products. *bioRxiv* 2020.05.19.104752 (2020) doi:10.1101/2020.05.19.104752.

119. Medema, M. H. *et al.* AntiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **39**, W339 (2011).

120. Blin, K., Shaw, S., Kautsar, S. A., Medema, M. H. & Weber, T. The antiSMASH database version 3: increased taxonomic coverage and new query features for modular enzymes. *Nucleic Acids Res.* **49**, D639–D643 (2021).

121. Kalkreuter, E., Pan, G., Cepeda, A. J. & Shen, B. Targeting Bacterial Genomes for Natural Product Discovery. *Trends Pharmacol. Sci.* **41**, 13 (2020).

122. Desriac, F. *et al.* Antimicrobial Peptides from Marine Proteobacteria. *Mar. Drugs 2013, Vol. 11, Pages 3632-3660* **11**, 3632–3660 (2013).

123. Mohamed, N. M. *et al.* Diversity and quorum-sensing signal production of Proteobacteria associated with marine sponges. *Environ. Microbiol.* **10**, 75–86 (2008).

124. Timmermans, M. L., Paudel, Y. P. & Ross, A. C. Investigating the Biosynthesis of Natural Products from Marine Proteobacteria: A Survey of Molecules and Strategies. *Mar. Drugs 2017, Vol. 15, Page 235* **15**, 235 (2017).

125. Guo, H. *et al.* Natural products and morphogenic activity of γ-Proteobacteria associated with the marine hydroid polyp Hydractinia echinata. *Bioorg. Med. Chem.* **25**, 6088–6097 (2017).

126. van Bergeijk, D. A., Terlouw, B. R., Medema, M. H. & van Wezel, G. P. Ecology and genomics of Actinobacteria: new concepts for natural product discovery. *Nat. Rev. Microbiol. 2020 1810* **18**, 546–558 (2020).

127. Jose, P. A., Maharshi, A. & Jha, B. Actinobacteria in natural products research: Progress and prospects. *Microbiol. Res.* **246**, 126708 (2021).

128. Valliappan, K., Sun, W. & Li, Z. Marine actinobacteria associated with marine organisms and their potentials in producing pharmaceutical natural products. *Appl. Microbiol. Biotechnol. 2014 9817* **98**, 7365–7377 (2014).

129. Manivasagan, P. *et al.* Marine actinobacteria: An important source of bioactive natural products. *Environ. Toxicol. Pharmacol.* **38**, 172–188 (2014).

130. Axenov-Gribanov, D. V. *et al.* Cultivable Actinobacteria First Found in Baikal Endemic Algae Is a New Source of Natural Products with Antibiotic Activity. *Int. J. Microbiol.* **2020**, (2020).

131. Poorinmohammad, N., Bagheban-Shemirani, R. & Hamedi, J. Genome mining for ribosomally synthesised and post-translationally modified peptides (RiPPs) reveals undiscovered bioactive potentials of actinobacteria. *Antonie van Leeuwenhoek 2019 11210* **112**, 1477–1499 (2019).

132. Malit, J. J. L., Wu, C., Liu, L.-L. & Qian, P.-Y. Global Genome Mining Reveals the Distribution of Diverse Thioamidated RiPP Biosynthesis Gene Clusters. *Front. Microbiol.* **12**, (2021).

133. Gomes, K. M., Duarte, R. S. & Bastos, M. do C. de F. Lantibiotics produced by Actinobacteria and their potential applications (a review). *Microbiology* **163**, 109–121 (2017).

134. Walker, M. C. *et al.* Precursor peptide-targeted mining of more than one hundred thousand genomes expands the lanthipeptide natural product family. doi:10.1186/s12864-020-06785-7.

135. Kloosterman, A. M. *et al.* Expansion of RiPP biosynthetic space through integration of pan-genomics and machine learning uncovers a novel class of lanthipeptides. *PLOS Biol.* **18**, e3001026 (2020).

136. Ren, H., Shi, C., Bothwell, I. R., Donk, W. A. van der & Zhao, H. Discovery and Characterization of a Class IV Lanthipeptide with a Nonoverlapping Ring Pattern. *ACS Chem. Biol.* **15**, 1642–1649 (2020).

137. Zhang, Q., Doroghazi, J. R., Zhao, X., Walker, M. C. & van der Donk, W. A. Expanded natural product diversity revealed by analysis of lanthipeptide-like gene clusters in Actinobacteria. *Appl. Environ. Microbiol.* **81**, 4339–4350 (2015).

138. Diego Rodríguez-Hernández *et al.* Actinobacteria associated with stingless bees biosynthesize bioactive polyketides against bacterial pathogens. *New J. Chem.* **43**, 10109–10117 (2019).

139. Benndorf, R. *et al.* Natural Products from Actinobacteria Associated with Fungus-Growing Termites. *Antibiot. 2018, Vol. 7, Page 83* **7**, 83 (2018).

140. Gohain, A. *et al.* Phylogenetic affiliation and antimicrobial effects of endophytic actinobacteria associated with medicinal plants: prevalence of polyketide synthase type II in antimicrobial strains. *Folia Microbiol. 2019 644* **64**, 481–496 (2019).

141. Meena, B., Anburajan, L., Vinithkumar, N. V., Kirubagaran, R. & Dharani, G. Biodiversity and antibacterial potential of cultivable halophilic actinobacteria from the deep sea sediments of active volcanic Barren Island. *Microb. Pathog.* **132**, 129–136 (2019).

142. Zhang, Y.-M. *et al.* Anandins A and B, Two Rare Steroidal Alkaloids from a Marine Streptomyces anandii H41-59. *Mar. Drugs 2017, Vol. 15, Page 355* **15**, 355 (2017).

143. Nam, S.-J. *et al.* Actinobenzoquinoline and Actinophenanthrolines A–C, Unprecedented Alkaloids from a Marine Actinobacterium. *Org. Lett.* **17**, 3240–3243 (2015).

144. Kamala, K., Sivaperumal, P., Gobalakrishnan, R., Swarnakumar, N. S. & Rajaram, R. Isolation and characterization of biologically active alkaloids from marine actinobacteria Nocardiopsis sp. NCS1. *Biocatal. Agric. Biotechnol.* **4**, 63–69 (2015).

145. Primahana, G. *et al.* Nonocarbolines A–E, β-Carboline Antibiotics Produced by the Rare Actinobacterium Nonomuraea sp. from Indonesia. *Antibiot. 2020, Vol. 9, Page 126* **9**, 126 (2020).

146. Nithya, K. *et al.* Desert actinobacteria as a source of bioactive compounds production with a special emphases on Pyridine-2,5-diacetamide a new pyridine alkaloid produced by Streptomyces sp. DA3-7. *Microbiol. Res.* **207**, 116–133 (2018).

147. Heidari, B. & Mohammadipanah, F. Isolation and identification of two alkaloid structures with radical scavenging activity from Actinokineospora sp. UTMC 968, a new promising source of alkaloid compounds. *Mol. Biol. Reports 2018 456* **45**, 2325–2332 (2018).

148. Wang, H., Fewer, D. P. & Sivonen, K. Genome Mining Demonstrates the Widespread Occurrence of Gene Clusters Encoding Bacteriocins in Cyanobacteria. *PLoS One* **6**, e22384 (2011).

149. Micallef, M. L., D'Agostino, P. M., Sharma, D., Viswanathan, R. & Moffitt, M. C. Genome mining for natural product biosynthetic gene clusters in the Subsection V cyanobacteria. *BMC Genomics 2015 161* **16**, 1–20 (2015).

150. Leikoski, N. *et al.* Genome Mining Expands the Chemical Diversity of the Cyanobactin Family to Include Highly Modified Linear Peptides. *Chem. Biol.* **20**, 1033–1043 (2013).

151. Galica, T., Hrouzek, P. & Mareš, J. Genome mining reveals high incidence of putative lipopeptide biosynthesis NRPS/PKS clusters containing fatty acyl-AMP ligase genes in biofilm-forming cyanobacteria. *J. Phycol.* **53**, 985–998 (2017).

152. Larsen, J. S., Pearson, L. A. & Neilan, B. A. Genome Mining and Evolutionary Analysis Reveal Diverse Type III Polyketide Synthase Pathways in Cyanobacteria. *Genome Biol. Evol.* **13**, (2021).

153. PM, S. *et al.* Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 1053–1058 (2013).

154. Popin, R., Alvarenga, D., Castelo-Branco, R., Fewer, D. & Sivonen, K. Mining of Cyanobacterial Genomes Indicates That Plasmids Are Involved in the Production of Natural Products. doi:10.21203/rs.3.rs-121751/v1.

155. David Prihoda *et al.* The application potential of machine learning and genomics for understanding natural product diversity, chemistry, and therapeutic translatability. *Nat. Prod. Rep.* **38**, 1100–1108 (2021).

156. Hrab, P. *et al.* Complete genome sequence of Streptomyces cyanogenus S136, producer of anticancer angucycline landomycin A. *3 Biotech 2021 116* **11**, 1–10

(2021).

157. Yamany, A. Mining selected metagenomes/metatranscriptomes for biosynthetic gene clusters and antimicrobial resistance gene. (American University in Cairo, 2021).

158. Bidaisee, S. & Macpherson, C. N. L. Zoonoses and one health: A review of the literature. *J. Parasitol. Res.* **2014**, (2014).

159. Berrian, A. M. *et al.* A community-based One Health education program for disease risk mitigation at the human-animal interface. *One Heal.* **5**, 9–20 (2018).

160. Cantas, L. & Suer, K. Review: The Important Bacterial Zoonoses in â€œOne Healthâ€ Concept. *Front. Public Heal.* **2**, 1–8 (2014).

161. McFall-Ngai, M. *et al.* Animals in a bacterial world, a new imperative for the life sciences. *Proc. Natl. Acad. Sci.* **110**, 3229–3236 (2013).

162. Gilbert, J. A. *et al.* Microbiome-wide association studies link dynamic microbial consortia to disease. *Nature* vol. 535 94–103 (2016).

163. Zhang, T., Zhang, X. X. & Ye, L. Plasmid metagenome reveals high levels of antibiotic resistance genes and mobile genetic elements in activated sludge. *PLoS One* **6**, 26041 (2011).

164. Carattoli, A. Plasmids and the spread of resistance. *International Journal of Medical Microbiology* vol. 303 298–304 (2013).

165. Actis, L. A. *et al.* Characterization of anguibactin, a novel siderophore from Vibrio anguillarum 775(pJM1). *J. Bacteriol.* **167**, 57–65 (1986).

166. Lee, C. Te *et al.* The opportunistic marine pathogen Vibrio parahaemolyticus becomes virulent by acquiring a plasmid that expresses a deadly toxin. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 10798–10803 (2015).

167. *Leaflet CHROMagar TM ESBL TM For overnight detection of Gram-negative bacteria producing Extended Spectrum Beta-Lactamase CHROMagar TM ESBL For overnight detection of Gram-negative bacteria producing Extended Spectrum Beta-Lactamase Medium Performance*.

168. Rameshkumar, M. R., Vignesh, R., Swathirajan, C. R., Balakrishnan, P. & Arunagirinathan, N. Chromogenic agar medium for rapid detection of extended-spectrum β-lactamases and Klebsiella pneumoniae carbapenemases producing bacteria from human immunodeficiency virus patients. *Journal of Research in Medical Sciences* vol. 20 1219–1220 (2015).

169. Bimboim, H. C. & Doly, J. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**, 1513–1523 (1979).

170. Feliciello, I. & Chinali, G. A modified alkaline lysis method for the preparation of highly purified plasmid DNA from Escherichia coli. *Anal. Biochem.* **212**, 394–401 (1993).

171. Wang, W. *et al.* Application of metagenomics in the human gut microbiome. **21**, 803–814 (2015).

172. Jones, B. V & Marchesi, J. R. Transposon-aided capture (TRACA) of plasmids

resident in the human gut mobile metagenome. **4**, (2007).

173. Norman, A., Hansen, L. H. & Sørensen, S. J. Conjugative plasmids : vessels of the communal gene pool. 2275–2289 (2009) doi:10.1098/rstb.2009.0037.

174. Froquet, R., Lelong, E., Marchetti, A. & Cosson, P. Dictyostelium discoideum: A model host to measure bacterial virulence. *Nat. Protoc.* **4**, 25–30 (2009).

175. Klapper, M., Gçtze, S., Barnett, R., Willing, K. & Stallforth, P. Natural Products Bacterial Alkaloids Prevent Amoebal Predation Angewandte. 8944–8947 (2016) doi:10.1002/anie.201603312.

176. Stallforth, P. *et al.* A bacterial symbiont is converted from an inedible producer of bene fi cial molecules into food by a single mutation in the gacA gene. **110**, 14528–14533 (2013).

177. Miethke, M. & Marahiel, M. A. Siderophore-Based Iron Acquisition and Pathogen Control. **71**, 413–451 (2007).

178. Neilands, B. Universal Chemical Assay for the Detection Determination of Siderophores'. **56**, 47–56 (1987).

179. Brady, S. F. Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules. **2**, 1297–1305 (2007).

180. Craig, J. W. *et al.* Expanding Small-Molecule Functional Metagenomics through Parallel Screening of Broad-Host-Range Cosmid Environmental DNA Libraries in Diverse Proteobacteria □ †. **76**, 1633–1641 (2010).

181. Chang, F. & Brady, S. F. Discovery of indolotryptoline antiproliferative agents by homology-guided metagenomic screening. **2012**, (2013).

182. Kearse, M. *et al.* Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).

183. Bauman, C. A., Barkema, H. W., Dubuc, J., Keefe, G. P. & Kelton, D. F. Identifying management and disease priorities of Canadian dairy industry stakeholders. *J. Dairy Sci.* **99**, 10194–10203 (2016).

184. de Graaf, F. K., Spanjaerdt Speckman, E. A. & Stouthamer, A. H. Mode of action of a bacteriocin produced by Enterobacter cloacae DF 13. *Antonie Van Leeuwenhoek* **35**, 287–306 (1969).

185. De Graaf, F. K. & Klaasen-Boor, P. *Volume 40, number 2 FEBS LETTERS PURIFICATION AND CHARACTERIZATION OF THE CLOACIN DF13 IMMUNITY PROTEIN*. (1974).

186. Netz, D. J. A. *et al.* Molecular characterisation of aureocin A70, a multi-peptide bacteriocin isolated from Staphylococcus aureus. *J. Mol. Biol.* **311**, 939–949 (2001).

187. Coelho, M. L. V., Fleming, L. R. & Bastos, M. do C. de F. Insights into aureocin A70 regulation: Participation of regulator AurR, alternative transcription factor σB and phage φ11 regulator cI. *Res. Microbiol.* **167**, 90–102 (2016).

188. Coelho, M. L. V., Coutinho, B. G., Cabral da Silva Santos, O., Nes, I. F. & Bastos, M. do C. de F. Immunity to the Staphylococcus aureus leaderless four-peptide bacteriocin aureocin A70 is conferred by AurI, an integral membrane protein. *Res. Microbiol.* **165**, 50–59 (2014).

189. Madden, T. *The BLAST Sequence Analysis Tool*.

190. Sánchez-Osuna, M., Cortés, P., Barbé, J. & Erill, I. Origin of the mobile di-hydropteroate synthase gene determining sulfonamide resistance in clinical isolates. *Front. Microbiol.* **10**, 3332 (2019).

191. Bush, K. & Bradford, P. A. Epidemiology of β-lactamase-producing pathogens. *Clinical Microbiology Reviews* vol. 33 (2020).

192. Sawa, T., Kooguchi, K. & Moriyama, K. Molecular diversity of extended-spectrum β-lactamases and carbapenemases, and antimicrobial resistance. *Journal of Intensive Care* vol. 8 (2020).

193. Bush, K. Past and present perspectives on β-lactamases. *Antimicrobial Agents and Chemotherapy* vol. 62 (2018).

194. Tooke, C. L. *et al.* β-Lactamases and β-Lactamase Inhibitors in the 21st Century. *Journal of Molecular Biology* vol. 431 3472–3500 (2019).

195. Briñas, L., Zarazaga, M., Sáenz, Y., Ruiz-Larrea, F. & Torres, C. β-lactamases in ampicillin-resistant Escherichia coli isolates from foods, humans, and healthy animals. *Antimicrob. Agents Chemother.* **46**, 3156–3163 (2002).

196. Chopra, I. & Roberts, M. Tetracycline Antibiotics: Mode of Action, Applications, Molecular Biology, and Epidemiology of Bacterial Resistance. *Microbiol. Mol. Biol. Rev.* **65**, 232–260 (2001).

197. Akiba, T., Koyama, K., Ishiki, Y., Kimura, S. & Fukushima, T. ON THE MECHANISM OF THE DEVELOPMENT OF MULTIPLE-DRUG-RESISTANT CLONES OF SHIGELLA. *Jpn. J. Microbiol.* **4**, 219–227 (1960).

198. Roberts, M. C. Update on acquired tetracycline resistance genes. *FEMS Microbiology Letters* vol. 245 195–203 (2005).

199. Grossman, T. H. Tetracycline antibiotics and resistance. *Cold Spring Harb. Perspect. Med.* **6**, a025387 (2016).

200. McMurry, L., Petrucci, R. E. & Levy, S. B. Active efflux of tetracycline encoded by four genetically different tetracyline resistance determinants in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **77**, 3974–3977 (1980).

201. Salyers, A. A., Speer, B. S. & Shoemaker, N. B. New perspectives in tetracycline resistance. *Molecular Microbiology* vol. 4 151–156 (1990).

202. Thaker, M., Spanogiannopoulos, P. & Wright, G. D. The tetracycline resistome. *Cellular and Molecular Life Sciences* vol. 67 419–431 (2010).

203. Hasan, B. *et al.* Antimicrobial drug-resistant escherichia coli in wild birds and free-range poultry, Bangladesh. *Emerg. Infect. Dis.* **18**, 2055–2058 (2012).

204. Ahmed, A. M., Shimamoto, T. & Shimamoto, T. Molecular characterization of multidrug-resistant avian pathogenic Escherichia coli isolated from septicemic

broilers. *Int. J. Med. Microbiol.* **303**, 475–483 (2013).

205. Enany, M. E. *et al.* The occurrence of the multidrug resistance (MDR) and the prevalence of virulence genes and QACs resistance genes in E. coli isolated from environmental and avian sources. *AMB Express* **9**, 1–9 (2019).

206. Azam, M., Mohsin, M., Sajjad-ur-Rahman & Saleemi, M. K. Virulence-associated genes and antimicrobial resistance among avian pathogenic Escherichia coli from colibacillosis affected broilers in Pakistan. *Trop. Anim. Health Prod.* **51**, 1259–1265 (2019).

207. Azam, M. *et al.* Genomic landscape of multi-drug resistant avian pathogenic Escherichia coli recovered from broilers. *Vet. Microbiol.* **247**, (2020).

208. Kimura, A. H. *et al.* Characterization of multidrug-resistant avian pathogenic Escherichia coli: an outbreak in canaries. *Brazilian J. Microbiol.* **52**, 1005–1012 (2021).

209. Parker, D., Sniatynski, M. K., Mandrusiak, D. & Rubin, J. E. Extended-spectrum β-lactamase producing Escherichia coli isolated from wild birds in Saskatoon, Canada. *Lett. Appl. Microbiol.* **63**, 11–15 (2016).

210. Ritchie, L. E., Steiner, M. & Suchodolski, J. S. Assessment of microbial diversity along the feline intestinal tract using16S rRNA gene analysis. (2008) doi:10.1111/j.1574-6941.2008.00609.x.

211. Desai, A. R., Musil, K. M., Carr, A. P. & Hill, J. E. Characterization and quantification of feline fecal microbiota using cpn60 sequence-based methods and investigation of animal-to-animal variation in microbial population structure. *Vet. Microbiol.* **137**, 120–128 (2009).

212. Handl, S., Dowd, S. E., Garcia-Mazcorro, J. F., Steiner, M. & Suchodolski, J. S. Massive parallel 16S rRNA gene pyrosequencing reveals highly diverse fecal bacterial and fungal communities in healthy dogs and cats. doi:10.1111/j.1574-6941.2011.01058.x.

213. Tun, H. M. *et al.* Gene-centric metagenomics analysis of feline intestinal microbiome using 454 junior pyrosequencing. *J. Microbiol. Methods* **88**, 369–376 (2012).

214. Bermingham, E. N. *et al.* The Fecal Microbiota in the Domestic Cat (Felis catus) Is Influenced by Interactions Between Age and Diet; A Five Year Longitudinal Study. *Front. Microbiol.* **9**, 1231 (2018).

215. Lubbs, D. C., Vester, B. M., Fastinger, N. D. & Swanson, K. S. Dietary protein concentration affects intestinal microbiota of adult cats: a study using DGGE and qPCR to evaluate differences in microbial populations in the feline gastrointestinal tract. *J. Anim. Physiol. Anim. Nutr. (Berl).* **93**, 113–121 (2009).

216. Fischer, M. M. *et al.* Effects of obesity, energy restriction and neutering on the faecal microbiota of cats. *Br. J. Nutr.* **118**, 513–524 (2017).

217. Mao, S., Zhang, M., Liu, J. & Zhu, W. Characterising the bacterial microbiota across the gastrointestinal tracts of dairy cattle: Membership and potential function. *Sci. Rep.* **5**, 1–14 (2015).

218. Clemmons, B. A., Voy, B. H. & Myer, P. R. Altering the Gut Microbiome of Cattle: Considerations of Host-Microbiome Interactions for Persistent Microbiome Manipulation. *Microbial Ecology* vol. 77 523–536 (2019).

219. Rojas, C. A., Ramírez-Barahona, S., Holekamp, K. E. & Theis, K. R. Host phylogeny and host ecology structure the mammalian gut microbiota at different taxonomic scales. *Anim. Microbiome* **3**, 33 (2021).

220. Shintani, M., Sanchez, Z. K. & Kimbara, K. Genomics of microbial plasmids: Classification and identification based on replication and transfer systems and host taxonomy. *Frontiers in Microbiology* vol. 6 242 (2015).

221. Dos Santos Nascimento, J., Dos Santos, K. R. N., Gentilini, E., Sordelli, D. & De Freire Bastos, M. D. C. Phenotypic and genetic characterisation of bacteriocin-producing strains of Staphylococcus aureus involved in bovine mastitis. *Vet. Microbiol.* **85**, 133–144 (2002).

222. Ceotto, H., Nascimento, J. dos S., Brito, M. A. V. de P. & Bastos, M. do C. de F. Bacteriocin production by Staphylococcus aureus involved in bovine mastitis in Brazil. *Res. Microbiol.* **160**, 592–599 (2009).

223. Brito, M. A. V. P., Somkuti, G. A. & Renye, J. A. Production of antilisterial bacteriocins by staphylococci isolated from bovine milk. *J. Dairy Sci.* **94**, 1194–1200 (2011).

224. Kav, A. B. *et al.* Insights into the bovine rumen plasmidome. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 5452–5457 (2012).

225. Gumpert, H. *et al.* Transfer and persistence of a multi-drug resistance plasmid in situ of the infant gut microbiota in the absence of antibiotic treatment. *Front. Microbiol.* **8**, 1852 (2017).

226. Aviv, G., Rahav, G. & Gal-Mor, O. Horizontal transfer of the Salmonella enterica serovar infantis resistance and virulence plasmid pESI to the gut microbiota of warm-blooded hosts. *MBio* **7**, (2016).

227. Maeusli, M. *et al.* Horizontal Gene Transfer of Antibiotic Resistance from Acinetobacter baylyi to Escherichia coli on Lettuce and Subsequent Antibiotic Resistance Transmission to the Gut Microbiome . *mSphere* **5**, (2020).

228. Liao, H. *et al.* Horizontal gene transfer and shifts in linked bacterial community composition are associated with maintenance of antibiotic resistance genes during food waste composting. *Sci. Total Environ.* **660**, 841–850 (2019).

229. Bonham, K. S., Wolfe, B. E. & Dutton, R. J. Extensive horizontal gene transfer in cheese-associated bacteria. *Elife* **6**, (2017).

230. Zhou, Z. C. *et al.* Prevalence of multi-resistant plasmids in hospital inhalable particulate matter (PM) and its impact on horizontal gene transfer. *Environ. Pollut.* **270**, 116296 (2021).

231. Evans, D. R. *et al.* Systematic detection of horizontal gene transfer across genera among multidrug-resistant bacteria in a single hospital. *Elife* **9**, (2020).

232. Jiang, X., Hall, A. B., Xavier, R. J. & Alm, E. J. Comprehensive analysis of chromosomal mobile genetic elements in the gut microbiome reveals phylum-level

niche-adaptive gene pools. *PLoS One* **14**, e0223680 (2019).

233. Bosch, T. *et al.* Outbreak of NDM-1-producing klebsiella pneumoniae in a Dutch hospital, with interspecies transfer of the resistance plasmid and unexpected occurrence in unrelated health care centers. *J. Clin. Microbiol.* **55**, 2380–2390 (2017).

234. Mathers, A. J. *et al. Klebsiella quasipneumoniae Provides a Window into Carbapenemase Gene Transfer, Plasmid Rearrangements, and Patient Interactions with the Hospital Environment*. http://aac.asm.org/ (2019).

235. Arias-Andres, M., Klümper, U., Rojas-Jimenez, K. & Grossart, H. P. Microplastic pollution increases gene exchange in aquatic ecosystems. *Environ. Pollut.* **237**, 253–261 (2018).

236. Caro-Quintero, A., Ritalahti, K. M., Cusick, K. D., Löffler, F. E. & Konstantinidis, K. T. The chimeric genome of sphaerochaeta: Nonspiral spirochetes that break with the prevalent dogma in spirochete biology. *MBio* **3**, (2012).

237. Caro-Quintero, A. & Konstantinidis, K. T. Inter-phylum HGT has shaped the metabolism of many mesophilic and anaerobic bacteria. *ISME J.* **9**, 958–967 (2015).

238. Panda, A., Drancourt, M., Tuller, T. & Pontarotti, P. Genome-wide analysis of horizontally acquired genes in the genus Mycobacterium. *Sci. Rep.* **8**, 1–13 (2018).

239. Loy, A. *et al.* Lifestyle and Horizontal Gene Transfer-Mediated Evolution of Mucispirillum schaedleri, a Core Member of the Murine Gut Microbiota Downloaded from. **2**, 171–187.

240. Bodor, A. *et al.* Challenges of unculturable bacteria: environmental perspectives. *Rev. Environ. Sci. Bio/Technology 2020 191* **19**, 1–22 (2020).