Mining bacterial genomes for novel antimicrobials



CONNOR SHARP

KELLOGG COLLEGE

UNIVERSITY OF OXFORD

A thesis submitted in partial fulfilment of the degree of

DOCTOR OF PHILOSOPHY

TRINITY 2017

supervised by

Prof. Colin KLEANTHOUS

Prof. Martin MAIDEN

Abstract

Bacteria have evolved an arsenal of antimicrobial peptides and proteins to compete with each other. Bacteriocins are diffusible toxins released into the environment and differ from contact-dependent inhibitors (CDI) or type six secretions systems (T6SS) which require cell-cell contact. Protein bacteriocins are potent narrow-spectrum toxins with a modular domain organisation, making them ideal candidates for developing future antimicrobials. Though progress is being made to exploit protein bacteriocins, current limitations in our understanding of protein bacteriocin biology limits their possible use as antimicrobials. Firstly, due to high sequence diversity and similarity to other competition systems our understanding of the diversity and prevalence of protein bacteriocins in bacteria is poor. Secondly, high levels of resistance have been reported with tested under lab conditions. Finally, how protein bacteriocins translocate into the cell is poorly understood. The aim of this study is to answer a sequence of questions relating to these central limitations in our knowledge of protein bacteriocins. A bioinformatic pipeline was developed which exploits Hidden Markov Models to unambiguously identify nuclease bacteriocins and revealed a diverse family of proteins present in >2000 strains throughout the *Enterobacteriaceae* family and *Pseudomonas* spp. A novel grouping scheme is devised for NBs based on conserved structural organisation. Investigation of non-specific colicin resistance mechanism revealed an O-antigen dependent resistance involving density dependent blocking of the receptor proteins in the outer membrane. This resistance mechanism was sensitive to changes in environment and could be overcome by additives. Finally, a conserved motif, the 'DPY' motif, was identified across the nuclease bacteriocin family and effectors of the T6SS. While the role of the DPY motif is still unclear, it is associated with toxins which cross the inner membrane.

Acknowledgements

First and foremost I thank my supervisors, Professor Colin Kleanthous and Professor Martin Maiden, for their invaluable guidance, support and advice during my four years at Oxford. Thanks must go to Renata Kaminska for showing me how to purify proteins, James Bray for showing me the inner workings of the pubMLST database, and to Nick Housden for generating some of the constructs used in this work and being an irreplaceable mentor in the lab.

I would also like to thanks my collaborators at the Wellcome Trust Sanger Institute: Professor Julian Parkhill for allowing me to work in his group, Amy Cain for assistance with the BIOLOG, Christine Boinett for establishing the TraDIS library in UPEC ST131 and Teemu Kallonen for pointing me in the right way with pangenome analysis.

Charlie 'Figs' Bury, who has been a constant companion since my first python bug and always been there for a helpful coffee. For her help setting up the microscopy, I would like to thank Marie-Louise Francis and apologise for pronouncing her name wrong for a good half a year. All members of the Kleanthous lab for letting me chat about Markov Models for far too long during lab meetings, offering advice and helping me celebrate 3 birthdays.

The support of my partner, Yana Dem'yanenko, who put up with the complaining, ranting and late nights that are inevitable with a thesis. Thank you for your all of your help, for keeping me sane and telling me I can't just eat wraps for every meal.

Finally, I am grateful to the Systems Biology program of the EPSRC funded Doctoral Training Centre for funding the studentship that made all of this work possible.

Publications

Details of the following work have also been published in:

- **Sharp C**, Bray J, Housden NG, Maiden MCJ, Kleanthous C. Diversity and distribution of nuclease bacteriocins in bacterial genomes revealed using Hidden Markov Models. (Orengo CA, ed.) PLoS Computational Biology. 2017;13(7):e1005652.
- Joshi A, Grinter R, Josts I, Chen S, Wojdyla J A, Lowe E D, Kaminska R, **Sharp C**, et al. (2015). Structures of the Ultra-High-Affinity Protein-Protein Complexes of Pyocins S2 and AP41 and Their Cognate Immunity Proteins from *Pseudomonas aeruginosa*. Journal of Molecular Biology, 427(17), 2852-2866.
- Rajasekar K V, Zdanowski K, Yan J, Hopper JTS, Francis, M-L R, Seepersad C, **Sharp C**, et al. (2016). The anti-sigma factor RsrA responds to oxidative stress by reburying its hydrophobic core. Nature Communications, 7, 12194.

Abbreviations

aa ..... Amino Acid

AF647 ..... AlexaFluor 647

AFM ..... Atomic force microscopy

BRP ..... Bacteriocin release protein

CAMP ..... Cationic Antimicrobial Peptide

C-domain ..... Cytotoxic domain

CDI ..... Contact Dependent Inhibition

CDS ..... Contact Dependent Signalling

CLB ..... Colicin-like Bacteriocin

DNA ..... Deoxyribonucleic acid

DPY ..... Aspartate, Proline, Tyrosine motif

DTT ..... Dithiothreitol

EDTA ..... Ethylenediaminetetraacetic acid

HGT ..... Horizontal Gene Transfer

His_6 ..... Histidine tag

HMM ..... Hidden Markov Model

HNH ..... Histidine Asparagine Histidine DNase motif

IM ..... Inner Membrane

IPTG ..... Isopropyl-β-D-thiogalactopyranoside

IUTD ..... Intrinsically Unstructured Translocation Domain

kDa ..... kiloDalton

kdo ..... 3-Deoxy-D-manno-oct-2-ulosonic acid

LPS ..... Lipopolysaccharide

MEME ..... Multiple (Expectation maximisation) for Motif Elicitation

MERV ..... Multiple Effector Translocation Vgr

MIC ..... Minimal Inhibitory Concentration

MRSA ..... Methicillin resistant *Staphylococcus aureus*

MMBL ..... Monocot Mannose Binding Lectin

NB ..... Nuclease bacteriocin

OBS ..... Omp Binding Site

OM ..... Outer Membrane

ORF ..... Open reading frame

PAAR ..... Proline Alanine Alanine Arginine

PAGE ..... Polyacrylamide gel electrophoresis

PCR ..... Polymerase chain reaction

pdb ..... Protein databank

PFB ..... Pore-forming Bacteriocin

pHMM ..... profile Hidden Markov Model

PM ..... Phenotypic Microarray

PMF ..... Proton Motive Force

PMSF ..... Phenylmethylsulphonyl fluoride

PMT ..... Polymorphic Toxin

ppHMM ..... paired profile Hidden Markov Model

PTM ..... Post-translational modification

PTS ..... Polymorphic Toxin System

R-domain ..... Receptor domain

rMLST ..... ribosomal Multi-Locus Sequence Typing

RMSD ..... Root mean squared deviation

RNA ..... Ribonucleic acid

RPS ..... Rock-paper-scissors model

rRNA ..... ribosomal Ribonucleic acid

SDS ..... Sodium dodecyl sulphate

ssDNA ..... Single Stranded DNA]

SNP ..... Single Nucleotide Polymorphism

SRA ..... Short Read Archive

T-domain ..... Translocation domain

T6SS ..... Type Six Secretion System

TBE ..... Tris-borate-EDTA buffer

TEMED ..... N, N, N', N'-tetramethyl ethylenediamine

TMHMM ..... Transmembrane Hidden Markov Model

TPS ..... Two Partner system

tRNA ..... transfer Ribonucleic acid

UPEC ..... Uropathogenic *Escherichia coli*

UTI ..... Urinary Tract Infection

WT ..... Wild Type

Contents

1 Introduction: Bacterial competition, competitive agents and colicins 25

1.1 Preface . . . 26

1.2 Social bacteria: co-operation, spite and greenbeards . . . 26

1.3 Polymorphic toxin systems . . . 28

1.4 Contact dependent inhibition . . . 29

1.4.1 Contact dependent signalling . . . 31

1.5 Type Six Secretion System . . . 32

1.5.1 T6SS structure . . . 32

1.5.2 Effectors of the T6SS . . . 34

1.5.3 Possible roles of the T6SS . . . 35

1.6 Bacteriocins . . . 36

1.7 Gram-positive bacteriocins . . . 36

1.7.1 Class I bacteriocins . . . 37

1.7.2 Class II bacteriocins . . . 37

1.7.3 Class III bacteriocins (the bacteriolysins) . . . 39

1.8 Microcins . . . 39

1.8.1 Microcin genetic organisation and function . . . 40

1.9 Tailocins - Phage tails and competition . . . 42

1.10 Lectin-like bacteriocins . . . 43

1.11 Colicins and colicin-like bacteriocins (CLBs) . . . 43

1.11.1 Overview . . . 43

1.11.2 Genetic organisation of CLBs 44
1.11.3 Colicin plasmids 45
1.11.4 Chromosomal CLBs 46
1.11.5 Colicin and CLB regulation 46
1.11.6 Colicin release by bacteriocin release protein 47
1.11.7 Lysis protein independent release 48
1.12 Colicin domain organisation 49
1.12.1 Receptor domain 49
1.12.2 Translocation domain 51
1.13 Crossing the inner membrane 52
1.14 Cytotoxic domains 55
1.14.1 DNase 55
1.14.2 rRNase 56
1.14.3 tRNase 57
1.15 Nuclease bacteriocin immunity proteins 60
1.16 Aims 63

2 Materials and Methods 64

2.1 Plasmids and strains 65
2.2 Whole Plasmid Site Directed Mutagenesis 67
2.3 Agarose gel electrophoresis 67
2.4 Transformation of Competent *E. coli* cells 68

2.5 Expression and purification of Colicin E9 mutants 68

2.5.1 Protein expression 68

2.5.2 Protein purification 69

2.5.3 Protein purification of ColE9 with engineered cysteine 69

2.5.4 Fluorescent labelling of Colicin E9 69

2.6 Calculation of protein concentration 70

2.7 Solid media cytotoxic assay 70

2.8 Liquid media killing assays 71

2.9 Chloroform release of induced plasmid 71

2.10 SDS polyacrylamide gel electrophoresis 71

2.11 Purification and visualisation of LPS 72

2.12 Microscopy using fluorescently labelled Colicin E9 72

2.13 Transposon Direction Insertion Sequencing (TraDIS) to identify genes involved in resistance 73

2.13.1 Exposing a TraDIS library to colicin E9 73

2.13.2 Analysis of insertions using the BioTraDIS pipeline 73

2.14 BIOLOG Phenotypic Microarray analysis 74

2.15 SNP analysis of resistance strains from BIOLOG analysis 75

2.16 Resistance testing of diverse UPEC strains 75

2.17 Genomic databases 76

2.18 Plasmid or chromosomal association 77

2.18.1 Assigning plasmid or chromosomal association using Plasmid Finder 77

2.18.2 Assigning plasmid or chromosomal association using a database of all known plasmid sequences 77

2.19 Pore forming pipeline 78

2.20 Identifying Ton and Tol operons in bacterial genomes 79

2.21 Clustering CLB sequences 80

2.21.1 Removing cytotoxic domains from sequences 80

2.21.2 CLANS analysis of CLB proteins 80

2.21.3 Prediction of Coiled-coil domains 81

2.21.4 Prediction of Disordered regions 81

2.22 Pangenome analysis of *E. coli* strains 81

2.22.1 Assembly of genomes from European Nucleotide Archive-Short Read Archive 81

2.22.2 Pangenome analysis using Roary 81

2.22.3 Statistical analysis of gene association 82

2.23 Translation of DNA sequences 82

2.24 Open reading frame identification 82

2.25 Identifying conserved domains using HMMer 82

2.26 Generating new HMM profiles 83

2.27 Clustering TR domains 83

2.28 Identifying pyocin receptor genes in bacterial genomes 83

2.29 Identification of motifs using MEME 83

2.30 Identifying physiological properties of cytotoxic domains 84

3 Chapter 3: Bioinformatic analysis of nuclease bacteriocins 85

3.1 Introduction 86

3.1.1 Competition systems 87

3.1.2 Colicins and colicin-like bacteriocins 87

3.1.3 Genetic organisation of CLB operons 90

3.1.4 Domain organisation of CLBs 91

3.1.5 Hidden Markov models 93

3.1.6 Bioinformatics of nuclease bacteriocins 94

3.1.7 Aims 95

3.2 Results 95

3.2.1 Developing a Hidden Markov Model based pipeline to identify nuclease bacteriocins 95

3.2.2 Non-Colicin like bacteriocin proteins identified and removed by the pipeline 102

3.2.3 Distribution of Nuclease Bacteriocin 102

3.2.4 CLBs are found in a variety of environments 103

3.2.5 Ton and Tol complexes are distributed throughout Gram negative bacteria 104

3.2.6 Chromosomal or plasmid association of NB genes 108

3.2.7 Delineation of C-terminal cytotoxic domains from CLB sequences . . 109

3.2.8 Domain organisation of NBs 111

3.2.9 Cytotoxic activity of a predicted NB from *P. aeruginosa* 116

3.2.10 Pangenome analysis of colicinogenic bacteria 118

3.2.11 Testing association between NBs and receptors 123

3.2.12 Relation between charge and cytotoxic domain 125

3.2.13 Identification of positive selection in bacterial cytotoxic domains . . 127

3.2.14 Identification of pore forming bacteriocins 128

3.2.15 Using the ppHMM pipeline to identify non-NB proteins 130

3.3 Discussion 131

4 Chapter 4: Non-specific colicin resistance of uropathogenic *E. coli* 137

4.1 Introduction 138

4.1.1 The outer membrane 138

4.1.2 Uropathogenic *E. coli* ST131 140

4.1.3 Transposon-directed insertion site sequencing (TraDIS) 141

4.1.4 Phenomic microarray and BIOLOG 142

4.1.5 Aims 143

4.2 Results 144

4.2.1 Observed resistance of UPEC ST131 144

4.2.2 Resistance is independent of immunity proteins or receptor mutations 145

4.2.3 EDTA significantly lowers the MIC of colicins for UPEC ST131 146

4.2.4 Resistance of multiple UPEC strains 147

4.2.5 TraDIS analysis reveals the role of O-antigen in non-specific colicin resistance . . . 147

4.2.6 Resistance is independent of metabolism . . . 156

4.2.7 pH and additives can drastically alter sensitivity . . . 157

4.2.8 SNP analysis of resistant strains from BIOLOG plates . . . 157

4.2.9 Non-specific colicin resistance correlates with the presence and density of O-antigen . . . 160

4.2.10 Media additives affect Colicin E9 susceptibility/tolerance of UPEC strains . . . 162

4.2.11 LPS staining revealed differences in concentration of LPS between L9 and L5 . . . 164

4.2.12 Fluorescence microscopy shows O-antigen dependent blocking of the colicin receptor . . . 165

4.3 Discussion . . . 168

4.3.1 UPEC ST131 is resistant to multiple colicins . . . 168

4.3.2 UPEC ST131 resistance is dependent on the O-antigen . . . 168

4.3.3 Non-specific colicin resistance is heavily influenced by environmental conditions . . . 169

4.3.4 O-antigen blocks colicin receptor binding . . . 170

4.3.5 Future work . . . 171

5 Chapter 5: Investigating the role the DPY motif in nuclease bacteriocins 173

5.1 Introduction . . . 174

5.1.1 Translocation across the inner membrane . . . 174

5.1.2 Aims 175

5.2 Results 176

5.2.1 Identification of a domain using MEME 176

5.2.2 Absence of the DPY in pore forming bacteriocins 177

5.2.3 Structural alignment 178

5.2.4 Scanning the Multispecies rMLST database for the Pyocin_S domain 179

5.2.5 Competition system utilising the Pyocin_S domain 180

5.2.6 Identifying NBs using the pyocin_S domain 183

5.2.7 Mutation of DPY residues 185

5.2.8 Incorporating *p*-benzoylphenylalanine into the DPY motif 191

5.3 Discussion 193

5.3.1 Characterisation of the pyocin_S domain and the DPY motif 193

5.3.2 The pyocin_S domains is primarily found within Gammaproteobacteria 193

5.3.3 The DPY motif is present in T6SS effectors 194

5.3.4 The pyocin_S domain identify novel nuclease bacteriocins 194

5.3.5 The role of the DPY domain in translocaton 195

5.3.6 Future work 196

6 Chapter 6: General Discussion 198

6.1 Colicin-like NBs are present only within Gammaproteobacteria 199

6.2 NBs as potential antimicrobials 201

6.3 Non-specific resistance mechanisms 202

6.4 Roles of NBs within the environment 203

6.5 Translocation of Nuclease Bacteriocins 203

7 Appendix 234

7.1 Genome assembly in the pubMLST Multispecies isolate database 259

7.2 Conditions tested by the BIOLOG Phenotypic microarray 259

List of Figures

1.1 Co-operation and competition of bacteria. 27

1.2 Operon structure of the two major classes of (CDI) proteins 30

1.3 Model of CDI translocation into target bacteria 31

1.4 T6SS structure 33

1.5 Operon Structure of Colicins 45

1.6 Domain organisation and structure of the archetypal colicin ColE9 in complex with its immunity protein Im9 49

1.7 Crystal structures of colicins 50

1.8 Structures of colicin R-domains interacting with receptor proteins 51

1.9 Crystal structure of Colicin E9 DNase domain bound to dsDNA 56

1.10 Crystal structure of ColE3 rRNase domain bound to the 70S ribosome . . . 57

1.11 Crystal structure of colicin D bound to its immunity protein 58

1.12 Crystal structure of ColE5 cytotoxic domain 59

1.13 Crystal structures of colicin tRNase domain and immunity protein complexes 61

1.14 Structural alignment of DNase immunity proteins 62

3.1 RPS model of CLB production promotes biodiversity 89

3.2 Operon structures of colicins 91

3.3 Overview of the pHMM strategy 93

3.4 Distribution of intergenic regions between cytotoxic HMMs and their associated immunitity HMM profiles 99

3.5 Profile pair Hidden Markov model approach 100

3.6 Nuclease bacteriocin identification and validation pipeline 101

3.7 Distribution of Nuclease bacteriocins across the bacterial kingdom 103

3.8 Distribution of different cytotoxic domains within major NB containing species104

3.9 Distribution of NBs across γ-proteobacteria 105

3.10 Proportions of CLB gene containing bacteria across different environments 106

3.11 Distribution of the Ton and Tol operons throughout the bacterial kingdom . 107

3.12 Association of NB containing contigs to plasmid sequences 109

3.13 Removing C-terminal cytotoxic domains using pHMM 110

3.14 Prediction of disorder and coils 111

3.15 NB gene clustering and domain organisations 113

3.16 Phylogenetic analysis of NB TR domains identified within *Pseudomonas spp.* 117

3.17 Cytotoxic activity of a novel predicted pyocin 118

3.18 Confoudning effects of phylogeny on identifying gene association 119

3.19 Distribution of total and conserved genes 120

3.20 Gene association as revealed by pangenome analysis 121

3.21 Association between pyocins and their receptors 124

3.22 Predictions of charge and isoelectric point for cytotoxic domains 126

3.23 Prediction of positively selected sites for HNH DNase cytotoxic domains . . 127

3.24 Prediction of TM helicies in pore forming bacteriocins 129

3.25 Crystal structure of the pore forming colicin, ColA 130

4.1 Model structure of core LPS in *E. coli* 139

4.2 Schematic of a standard TraDIS experiment 142

4.3 Killing assays of UPEC ST131 and colicin sensitive *E. coli* JM83 144

4.4 Plate killing assays of *E.coli* using a variety of colicins hijacking different proteins within the target cell 146

4.5 Increasing concentration of EDTA causes an increase in sensitivity towards colicins in UPEC ST131 146

4.4 TraDIS insertion plots 152

4.5 KEGG pathway analysis of genes identified as beinginvolved with resistance by TraDIS analysis 154

4.6 KEGG pathway analysis of genes identified as beinginvolved with resistance by TraDIS analysis 155

4.7 Example of BIOLOG output. 156

4.8 Phenotypic microarray exploring the sensitivity of UPEC ST131 with variying carbon sources 157

4.9 Phenotypic microarray plate (PM9) evaluating the effect of osmolarity on resistance 158

4.10 Phenotypic microarray plate (PM10) evaluating the effect of pH on resistance159

4.11 Plate killing assay using a variety of colicins against O-antigen mutants . . . 161

4.12 Changes in Colicin E9 susceptibility due to O-antigen 161

4.13 Addition of Tris and urea can increase sensitivity towards Col E9 in strains with an O-antigen 163

4.14 LPS straining shows a distinct difference in the density of O-antigen between L5 and L9 164

4.15 Labelling receptor proteins in the OM using flourescently labelled colicin E9 166

4.16 Fluorescence microscopy reveals O-antigen density dependent colicin receptor blocking 167

5.1 Logo plot of the DPY motif 176

5.2 Alignment of the pyocin_S domain from multiple bacteriocins 177

5.3 Location of the DPY motif in ColE9 178

5.4 DPY motif is part of a larger structural fold conserved in bacteriocin translocation domains 178

5.5 Position of the DPY motif in NBs 179

5.6 Taxonomic distribution of the DPY motif within *Gammaproteobacteria* . . . 180

5.7 Pfam domains associated with the pyocin_S domain 182

5.8 Activity of a novel pyocin cytotoxic domain 184

5.9 Alignment of a novel cytotoxic domain 185

5.10 Distribution of predicted NB genes amongst the most abundant species . . 186

5.11 Expression of three ColE9 DPY analine mutants 186

5.12 Example purification of a colicin E9Im9His_6 187

5.13 Killing activity assay by chloroform lysis 187

5.14 Killing assay using purified colicin E9 DPY mutants 188

5.15 Mutation the Y285 residue within the DPY domain 188

5.16 Killing assay for DPY mutants of ColD 189

5.17 Test expression of pyocin S2ImS2 190

5.18 Photo-activatable crosslinking of p-Benzoylphenylalanine 191

5.19 Test expression of unnatural amino acid containing ColE9 DPY mutants . . 192

List of Tables

1.1 Divisions of Class II bacteriocins 38

1.2 Colicin classification 53

1.3 Pyocin classification 54

2.1 Bacterial strains 65

2.2 Media 66

2.3 Description of plasmids used in this study 66

2.4 Reagent volumes used for whole plasmid site directed mutagenesis 67

2.5 Thermal cycling conditions 67

2.6 Phenotypic microarray plates used for BIOLOG analysis 74

2.7 Strains of uropathogenic *E. coli* used for sensitivity testing 76

2.8 Pore forming bacteriocins used as controls for TMHMM position. 79

2.9 Pfam profile HMMs used to identify proteins of the Ton and Tol operons . . 79

3.1 Nomenclature used to describe different classes of bacteriocins 86

3.2 Pfam profiles of cytotoxic domains and immunity proteins 96

3.3 Genes with a significant association to colicins 122

4.1 Colicins used to test UPEC ST131 145

4.2 MIC testing of 10 UPEC strains 147

4.3 Genes highlighted by TraDIS as providing a fitness benefit in the presence of colicins 149

4.4 Translocation protein genes identified by TraDIS 153

CHAPTER 1

Introduction: Bacterial competition, competitive agents and colicins

1.1 Preface

Bacteria exist in complex communities across a wide range of environments. Within these environments limitation of nutrients such as iron, amino acids and carbon sources drives fierce bacterial competition, necessary for survival. Bacteria can try and exclude competitors through exploitative competition; limiting a competitor by removing a nutrient from the environment, or interference competition; direct antagonism through aggression. In the bacterial kingdom interference competition can involve the production of cytotoxic proteins. Bacteria have evolved an arsenal of protein competition systems to compete with each other. Three major systems have been described: Type Six Secretion Systems (T6SS), Contact Dependent Inhibition (CDI) and the bacteriocins. Bacteriocins are unique in this group as they are the only system that does not require cell-to-cell contact and are instead diffusible elements released into the environment. Bacteriocins were first described in 1925 (Gratia, 1925), yet there are still many unanswered questions about this family of ribosomally-encoded peptides and proteins.

1.2 Social bacteria: co-operation, spite and greenbeards

Social behaviours present an intriguing problem that seems to contradict traditional Darwinian evolution; how could a trait that is costly to an individual but beneficial to others have evolved by natural selection? The answer was proposed by Hamilton in 1964. He suggested that whilst genes could increase their abundance through survival and reproduction, they could indirectly impact the transmission of copies within other individuals (Hamilton, 1964). As related individuals are more likely to share more genes, cooperation between related individuals is favoured. The inclusive fitness of an individual is therefore the sum of the direct fitness (selection of a gene by increasing the fitness of its bearer) and indirect fitness (selection of a gene by increasing the fitness of other individuals which carry it). This can be written elegantly as a simple equation (Equation 1.2.1). If the fitness cost of the trait to the actor (c) is less than the benefit to the recipient (b) times the relatedness between the two individuals (r) then the trait will be positively selected.

$$rb - c > 0 \tag{1.2.1}$$

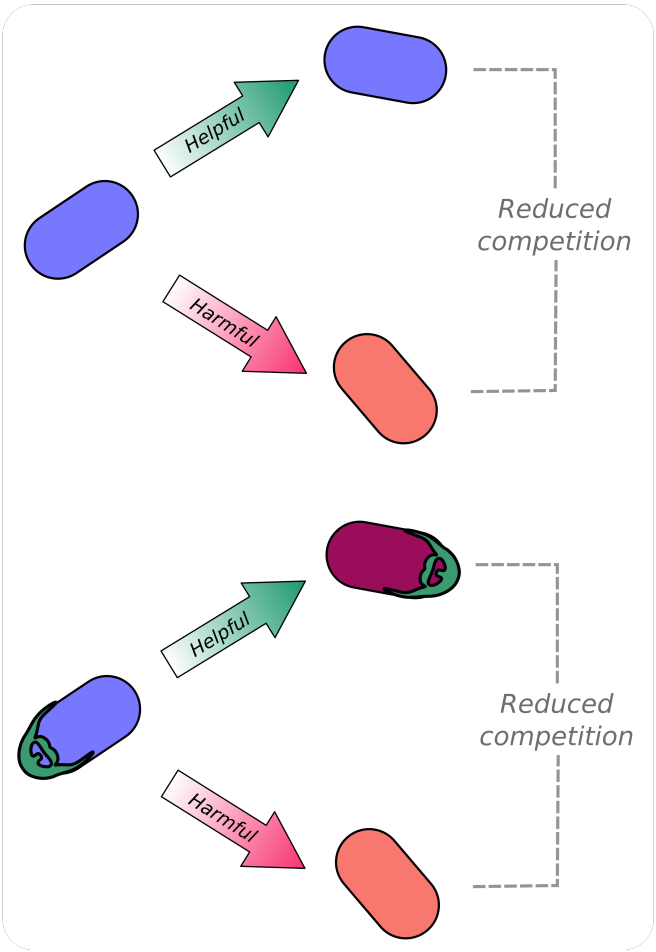


Figure 1.1: **Co-operation and competition of bacteria.** *Top*: **Kin selection** involves helping related bacteria even at a cost to yourself, as they are more likely to share your genes. *Below*: **Kind selection** or the greenbeard effect involves a recognisable trait which allows bacteria to co-operate or compete irrespective of relatedness (Adapted from (West and Gardner, 2010)).

Not only can bacteria co-operate with related individuals, they can also exhibit antagonistic behaviour towards those that are unrelated, even if this behaviour incurs a cost to the actor; a phenomenon known as spite (Hamilton, 1970). This behaviour still satisfies Hamilton's equation as *r* and *b* would both be negative in unrelated bacteria and negative benefit (Grafen, 1985). Harmful behaviour can also be explained by the addition of a second recipient of spiteful behaviour. If the actor harms an unrelated bacteria through a spiteful action, it has effectively reduced competition for a related neighbour (Lehmann *et al.*, 2006; Wilson, 1975). Bacteria exhibiting beneficial behaviour towards related bacteria (as they are more likely to share common genes) and antagonistic be-

haviour towards unrelated bacteria is called kin selection (Figure 1.1).

Indirect fitness does not necessarily involve kinship. If a single gene, or group of genetically linked genes, encodes a recognisable trait it would allow cooperation of carriers of this gene regardless of the underlying kinship (Hamilton, 1975). This idea was termed kind selection or the 'Greenbeard effect' after Dawkins, who proposed a hypothetical gene which causes individuals to grow green beards and act co-operatively towards others with green beards (Dawkins, 1976). Kind selection can act co-operatively, helping out your fellow greenbeards, or antagonistically, harming those around you without greenbeards. The dynamics of kin and kind selection have driven the selection of microbial competition even when it occurs at a cost to the cell (West and Gardner, 2010) (Figure 1.1).

1.3 Polymorphic toxin systems

Polymorphic toxins (PMT) systems are a recently described family of proteinaceous toxins characterised by modular multidomain toxins with an encoded immunity protein. Polymorphic toxins share a typical architecture with N-terminal domain/s involved in secretion and a C-terminal toxin domain (Jamet and Nassif, 2015). What makes these proteins so interesting is that individual N-terminal domains can be found in combination with different C-terminal toxin domains. Even within strains of the same species, N-terminal domains can be associated with a variety of different cytotoxic domains. Once the C-terminal toxin domain has been delivered into the target cell, cells which do not encode the cognate immunity protein will be killed whilst those with an immunity gene will survive. These toxin/immunity domains act as 'greenbeards' with bacteria that share the same toxin/immunity proteins (Danka *et al.*, 2017). Many PMT systems exist across the bacterial kingdom (Jamet and Nassif, 2015). Below, three of the main competition systems, which are all arguably members of the PMT family, are described: contact dependent inhibition (CDI), type six secretion systems (T6SS) and colicin-like bacteriocins (CLB).

1.4 Contact dependent inhibition

Contact dependent inhibition (CDI) was first observed in 2005 when a strain of *E. coli* (EC93) was reported to inhibit growth of *E. coli* K-12 (Aoki *et al.*, 2005). Unlike the diffusible colicins, this inhibition required direct contact between two cells. Inhibition was dependent on two genes, *cdiA* and *cdiB*, which are members of the two-partner secretion system (TPS), a subclass of the type V secretion systems (Vb). Introduction of the *cdiBAI* locus into *E. coli* confers the CDI phenotype. *cdiB* encodes a β-barrel outer membrane Omp85 family transport protein (Aoki *et al.*, 2005). The *cdiA* gene encodes a large protein (>3000 aa, 180-630 kDa) predicted to be β-helical that extends away from the cell surface (Aoki *et al.*, 2005). The C-terminus of CdiA encodes a toxin domain hence CdiA toxins have earned the name 'toxins on a stick' (Aoki *et al.*, 2011). CDI systems form two major classes: the *cdiBAI* system which is present in many genera of α, β and γ proteobacteria, and the *bcpAIOB* system which is only present within *Burkholderia* and a small number of closely related genera (Anderson *et al.*, 2012; Aoki *et al.*, 2010). *bcpAIOB* operons differ from the *cdiBAI* operons in three aspects. Firstly, the genes which encode β-barrel outer membrane proteins are encoded at different ends of the operon: 5' for *cdiB* and 3' for *bcpB* . Secondly, within the CdiA toxin the conserved region is separated from the C-terminal domain by a conserved, VENN or Nx(E/Q)LYN motif, for *cdiA* or *bcpA* toxins, respectively (Figure 1.2). Finally *bcpAIOB* operons encode a small lipoprotein, BcpO, which is essential for CDI activity and predicted to localise in the inner membrane (Anderson *et al.*, 2012).

BamA was identified as the receptor protein of $CdiA^{EC93}$ (Aoki *et al.*, 2008). However, while BamA is an essential protein within Gram-negative bacteria and sequence diversity between species is very low, receptor domains appear species specific (Noinaj *et al.*, 2013; Ruhe *et al.*, 2013; Voulhoux *et al.*, 2003). Similarly, $CdiA^{EC536}$ was found to utilise heterotrimeric OmpC-OmpF as a receptor (Beck *et al.*, 2016). The diversity within OmpC, particularly the extracellular loops, means $CdiA^{EC536}$ does not recognise the majority of *E. coli* even though it can target *Enterobacter cloacae* (Beck *et al.*, 2016). Recent biochemical analysis has identified the receptor binding domain as well as an additional receptor for

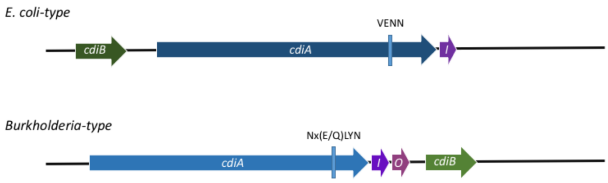


Figure 1.2: **Operon structure of the two major classes of (CDI) proteins**. *cdiA*/*bcpA* encodes a large β-helical protein that extends away from the cell. *cdiB*/*bcpB* encodes an outer membrane translocation protein. *cdiI*/*bcpI* encodes a small immunity protein. *bcpO* encodes a small lipoprotein of unknown function

CdiA^{STEC031}, the outer membrane protein Tsx (Ruhe *et al.*, 2017). The central position of the receptor binding domain (between R1358 and F1646 in CdiA^{EC93}) presents an intriguing question for translocation across the outer membrane, a process poorly understood for CDIs. For binding to the OM receptor, the CdiA toxin must be kinked to allow the binding domain to interact with its receptor (Danka *et al.*, 2017). Binding of the receptor causes translocation across the outer membrane via an unknown process. A proposed model suggests cleavage occurs here to leave a peptide containing the cytotoxic domain and a translocation domain (CdiA-CT) (Danka *et al.*, 2017). The T- (translocation) domain is predicted to interact with the inner membrane by binding a cytoplasmic membrane protein such as an ABC transporter membrane permease (Willett *et al.*, 2015). As the proton motive force (PMF) is essential for CDI activity, it is hypothesised that this brings the cytotoxic domain in close contact with the inner membrane where translocation via the PMF occurs (Ruhe *et al.*, 2014; Willett *et al.*, 2015) (Figure 1.3).

After cleavage and translocation into the cytoplasm not all cdiA-CTs are active (Diner *et al.*, 2012). CdiA-CT^{UPEC536} kills cells by cleavage of tRNA, however, it does not possess any tRNase activity *in vitro* (Diner *et al.*, 2012). tRNase activity was found to be dependant on CysK (*O*-acetylserine sulfhydrylase) activity and mutants lacking CysK were completely resistant to CdiA^{UPEC536} (Diner *et al.*, 2012). CysK also increases the affinity of CdiA-CT^{UPEC536} to its immunity protein (Johnson *et al.*, 2016; Kaundal *et al.*, 2016).

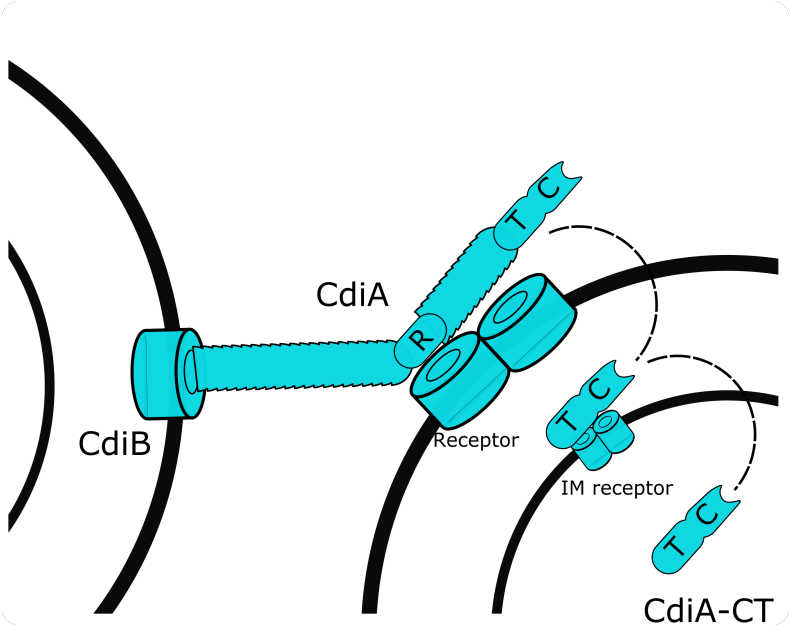


Figure 1.3: **Model of CDI translocation into target bacteria**. CdiB is an outer membrane protein which translocates the toxin CdiA out of the producing cell. The receptor binding domain interacts with an outer membrane protein of the target cell which triggers translocation through its outer membrane. Cleavage of the CdiA protein generates a peptide CdiA-CT. The T sub-domain binds proteins on the inner membrane which allows translocation across the inner membrane via the PMF.

1.4.1 Contact dependent signalling

CDI systems exhibit a clear inhibitory effect on bacteria that have the necessary receptor and translocation proteins but lack the immunity protein, highlighting their role as a competition system. TPS proteins, including CDIs, have also been shown to play a role in cell-cell adhesion and biofilm formation, which are co-operative traits (Menozzi *et al.*, 1994; Rojas *et al.*, 2002). This can be independent of any catalytic activity (Ruhe *et al.*, 2015). In *Burkholderia thailandensis*, an active BcpA-CT was shown as essential for biofilm development and induced transcriptional changes in the target cell (Anderson *et al.*, 2012; Garcia *et al.*, 2013). Interestingly, whilst catalytic activity of the toxin domain was shown to be necessary, CDI activity was not essential as expression of an immunity protein did not prevent biofilm development (Garcia *et al.*, 2013). This phenomenon was named contact-dependent signalling (CDS) (Danka *et al.*, 2017). Danka proposed four possibilities for how transcriptional changes could be induced by CDS. 1) CdiA-CT/CdiI causes non-lethal DNA nicking, 2) CdiA-CT/CdiI still retains DNA binding and acts as a transcriptional regulator, 3) CdiA-CT/CdiI binds and maybe cleaves small RNAs and 4) CdiA-CT/CdiI interacts with secondary nucleotide messengers (Danka *et al.*, 2017).

CDIs could potentially fulfil the role of both helping and harmful greenbeards, exacting CDI against those bacteria which do not contain the correct immunity protein and co-operating (CDS) with those that do (Danka *et al.*, 2017). The levels of specificity required, however, beg an interesting question about which occurs in nature. For such CDI to occur, target cells must have the correct outer membrane receptor, inner membrane receptor, a functional CysK but also lack the correct immunity protein. CDI systems are also often present on large genomic islands (Ruhe *et al.*, 2016). CDIs could therefore be acting to stabilise the island as any cells that lose the island would be vulnerable to the CDI systems of those cells around it (Ruhe *et al.*, 2016).

1.5 Type Six Secretion System

The type six secretion system (T6SS), originally identified in *Vibrio cholerae*, is broadly present in almost a quarter of Gram negative bacteria (Bingle *et al.*, 2008; Pukatzki *et al.*, 2006). T6SS forms a long needle and sheath attached to the outer membrane of the producing cell and has many structural similarities to the T4 phage contractile tail (Leiman *et al.*, 2009; Pell *et al.*, 2009). Quick contraction of the sheath (estimated 5 ms (Basler *et al.*, 2012)) forces the needle out which punctures the target cell and delivers proteinaceous effectors into prokaryotic or eukaryotic cells (Bingle *et al.*, 2008). T6SS has been associated with many roles in virulence and immunomodulation of eukaryotic cells (Chow and Mazmanian, 2010; Hachani *et al.*, 2016; Ma *et al.*, 2009), however, they also have a major role in interbacterial competition (Alcoforado Diniz and Coulthurst, 2015; MacIntyre *et al.*, 2010).

1.5.1 T6SS structure

The T6SS has both structural and mechanistic homology to the T4 phage contractile tail (Leiman *et al.*, 2009; Pell *et al.*, 2009) 1.4. Both puncture membranes through contraction of a sheath and extension of a central tube, however, whilst the phage tail is a one hit use, the T6SS can be recycled and puncture the membrane numerous times (Basler and Mekalanos, 2012). The central tube is made of hexamers of Hcp (hemolysin-coregulated

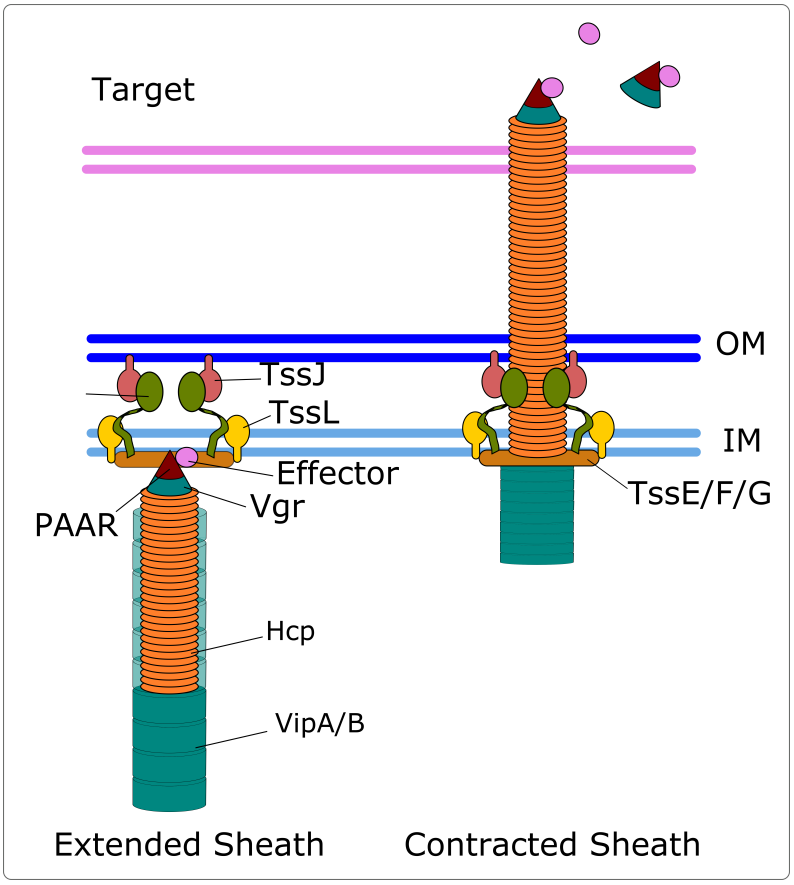


Figure 1.4: **Structure of the T6SS.** T6SS forms a sheath of stacked hexameric rings of Hcp. The top of the Hcp tube is tipped with Vgr/PAAR proteins that help penetrate the target cell. A sheath of VipA/B contracts forcing the Hcp tube our of the cell penetrating the target. Effectors fused to Vgr/PAAR are delivered into the cell.

protein) stacked together with a 40 Å diameter lumen allowing effectors to pass through (Ballister *et al.*, 2008; Mougous *et al.*, 2006; Osipiuk *et al.*, 2011). Hcp and its orthlogs have homology to gp19 of the T4 phage (Leiman *et al.*, 2009; Pell *et al.*, 2009). The Hcp tube is tipped with a trimer of Vgr (Valine-glycine-repeat) proteins which have homology to the needle tip or tail spike of the T4 tail (Hachani *et al.*, 2011; Leiman *et al.*, 2009). The Vgr tip is often associated with Proline-Alanine-Alanine-aRginine (PAAR) domain proteins which act to sharpen the tip (Shneider *et al.*, 2013). The T6SS polymerises on a baseplate of TssE, TssF, TssG, and TssK which assemble with Vgr proteins (Brunet *et al.*, 2015; Durand *et al.*, 2015). The baseplate is recruited to the membrane by the outer membrane TssJ and the inner membrane TssM and TssL proteins (Brunet *et al.*, 2015). In the extended form, the Hcp tube sits within a sheath of VipA/B (TssB/C) which has similarity to the T4 gp18 sheath (Kube *et al.*, 2014). Hcp has been shown to interact with the VipA sheath. The extended sheath can stretch the entire width of the cell but contracts to half the size which powers the delivery of effectors into the target cell (Chang *et al.*, 2017). The sheath rings collapse and more compact structures are stabilised by new charge interactions (Kube *et al.*, 2014).

1.5.2 Effectors of the T6SS

T6SSs can utilize a variety of effectors for both eukaryotic and prokaryotic targets (Hachani *et al.*, 2016; Russell *et al.*, 2013). Cargo effectors are recruited by interaction with Hcp, Vgr or PAAR proteins (Alcoforado Diniz *et al.*, 2015; Silverman *et al.*, 2013). Hcp has been shown to act as a chaperone for a variety of effectors and promote their stability in the Hcp tube (Silverman *et al.*, 2013). The same occurs for the Vgr proteins, which are often encoded adjacent to effector proteins (Barret *et al.*, 2011; Liang *et al.*, 2015). 'Evolved' effectors or specialized effectors are structural proteins of the T6SS which contain cytotoxic domains at their C-terminus (Brooks *et al.*, 2013; Ma *et al.*, 2017; Shneider *et al.*, 2013). VgrG1 from *V. cholerae* V52 has a typical Vgr domain at the N-terminal and a C-terminal domain with actin crosslinking activity (Ma *et al.*, 2009). Examples of Hcp evolved effectors have been identified in *Enterobacteriaceae* and many of the uropathogenic specific proteins (Usp) have an N-terminal Hcp domain (Ma *et al.*,

2017; Nipič *et al.*, 2013). For both these examples of evolved Hcp effectors, evidence of T6SS dependent secretion is missing. Finally, PAAR domains have been shown to also be associated with effectors, again covalently attached at the C-terminus (Shneider *et al.*, 2013). This has generated the **M**ultiple **E**ffector T**R**anslocation **V**gr (MERV) model which has been adapted to include the possible presence of Hcp evolved effectors (Ho *et al.*, 2014; Shneider *et al.*, 2013). Here, effectors can be covalently linked or chaperoned to Vgr, PAAR domain or Hcp proteins.

1.5.3 Possible roles of the T6SS

There is a large body of evidence to suggest a role in bacterial competition for the T6SS. T6SS have been shown to produce toxins that target bacteria and T6SS has been shown to increase fitness for multiple species during competition, suggesting a role in competition (Hood *et al.*, 2010; Murdoch *et al.*, 2011; Schwarz *et al.*, 2010). The evolution of *Burkholderia mallei* to an obligate pathogen, where it would face reduced competition, was associated with the loss of the T6SS (Schwarz *et al.*, 2010). The presence of multiple T6SS with differential regulation also suggests two possible roles for the T6SS. H1-T6SS of *P. aeruginosa* is repressed by quorum sensing at high cell density, suggesting it is expressed when cells using the same quorum signal are low in number, and therefore, may be used to colonise a new surface (Lesic *et al.*, 2009). A second system within *P. aeruginosa*, H2-T6SS, is induced in high cell density and could be used to defend an established colony (Lesic *et al.*, 2009). T6SS mediated killing of competitors has also been modelled to increase the benefits of public goods production. In a simplified system McNally *et. al.* used two strains with antagonistic T6SS effectors and showed rapid phase separation creating areas of identical cells. Within these occluded zones production of public goods is favourable as they will be exclusively shared with self. Indeed, a bioinformatic analysis showed a significant correlation with T6SS and secretome size (using secretome size as a measure of capacity for public goods production) (McNally *et al.*, 2017).

T6SS may also act to enforce social behaviour and prevent population collapse by cheaters (Diggle *et al.*, 2007). Cheaters are cells which benefit from co-operative behaviour without being burdened with the cost of producing the public good. As many T6SS and pub-

lic goods are co-regulated, down-regulating a public good will reduce fitness by down-regulation of the T6SS and immunity proteins (Russell *et al.*, 2014). T6SSs could have a role in cell-cell signalling in a manner analogous to the CDS described above (Russell *et al.*, 2014). Cells infected by phage could also be targeted by the T6SS. One benefit of T6SS assisted suicide is that T6SS effectors do not require cell machinery and would therefore be less vulnerable to phage inhibition (Russell *et al.*, 2014).

1.6 Bacteriocins

The Polymorphic toxins include a wide variety of proteinaceous toxins that have a modular architecture, can use different cytotoxic domains and frequently recombine toxin domains within the same N-terminal architecture. So far we have discussed two of the most widely studied contact dependent toxins. Bacteriocins are diffusible proteinaceous toxins which have cytotoxic activity against other bacteria, usually those from the same or related species. A subclass of the bacteriocins which we term the colicin-like bacteriocins (CLBs) has all the attributes which qualify them as polymorphic toxins (Jamet *et al.*, 2015). It is this subclass of bacteriocins which are the focus of this work. Before discussing CLBs it is important to mention other toxins such as the lantibiotics and microcins which are bacteriocins but are not members of the polymorphic toxins .

1.7 Gram-positive bacteriocins

Bacteriocins are a diverse family produced in all major linages and therefore, different classification schemes exist (Cotter *et al.*, 2013; Dimov *et al.*, 2005). For simplicity we use the scheme devised by Cotter and separate Gram-positive bacteriocins into three classes as discussed below. For Gram-negative bacteriocins we simply divide them into the small peptide microcins, bacteriocins that resemble phage tails and the larger protein bacteriocins.

1.7.1 Class I bacteriocins

Class I bacteriocins are characterised as small peptides, <5kDa, which undergo extensive post-translational modification to introduce unnatural amino acids such as lanthionine, β-methyllanthionine and dehydrated amino acids, earning them the name lantibiotics (Yang *et al.*, 2014). This class includes the well characterised Nisin which has been used as a food preservative for over 40 years (Delves-Broughton *et al.*, 1996; Shin *et al.*, 2016). Lantibiotics are synthesized as precursor peptides which are modified using the enzymes LanB and LanC for Class I lantibiotics, or a single enzyme LanM for Class II lantibiotcs. For Class I lantibiotics, LanB dehydrates serine and threonine residues before LanC cyclizes dehydrated residues in a reaction with cysteine. Processed lantibiotics are transported by LanT before an extracellular protease LanP cleaves the leader sequence from the peptide (Chatterjee *et al.*, 2005; Marsh *et al.*, 2010; van der Donk and Nair, 2014). With the requirement of such extensive modification and export machinery, lantibiotics are encoded as large clusters; for nisin this consists of 11 genes across three operons, *nisABTCIPRKFEG* (Willey and van der Donk, 2007). Group A lantibiotics are linear peptides that tend to kill via forming membrane pores whilst group B are globular peptides that can inhibit enzymes through binding to lipid substrates. However exceptions exist; Nisin has been shown to bind lipid II (a precursor of peptidoglycan) and insert itself into the membrane to form a pore. This high affinity binding to lipid II also acts to prevent cell wall synthesis by sequestering lipid II away from the division site (Wiedemann *et al.*, 2001). This dual mode of action may be a reason for the low resistance levels observed even after 40 years of use as a food preservative. Other potential medical uses of lantibiotics include, lacticin 3147 for MRSA treatment, Duramycin for inflammation, and Epidermin for the treatment of acne (Cotter *et al.*, 2005).

1.7.2 Class II bacteriocins

Class II bacteriocins are characterised as small, heat stable, positively charged peptides that do not undergo modification and do not contain unnatural amino acids. Class II bacteriocins are further divided into four classes (Cotter *et al.*, 2005) Table 1.1.

Table 1.1: Divisions of Class II bacteriocins

Sub-Class	Feature	Examples
IIa	Conserved YGXGV motif	Pediocin PA-1
IIb	Two peptide system	lactacin F
IIc	Cyclic peptides	Enterocin
IId	linear non-pediocin-like	epidermicin

Class IIa bacteriocins

Class IIa bacteriocins are the pediocin-like bacteriocins; well conserved linear peptides with an N-terminal 'pediocin box' Y-G-X-G-V, where X can represent any amino acid) (Eijsink *et al.*, 1998). These cationic and amphiphatic or hydrophobic peptides cause cytotoxicity through permeablizing bacterial membranes (Chen *et al.*, 1997). They are the most well studied of the class II bacteriocins due to their potential as antimicrobial agents. Pediocin Pa-1 has potential shown as a food preservative with strong activity against *Listeria* (Hechard and Sahl, 2002).

Class IIb bacteriocins

Class IIb bacteriocins represent the two peptide bacteriocins. These bacteriocins are unique as they require two peptides in equal concentration for optimal activity. Indeed many class IIb bacteriocins display no cytotoxic activity without their partner. Like the class IIa bacteriocins these peptides are also mostly cationic and amphiphatic or hydrophobic peptides which cause cell death by forming pores in the membrane (Garneau *et al.*, 2002).

Class IIc bacteriocins

Class IIc bacteriocins are cyclic peptides slightly larger than other class II bacteriocins (3.4-7.2 kDa) with their N-terminus and C-terminus covalently linked. They also form pores in the cell membrane and cause cell death by disrupting the PMF (Maqueda *et al.*,

2008). A small number of class IIc bacteriocins can contain D-amino acids (Kawai *et al.*, 2004).

Class IId bacteriocins

Class IId bacteriocins are a miscellaneous class of single linear peptides that do not represent pediocin-like bacteriocins (Cotter *et al.*, 2005). These include bacteriocins such as Lactococcin A, a 54 residue bacteriocin which acts by disrupting the PMF by pore formation (Varcamonti *et al.*, 2001).

1.7.3 Class III bacteriocins (the bacteriolysins)

Class III bacteriocins or as they are named in the Cotter nomenclature, the bacteriolysins, differ greatly from other Gram-positive bacteriocins. They are large proteins (>30 kDa) which induce cell death by degrading the bacterial cell wall. They have a recognizable domain structure but differ from the large Gram-negative bacteriocins as they have a cytotoxic domain at their N-terminus and do not encode immunity proteins (Simmonds *et al.*, 1997). Class III bacteriocins can further be subdivided into two groups, group A proteins have lytic activity whilst the group B are non-lytic proteins (Yang *et al.*, 2014).

1.8 Microcins

Gram-negative bacteriocins can be divided into two main classes: the microcins (small peptides) and large protein bacteriocins.

Microcins, named to distinguish them by size from the much larger colicin-type bacteriocins, are small (<10 kDa) hydrophobic peptides and stable to heat, pH and proteases (Duquesne *et al.*, 2007). Microcins are a diverse family and in a similar manner to the Gram-positive bacteriocins, multiple classification systems exist. Here we use the system proposed by Duquesne and colleagues where microcins are classified into two main groups: small <5 kDa peptides which undergo extensive post translational modification

(class I) and larger >5 kDa peptides (class II) (Duquesne *et al.*, 2007). Class II peptides may also be further subdivided into two categories based on disulfide bonds (class IIa) or other C-terminal post-translational modification (Duquesne *et al.*, 2007).

1.8.1 Microcin genetic organisation and function

Microcins are encoded by a series of genes across a single or multiple operons. The genetic organisation of microcins is semi-conserved across the family. The pre-microcin is encoded by a single gene, with an immunity gene normally within the same operon to prevent death of the producing cell. Genes for post-translational modification and export are also included in a single cluster with a single or multiple operons. Nomenclature for microcin genes is inconsistent, particularly between classes. For example, class I pre-microcins are encoded by a single gene *mcxA* where *x* represents the unique microcin, however in class II microcins, export genes are named *mcxA*.

Class I microcins such as MccB17 and MccC7/C51, are encoded on large (40-70 kb) mono-copy plasmids (Baquero *et al.*, 1978; Novoa *et al.*, 1986). For MccB17, the precursor is encoded by a single gene *mcbA*, and three genes are responsible for the post-translational modification (PTM) (Genilloud *et al.*, 1989). Two genes encode proteins which are involved in both secretion and offering limited immunity; *mcbE* and *mcbF*. *mcbG* encodes a protein which provides full immunity. MccB17 is highly modified, 14/43 residues are post-translationally modified (Yorgey *et al.*, 1994). Modifications form multiple five membered heteroaromatic rings along the backbone. Mutations in the outer membrane porin OmpF and its regulator OmpR provide resistance to bacteria suggesting that MccB17 uses these proteins to translocate through the outer membrane (Lavina *et al.*, 1986). The mechanisms of action for class I microcins are unique to each peptide; MccB17 appears to inhibit DNA gyrase (Heddle *et al.*, 2001) whereas MccC7 inhibits protein synthesis (Garcia-Bustos *et al.*, 1985).

Class IIa microcins are larger peptides, 5-10 kDa, which contain disulfide bonds but are not as extensively post-translationally modified. The most extensively studied class IIa microcin is the MccV (previously and more commonly known as colicin V) which is en-

coded on a large >80kb low copy number plasmid . MccV, its export and immunity is encoded by 4 genes on a 4.2 kb region which contains 2 operons (Waters and Crosa, 1991). The first operon encodes the precursor proteins and the immunity gene. Two genes in the second operon encode the well studied export system. *cvaB* encodes an ATP binding cassette whilst *cvaA* encodes an inner membrane anchored protein which extends into the periplasm. Export also relies on the outer membrane protein TolC (Gilson *et al.*, 1990). Colicin V causes cytotoxicity by depolarising the inner membrane. It binds to the inner membrane protein SdaC (Gerard *et al.*, 2005).

Class IIb are chromosomally encoded microcins encoded on large DNA fragments involving numerous genes with complex transcription. MccE492 relies on ten genes for production, immunity and export, which are transcribed in six transcriptional units (Lagos *et al.*, 2001). Many class IIb C-terminal regions are rich in serine residues. MccE492 undergoes post-translational modification at the C-terminal region to mimic siderophores (Thomas *et al.*, 2004). MccE492 is suspected to use structural mimicry to bind its receptor which can be one of three siderophore receptors; FepA, Cir or Fiu (Duquesne *et al.*, 2007; Thomas *et al.*, 2004).

Pseudomonads are not capable of the elaborate PTM required for production of the majority of microcin classes, however, a gene cluster which could provide the capacity to produce certain microcins has been identified in certain pathovars of *P. syringae* (Arnison *et al.*, 2013). Though direct production of a microcin by these pathogens has not been observed, two microcins with activity towards *P. syringae* and *P. aeruginosa* were produced an *E. coli* strain containing gene clusters (Arnison *et al.*, 2013).

The potent activity of microcins would appear to make them an ideal target for the development of future antimicrobials, however, the complex PTM needed for their activity limits their usefulness for two major reasons. Firstly, the large scale production of microcins would be difficult. Secondly the extensive PTMs, which affect activity, suggest designing of microcins to target new strains or receptors would not be trivial.

1.9 Tailocins - Phage tails and competition

Unlike the microcins and other small molecular weight bacteriocins, tailocins are high molecular weight bacteriocins which share homology to the bacteriophage tail and the T6SS (Ghequire and De Mot, 2015a; Nakayama *et al.*, 2000). Unlike the T6SS, tailocins are secreted into the environment in an SOS dependent manner (Matsui *et al.*, 1993) and cause death of the target cell by puncturing and depolarizing the cell membrane rather than the delivery of a cytotoxic effector (Uratani and Hoshino, 1984). Two major classes of tailocin exist, the rigid R-type or the flexible F-type which share homology to the P2 and lamba type phages, respectively (Nakayama *et al.*, 2000). The majority of studies on tailocins have used *Pseudomonas aeruginosa* which is capable of producing both classes (R-and F-type pyocins) (Michel-Briand and Baysse, 2002). The tailocins closely resemble phage tails but similar to the T6SS syringe, lack the head and do not contain any DNA. The structure of the tail, again, involves a central tube with a hexameric disc shape surrounded by a sheath which is attached to a baseplate (Ge *et al.*, 2015). A central spike punctures the cell wall. Tail fibers attached to the baseplate are responsible for specificity and target LPS on prey cells (Williams *et al.*, 2008). Five R-type pyocins have been identified so far, R1-5 (Michel-Briand and Baysse, 2002) and are encoded by a single gene cluster in *P. aeruginosa* (Shinomiya *et al.*, 1983)). Tail fibers are encoded by a single gene *prf15* (though a chaperone exists *prf16*). Three F-type pyocins have also been identified (Ghequire and De Mot, 2014). Tailocins mostly have narrow target specificity, though broad spectrum tailocins have been identified in *Bukholderia cenocepacia* (Yao *et al.*, 2017) and producing bacteria are resistant to the tailocin that they produce. Interestingly, by swapping the tail fibers of R-type pyocins or fusing the C-terminal with phage tail fibers allows specific targeting of tailocins to clinically important bacteria such as uropathogenic *Escherichia coli* (UPEC), *Yersinia pestis* and Shigella toxin producing *E. coli* (STEC) (Williams *et al.*, 2008).

1.10 Lectin-like bacteriocins

A relatively recently discovered group of bacteriocins are the Lectin-like bacteriocins (Parret *et al.*, 2003). These proteins contain two carbohydrate binding domains of the monocot mannose-binding lectin (MMBL) family. Lectins are a diverse range of proteins found across prokaryotes and eukaryotes which mediate a variety of biological activities through binding carbohydrates. Mannose-binding B-lectins are found in numerous monocot plants and are characterised by the MMBL domain with the QXDXNXVXY carbohydrate binding motif (Ghequire *et al.*, 2012). LlpA (Lectin-like putidacin A) was the first MMBL bacteriocin identified, isolated from a banana rhizosphere isolate of *Pseudomonas putida* BW11M1 (Parret *et al.*, 2003). *Pseudomonas* LlpAs have been shown to bind D-rhamnose oligosaccharides of the common *Pseudomonas* antigen, which explain the genus level specificity (McCaughey *et al.*, 2014). Homologous proteins within *Xanthomonas* and *Burkholderia* also display a genus level specificity. Despite numerous studies and crystal structures of full length proteins, the mode of action of lectin-like bacteriocins is still unknown.

1.11 Colicins and colicin-like bacteriocins (CLBs)

1.11.1 Overview

Colicins are proteins produced by strains of *Escherichia coli* which have cytotoxic activity against closely related strains (Cascales *et al.*, 2007). Since the identification of the first colicin in 1925 (which was later categorised as microcin MccV (Gratia, 1925)), homologous proteins have been identified in a number of enteric bacteria as well as in the *Pseudomonadales*. These bacteriocins are named by the species or genus which produce them, followed by the suffix *-cin* (e.g. colicin (*E. coli*), klebicin (*Klebsiella pneumoniae*); and pyocin (*Pseudomonas aeruginosa*)). Whilst these CLB proteins have been known for many years, colicins remain the most well studied and most of what is known about this family has been elucidated from a handful of colicins (Kleanthous, 2010a). Colicins

and colicin-like bacteriocins (CLBs) share a number of properties which make them a promising target for novel antimicrobials (Behrens *et al.*, 2017; McCaughey *et al.*, 2016a). They are extremely potent antimicrobials; for colicins a single molecule can cause death of the target cell (Shannon and Hedges, 1967). Secondly, they are highly species specific which would reduce off-target effects, such as dysbiosis, by specifically targeting a pathogen without having an effect on the host's microbiota (Kageyama *et al.*, 1996). Unlike the peptide microcins and many of the Gram-positive bacteriocins, CLBs are not post-translationally modified and have relatively simple genetic organisation. CLB proteins also have a modular structure that allows for genetic manipulation and the potential to design novel antimicrobials to target individual pathogens. The archetype colicin structure includes three domains; translocation (T), receptor-binding (R) and the cytotoxic (C) domain. CLBs hijack target cell proteins to translocate into the cell, first binding to an outer membrane receptor using the receptor domain before translocating into a cell using the translocator domain (Cascales *et al.*, 2007).

1.11.2 Genetic organisation of CLBs

Compared to other competition systems and non-CLB bacteriocins, CLBs have a simpler genetic organisation which can consist of only 2-3 genes dependent on the type of CLB and its cytotoxic activity (Riley, 1993a,b). The CLB toxin is encoded by a single gene (*cxa*) and to prevent death of the producing cell, an immunity gene is also encoded (*cxi*). For nuclease CLBs, the immunity gene is encoded immediately downstream, producing a single transcript of both toxin and immunity gene with a terminator directly after the immunity gene (Soong *et al.*, 1994; van den Elzen *et al.*, 1980) (Figure 4.5). A second internal promoter in the toxin acts to ensure there is always a greater amount of immunity protein than toxin (Chak and James, 1985; Soong *et al.*, 1994). If the cytotoxic domain acts in the periplasm (pore-forming or enzymatic), the immunity gene is not immediately downstream of the operon and in many cases encoded on the opposite strand to the toxin (Lloubes *et al.*, 1986). Colicins are divided into two groups (A and B) dependent on proteins involved in translocation and cross-resistance (Davies and Reeves, 1975a,b). For group A colicins, a small gene downstream of the immunity gene encodes for a lysis

protein which causes cell lysis and release of the bacteriocin. The lysis gene does not have a promoter so must rely on a complete read-through from the colicin promoter meaning it is always present in lower concentrations than the toxin. A termination site downstream of the immunity gene ensures that levels of lysis gene transcription are low until induction. Group B colicins do not encode a lysis gene and recent evidence suggests that these bacteriocins hijack phage lysis genes for release (Christenson and Gordon, 2009; Nedialkova *et al.*, 2015).

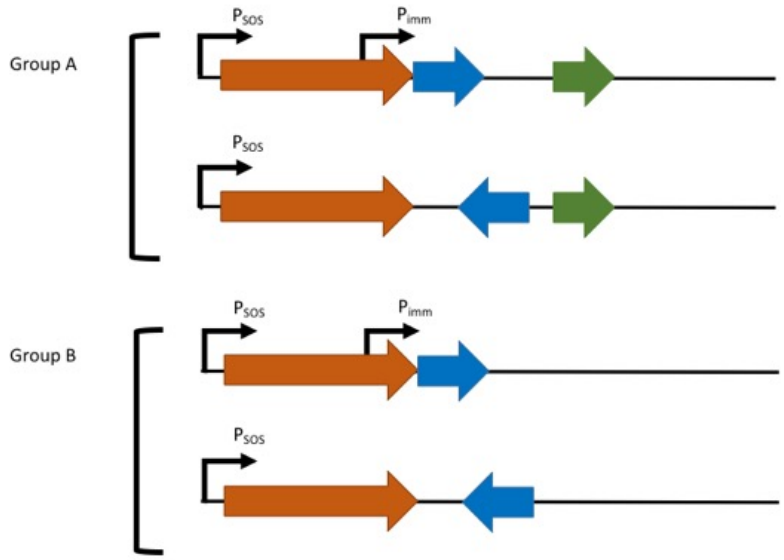


Figure 1.5: **Operon structure of group A and B colicins**. Toxin genes (*cxa*) are shown in brown, immunity genes (*imx*) in blue and lysis genes (*cxl*) in green. Promoters are labelled with arrows.

1.11.3 Colicin plasmids

Colicins are exclusively encoded on plasmids, pCol, of which two types exist. Type I colicin plasmids, which are small (6-10 kb) and mainly encode group A colicins, are present in high copy number and require a second plasmid for conjugation (Hardy *et al.*, 1973; Riley and Gordon, 1992). Type II plasmids are larger (40 kb) monocopy plasmids that encode group B bacteriocins . These plasmids are conjugative and do not require a second plasmid for HGT. Type II plasmids can also encode multiple colicin operons meaning strains would be able to produce multiple colicins. The pore-forming colicin E1 replicon is widely used in molecular biology for synthetic plasmids.

Plasmid-encoded CLBs have been identified in different *Enterobacteriaceae*. The plasmid

of DNase klebicin B from *Klebsiella* is also encoded on plasmids. Klebicin B has high similarity to colicin plasmids and is suspected to be a member of the ColE1 replicon family (Riley *et al.*, 2001).

1.11.4 Chromosomal CLBs

The pyocins of *Pseudomonas aeruginosa* differ from the colicins as they are exclusively chromosomal (Michel-Briand and Baysse, 2002). Chromosomal CLBs are also found in *Serratia marcescens* (Guasch *et al.*, 1995) It is unknown how chromosomal encoding affects the biology of pyocins though presumably pyocins genes could not be lost as easily as the colicinogenic plasmids and could not undergo horizontal gene transfer (HGT) via conjugation.

1.11.5 Colicin and CLB regulation

Colicin production is a costly trait. Not only do cells invest in protein production, release of colicin via cell lysis makes it a lethal trait to the producing cell (Herschman and Helinski, 1967). Because of this, colicin gene expression is under tight regulation by LexA and the SOS response. In *E. coli*, the SOS response is the general stress response triggered by the accumulation of ssDNA, a sign of DNA damage and stress. Under normal conditions LexA binds to the SOS box and acts as a gene repressor. As ssDNA accumulates due to damage, recA increases and causes the autocleavage of LexA and de-repression of the SOS regulon (Baharoglu and Mazel, 2014). A study of the promoter sequences of CLB sequences from enteric bacteria showed the majority of CLBs (except CloDF13) are regulated by two dual overlapping LexA boxes (Gillor *et al.*, 2008).

The pyocins of *Pseudomonas* are regulated by between 2 and 4 repeating motifs within a P box. Whilst the P box does not have sequence similarity with the SOS box, it has been found to be regulated by RecA and two regulator proteins; the positive regulator PrtN and the repressor PrtR (Matsui *et al.*, 1993; Michel-Briand and Baysse, 2002). Under non-inducing conditions PrtN is repressed by PrtR. Under SOS response inducing conditions, *recA* is induced and RecA protein cleaves the repressor PrtR, leaving PrtN free to bind

P-boxes and cause pyocin expression.

As well as control by the SOS response, colicins are also regulated by nutrient levels. The lysis gene for colicin E2 is repressed by the carbon storage regulator protein (CsrA) (Hol *et al.*, 2014). Mutation of the CsrA gene results in a 5-fold increase in lysis protein. Colicin E1 is also regulated by the cyclic AMP-cyclic AMP receptor protein complex, which is proposed to interact with LexA and delay the de-repression of colicin expression (Salles *et al.*, 1987).

Even though colicins are very tightly regulated a small sub-population, between $0.5-3\%$ of planktonic *E. coli* and $7-9\%$ of sessile populations (Majeed *et al.*, 2015), bears the burden of colicin expression (Bayramoglu *et al.*, 2017). Previous studies have shown that within a population, genes within the SOS regulon are spontaneously expressed due to random mutations triggering the SOS response (Kamenšek *et al.*, 2010). Switching between a colicin expressing and a colicin silencing state offers an overall increase in the fitness of the population in response to competition (Bayramoglu *et al.*, 2017).

1.11.6 Colicin release by bacteriocin release protein

CLBs do not have a dedicated secretion system and act as diffusible molecules in the environment (Cascales *et al.*, 2007). For group A colicins, a single gene (*cxl* or *kil*) encodes the lysis or bacteriocin release protein (BRP). BRPs are small lipoproteins (4.5 kDa) produced as precursor proteins with a 15-22 amino signal sequence (VanderWal *et al.*, 1995). The lipobox LXYC (where X is A/S/V and Y is A/G) contains the signal peptide cleavage site and is modified by acylation (Sankaran and Wu, 1994). The precursor protein signal sequence drives BRPs to the inner membrane where they are translocated to the outer leaflet of the inner membrane in a Sec-dependent or Sec-independent manner (according to the type of BRP) (Cavard, 1992; Oudega *et al.*, 1993). Here, a diglyceride is added to the sulfydryl group of the cysteine by the lipoprotein glyceryl-transferase Lgt before LspA cleaves the signal peptide leaving a lipoprotein that is then N-acylated by Lnt to produce the mature BRP protein (Cascales *et al.*, 2007). The BRP protein is thought to be recognised by the Lol system and shuttled to the outer membrane. All BRPs are found in

both the inner and outer membrane apart from the BRP for colicin N which is restricted to the inner membrane (Cole *et al.*, 1985; Pugsley, 1988). Studies on the lysis protein of ColE7 identified its BRP is located in the inner and outer membrane 45 minutes post-induction, however, the majority of protein was found in the outer membrane after 135 minutes suggesting this is due to slow processing (Lin *et al.*, 2009).

BRPs cause the phenomenon of quasilysis, a sharp decrease in optical density (OD) 60 minutes after induction without observing lysed bacteria or membrane fragments (Jakes and Model, 1979). Quasilysis is not observed in bacteria lacking the *pldA* gene which encodes the outer membrane phospholipase A (OmpLA) (Pugsley and Schwartz, 1984). OmpLA is thought to cause the formation of lysophospholipids which act as detergents and cause permeability of the outer membrane (Pugsley and Schwartz, 1984). Work on the colicin E7 BRP suggested both inner and outer membrane permeability is required for colicin release (Lin *et al.*, 2009).

1.11.7 Lysis protein independent release

As previously described, group B colicins and many CLBs do not encode a lysis protein yet they do not rely on a traditional secretion system for toxin release. Nedialkova demonstrated that release of the group B colicin, ColIb, in *Salmonella enterica* serovar Typhimurium was dramatically increased by temperate phages in a SOS dependent manner (Nedialkova *et al.*, 2015). ColIb release was observed in bacteria cured of prophage, however, bacterial genomes are highly populated with phage genes (up to 20%) and the release of ColIb could be explained by a remnant phage lysis gene. Pyocins of *Pseudomonas* also lack an identified lysis gene but it is yet to be seen how widespread phage gene hijacking is. A mathematical model established that CLBs could 'piggy back' on phage genes and become established within a population, however the system was sensitive to cheaters, cells that do not secrete colicin but benefit from the colicins produced by neighbouring bacteria (van Raay and Kerr, 2016).

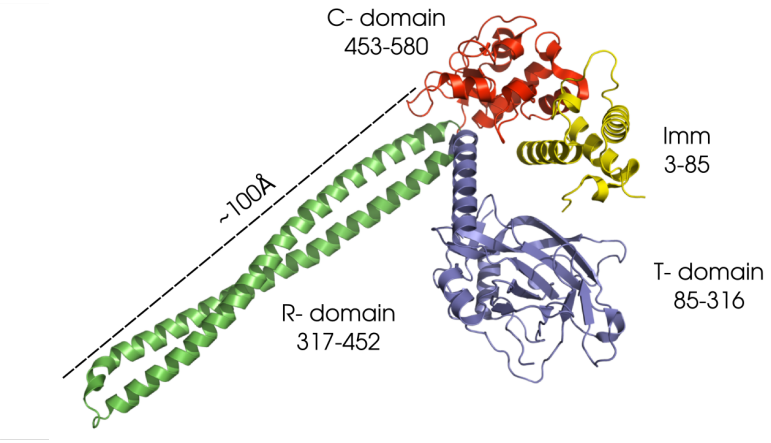


Figure 1.6: **Domain organisation and structure of the archetypal colicin Col E9 in complex with its immunity protein Im9.** Crystal structure of Colicin E9 (PDB: 5ew5) with tripartite domain organisation.

1.12 Colicin domain organisation

Gram-negative bacteria are surrounded by two membranes, the phospholipid inner membrane (IM) and the asymmetric outer membrane (OM) which consists of a phospholipid inner leaflet and a lipopolysaccharide outer leaflet. To have a cytotoxic effect on the target cell, CLBs must pass through one or both of these membrane barriers. To enter a cell, CLBs use different domains to parasatize target cell proteins. The majority of studies into the translocation of CLBs have focussed on colicins which in general have three domains; a receptor binding (R) domain, a translocation (T) domain and a cytotoxic (C) domain (Figure 1.6). While this is generally thought of as the typical colicin domain organisation, duplications (Arnold *et al.*, 2009) and merged translocation/receptor domains exist (Devanathan and Postle, 2007) (Figure 1.7)

1.12.1 Receptor domain

The first stage of CLB translocation is the binding to an outer membrane receptor on the target cell by the receptor domain. CLB receptor domains hijack a wide array of outer membrane proteins, including iron transporters (Cir, FhuA, FepA), vitamin receptors (BtuB), nucleotide receptors (Tsx) and porins (OmpF, OmpW) (Table 1.2). Colicin E9 and E3 both bind their receptor, BtuB, with a high affinity (K_d ~2 nM) through an interaction with their coiled-coil receptor domain. The high affinity binding of colicins to outer

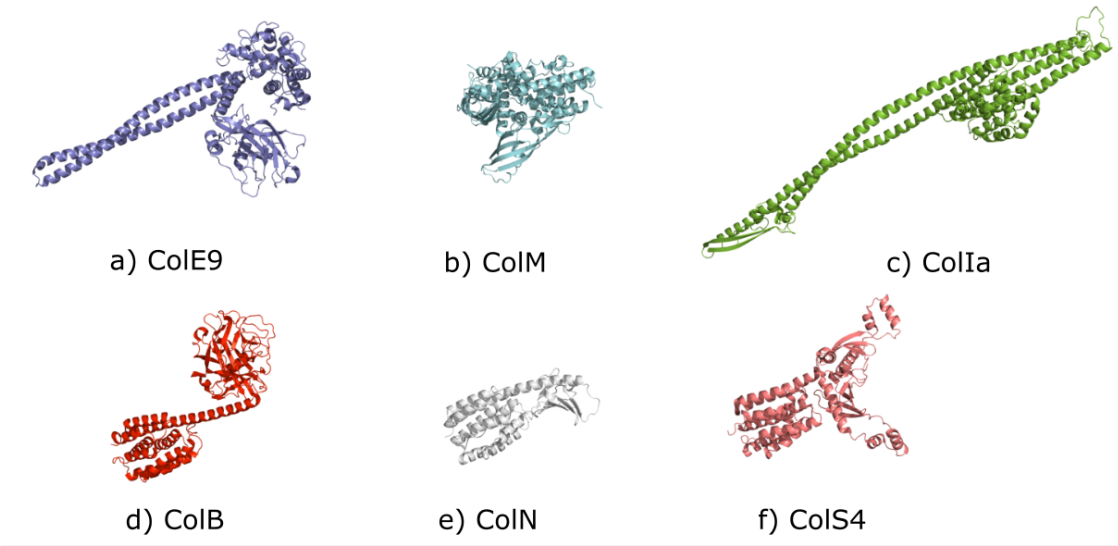


Figure 1.7: **Crystal structures of colicins** The only full length crystal structures available for CLBs come from colicins. Colicins exist in a variety of shapes and sizes from the compact ColM and the dumbbell ColB to the highly elongated ColIa. *a)* ColE9 (pdb:5ew5), *b)* ColM (pdb:3da3), *c)* ColIa (pdb:1cII), *d)* ColB (pdb:1rh1), *e)* ColN (pdb:1a87), *f)* ColS4 (pdb:3few)

membrane receptors has made them a powerful tool for investigating the organisation of the outer membrane (Rassam *et al.*, 2015).

The receptor proteins for a small number of pyocins have also been identified. It was observed that iron limited conditions increase the cytotoxic activity of five pyocins; S1 (Sano *et al.*, 1993b), S2 (Ohkawa *et al.*, 1980), S3 (Duport *et al.*, 1995), S4 (Elfarash *et al.*, 2012) and pyocin PmnH (Ghequire *et al.*, 2017), which suggested that these pyocins target iron receptors. While the receptor for pyocin S1 is still unknown, the others have been confirmed to use iron transporters, the pyoverdine receptor FpvA or the pyochelin receptor FptA (Table 1.3). Evidence also suggests that pyocins bind to the Common Polysaccharide Antigen (CPA) which helps increase pyocin concentration on the surface of the cell and enhances killing (McCaughey *et al.*, 2016b). After binding to the cell surface via the receptor, CLBs can recruit a second protein (either an additional protein or a second copy of the receptor molecule) which acts as the translocator protein.

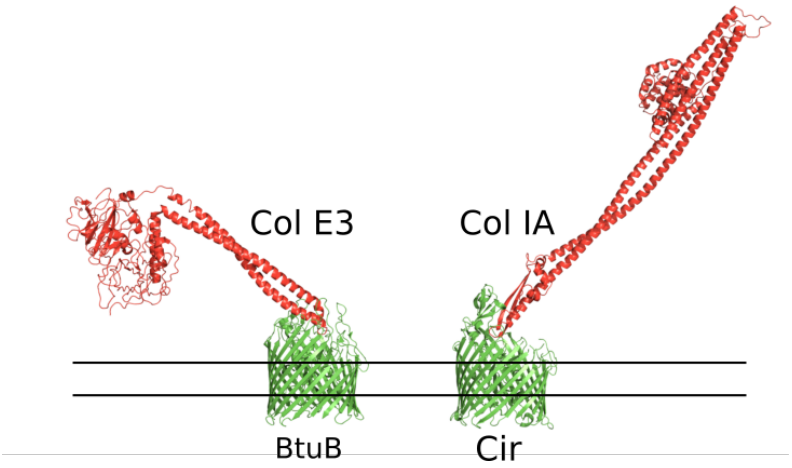


Figure 1.8: **Structures of colicin R-domains interacting with receptor proteins.** Colicins E3 (PDB: 1JCH) and Ia (PDB: 1CII) bound to their receptors, BtuB (PDB: 1UJW) and Cir (PDB: 2HDF). Both proteins interact with their receptors by elongated coiled-coil domains.

1.12.2 Translocation domain

Passage through the OM is an intriguing issue for a protein which can be ~80kDa. As the OM is not energized, CLBs rely on contacting periplasmic PMF-coupled complexes for translocation (Ton or Tol). These PMF-coupled complexes are not only responsible for colicin translocation, they also cause dissociation of the extremely high affinity immunity protein at the cell surface (Vankemmelbeke *et al.*, 2009). Though group A colicins bind receptor proteins that can translocate small molecules into the cell, they require a separate protein to pass through the OM barrier. For example, colicins E2 and E3 both bind BtuB, however, binding does not cause any significant movement of the BtuB plug domain meaning the colicins cannot pass (Kurisu *et al.*, 2003; Sharma and Cramer, 2007). The receptor domain also places the colicin at a 45° angle to the BtuB suggesting involvement of a second protein. Group A colicins recruit the abundant outer membrane porin OmpF to act as a translocation protein using the intrinsically unstructured translocation domain (IUTD), deletion of which abolishes OmpF recruitment (Housden *et al.*, 2005). The IUTD of colicin E9 possesses two OmpF binding sites (OBS) which interact with two separate pores of the OmpF trimer. Located between these two OBS sites lies a TolB binding epitope (TBE) (Housden *et al.*, 2010). This epitope was shown to compete TolB away from its binding partner Pal (Loftus *et al.*, 2006). For group B colicins, a TonB box is present within the IUTD which is thought to mimic the interaction between TonB and

the plug domains of receptors. Apart from colicin Ia, no other TonB dependent group B colicin or pyocin has shown a reliance on a second outer membrane protein as a translocator.

1.13 Crossing the inner membrane

Pore forming and lipid II degrading CLBs do not need to cross the IM and cause cytotoxicity by depolarising the IM or degrading peptidoglycan precursors. For Nuclease Bacteriocins (NBs), their target lies within the cytoplasm and the IM presents a final barrier. Crossing the inner membrane requires a cleavage step, as shown for many NBs where only the cytotoxic domain is identified in the cytoplasm (Chauleau *et al.*, 2011). The outer membrane protease OmpT has been associated with cleavage of NBs, however, cells with a OmpT deletion are still sensitive to colicins (Cavard and Lazdunski, 1990). Futher work provided evidence for a role of the inner membrane protease FtsH in colicin cleavage (Chauleau *et al.*, 2011; Walker *et al.*, 2007). It is likely that OmpT cleavage of colicins is actually a protective defence mechanism. NB cytotoxic domains have been shown to associate with the anionic phospholipids of the inner membrane in a charge dependent manner through electrostatic interactions (Walker *et al.*, 2007). As the housekeeping role of FtsH is the degradation of misfolded proteins, it is possible that the association of nuclease domains with the inner membrane causes them to be recognised by FtsH as misfolded and are targeted for degradation. The nuclease domain is then processed by FtsH and cleaved at a specific site but is not fully degraded (Mora and de Zamaroczy, 2014). How CLBs escape being fully degraded is still unknown but the ability of nuclease domains to quickly refold could serve a vital role. In tRNase cytotoxic domains, which lack the strong positive charge found in other cytotoxic domains, an additional interaction with the IM signal-peptidase, LepB is required (de Zamaroczy *et al.*, 2001). After passing the IM, NB cytotoxic domains then begin degrading their targeted nucleic acid. Unlike the cytotoxic domains of CDIs, NBs do not require any activating partner (Diner *et al.*, 2012).

Table 1.2: **Colicin classification.** Colicins are classified into two groups which utilise either the Tol (Group A) or Ton (Group B) systems.

Colicin	Group	Cytotoxic domain type	Receptor proteins	Translocation proteins
A	A	Pore	BtuB	OmpF, TolABQR
B	B	Pore	FepA	TonB, ExbBD
D	B	tRNase	FepA	TonB, ExbBD
E1	A	Pore	BtuB	TolC, TolAR
E2	A	DNase	BtuB	OmpF, TolABQR
E3	A	rRNase	BtuB	OmpF, TolABQR
E4	A	rRNase	BtuB	OmpF, TolABQR
E5	A	tRNase	BtuB	OmpF, TolABQR
E6	A	rRNase	BtuB	OmpF, TolABQR
E7	A	DNase	BtuB	OmpF, TolABQR
E8	A	DNase	BtuB	OmpF, TolABQR
E9	A	DNase	BtuB	OmpF, TolABQR
Ia	B	Pore	Cir	Cir, TonB, ExbBD
Ib	B	Pore	Cir	TonB, ExbBD
K	A	Pore	Tsx	OmpF, OmpA, TolABQR
M	B	Lipid II degradase	FhuA	TonB, ExbBD
N	A	Pore	OmpF/LPS	OmpF, TolAQR
S4	A	Pore	OmpW	OmpF, TolABQR
U	A	Pore	OmpA	OmpF, TolABQR
5	B	Pore	Tsx	TolC, TonB, ExbBD
10	B	Pore	Tsx	TolC, TonB, ExbBD

Table 1.3: **Pyocin classification.** Classification of colicin sequences, the cytotoxic domains they utilize and the receptor proteins they target (Ghequire and De Mot, 2014)(* indicates receptor protein is predicted from sequence homology).

Pyocin	Cytotoxic domain	N-terminal similarity	Receptor proteins
S1	HNH DNase	-	-
S2	HNH DNase	-	FpvAI
S3	Non-HNH DNase	-	FpvAII
S4	tRNase	-	FpvAI
S5	Pore	-	FptA
S6	rRNase	S1	-
S7	rRNase	S2	FpvAI*
S8	HNH-DNase	AP41	-
S9	HNH-DNase	-	-
S11	tRNase	S2	FpvAI
S12	tRNase	S3	FpvAII
S13	tRNase	S1	-
AP41	HNH DNase	-	-
M1	Lipid II degradation	-	FiuA
M4	Lipid II degradation	-	FiuA

1.14 Cytotoxic domains

Cytotoxic domains of CLBs cause death of the target cell by degrading nucleic acid in the cytoplasm, depolarising the inner membrane, or degrading the precursors for peptidoglycan. Nuclease CLBs can target three types of nucleic acid: DNA, rRNA or tRNA. NB cytotoxic domains and their modes of action are reviewed below.

1.14.1 DNase

The majority of NBs which degrade DNA utilize a HNH-motif nuclease domain. These nucleases are structurally similar to the $\beta\beta\alpha$ motif and common across all kingdoms of life. They are found in the active sites of multiple enzymes such as: Type II restriction enzymes, homing endonucleases, mismatch repair enzymes and CRISPR/Cas-9 (Jiang *et al.*, 2013; Keeble and Kleanthous, 2005; Shen *et al.*, 2004). HNH utilising NBs contain a ~30 amino acid motif ($HHX_{14}NX_8HX_3H$) at the catalytic core. Histidines in the core are involved in co-ordinating a single metal ion. Activity studies, crystal structures and X-ray fluorescence spectroscopy have shown the site can be occupied by either Mg^{2+} for colicin E9 or Ni^{2+} for pyocin AP41. The identity of the cation can change levels of activity and specificity for ssDNA or dsDNA (Joshi *et al.*, 2015; Pommer *et al.*, 2001). The HNH DNase motif binds to the minor groove of DNA causing distortion of the DNA backbone. This distortion positions a scissile phosphodiester bond near the divalent cation (Hsia *et al.*, 2004). It is unknown how DNA nicking causes cell death as bacteria have repair systems induced by the SOS response. It is possible that the nicking of HNH DNases overcomes the repair mechanisms or it could cause random lethal dsDNA breaks.

Though most DNase NBs use the HNH motif, a non-HNH DNase has been identified in pyocin S3 and is present in T6SS effectors in *K. pneumoniae* (Duport *et al.*, 1995). How these domains cause cleavage of DNA is unknown.

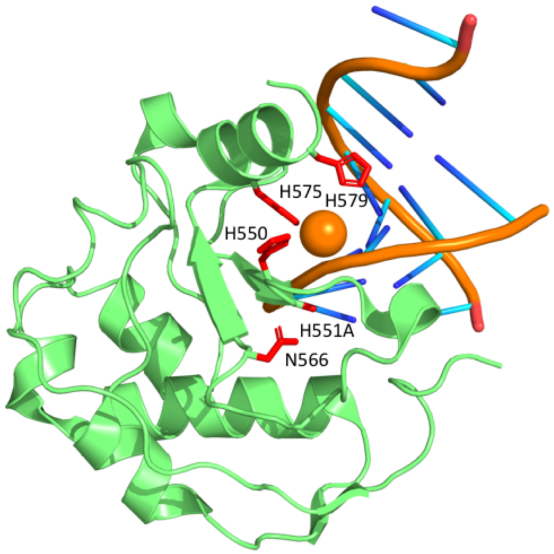


Figure 1.9: **Crystal structure of Colicin E9 DNase domain bound to dsDNA.** ColE9 DNase domain with H551A mutation. HNH motif residues are shown in red. Three histidines coordinate the Mg^{2+} ion (PDB:1V14). Residue numbering is for full length ColE9.

1.14.2 rRNase

The rRNase mechanisms of CLB have been elucidated using colicin E3, the first of the rRNase colicins to be isolated. ColE3 cleaves a bond between A1493/G1494 in helix 44 of the 16S rRNA at the A-site of the ribosome (Bowman *et al.*, 1971). A crystal structure of the ColE3 rRNase domain bound to the 70S ribosome of *Thermus thermophilus* (PDB: 4V5K) (Ng *et al.*, 2010) showed ColE3 binding and suggested a mechanism where ColE3 residues His513 and Glu517 form an acid-base pair inducing 2' OH-induced hydrolysis, stablised by Asp55 and Glu60 (Ng *et al.*, 2010). An additional residue, Arg90, was also shown to significantly reduce activity when mutated, however, it appears to be uninvolved in the cleavage step (Walker *et al.*, 2004) Cleavage of 16S rRNA decreased the stability of codon-recognition complexes in the elongation step and greatly increases the rate of tRNA rejection (Ng *et al.*, 2010). The decrease in rate of initiation and elongation causes the build up of short truncated peptides which ultimately leads to cell death.

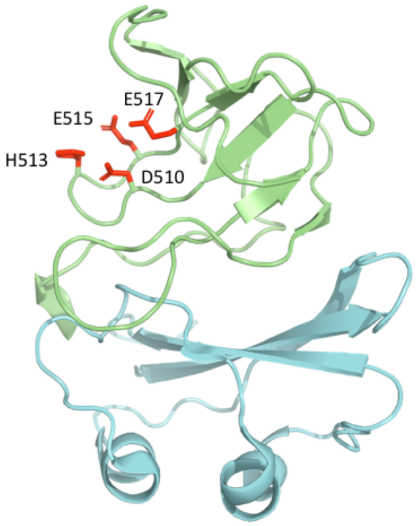


Figure 1.10: **Crystal structure of the ColE3 rRNase domain (green) bound to its immunity protein (blue).** Residues predicted to be involved in cytotoxic activity are shown in red. The active site is on the opposite face to the immunity binding region. (PDB:1JCH)

1.14.3 tRNase

Two classes of tRNase cytotoxic domains have been reported in NBs, both cleave a single phosphodiester bond in specific tRNAs to prevent protein synthesis and cause death of the cell. ColE5-like tRNases (also found in pyocin S4) cleave the Q(G)34pU35p dinucleotide in the anticodon triplet of $tRNA^{Tyr}$, $tRNA^{His}$, $tRNA^{Asn}$ and $tRNA^{Asp}$ (Ogawa *et al.*, 2006). ColD-like tRNases (also found in pyocin S11-13 (SD1-SD3 (McCaughey *et al.*, 2016b))) cleave three of the four $tRNA^{Arg}$ (Tomita *et al.*, 2000). All ribonucleases, apart from colicin E5, contain at lease one histidine residue in the catalytic core. Colicin D cytotoxic domains contains two histidines and mutation of His611Tyr abolished activity (Tomita *et al.*, 2000). His611 may form a triad with Lys608, Lys610 or Ser677 (Yajima *et al.*, 2004).

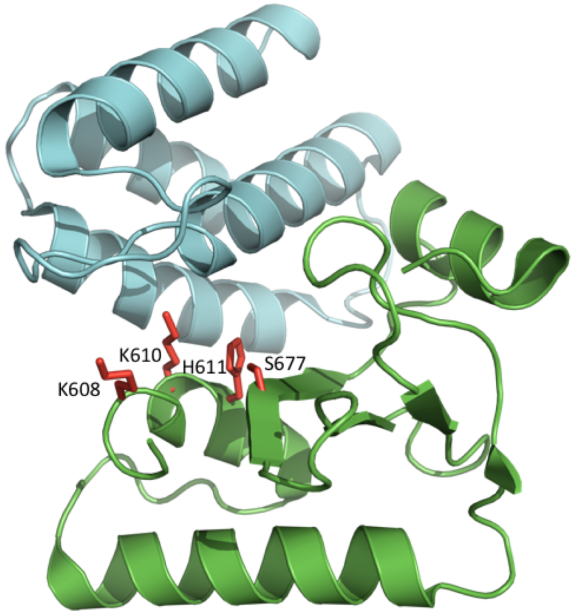


Figure 1.11: **Crystal structure of colicin D bound to its immunity protein** Colicin D cytotoxic domain with residues predicted to be involved in catalysis (Yajima *et al.*, 2004). (PDB: 1V74)

Colicin E5 is unique amongst tRNases as it has no histidines involved in catalysis. Alanine scanning of basic/acidic residues in the cytotoxic domain suggested two Asp/Arg pairs are involved in recognition (D97/R99) and cleavage (D46/R48) of tRNA molecules (Lin *et al.*, 2005).

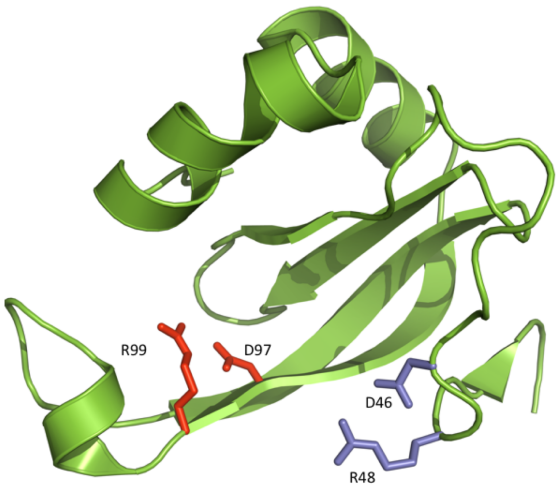
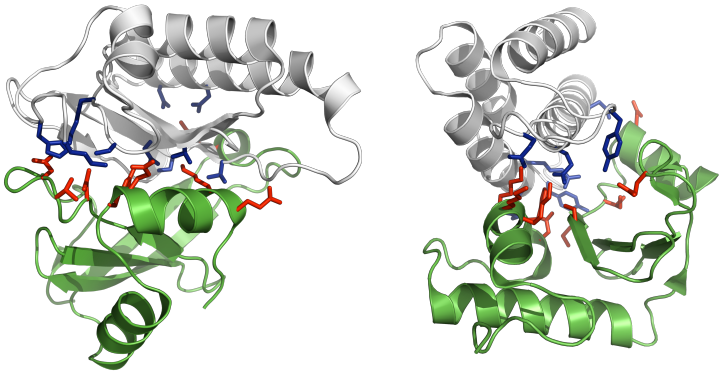


Figure 1.12: **Crystal structure of ColE5 cytotoxic domain** tRNase cytotoxic domain of colicin E5 (PDB:2A8K). Alanine scanning proposed two Asp/Arg pairs that are involved in recognition (blue) or catalysis (red) of tRNA molecules.

1.15 Nuclease bacteriocin immunity proteins

Nuclease bacteriocins are potent toxins. To prevent death of the NB producing cell an immunity protein is encoded which binds the cytotoxic domain with extremely high affinity. Binding between cytotoxic domains and immunity proteins has been widely studied as a model for specificity in protein-protein interactions (Kuhlmann *et al.*, 2000; Li *et al.*, 2004; Papadakos *et al.*, 2012). Interestingly, immunity proteins bind the cytotoxic domain at an exosite away from the active site (Kleanthous and Walker, 2001). Exosite binding of immunity protein blocks substate binding. Such a mechanism is rarely observed as it leaves an exposed catalytic site which could retain activity. However, exosite binding allows for the evolution of specificity in the immunity binding region without introducing mutations in the active site that could impact activity. Both tRNase (ColE5 and ColD) immunity proteins mimic the tRNA substrate and bind directly to the active site to block binding of the tRNA substrate (Graille *et al.*, 2004; Luna-Chavez *et al.*, 2006). The ColD/ImmD interface has a buried surface area of 1920 Å^2 and the interaction is largely driven by the complementary charges of a positive patch on the ColD tRNase and a negative patch on the immunity which form three of seven salt bridges. Mutation of residues involved in the salt bridges completely abolishes protection by ImmD (Graille *et al.*, 2004). A second region in the α2 helix also abolished binding when mutated, R57 and F52 which interact with R651 and W679, respectively (Graille *et al.*, 2004). Similarly the immunity/ColE5 cytotoxic domain interaction relies on complementary charges with a buried surface area of 2425 Å^2 (Figure 1.13).

DNase immunity proteins have been the most intensely studied of the CLB immunity interactions (Kleanthous and Walker, 2001; Papadakos *et al.*, 2012) (Figure 1.14). A dual recognition mechanism has been suggested for DNase/Imm complexes (Kleanthous and Walker, 2001). DNases from colicin and pyocins contain conserved aromatic residues within αIII before a helix breaking proline and mutation of these aromatic residues decreases immunity binding (Joshi *et al.*, 2015; Kuhlmann *et al.*, 2000). This contact is conserved across all CLB DNase/Imm complexes. Specificity within DNase/Imm complexes arises from interactions with the αII helix. Cognate (E9-Im9) and non-cognate (E9-Im2)



a) Colicin E5 b) Colicin D

Figure 1.13: **Crystal structures of colicin tRNase domain and immunity protein complexes.** tRNase immunity proteins bind directly to the active site using complementary charges. tRNase domains are shown in green with residues involved in the interaction coloured red. Immunity proteins are shown in white with interacting residues coloured in blue, *a*) Colicin E5 (PDB: 2FHZ) *b*) Colicin D (PDB: 1v74)

complexes have been studied. Non-cognate complexes have a several orders of magnitude difference in binding (10^{-4}-10^{-10}M for non-cognate and 10^{-14}M for cognate complexes). These differences in affinity are attributed to frustration of non-optimal contacts (Meenan *et al.*, 2010).

The polymorphism observed in the αII helix is due to strong selective pressure. A two step evolution process has been proposed to explain the polymorphism and how new colicin/immunity pairs are generated (Riley and Wertz, 2002). First, a mutation occurs within the immunity protein which results in broadened immunity specificity. This increases the fitness of the producing cell as it is now immune to the ancestral toxin and similar CLBs. This is then followed by a complementary mutation in the CLB cytotoxic domain, meaning the ancestral immunity protein is no longer recognised. This new CLB/Imm pair will rapidly displace the bacteria producing the ancestral CLB. This new CLB/Imm pairing will of course ultimately be replaced by a further novel CLB. Binding of the immunity at an exosite allows this diversification without affecting the catalytic function of the enzyme.

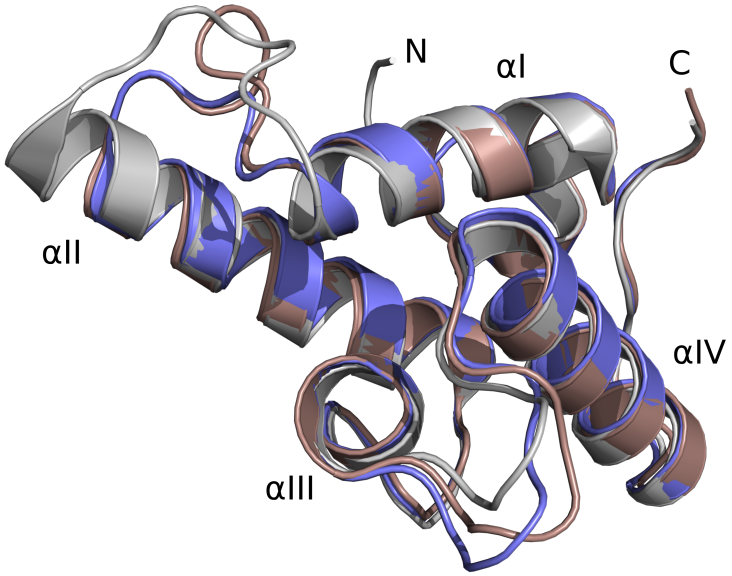


Figure 1.14: **Structural alignment of DNase immunity proteins.** Immunity proteins for three DNase immunity proteins: ColE9 (blue, PDB:1EMV), Pyocin S2 (pink, PDB:4QKO) and Pyocin S4 (grey, PDB:4UHP)

1.16 Aims

Although colicins have been studied for over 90 years many questions still remain. As the need for new antibiotics becomes ever more pressing, these unknowns are hurdles limiting the use of colicin-like bacteriocins as antibiotics.

As CLBs have a narrow spectrum of activity and tend to only have activity against closely related species, it is important to understand the distribution of bacteria which encode CLB genes and the diversity of the CLB sequences. Chapter 3 of this work will aim to identify the distribution and diversity of CLB proteins using a bespoke high-throughput bioinformtic pipeline, before using the data generated from this pipeline to answer questions about the biology of the CLB family.

Resistance to CLBs is often thought to solely rely on the presence of an immunity protein or the absence of a receptor protein. However, resistance has been observed independent of those mechanisms. If CLBs were ever to be used in the clinic it is important to understand the mechanisms behind this resistance and whether or not it is a consequence of the laboratory set-up. A mixture of high throughput genetic/phenotypic analysis and microscopy are applied to elucidate the underlying mechanism of resistance.

Finally, the work will test the hypothesis generated by the bioinformatic work to address questions on the mode of action of CLB proteins, particularly the translocation of CLBs across the inner membrane.

CHAPTER 2

Materials and Methods

2.1 Plasmids and strains

Table 2.1: **Bacterial strains** Details of the bacterial strains used in this study (***** (Browning *et al.*, 2013))

Species	**Strain Name**	**Genotype**
E. coli	BL21 (DE3)	*fhuA2* [lon] *ompT gal (λ)* [dcm] Δ*hsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*
E. coli	NEB 5α	*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 ϕ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*
E. coli	MG1655	*F- lambda- ilvG- rfb-50 rph-1*
E. coli	JM83	*F- ara Δ(lac-proAB) rpsL (StrR)[ϕ80 dlacΔ(lacZ)M15] thi*
E. coli	L9*	MG1655 pJP5603/wbbL inserted into rfb gene cluster
E. coli	L5*	MG1655 pJP5603/wbbL inserted into rfb gene cluster
E. coli	UPEC ST131 (NCTC 13441)	CTX-M-15 ESBL
P. aeruginosa	PAO1	
P. aeruginosa	PA UCBPP- PA14	
P. aeruginosa	PA17	

Table 2.2: **Media**. Details of key media used in this study

Name	Composition	Usage
LB	170 mM NaCl, 5 g/l Yeast extract, 10 g/l Tryptone pH 7.2	Used for generic growth purposes. Is the medium used in all experiments unless stated otherwise
M9	43 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.5 mM NaCl, 18 mM NH_4Cl pH 7.2	Used in liquid killing assays
MMTE	MOPS Minimal media KPO_4, 1 mM EDTA, 100 mM Tris-Hcl pH 8 with 0.2% glucose	Used in the growth of TraDIS libraries
MMT	MOPS Minimal media KPO_4, 100 mM Tris-Hcl pH 8 with 0.2% glucose	Used in the growth of Tradis libraries and to calculate the MIC of UPEC strains
LB+Urea	LB, 150 mM Urea pH 7.2	Used in liquid growth killing assays
LB+Tris	LB, 150 mM Tris-HCl pH 7.2	Used in liquid growth killing assays
LB+NaCl	LB + an additional 150 mM NaCl pH 7.2	Used in liquid growth killing assays

Table 2.3: **Description of plasmids used in this study**

Name	Description
pCS4	pET21d containing full length Colicin E9 Im9 $(His)_6$ with Amp^R
pCGS1	pCS4 with ColE9 R185A
pCGS2	pCS4 with ColE9 D270A
pCGS3	pCS4 with ColE9 Y285A
pCGS4	pCS4 with ColE9 R185AD270A
pCGS5	pCS4 with ColE9 R185AY285A
pCGS5	pCS4 with ColE9 D270AY285A
pCGS6	pCS4 with ColE9 K469CH551A
pCGS7	pCS4 with ColE9 K469CY285A
pEE01	Colicin B $(His)_6$
pColD	Colicin D
pPW06	pET21 with full length PyocinS2 Im2

2.2 Whole Plasmid Site Directed Mutagenesis

Single residue mutations were introduced using Whole Plasmid Site Directed Mutagenesis and mutagenic primers. Reaction mixtures were as described in Table 2.4 and thermal cycling was performed as described in Table 2.5 using the eppendorf Mastercycler Personal.

Table 2.4: **Reagent volumes used for whole plasmid site directed mutagenesis**

Reagent	**Volume (µl)**
MilliQ water	33
5x HF buffer	10
Plasmid template 100ng/µl	1
F primer 125ng/µl	1
R primer 125ng/µl	1
pfu turbo polymerase	1
dNTP (10mM)	1
DMSO	2
Total	50

Table 2.5: **Thermal cycling conditions.** Thermal cycling conditions used for who plasmid mutagenesis. Steps 2-4 were repeated for 16-18 times dependent on yield.

Step	**Temperature** °C	**Time (min)**
1	98	2
2 Repeated	98	1
3 Repeated	55	30
4 Repeated	72	4
5	72	5

2.3 Agarose gel electrophoresis

Electrophoresis grade agarose (Invitrogen) was dissolved in 0.5x Tris-Borate-EDTA (TBE) buffer to give a 0.5% (w/v) agarose gel. SYBR safe DNA gel stain was added (5µl in 50ml gel) and gels were run at a constant voltage of 80V for approximately 40 min or until

appropriate resolution has been achieved.

2.4 Transformation of Competent *E. coli* cells

10 µl of competent NEB 5α or NEB BL21 (DE3) cells were thawed on ice for 5 minutes. 2 µl (~100 ng/µl) of plasmid DNA was added to cells and gently mixed. Cells were then kept on ice for 30 minutes before heat shock for 25s or 60s for NEB 5α or NEB BL21 cells, respectively. Cells were then placed on ice for a further 5 minutes before recovery in SOC medium at 37 °C for one hour. Cells were then spread onto agar containing the appropriate antibiotic and grown overnight. Colonies that grew on the selective media were cultured in 10 ml of LB broth containing appropriate antibiotic. Plasmids extracted using the QIAprep Spin Miniprep kits. Total DNA concentration was calculated using the Nanodrop ND1000 Spectrophotometer. Plasmid sequences were validated by sequencing using Genewiz express sequencing and analysed using snapgene (http://www.snapgene.com).

2.5 Expression and purification of Colicin E9 mutants

2.5.1 Protein expression

E. coli BL21 (DE3) cells transformed with pET21d plasmids containing colicin E9 mutants were cultured overnight at 37 °C with shaking in 50 ml LB media with 100µM ampicillin. These cultures were used to inoculate (1:100) 5 l of LB media with 100 µM ampicillin. Cells were incubated for approximately 2.5 hours to reach an OD_{600} ~0.6 before induction by addition of 1mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were incubated for a further 3 hours before pelleting by centrifugation (4500 rpm (Beckman JLA-8.1000 rotor) for 15 minutes at 4 °C). Cells were resuspended in 35 ml of phosphate buffer pH 7 before pelleting by centrifugation (Sigma 19776 rotor 4,500rpm for 20 minutes at 4 °C) and frozen at -20 °C.

2.5.2 Protein purification

Cell pellets were resuspended in 35 ml binding buffer (20 mM Tris-HCl, 500mM NaCl, 5mM Imidazole pH 7.5) with 1mM of the protease inhibitor phenylmethane sulfonyl fluoride (PMSF). Cells were sonicated on ice for 2 x 90s with mixing by inversion in between (3 second pulses / 7 seconds rest). Insoluble cell debris was removed through centrifugation (Sigma 19776 rotor 12500 rpm for 30 minutes at 4 °C) and supernatant passed through a 0.45 µm filter. Protein was purified using a 5 ml HisTrap HP column (GE Healthcare) equilibrated in binding buffer. Bound protein was eluted using a 0 to 250 mM imidazole gradient over 10 column volumes. 1 ml fractions were collected and visualised for protein on a 15% SDS-PAGE gel as described in 2.10. Fractions that contained protein of interest were pooled and dialysed overnight against 25 mM Tris-HCl pH 7.5, 150 mM NaCl using a 12-14 kDa dialysis membrane at 4 °C. Protein was further purified by size exclusion chromatography using a HiLoad 26/60 Superdex 200 column (GE Healthcare) equilibrated in 25 mM Tris-HCl pH 7.5, 150 mM NaCl. Fractions were examined again using a 15% SDS-PAGE gel and fractions which contained protein of interest were pooled.

2.5.3 Protein purification of ColE9 with engineered cysteine

ColE9 H551A K469C in complex with Im9His$_6$ was purified as described in 2.5.2 with the exception of 10 mM DTT in all buffers after the HisTrap column to prevent dimerisation due to intermolecular disulfide bond formation.

2.5.4 Fluorescent labelling of Colicin E9

Colicin E9 H551A K469C was labelled with AlexaFluor^{TM} 647 maleimide (AF647) (Rassam *et al.*, 2015). AF647 fluorophore with an exposed maleimide group is covalently linked to the exposed sulfhydryl of the engineered cysteine to form a thioether bond and a stable conjugate. Colicin E9 H551A K469C in 10mM DTT was loaded onto a 5 ml HiTrap desalting column (GE Healthcare) to remove DTT and eluted in 0.5 ml fractions. Fractions which contained protein were pooled and concentrated to ~1ml. AF647 was added to

the protein at a molar ratio of 10:1 and incubated at room temperature for 1 hour with rotary inversion. The reaction was quenched with addition of 2mM DTT. Excess dye was removed by passing the protein through a Superdex 75 10/300 column (GE Healthcare). Fractions were visualised on a 15% SDS-PAGE gel and fractions which contained protein pooled.

2.6 Calculation of protein concentration

Protein concentration was calculated using a biophotometer (Eppendorf) measuring A_{280nm} and using the Beer-Lambert Law (Equation 2.6.1, where *A* = Absorbance at 280 nm, *L* = Path length, C = Concentration, ϵ = Extinction coefficient).

$$A = \epsilon LC \tag{2.6.1}$$

Extinction co-efficients for proteins were predicted using Equation 2.6.2 with coefficients Cys: 125 $M^{-1}cm^{-1}$, Tyr: 1490 $M^{-1}cm^{-1}$, Trp: 5500 $M^{-1}cm^{-1}$ (Pace *et al.*, 1995).

$$\epsilon_{280} = (N_{cys} \times \epsilon_{280nm\ of\ cys}) + (N_{tyr} \times \epsilon_{280nm\ of\ tyr}) + (N_{trp} \times \epsilon_{280nm\ of\ trp}) \tag{2.6.2}$$

2.7 Solid media cytotoxic assay

Killing assays were performed by inoculating soft LB agar (0.7%) with *E. coli* and overlaid on LB agar. Serial dilutions (3 fold from 10 µM to 57 pM) were prepared using 20mM Tris-HCl pH 7 and 2 µl spotted onto the agar. Plates were incubated overnight at 37 °C and scored on presence of a zone of clearance in the soft LB agar.

2.8 Liquid media killing assays

Cultures of the strains *E. coli* MG1655, *E. coli* L5, *E. coli* L9, and UPEC ST131 were cultured overnight in LB. Overnight cultures were used to inoculate (1:100) 10 ml fresh medium supplemented as appropriate and incubated at 37 °C (Table 2.2). When cultures reached OD_{600} ~0.4 they were diluted (1:3, 200µl volume) in a flat bottomed 96 plate in triplicate containing either no colicin or 50nM ColE9Im9$(His)_6$. Growth was measured as an increase in OD_{600} using the Clariostar High performance Microplate reader.

2.9 Chloroform release of induced plasmid

E. coli BL21 cells transformed with a plasmid containing an inducible colicin were stabbed on to an LB agar plate containing selective antibiotic and IPTG. Cells were incubated overnight at 37 °C. Cells were lysed by placing the plate upside-down onto chloroform soaked paper for 15 minutes. Lysed cells were overlayed with sensitive *E .coli* JM83 cells in soft agar and cytotoxic activity was detected by zones of clearance.

2.10 SDS polyacrylamide gel electrophoresis

Samples were boiled for 5 minutes in 4x sample buffer to give a final concentration 50mM Tris-HCl pH 6.8, 2.5% β-mercaptoethanol, 2% SDS, 0.1% bromophenol blue, 10% glycerol. Between 5-10 µl of boiled sample loaded onto a 12 % SDS page gel (12 % Bis-Acrylamide, 375mM Tris-HCl pH 8.8, 0.1% SDS, 0.08 % ammonium persulfate (APS), 0.13%, N,N,N',N'-Tetramethylethane-1,2-diamine (TEMED)) with a 5 % stacking gel (5 % Bis-Acrylamide, 125mM Tris-HCl pH 6.8, 0.1 % SDS, 0.06 % APS, 0.13 % TEMED).

Gels were run at 30mA for 30 minutes in a buffer of 23mM Tris HCl, 193 mM glycine, 10% SDS. Molecular marker of the range 30kDa - 200kDa (Sigma) was used to approximate the weight of separated proteins. Proteins were stained using 0.2% Comassie brilliant

blue in 10% (v/v) acetic acid and 50% (v/v) ethanol for approximately 40 minutes before destaining in a 10 % (v/v) acetic acid, 10% (v/v) ethanol solution.

2.11 Purification and visualisation of LPS

Cells were grown in LB media with or without additives (Table 2.2) until late log phase OD_{600} 1. Cells were pelleted and resuspended in PBS to give a cell suspension of OD_{600} 1.5. Cell pellet from 500 µl of sample was resuspended in 100 µl of distilled water with an equal volume of 2x SDS dye (50mM Tris-HCl pH 8, 2.5% β-mercaptoethanol, 2 % SDS, 0.1 % bromophenol blue, 10 % glycerol) and boiled for 10 minutes before incubation overnight at 37 °C with 50 µg/ml proteinase K. After incubation, an additional volume of proteinase K equal to the first was added and incubated at 55 °C for 3 hr. 5 µl of sample was analysed on a Novex^{TM} 4-20% glycine gel (Invitrogen). LPS from *E. coli* serotype O55:B5 was used as a marker. LPS bands were visualised using Pro-Q ® Emerald 300 Lipopolysaccharide Gel Stain Kit.

2.12 Microscopy using fluorescently labelled Colicin E9

E. coli K-12 MG1655, *E. coli* DFB1655 L5, *E. coli* DFB1655 L9 and UPEC ST131 were incubated overnight in LB media at 37 °C. Overnight cultures were used to inoculate (1:100) LB media with additives (Table 2.2). Cells were incubated at 37 °C until they reached OD_{600} 0.4. 1600 µl of each sample was pelleted by centrifugation at 7000 x *g* for 3 minutes. Cells were resuspended in 200 µl of 1.5 µM ColE9AF647 Im9His_6, 20 µM Hoechst stain for visualising DNA and incubated with rotation for 30 minutes at room temperature. Cells were washed three times in 400 µl of media supplemented with the additive of interest. Cells were resuspended in approximately 200 µl of media and 10 µl of cells placed onto a 1% agarose pad. Fluorescence was visualised by widefield microscopy using the DeltaVision OMX V3 system. Images were taken using a 50 ms exposure and exciting cells at 405 nm wavelength for the Hoechst stain followed by 642 nm for excitation of AF647. Images were processed using Fiji image processing software (Schindelin

et al., 2012). Experiments were performed in duplicate with N>10 cells used per each condition.

2.13 Transposon Direction Insertion Sequencing (TraDIS) to identify genes involved in resistance

2.13.1 Exposing a TraDIS library to colicin E9

E. coli NCTC13441 TraDIS library was incubated overnight at 37 °C in MMTE media (MOPS Minimal media KPO_4, 1mM EDTA, 100mM Tris-Hcl pH8 with 0.2% glucose) or MMT (MOPS Minimal media KPO_4, 100mM Tris-Hcl pH8 with 0.2% glucose) with 0.5 nM ColE9Im9His_6 or without colicin as a control. Cultures were then used to inoculate identical media 1:10 which is then cultured overnight at 37 °C. DNA was extracted from the passaged culture and sequenced using the Illumina Hiseq2500 platform.

2.13.2 Analysis of insertions using the BioTraDIS pipeline

The BioTraDIS pipeline is pearl library offering multiple tools to analyse insertions (Barquist *et al.*, 2016). BAM files from sequencing runs were converted to FASTq files containing sequence reads and the transposon tag. Only reads containing the transposon tag were used for further analysis. The transposon tag ('TAAGAGAC') was removed from reads and the sequence mapped to a reference genome (UPEC ST131- NCTC13441) using SMALT2. 20 reads per insertion was considered significant and reads within 10% of the 3' end of an open reading frame ignored, as this has been shown to not interrupt protein function (Griffin *et al.*, 2011). Index parameters were altered for reads of different length (read length less than 70bp: -k 13 -s 4, 70bp-100bp: -k 13 -s 6, greater than 100bp: -k 20 -s 13). Reads with multiple best mappings were reported as unmapped and read identity was set at 96%. For each gene an insertion index was calculated which normalizes the number of insertions against the length of the gene. Insertion indexes give a bimodal distribution of essential and non-essential genes calculated using likelihood

ratios. The tradis_compairsons R script calculates the fold change of insertions indexes between two conditions. A log fold change >2 with a q value < 0.00005 was used as a significant threshold.

2.14 BIOLOG Phenotypic Microarray analysis

BIOLOG Phenotypic Microarrays (PM) monitor respiration of bacteria across a variety of growth conditions through measuring the reduction of a tetrazolium based dye (Bochner *et al.*, 2001).

MIC of ColE9 for UPEC ST131 was established in the two different media used in BIOLOG PM, 0.37 µM with IF-0 (a minimal media) and 41nM for IF-10 (a rich media). Four PM plates were selected to test the effect of growth conditions on resistance (Table 2.6). UPEC ST131 was grown overnight in MMTE medium at 37 °C with shaking and used to inoculate IF-0 and IF-10 media with 1/3 MIC ColE9Im9His$_6$. As a control, UPEC ST131 was used to inoculate IF-0 and IF-10 media containing an inactive colicin, which lacks a region at the N-terminus responsible for translocation and a disulfide bond (NGH08 µM δ1-60 Y324C, L447C ColE9Im9His$_6$) which does not allow the conformational change needed for translocation. Inoculated PM plates were observed for 48 hours using the OmniLog incubator/reader measuring absorbance every 15 minutes.

Table 2.6: **Phenotypic microarray plates used for BIOLOG analysis.** Four PM plates were used to test the effect of growth conditions on the resistance of UPEC ST131 towards Colicin E9.

Plate	Media	Comments
PM01	IF0	Tests a range of carbon sources
PM02a	IF0	Tests a range of carbon sources
PM09	IF10	Tests for the effects of osmolytes
PM10	IF10	Tests a range of pH with different additives

BIOLOG data was analysed using the DUCTAPE phenotypic microarray analysis suite (Galardini *et al.*, 2014). DUCTAPE analyses respiration curves by measuring nine parameters (min, max, height, plateau, slope, lag, area, v, y0) and performing K-means clustering to assign each curve an Activity Index (AI). An elbow test was performed and eight clus-

ters selected as optimal clustering. Differences in AI between the active and inactive colicin treated cultures were calculated.

2.15 SNP analysis of resistance strains from BIOLOG analysis

Resistance mutations arose in wells of the BIOLOG PM plates (a single well out of 3 replicates showed growth). Genomic analysis for Single Nucleotide Polymorphisms (SNP) which explain resistance was performed. Four cultures from BIOLOG wells which had suspected resistance mutations were cultured overnight in MMT media at 37 °C with shaking. DNA was extracted using the Wizard® genomic DNA extraction kit and sequenced using the Illumina Hiseq2500 platform. Sequence reads were mapped against the reference genomes (UPEC ST131 NCTC13441) using SMALTv2.0 using parameters identical to the TraDIS analysis. Detection of variants was performed using samtools mpileup with the parameters -d 1000 -DSugBf to generate the BCF file of variant sites. Raw SNPs were passed to bcftools and SNPs with mapping quality < 30 ignored. For each position the majority base call was required to be in $\geq$ 75% of reads with a minimum sequence depth of 2 reads on each strand.

2.16 Resistance testing of diverse UPEC strains

Ten *E. coli* strains from clinical cases of Urinary Tract Infection (UTI) were tested for sensitivity towards colicin E9.

Cells were grown in MMT overnight at 37 °C with shaking. Cells were diluted in MMTE to reach McFarland standard and swabbed onto a plate of MMTE agar. 2 µl serial dilutions of ColE9Im9His_6 in MMT were added to the plate and incubated overnight at 37 °C. Sensitivity was marked as zones of clearance on the plate.

Table 2.7: **Strains of uropathogenic *E. coli* used for sensitivity testing** Diverse strains of uropathogenic *E. coli collected from different hosts were tested for sensitivity to colicin E9*

Strain	Sequence Type	Serotype	Host	Year of Isolation
17424	10	nt	Pig	2008
8130	10	O32:H34	Dog	2003
15010	12	O18:H6	Dog	2005
6745	14	O18:H5/11:K5	Human	2003
7920	14	O75:nt	Human	unknown
13836	48	nt	Human	2001
13843	69	nt	Human	2001
14967	73	O74:H1	Dog	2005
9258	73	O6:K2:H	Dog	2004
13832	95	O2:H4	Dog	2001

2.17 Genomic databases

Nucleotide fasta files are freely available for PATRIC (Wattam *et al.*, 2014) (ftp.patricbrc.org/patric2 accessed 14/05/2014) and NCBI assembled bacteria at (ftp.ncbi.nlm.nih.gov/genomes/ASSEMBLYBACTERIA accessed 14/05/2015). We also used the PubMLST Multispecies isolate database (Ribosomal Multilocus Sequence Typing (rMLST database)) (https://pubmlst.org/rmlst/ accessed 10/12/2014) (Jolley *et al.*, 2012). We subjected all databases to the nuclease bacteriocin identification pipeline, however due to the overwhelming size of the multispecies database (~10 fold larger than PATRIC) results for overall NB presence are only shown for the multispecies database. Each database was used for a specific purpose. PATRIC has a high level of associated metadata and was used to calculate environmental association. PATRIC databases were also used to identify pore forming bacteriocins (2.19). NCBI assembled bacterial genomes were used to identify Ton and Tol operons as the genomes are fully assembled. The multispecies database contains NCBI nucleotide database (finished genomes) and community contributed NCBI assembled genomes (NCBI assembled database). It also contains publicly available raw read data from the European Nucleotide Archive-Short Read Archive (ENA-SRA). The PubMLST Multispecies isolate database was established with the aim to

collect the widest range of genomes from all bacterial classes that have been sequences by NGS. It aims to reflect the current knowledge of bacterial genomes and does not explicity bias itself towards any subset of bacteria. It is however subject to any bias of the sequences deposited into the ENA-SRA. Details of how sequences were assembled are included in the appendix.

2.18 Plasmid or chromosomal association

2.18.1 Assigning plasmid or chromosomal association using Plasmid Finder

The Plasmid finder database (accessed 29/03/2017) was downloaded and indexed into a blast database for local use (Carattoli *et al.*, 2014). Contigs which contained a NB were extracted from genomes and blastn performed against the Plasmid finder database. If a contig had a hit to the Plasmid finder database with E-value lower than $5x10^{-5}$ the contig was assigned as having a plasmid locus. Chromosomal association was considered to be a contig that had no hit to the plasmid finder database or a hit with E-value greater than $5x10^{-5}$. To increase our ability to detect plasmid sequences we added a conserved region for pColE1 replicons (Rijavec *et al.*, 2007).

2.18.2 Assigning plasmid or chromosomal association using a database of all known plasmid sequences

Plasmid sequences were taken from ftp://ftp.ncbi.nlm.nih.gov/genomes/Plasmids (accessed 20/02/2015 now archived at ftp://ftp.ncbi.nlm.nih.gov/genomes/archive/oldrefseq/Plasmids/) and indexed into a local blast database. NB containing contigs were extracted from genomes and blastn performed to show alignment with the plasmid database. For each contig the total proportion of the contig that aligned to a plasmid with E-value lower than $5x10^{-5}$ was summed and each contig given a percentage overlap to plasmid sequences. For each of the genus, the distribution was presented as a 'beeswarm' plot.

2.19 Pore forming pipeline

To scan genomes for pore forming bacteriocins (bacteriocins which have a cytotoxic effect through ionophore activity) a novel pipeline utilising Trans Membrane Hidden Markov Models (TMHMM) was used (Krogh *et al.*, 2001). The pipeline makes three assumptions about properties of pore forming bacteriocins.

1. Pore forming bacteriocins contain only 1 region of transmembrane helicies.
2. This region of TM helix is found at the C-terminus
3. Pore forming bacteriocins do not contain any enzymatic motifs or signalling domains.

As the position of the transmembrane helices within an ORF is necessary for determining if a sequence is a pore forming bacteriocin (PFB), analysis was performed on annotated genomes from the PATRIC database. To reduce computational time ORF predictions for each genus tested were clustered for analysis by cd-hit at a 95% sequence identity. A predicted ORF from each cluster was analysed using TMHMM2.0. To test for the expected position of TMHMM in PFBs, characterised PBs were accessed from NCBI and used as test database (Table 2.8). Proteins were scanned for the position of transmembrane helices and the position of the TM region was given as a percentage of sequence length.

Cutoffs were set so TMHMM predictions over 0.8 must be found in the final 2/7th of a sequence and not in the first 3/7th of a sequence. Sequences that passed this rule were then scanned against the Pfam database as previously described for the NB pipeline and sequences that had profiles other than Colicin (PF01024.14) or Cloacin (PF03515.9) were rejected.

Table 2.8: Characterised pore forming bacteriocins from three species were used to establish cut offs for the position of the TM region.

Pore forming bacteriocin	Accession
Colicin B	YP_003829044.1
Colicin K	Q47502.1
Colicin E1	NP_040356.1
Colicin N	P08083.1
Marcescin N28b	CAA44310.1
Colicin Ia	WP_001582575.1
Colicin U	CAA72509.1
Pyocin S5	NP_249676.1

2.20 Identifying Ton and Tol operons in bacterial genomes

2785 fully assembled genomes with ORFs predicted by Glimmer v.3.0 were downloaded from NCBI (ftp.ncbi.nlm.nih.gov/genomes/ASSEMBLY_BACTERIA (accessed 14/05/2015)). PfamA.v27 profiles and HMMer V3.1 (Eddy, 2011) were used to identify the conserved regions of proteins within Ton/Tol operons (Table 2.9).

Table 2.9: **Pfam profile HMMs used to identify proteins of the Ton and Tol operons**.HMM profiles were used to identify the conserved regions of proteins within Ton/Tol operons

PFAM accession	Name	Description
PF03544.9	TonB_C	C-terminal region of TonB. Interacts with TonB box.
PF01618.11	MotA ExbB	MotA/TolQ/ExbB proton channel family
PF02472.11	ExbD	Biopolymer transport protein ExbD/TolR/MotB
PF07676.7	PD40	WD40-like Beta Propeller Repeat associated with TolB

HMMscan was used to identify profile matches in open reading frames of genomes as predicted by Glimmer. Within individual genomes regions of interest were identified

by clustering the co-ordinates of ORFs highlighted as potential Ton/Tol proteins with a threshold of 2000 bases. Simple rules were implemented by custom python scripts to distinguish between operons of Ton, Tol and Mot. Tol operons were defined as regions of interest containing hits to profiles ExbD, MotA_ExbB and PD40. Ton operons were defined as containing ExbD, MotA_ExbB and TonB_C but without a PD40 repeat associated with TolB.

2.21 Clustering CLB sequences

2.21.1 Removing cytotoxic domains from sequences

The modular domain organisation of CLBs, as well as many other polymorphic toxin proteins, is well conserved. C-terminal domains contain the cytotoxic region whilst the N-terminal domains are involved in translocation. Whilst in Colicin E9 the translocation and receptor domains are well defined, in many other colicins and CLBs from other species they are not well delineated and in the case of pyocins are rearranged. To investigate the diversity of the translocation and receptor domains independently of the cytotoxtic, domains associated with receptor binding and translocation were separated from cytotoxic domains by removing sequence that overlapped with the Pfam cytotoxic domain (as predicted using HMMer3.1.b) to leave what we term the translocation/receptor domain (TR).

2.21.2 CLANS analysis of CLB proteins

Sequences were clustered using CLANS which uses a blast all approach to calculate the forces between nodes in a Fruchterman-Reingold graph layout algorithm (Frickey and Lupas, 2004). Clusters were calculated using a network approach (each sequence emits the clusters it is linked to weighted by the negative log of the p-value between two sequences). Only high scoring pair hits with $p<1x10^{-10}$ were used in force calculation.

2.21.3 Prediction of Coiled-coil domains

Left handed coiled-coil regions were predicted using ncoils and the MTK database with a 28 residue window (Lupas *et al.*, 1991). ncoils compares protein sequences to a database of coiled-coil regions and generates a sore for each residue.

2.21.4 Prediction of Disordered regions

Disorder regions were predicted using IUPRED (Dosztanyi *et al.*, 2005). IUPRED was used with long disorder detection. To be classified as disorder a region must contain $\geq$ 30 residues predicted greater than 0.5. Regions were identified using custom python scripts and the intertools package.

2.22 Pangenome analysis of *E. coli* strains

2.22.1 Assembly of genomes from European Nucleotide Archive-Short Read Archive

357 *E. coli* genomes (166 colicinogenic strains and 191 non-colicinogenic strains) were accessed from the ENA-SRA. Fastq reads were assembled via a Velvet based assembly and improvement pipeline as described (Page *et al.*, 2016). Briefly reads were assembled using VelvetOptimiser v2.2.5 and Velvet. For each sample sequences with the best N50 were improved via scaffolding using SSPACE before sequences gaps were filled by GapFiller. Annotation was performed using Prokka.

2.22.2 Pangenome analysis using Roary

Roary is a popular pangenome software for bacterial genomes. It offers a significant increase in speed on other softwares primarily due to a iterative clustering step before the computationally expensive all against all comparison (Page *et al.*, 2015). Roary was performed on gff files as generated by Prokka with a similarity cut off of 95%. Phylogeny

was calculated from a core gene alignment produced by Roary using the software RAxML (WAG model with Gamma correction) (Stamatakis *et al.*, 2005).

2.22.3 Statistical analysis of gene association

Population stratification was calculated using Bayesian analysis of population structure (BAPS) clustering (Corander *et al.*, 2004). Genes which showed no statistical association to colicinogenic genomes by a Fisher's exact test were removed from further analysis. Clusters predicted by BAPS were used as the strata for a Cochran-Mantel-Haenzsel test implemented in R.

2.23 Translation of DNA sequences

DNA sequences of bacterial genomes were translated to protein sequences using transeq translating using a bacterial codon table (table 11) and calculating all six open reading frames.

2.24 Open reading frame identification

Open reading frames were identified using Prodigal v2.6.1 in metagenomic mode ignoring any open reading frames less than 200 nt.

2.25 Identifying conserved domains using HMMer

HMMscan was used to identify conserved domains in either proteins sequences or proteins translations of entire bacterial genomes. Profiles hidden markov models were either made from alignments of protein sequences or selected from Pfam 27.A .

2.26 Generating new HMM profiles

HMM profiles were generated from sequence alignments produced using Clustal Omega and trimmed to a conserved sequence. Profiles were built using HMMbuild from the HMMer3.1b suite (Eddy, 2011).

2.27 Clustering TR domains

TR domains (as calculated by 2.21.1) were clustered using cd-hit, a greedy clustering algorithm that clusters sequences based on global similarity.

2.28 Identifying pyocin receptor genes in bacterial genomes

To investigate the association between genes of pyocins and *fpvA* receptors genes (fpvAI, fpvAII and fpvAIII) genes were identified within bacterial genomes from the rMLST multispecies database. Genes were identified using BLASTN (word size: 11, reward: 2, penalty: -3, gap open: 5, gap extend: 2). Results were filtered by >90% sequence identity and length.

2.29 Identification of motifs using MEME

Multiple Em for Motif Elicitation (MEME) was used to identify conserved motifs within proteins (Bailey *et al.*, 2009). Cytotoxic domains were removed from sequences using the method stated in Section 2.21.1. MEME had parameters of identifying 1 or more repeats of a motif in a sequence, with a motif width between 6 and 300 residues.

2.30 Identifying physiological properties of cytotoxic domains

The EMBOSS program pepstats was used to was used to calculate charge and isoelectric point of cytotoxic domains (Rice *et al.*, 2000). Cytotoxic domains were identified as described in Section 2.21.1.

CHAPTER 3

Chapter 3: Bioinformatic analysis of nuclease bacteriocins

3.1 Introduction

Antibiotics have revolutionised and become indispensable for the treatment of bacterial infections. The impact of antibiotics is, however, being undermined by the development of antibiotic resistant bacteria. The spread of resistance has greatly outpaced the development of new classes of antibiotics leading to the antibiotic resistance crisis. Current generations of antibiotics are often broad spectrum antibiotics, with the potential to target a wide array of bacteria, which can cause off target effects and disrupt the native microflora which is increasingly being linked to major health roles (Francino, 2015). As this work will discuss peptide bacteriocins and protein bacteriocins which have overlapping terminology and nomenclature, the nomenclature used throughout is listed below (3.1).

Table 3.1: **Nomenclature used to describe different classes of bacteriocins**

Term	**Acronym**	**Description**
Bacteriocins	-	Broad term encompassing all diffusible proteinaceous agents of bacterial competition
Protein bacteriocins	PB	All protein bacteriocins from **Gram-negative** bacteria, including colicin-like bacteriocins and taliocins
Colicin-like bacteriocins	CLBs	Gram-negative proteinaceous bacteriocins with a modular domain organisation, no secretion system and no structural homology to phage tails (tailocins)
Nuclease bacteriocins	NBs	CLBs which carry a nuclease enzymatic motif within the cytotoxic domain
Pore forming bacteriocins	PFBs	CLBs which a pore forming region as the cytotoxic domain

Bacteriocins are proteinaceous agents of bacterial competition. They possess potent antimicrobial activity and species specificity, two desirable traits for the next generation of antimicrobials. Peptide bacteriocins have received much focus. Indeed nisin, a peptide bacteriocin produced by *Lactococcus lactis* ssp. *lactis*, has been used commercially as a

food preservative since the 1950's (Cotter *et al.*, 2005). Protein bacteriocins have often been overlooked as potential antimicrobials but they have a number of advantages over their peptide cousins. Colicin-like bacteriocins (CLBs) are highly potent antimicrobials, target a narrow spectrum, do not have post translational modifications (which are extensive in certain peptide bacteriocins), and they have a modular domain structure which allows relatively easy engineering and the future possibility of designer antimicrobials (Cascales *et al.*, 2007).

3.1.1 Competition systems

CLBs can be classified as a member of the polymorphic toxin family (PMT), a family of protein toxins within bacteria. PMT are characterised by an N-terminal delivery mechanism and a C-terminal cytotoxic domain (Jamet and Nassif, 2015). PMT encompasses many of the well studied competition systems; CLB, Type 6 Secretion System effectors (T6SS) and contact dependent inhibitors (CDI) (Reviewed in sections 1.5 and 1.4 respectively). One defining trait of PMTs is the ability to utilise different cytotoxic domains and recombination of cytotoxic domains with N-terminal delivery systems. Cytotoxic domains across PMTs share high sequence similarity and contain conserved enzymatic motifs. CLBs have unique properties that separate them from other PMTs. CLBs do not have a known dedicated secretion system and instead rely on quasi-lysis of the cell for release. CLBs are also diffusible toxins and do not require contact between producing and target cells.

3.1.2 Colicins and colicin-like bacteriocins

Colicins were first identified over 90 years ago (Gratia, 1925) and homologues have since been identified in multiple species including: *Pseudomonas aeruginosa* (pyocins), *Klebsiella pneumoniae* (klebicins) and *Yersinia pestis* (pesticins) (James *et al.*, 1987; Michel-Briand and Baysse, 2002; Rakin *et al.*, 1996). CLBs target cells by binding specific outer membrane proteins which gives them a narrow spectrum of activity, typically targeting closely related organisms. Both experimental and theoretical work have suggested

a role for bacteriocins in microbial competition. The well studied Rock Paper Scissors (RPS) model in which a colicin producing strain, a sensitive strain and a resistant strain all compete within an environment, has provided a model which supports the hypothesis that bacteriocins can act to maintain a diverse microbial community (Kerr *et al.*, 2002) (Figure 3.1), however the parameters of this model are not fully understood. It is still unclear what proportion of the population are CLB producing cells and how widespread is CLB resistance. The role of CLBs in microbial diversity becomes more apparent when the RPS model is expanded to include different structured environments. In well mixed cultures with freely diffusing cells, the burden of production is high but the benefits of toxin production are shared amongst the population. Introducing a structured environment restricts the benefits of CLB production while increasing the local concentration of CLB, thereby offering a greater advantage (Chao and Levin, 1981; Kerr *et al.*, 2002). Spatial dynamics can also maintain a population in which multiple CLB types are being produced (Majeed *et al.*, 2011). Experimental work into the RPS model has typically used artificial environments or simple two species murine models. How well the RPS framework can model an *in vivo* environment has yet to be established.

As well as their role in maintaining diverse microbial communities, CLBs could have a role in virulence; acting to displace native microflora and colonise a niche. *Salmonella enterica* Typhimurium utilises colicins to outcompete commensual bacteria during intestinal inflammation (Nedialkova *et al.*, 2014) and a survey of *Shigella sonnei* isolates across Vietnam identified a colicin plasmid as a factor which allowed a strain to become established within a population as the major cause of dysentery (Holt *et al.*, 2013). Extra intestinal *E. coli* strains show significant association between colicins and virulence factors (Micenková *et al.*, 2014a; Petkovsek *et al.*, 2012). These association studies which link colicins and certain pathotypes of *E. coli* use PCR based methods which focus on small groups of virulence factors and test for a simple association. As the computational power and the number of sequenced bacterial genomes are both increasing exponentially, pangenome studies are now a feasible method to study genome wide associations, testing for association between a trait such as CLB production and all genes across hundreds of genomes.

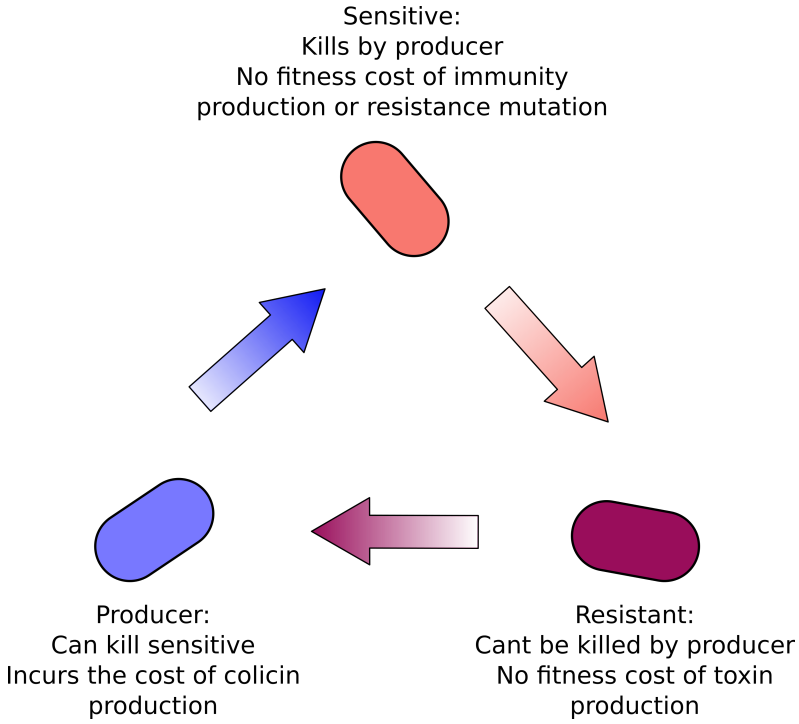


Figure 3.1: **RPS model of CLB production promotes biodiversity** The Rock-paper-scissors model divides bacteria into CLB producing, sensitive and resistant cells. Producing bacteria suffer the burden of CLB production but can kill sensitive cells, sensitive cells are killed by the producer but are not burdened by the cost of resistance e.g. mutation or production of an immunity protein, meaning they can outgrow resistant cells. Resistant cells are not burdened by toxin production and can out-compete producing cells but suffer the burden of resistance and are therefore out-competed by sensitive cells.

CLBs have the potential as the next generation of antimicrobials. Few studies, however, have explored this potential, particularly *in vivo*, though the few studies that have been performed have given promising results. In a murine model of *P. aeruginosa* lung infection, pyocin S5 was shown to be over 100-fold more potent than tobramycin, a common antibiotic with high activity against *P. aeruginosa* (McCaughey *et al.*, 2016a). As well as deploying CLB proteins as antimicrobials, CLB producing *E. coli* were used to inoculate a catheter and prevented colonization of uropathogenic *E. coli* (Trautner *et al.*, 2005).

3.1.3 Genetic organisation of CLB operons

CLBs have no post translational modification, therefore, compared to peptide bacteriocins they have a relatively simple genetic organisation. The CLB toxin is encoded by a single gene (*cxa*). For nuclease bacteriocins (NBs) this toxin is followed immediately downstream by an immunity protein (*cxi*). An SOS response dependent promoter leads to expression of a single transcript containing the open reading frames for both the NB and the immunity protein (Gillor *et al.*, 2008; Soong *et al.*, 1994). A second promoter within the NB ensures that immunity protein is always in excess of the toxin to prevent both self induced death and death from exogenous toxin (Soong *et al.*, 1994). Pore forming bacteriocins (PFB) have a different genetic organisation. As the immunity protein acts at the inner membrane, it is not encoded immediately downstream of the toxin and is often found on the opposite strand (Lloubes *et al.*, 1986). Further downstream of the immunity protein lies a lysis protein (in group A colicins). The lysis protein causes release of the CLB bacteriocin by quasilysis and eventually, death of the cell (Cascales *et al.*, 2007). Where a lysis protein is not present (such as group B colicins), lysis is believed to be caused by the hijacking of latent phage lysis genes, both of which are induced late in the SOS response and cause lysis of the cell (Nedialkova *et al.*, 2015). The location of CLB genes varies between species. Colicin genes are exclusively located on plasmids whereas pyocins are chromosomal (Michel-Briand and Baysse, 2002).

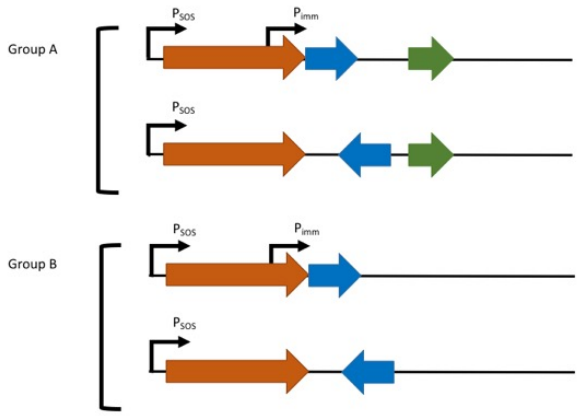


Figure 3.2: **Operon structures of colicins**

3.1.4 Domain organisation of CLBs

One of the major advantages for the engineering of CLBs is the modular domain structure. The majority of studies into CLB structure and its relation to translocation into target cells, have been primarily on colicins, particularly the 'E-type' colicins. Indeed, the only crystal structures of full length nuclease and pore forming CLBs are from colicins; the PFBs (Colla (Wiener *et al.*, 1997), ColB (Hilsenbeck *et al.*, 2004), ColS4 (Arnold *et al.*, 2009) and ColN (Vetter *et al.*, 1998)) the NBs (ColE3 (Soelaiman *et al.*, 2001) and ColE9 (Klein *et al.*, 2016)), and for the periplasmic murimidases (ColM (Zeth *et al.*, 2008), Syringacin M (Grinter *et al.*, 2012) and Pectocin M (Grinter *et al.*, 2014)). For simplicity, domains will be introduced by the archetypal E-type colicin domain organisation. The domain organisation is predicted to be rearranged in pyocins however more biochemical analysis need to be performed to provide evidence for the exact function of each domain (Michel-Briand and Baysse, 2002).

Colicins vary in size and sequence similarity, yet mostly share a common tripartite domain organisation; an N-terminal translocation (T) domain, a central receptor binding (R) domain and the C-terminal cytotoxic (C) domain. Colicins enter a target cell by binding an outer membrane receptor, typically a vitamin or iron receptor before, in the case of certain NBs, recruiting a second outer membrane protein for translocation. The translocation domain contains specific peptide epitopes which are threaded through the outer membrane translocator protein and contact either the Ton or Tol systems (Kleanthous,

2010b). In E-type colicins, the peptide epitopes are located within a disordered region known as the intrinsically unstructured translocation domain (IUTD) and contact the Tol system (Housden *et al.*, 2010). The Ton and Tol systems are PMF coupled complexes with homology to the flagella motor proteins, the Mot system (Zhai *et al.*, 2003).

In keeping with their classification as polymorphic toxin systems, the T and R domains of colicins are just a complex system for delivery of the toxin, the cytotoxic domain. CLBs can deliver one of five types of toxins into the target cell, a pore forming or a lipid II hydrolase which act within the IM/periplasm and a DNase, tRNase or rRNase which act within the cytoplasm. Nuclease bacteriocins cause death of the target cell by degrading nucleic acid. These cytotoxic domains are bound by highly specific and extremely high affinity immunity proteins and have been extensively studied as a model for affinity and specificity in protein-protein interactions (Kleanthous *et al.*, 1998; Kleanthous and Walker, 2001; Papadakos *et al.*, 2012). Unlike pore forming bacteriocins which use hydrophobic helices to insert into the membrane, nuclease bacteriocins contain conserved enzymatic motifs within their catalytic core. These conserved regions are often surrounded by regions of polymorphism which determine specificity in the interaction with immunity proteins and vary between species (Riley, 1993b). Five different NB families have been described: the HNH DNase, Non-HNH DNase, ColD-like tRNase, ColE5-like tRNase and rRNase. Each of these families have conserved enzymatic motifs, for example HNH DNases contain a 30 residue motif ($HHX_{14}NX_8HX_3H$) in the active site (also known as the $\beta\beta\alpha$ motif). The HNH motif is found in a multitude of enzymes; Type II restriction enzymes, homing endonucleases, mismatch repair enzymes and CRISPR/Cas9 (Jiang *et al.*, 2013; Keeble and Kleanthous, 2005; Shen *et al.*, 2004). Similar motifs have also been identified in the other NB cytotoxtic domain families (Lin *et al.*, 2005; Walker *et al.*, 2004; Yajima *et al.*, 2004). Polymorphic toxin systems utilise conserved enzymatic motifs to kill cells therefore domains with similarity to colicin cytotoxic domains are found across the PMT family.

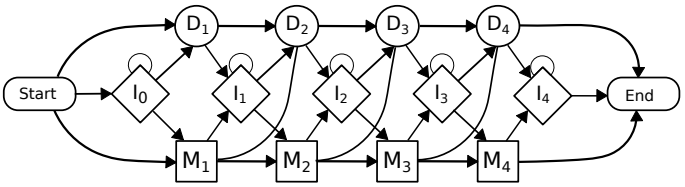


Figure 3.3: **Overview of the pHMM strategy** Observed states (residues) are dependent on an underlying series of states. Positions can either be in the matched state (M_k), insertion state (I_k) or deletion state (D_k)

3.1.5 Hidden Markov models

Observing sequence similarity greater than that expected by chance between two proteins is often used to define homology. In a basic view, if two proteins share similarity the simplest explanation is that the share a common ancestor. Residues within a protein sequence are not all under the same degree of selection. For instance, in HNH-DNase colicin domains, the histidine and asparagine residues of the HNH motif are highly conserved and invariant between species, whereas the majority of the domain is under positive selection and highly variable. Pairwise alignments methods often treat similarity across the protein as uniform and make no inferences about conserved or variable residues.

HMMs are statistical models where a series of observable events is dependent on an underlying series of hidden states. In terms of protein homology, the residue is the observed state. Profile Hidden Markov Models (pHMMs) allow position specific information on the conservation or variability to be used in homology searching. pHMMs take a multiple sequence alignment of known homologues and for each position in the alignment calculates the distribution of residues to define the 'match state'. The emission probability at any position *k* in the alignment is given as $e(x|M_k)$ where M_k is the frequency distribution of each residue at position k in the alignment. If the alignment position k contains more gaps than aligned residues, an insertion state I_k is used to represent the column. If position k has a small number of gaps it is represented by both a match state, M_k and deletion state D_k. pHMMs then work in a linear manner 'left-right' (N-terminal to C-terminal) and to calculate a score by aligning a protein sequence to the profile.

HMM methods had previously suffered from expensive computation of parametrisation and scoring. Recent improvements in algorithms and computation now mean a typical search using HMMer3.1 is equatable to a sensitive BLAST search (Eddy, 2011). HMMer3 has been widely adopted by protein databases such as Pfam and remains the gold standard for remote homology detection by HMMs (Finn *et al.*, 2010). The improvements in speed now allow HMMER to be used as a tool by researchers, even with large data sets.

3.1.6 Bioinformatics of nuclease bacteriocins

There have been multiple attempts to identify the distribution of CLBs within bacteria, using both experimental and bioinformatic approaches (Ghequire and De Mot, 2014; Gordon and O'Brien, 2006; Parret and De Mot, 2002). While these approaches can give us a valuable insight into the diversity of CLBs, they suffer from certain weaknesses. Experimentally, studies have mainly focussed on using PCR based methods to identify colicin genes, however, this approach has obvious limitations. PCR primers will detect only a small proportion of colicins and polymorphism in regions targeted by primers would cause false negatives (Gordon and O'Brien, 2006). Bioinformatic studies are limited by the low sequence conservation in bacteriocins and the relatively small number of bacterial genomes that have been sequenced. Studies using either approach have mainly focussed on investigating a single species/genus.

Nuclease bacteriocins present an intriguing bioinformatics challenge. As nuclease bacteriocins target different receptors and hijack different proteins for translocation the sequence conservation is low, even in bacteriocins that target the same species. Secondly, NBs domains undergo rearrangements and duplications. The cytotoxic domains of NBs also share homology with domains from other PMT proteins or enzymes. For example, HNH nuclease domains are used in a variety of repair mechanisms and phage proteins and as previously mentioned, cytotoxic domains are also shared between the polymorphic nuclease toxins. Finally, nuclease bacteriocins are only present in a small proportion of the population. Estimates for colicins suggest only 4% of the population may be able to produce colicin meaning that analysis of a single genome representing a species, is unlikely to identify an NB gene (Gordon and O'Brien, 2006). There are, however, traits

of NBs which can allow their identification. They utilize similar cytotoxic domains which contain enzymatic motifs, they always have an immunity protein downstream to protect the producing cell from death and they are always of a certain length as they must encode enough information to pass through both the inner and outer membranes of the target cell. All of these identifying characteristics were applied to a bioinformatic pipeline to develop a robust protocol for the unambiguous identification of nuclease bacteriocins. The central principle of the pipeline uses profile Hidden Markov Models to identify both the cytotoxic domain and the immunity protein. As the two proteins are tightly genetically linked, additional criteria applies a distance metric between the proteins, ensuring one if always downstream of the next, a search strategy we term a 'paired profile Hidden Markov Model' or ppHMM.

3.1.7 Aims

Before CLBs could be considered as potential antimicobials it is important to know the diversity and distribution of these proteins within bacteria. This chapter aims to develop a bioinformatic pipeline to unambiguously identify CLB genes in bacterial genomes and characterise the levels of diversity present. Using the sequence data generated by the pipeline, we aim to answer questions about the biology of CLBs which have not been possible to answer experimentally including the hypothesis that CLBs are exploited by pathogenic bacteria.

3.2 Results

3.2.1 Developing a Hidden Markov Model based pipeline to identify nuclease bacteriocins

Nuclease bacteriocins utilise a variety of enzymatic motifs to degrade nucleic acids. Five main cytotoxic motifs are reported associated with nuclease bacteriocins in the literature, two DNases and three RNases (two tRNases and one rRNase). To identify cytotoxic motifs, known examples of CLBs containing the different motifs and their cognate immu-

nity proteins (e.g. the immunity that lies directly downstream and provides protection to the encoded toxin), were scanned against Pfam 27a, a database of profile Hidden Markov Models. Where profiles could not be identified in the Pfam database, profiles were generated from alignments of cytotoxic domains from diverse competition systems (Table 3.2).

Table 3.2: **Profile HMMs of cytotoxic domains and immunity proteins.** The cytotoxic domains and immunity proteins contain conserved motifs which can be identified using profile Hidden Markov models.

Name	Description	Pfam accession
DNase/IMM		
Colicin-DNase	Cytotoxic domain of HNH DNase colicins	PF12629.2
Colicin Pyocin	HNH DNase immunity protein	PF01320.13
tRNase		
Colicin D	tRNase cytotoxic domain of colicin D	PF11429.3
Colicin-immun	Immunity protein for colicin D tRNases	PF09204.5
Colicin_C	tRNase cytotoxic domain od colicin E5	PF12106.3
ImmE5	Immunity protein for Colicin E5	PF11480.3
rRNase		
Cytotoxic	rRNase cytotoxic domain of colicin E3	PF09000.5
ColE3_Immunity	Immunity protein for E3 rRNase	-
Non-HNH DNase		
Non-HNH DNase	Cytotoxic domain of Pyocin S3	-
Non-HNH DNase	Immunity protein of Pyocin S3	-

pubMLST contains 100,154 genomes (accessed 10/12/2014). To identify NB genes in such a large dataset required the development of a high-throughput bioinformatic pipeline. Genomes were first translated to protein in all six frames using Transeq. Each of the six reading frames were scanned using HMMscan (HMMer3.1b) to identify matches to either the cytotoxic domain profile or the immunity profile. The strategy of scanning a six frame translation of a genome rather than specific open reading frames (ORFs) has two mayor advantages. Firstly, scanning six translations is an order of magnitude faster than running ORF finder software and then scanning each ORF. Secondly, immunity proteins are not readily identified by ORF software, whereas the immunity profile is easily identified. As the presence of an immunity profile is required as a validation step, missing immu-

nity ORFs would greatly increase false negatives. The output from HMMscan was parsed using custom python scripts. Further custom scripts create a python object for each hit with attributes for statistics given by the output of the HMMscan. The cytotoxic domains of CLBs are used in a variety of enzymes and other competition systems. NBs, however, always have cytotoxic domains at the C-terminus and have a cognate immunity protein immediately downstream of the toxin. Therefore, to remove false positives (proteins which contain the motif but are not NBs), the criterion that cytotoxic profiles must be found downstream of the cytotoxic domain was applied. The combination of cytotoxic profile and immunity profile is termed a 'profile pair'. For each contig which contains both profiles of a 'profile pair', the distance between the profile pairs was recorded. To determine the allowed distance between profiles within a true NB pair, both literature and the distribution of distances from the entire database were used. The inter-profile distance is measured as the distance between the end of the overlap between the cytotoxic profile and sequence and the beginning of the overlap between the immunity profile and sequence (Figure 3.4). For four out of five profile pairs, a distribution between 0-60 bps (0-20 amino acids) was observed. ColE5 tRNase domain and immunity protein showed a narrow distribution centred around 180 bps. No negative inter-profile distances were observed which suggests the cytotoxic profile and immunity profile are independent of each other.

To remove other PMTs which utilise NB cytotoxic domains, the ORFs which overlapped either the suspected NB or its immunity protein were identified using Prodigal. Immunity protein open reading frames were extracted, however, they were not necessary for the analysis. Before scanning the ORFs for domains which are atypical of NBs, ORFs were discarded by size. Custom python scripts take each open reading frame and only pass those which have length within range 350 aa $\leqslant l \geqslant$ 950 aa. NB ORFs were then scanned against the entire Pfam 27a database to identify domains involved in export or delivery. As these domains are often associated with the very N-terminus of proteins (such as the Hcp-1 domain of Usp), and ORF software can often make errors in predicting the 5'-end of ORFs, all predicted NBs were clustered with harsh identity parameters (sequence similarity 98% using cd-hit) and the longest ORF within that cluster used as the example sequence for all the NBs within that cluster. Not only does this ensure sequences are

passed as NBs because N-terminal domains are missed, it also dramatically decreased computational time. Scanning all potential NBs against the entire Pfam 27a database is computationally expensive, therefore, only scanning one representative of each cluster decreases this time. Domains associated with other competition systems or phages were removed using custom python scripts. Two additional pHMMs were added to remove false positives not identified by scanning with Pfam 27a. First a pHMM for the MIX domain identified within the T6SS effectors of *Vibrio* spp. and secondly, after analysing the proteins at the N-terminus of potential NBs for T6SS related domains, a group of sequences from *Klebsiella* was identified (Salomon *et al.*, 2014). These sequences all contained a conserved N-terminal domain. These sequences were aligned using Clustal omega, trimmed using Jalview and used to generate a pHMM.

A schematic of the profile-pair Hidden Markov Model strategy is presented in Figure 3.5. 100,154 genomes from the pubMLST multispecies rMLST database were analysed and 282,414 hits to either cytotoxic domains or immunity protein profiles were identified. After applying the inter profile distance criteria 14,586 profile pairs was identified suggesting approximately only 1 in 10 were hits to genuine cytotoxic immunity profile pairs. Applying criteria restricting ORF size and Pfam profiles reduced the number of predicted NBs to 3095 from 100,154.

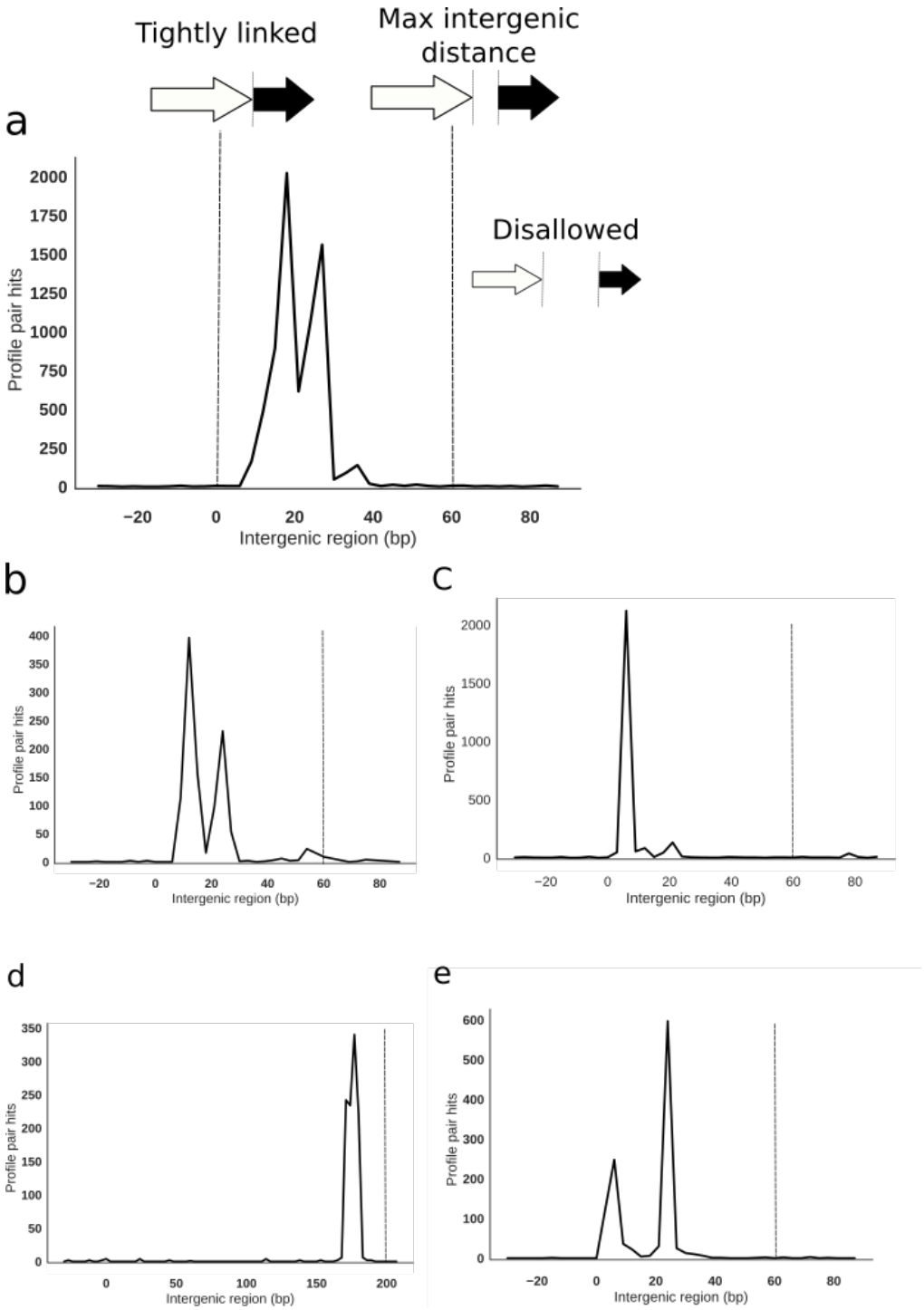


Figure 3.4: **Distribution of intergenic regions between cytotoxic HMMs and their associated immunitity HMM profiles**. Profile pairs with an intergenic region greater than 60 bp (or 200 bp for the case of ColE5 tRNases) were discounted from the analysis. ***a***) Distribution for the HNH DNase and associated immunity gene. ***b***) ColD tRNase, ***c***) Non-HNH DNase, ***d***) ColE5 tRNase, ***e***) rRNase

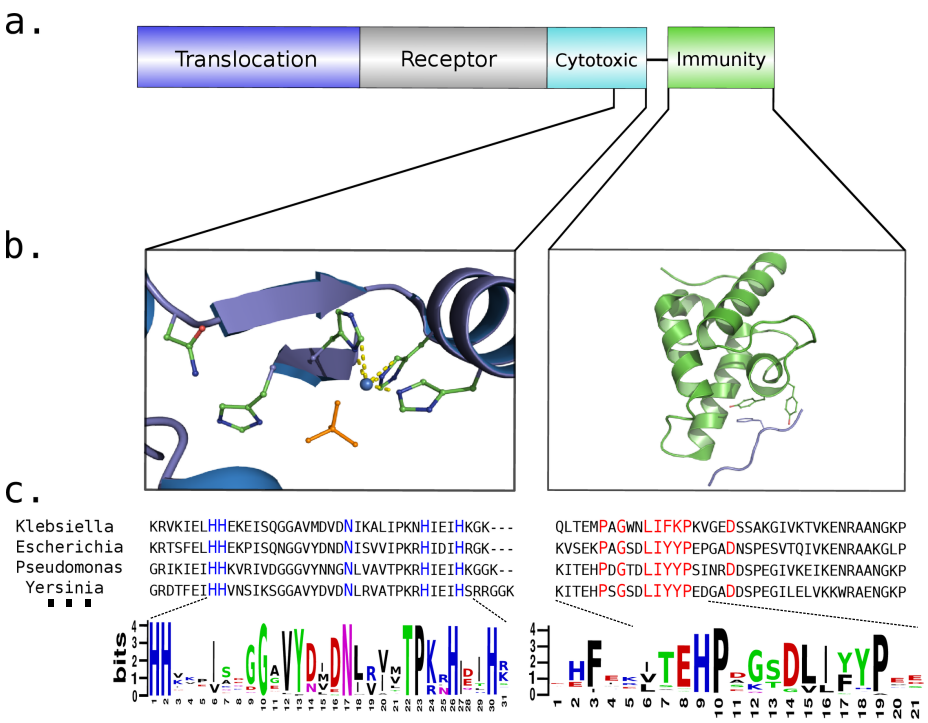


Figure 3.5: **Profile pair Hidden Markov model approach**. ***a.***Gene organisation of a typical nuclease bacteriocin/immunity protein from *E. coli*. ***b.*** *Left hand panel* Key conserved catalytic residues in the HNH motif of DNase bacteriocins. A divalent metal ion is coordinated by two histidine residues and an asparagine constrains the loop. The phosphate anion represents the position of the scissile phosphate in substrate DNA (PDB code, 1V14), *Right hand panel* the immunity protein (green) with the aromatic residues of α helix-III which is crucial for binding the DNase domain. ***c.*** Conserved residues are highlighted within the sequence alignment of probabilities of each residue in the motif are represented in a logo plot, only four sequences are shown as an example (HMM profiles are generated from alignments with a large number of sequences).

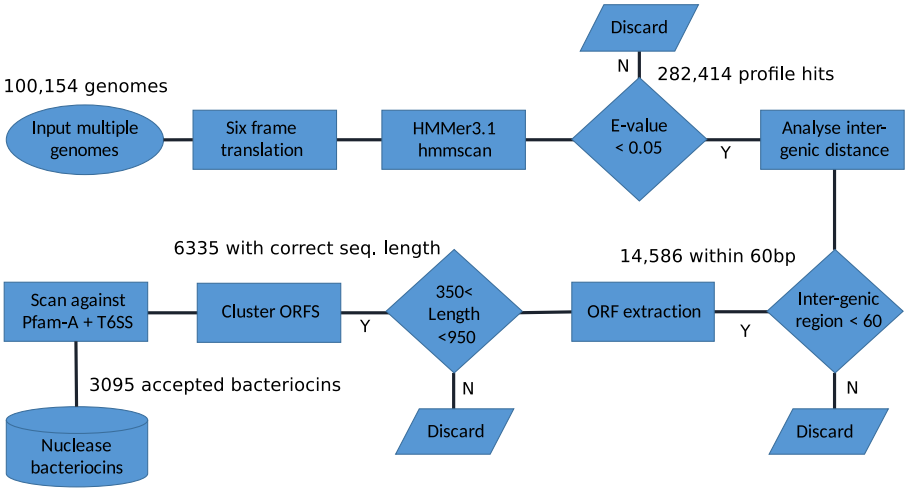


Figure 3.6: **Nuclease bacteriocin identification and validation pipeline.** Numbers indicate the accepted or rejected profile pairs at each decision point in the pipeline for sequences in the pubMLST multispecies database.

3.2.2 Non-Colicin like bacteriocin proteins identified and removed by the pipeline

Many PMTs and competition systems use a C-terminal cytotoxic domain followed immediately by an immunity protein to prevent death of the producer (Jamet and Nassif, 2015). PMTs which utilise a cytotoxic domain with homology to CLB cytotoxic domains were identified by the pipeline. 'Contaminants' of the database included Uropathogenic Specific Protein (Usp) produced by over 2000 *E. coli* and *Salmonella* strains. Usp has a cytotoxic domain with 40-50% similarity to the HNH-DNase domains of CLBs. Usp was not included as a NB for two reasons. Firstly, at the N-terminal of Usp is the domain T6SS_HCP (PF05638). Hcp proteins are both structural and exported by the T6SS. Whilst suspected Hcp domains with associated effectors have been described, no definitive evidence for their secretion by the T6SS has been shown (Ma *et al.*, 2017). Secondly, Usp has been shown to have activity against mammalian cells and therefore are unlikely to be a bacterial competition system (Nipič *et al.*, 2013). A non-HNH DNase T6SS effector was identified in over 50% *Klebsiella pneumoniae* genomes and only removed by examination of open reading frames N-terminal to the effector. A group of HNH-DNase domain containing proteins were identified in Gram positive *Bacillus* species linked to the SUKH domain (Syd, US22, Knr4 homology) superfamily of immunity proteins (Zhang *et al.*, 2011). The DNase domains of these bacteria were notably different to CLBs containing an extended region between the final two histidine residues, $HHX_{14}N_8HX_8H$. T6SS effectors were identified in *Vibrio* species containing the peptidoglycan binding LysM domain as well as the presence of the recently described MIX domain (Salomon *et al.*, 2014).

3.2.3 Distribution of Nuclease Bacteriocin

100,154 genomes within the pubMLST Multispecies rMLST database, covering >3,000 species, were subjected to the NB identification pipeline. 3094 bacteriocin genes were identified in 2479 isolates across 53 species.

NB genes were found in significant proportions in *Enterobacteriaceae* and *Pseudomonadaceae* but the proportion of a species that contained NB genes varied greatly; $<$5% of

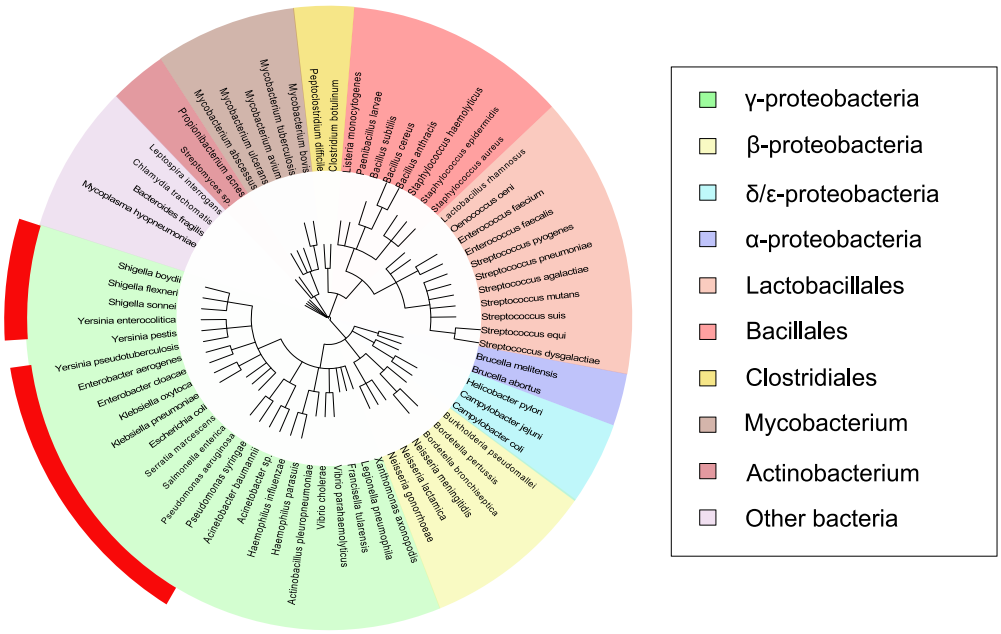


Figure 3.7: **Distribution of Nuclease bacteriocins across the bacterial kingdom.** Taxonomic tree representing the species within pubMLST that have over 100 genomes. Taxomony is constructed using NCBI commontree. Species that contained at least one identified NB are indicated by a red bar.

E. coli genomes encoded NBs compared to 85% for *Pseudomonas aeruginosa* and 31% for *Klebsiella pneumoniae*, which agrees with previous experimental work (Figure 3.8) (Gordon and O'Brien, 2006; Michel-Briand and Baysse, 2002; Riley and Gordon, 1996). *Pseudomonas* and *Klebsiella* isolates often encode multiple NBs which is predicted to have impacts on their evolution through increased rates of combination (as we discuss further in 3.2.13). *Klebsiella pneumonia* commonly possess both a DNase bacteriocin as well as the rRNase bacteriocin, CloacinDF13. This trend is also observed in *Pseudomonas* spp. and *Yersinia mollaretti* where four genomes encoded both a DNase and an rRNase NB.

3.2.4 CLBs are found in a variety of environments

The multispecies rMLST database, whilst expansive in terms of genomic data and species coverage, lacks any metadata, such as environmental niche of the isolate. To look at any association between environment and bacteriocinogenicity the PATRIC2 database was subjected to the NB pipeline. The PATRIC database contains 6256 genomes with associated environmental data. Over 1500 listed environments were grouped by key word

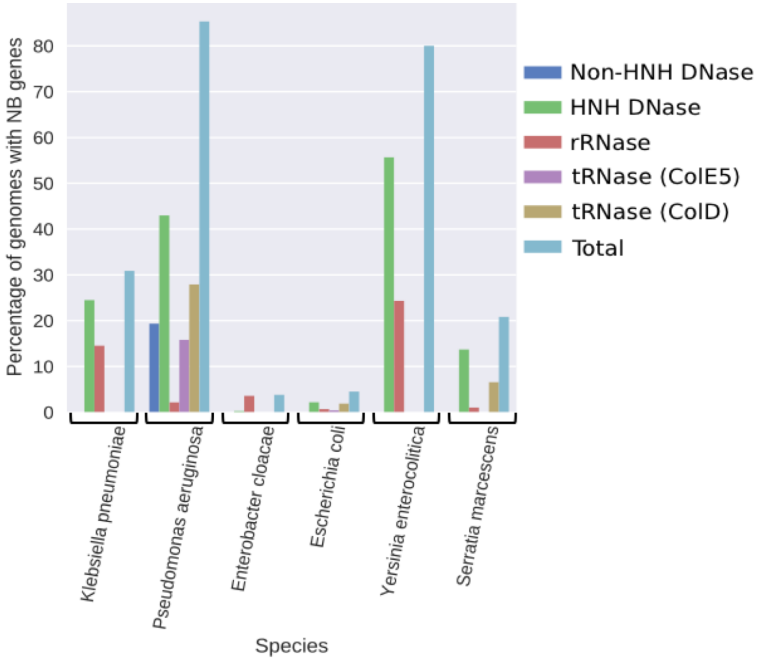


Figure 3.8: **Distribution of different cytotoxic domains within selected NB containing species** The presence of NBs and the different cytotoxic domains differs greatly between different species. 85% of *P. aeruginosa* contained an NB whilst only ~4% of *E. coli* genomes had an identifiable NB gene.

searching using custom python scripts. 120 environments were identified over 239 NB containing bacteria. Unsurprisingly, the majority of *E. coli* containing NBs were identified in the GI tract and the majority of *Pseudomonas* NB containing species were isolated from soil or plants. NB containing bacteria were also identified in more invasive environments such as the blood and urine of patients suffering from septicemia (*P. aeruginosa* VRFPA01 and VRFPA02) or urinary tract infection (*E. coli* UMEA-3342-1).

3.2.5 Ton and Tol complexes are distributed throughout Gram negative bacteria

The most surprising outcome of ppHMM search is that NBs (of the different cytotoxic domains used in the search) are found exclusively in γ-proteobacteria. Three hypotheses were generated to explain this taxonomic restriction: the environmental niche of bacteriocins, the machinery involved in their uptake or a failure of the profile Hidden Markov Models to detect distant homology. An environmental factor was dismissed as NB encoding bacteria were found in a variety of environmentally niches and were found abundant in soil, yet α-,β- and ϵ-proteobacteria are all more prevalent in soil (Figure

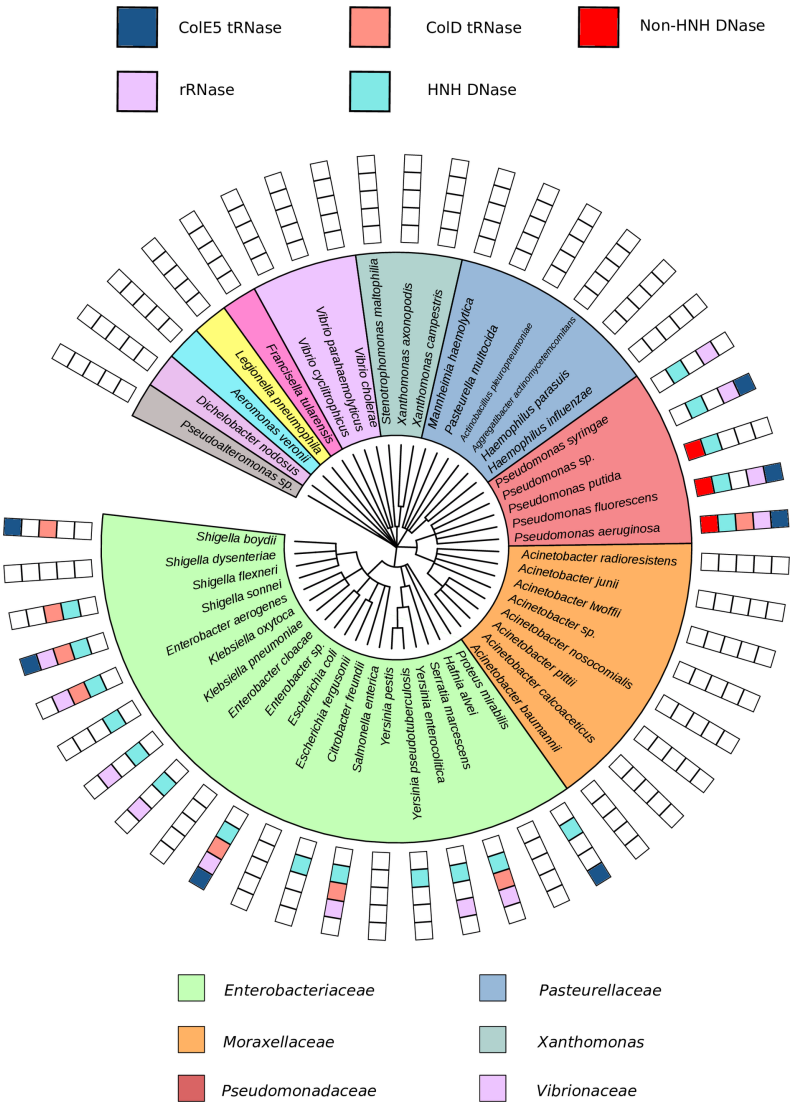


Figure 3.9: **Distribution of NBs across γ-proteobacteria** Taxonomic tree representing γ-proteobacteria which have over 15 genomes within the pubMLST. The color key represents the different cytotoxic domains. NBs are found throughout *Enterobacteriaceae* and *Pseudomonadaceae*

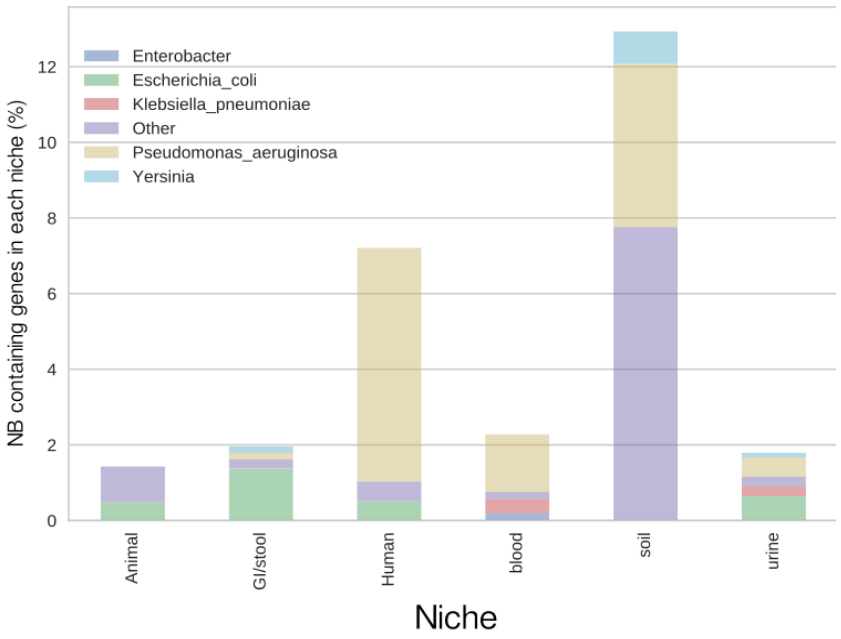


Figure 3.10: **Proportions of NB gene containing bacteria across different environments within the PATRIC2 database.** Environmental metadata is available for ~10% of isolates. Environments are split to show abundance of the most prevalent CLB containing bacteria (Sample sizes: Animal: 210, Blood: 527, Environmental: 641, GI/Stool: 1174, Human: 194, Lung: 385, Plant: 125, Soil: 116, Urine 778).

3.10) (Spain *et al.*, 2009). CLBs hijack PMF coupled proteins for translocation into the cell. Using a hidden Markov Model approach (Section 2.20) showed that Ton and Tol homologues are found throughout Gram negative bacteria (Figure 3.11). A small number of proteobacteria did not appear to have a Ton or Tol operon. The methodology used requires all the genes to be found within a small locus of a genome as Ton and Tol genes are normally found as an operon. Closer analysis of *Rickettsia* genomes showed that they do contain all the genes required for a Ton operon however they are split over two loci. Finally, the identified NBs could be restricted to γ-proteobacteria as the profile Hidden Markov Models may be limited to only identifying cytotoxic domains within

γ-proteobacteria, however, the ppHMMs identified hits to both cytotoxic and immunity profiles not only within proteobacteria but also in Gram positive bacteria.

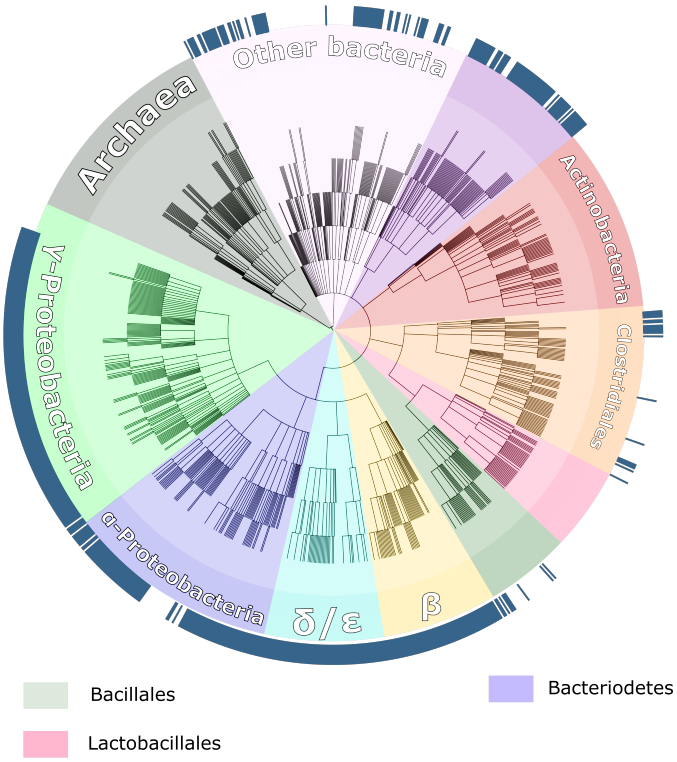


Figure 3.11: **Distribution of the Ton and Tol operons throughout the bacterial kingdom.** Taxonomic tree of assembled genomes within NCBI. Blue bars indicate a species where a Ton or Tol operon was identified. Taxonomy was taken from NCBI commontree.

3.2.6 Chromosomal or plasmid association of NB genes

The CLBs of *E. coli* (colicins) are plasmid encoded genes whereas the CLBs of *P. aeruginosa* (pyocins) are located on the chromosome. Due to the varying levels of assembly of the genomes it can be difficult to assign genes to chromosomal or plasmid localisation. To distinguish between chromosomal or plasmid encoded NBs genes two protocols were designed. The Plasmid finder database uses the Carattoli typing scheme for plasmids of the *Enterobacteriaceae*. The plasmid finder database identified replicons and conserved plasmid regions within NB containing contigs (Carattoli *et al.*, 2014). Surprisingly the Plasmid finder database could not identify a large proportion of colicins which based on literature are plasmid bourne elements. Addition of a conserved region identified in the pColE1 replicons allowed the typing of an additional large tRNase containing plasmid (Rijavec *et al.*, 2007). The Caratolli typing scheme used by the plasmid finder database does not identify non-*Enterobacteriacae* plasmids. To assign plasmid or chromosomal localisation to these species whose plasmids cannot be identified using the Caratolli typing scheme, a python script was designed and implemented to measure the percentage alignment of a NB containing contig to a database of known plasmids (Section 2.18).

As expected colicins and pyocins have a strong association with plasmid and chromosomes respectively (Figure 3.12). NBs from *Yersinia* spp. demonstrated little to no association to plasmids suggesting a chromosomal location. *Enterobacter* had a bimodel distribution suggesting both plasmid and chromosomal NBs. Using NCBI fully assembled genomes, it was confirmed that *Enterobacter* contains both plasmid and chromosomally encoded CLBs. Four genera were predicted to have chromosomally encoded NBs: *Pseudomonas*, *Serratia*, *Yersinia* and *Enterobacter*. A sequence with high identity to cloacin DF13 (CloDF13), a ~60 kDa rRNase NB originally identified in *Emterobacter cloacae*, was identified in multiple genera. MAUVE 2.0 was used to compare a plasmid in seven species that included transfer and mobilisation genes with >90% sequence identity to pCloDF13, which has been shown to be a broad range transmissible plasmid. pCloDF13 was identified in *Enterobacter cloacae*, *Enterobacter aerogenes*, *Escherichia coli*, *Klebsiella pneumoniae* and *Salmonella enterica*.

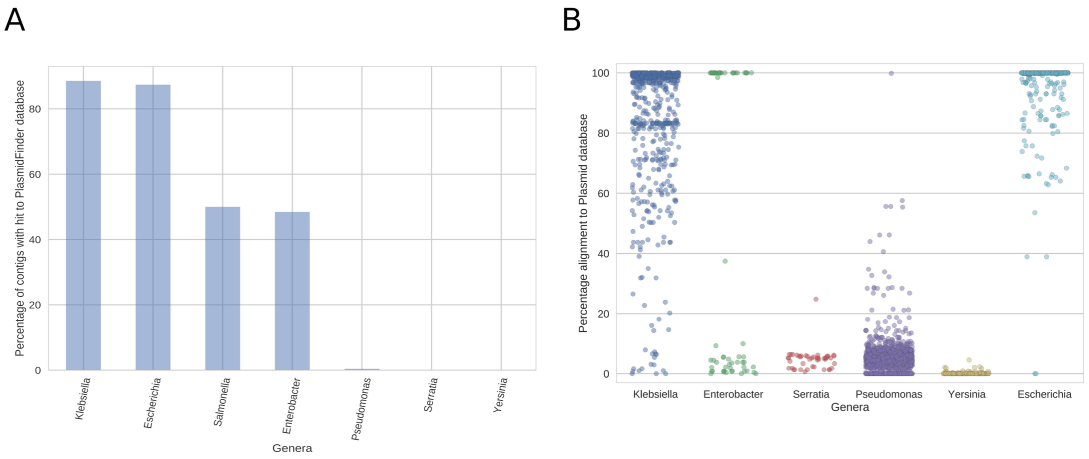


Figure 3.12: **Association of NB containing contigs to plasmid sequences *A***. NB containing contigs were aligned to the Plasmid Finder database Bars show the percentage of contigs where a conserved plasmid sequence could be identified. ***B***. Association of NB containing contigs to a database of known plasmid sequences. Each point represents an individual contig and the percentage of its association.

3.2.7 Delineation of C-terminal cytotoxic domains from CLB sequences

Polymorphic toxins share the trait of N-terminal domains associated with translocation and C-terminal cytotoxic domains. These cytotoxic domains can frequently undergo recombination and for nuclease bacteriocins are often well studied enzymes. The investigate the diversity of the translocation and receptor domains, cytotoxic domains were removed from NB sequences.

To remove cytotoxic domains, sequences were aligned to the cytotoxic domain profile used to originally identify them. Regions of sequences that aligned to the profile were removed using custom python scripts and the easl package. To test if this effectively removed the cytotoxic domain, colicins which have had their domains mapped were used. Applying this to colicin E9 (pdb: 5EW5) predicted the cytotoxic domain within 3 residues of the experimentally derived domain. Applying the same procedure to colicin E3, which uses a rRNase cytotoxic domain, predicted the cytotoxic domain within 9 residues of the experimentally derived domain (Figure 3.13).

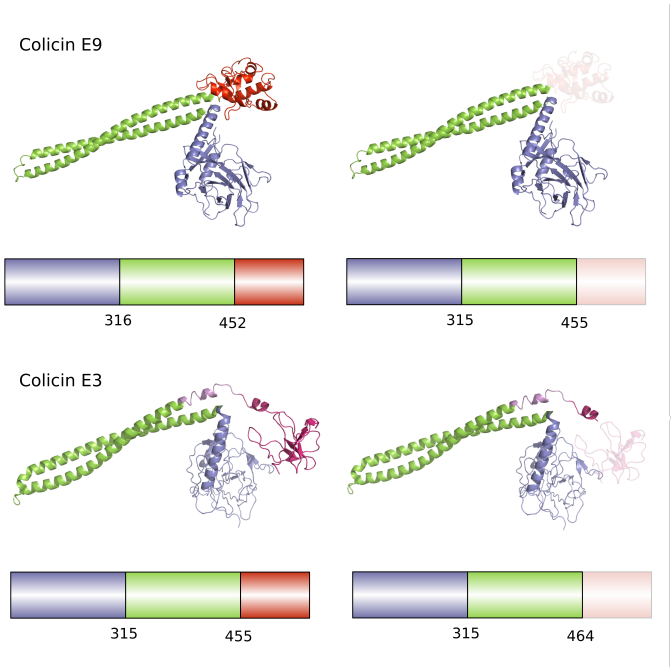


Figure 3.13: **Removing C-terminal cytotoxic domains using pHMM.** Cytotoxic domains were removed using pHMMs. By removing the sequence which overlaps with the cytotoxic pHMM we could accurately remove cytotoxic domains within 10 residues. *Left*: Domains predicted by structural and biochemical studies. *Right*: Predicted 'TR' region after removal of the cytotoxic domain.

3.2.8 Domain organisation of NBs

There has been no systematic *in silico* analysis on the diversity of NBs across the entire bacterial kingdom. The low sequence conservation of NBs (<20%) does not allow for a meaningful analysis by sequence alignment. To overcome this, protein sequences were analysed for structural elements and motifs associated with NBs. These features include: coiled-coil regions which are often present in the receptor domains of NBs, intrinsically disordered regions present in the IUTD of NBs, structural domains predicted by PFAM and the 'DPY' motif, a domain found to be conserved amongst NBs which is discussed in greater detail in Chapter 5 (Sano *et al.*, 1993a).

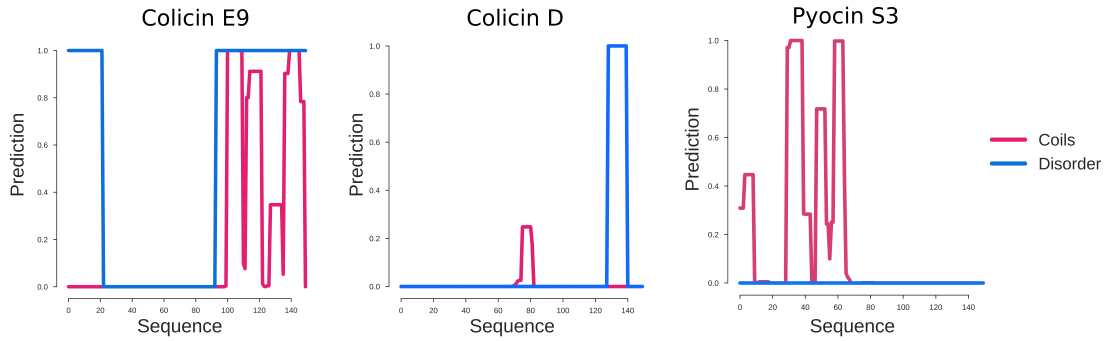


Figure 3.14: **Prediction of disorder and coils** Regions of disorder and left handed coiled-coil regions are predicted using either IUPRED or Coils, respectively. To avoid the effects of sequence length, all sequences were normalised into 150 discrete points. Positions of coiled-coil regions and the presence of N-terminal disorder differ between NB sequences and can provide the basis of the system for classification.

The well studied E-type nuclease colicins (ColE2,3,4,5,6,7,8,9) contain an N-terminal region of disorder and a central coiled-coil receptor domain. This translocation-receptor-cytotoxic domain organisation is often cited as the structure of NBs particularly for colicins but sequence analysis showed the majority of NBs, excluding those from some *Pseudomonas* species (2866: ~93%) could be divided into 8 different domain organisations (Figure 3.15. Six of the eight groups contain NBs from *Enterobacteriaceae* and two are specific to *P. aeruginosa*. These groups are not necessarily comprised by NBs from a single genus (e.g. group V includes *E. coli, Serratia marcescens, Salmonella enterica* and *Klebsiella pneumoniae*) but can contain a single species or genus (e.g. group VI contains only *Salmonella enterica* and group IV only *Yersinia* spp). Species themselves are not restricted to encoding NBs from a single group (*Klebsiella pneumoniae* encodes NBs from

groups V and VII). The sequence identity within individual groups is relatively high (40-90% average identity across a group), the sequence identity between groups is very low (<20%).

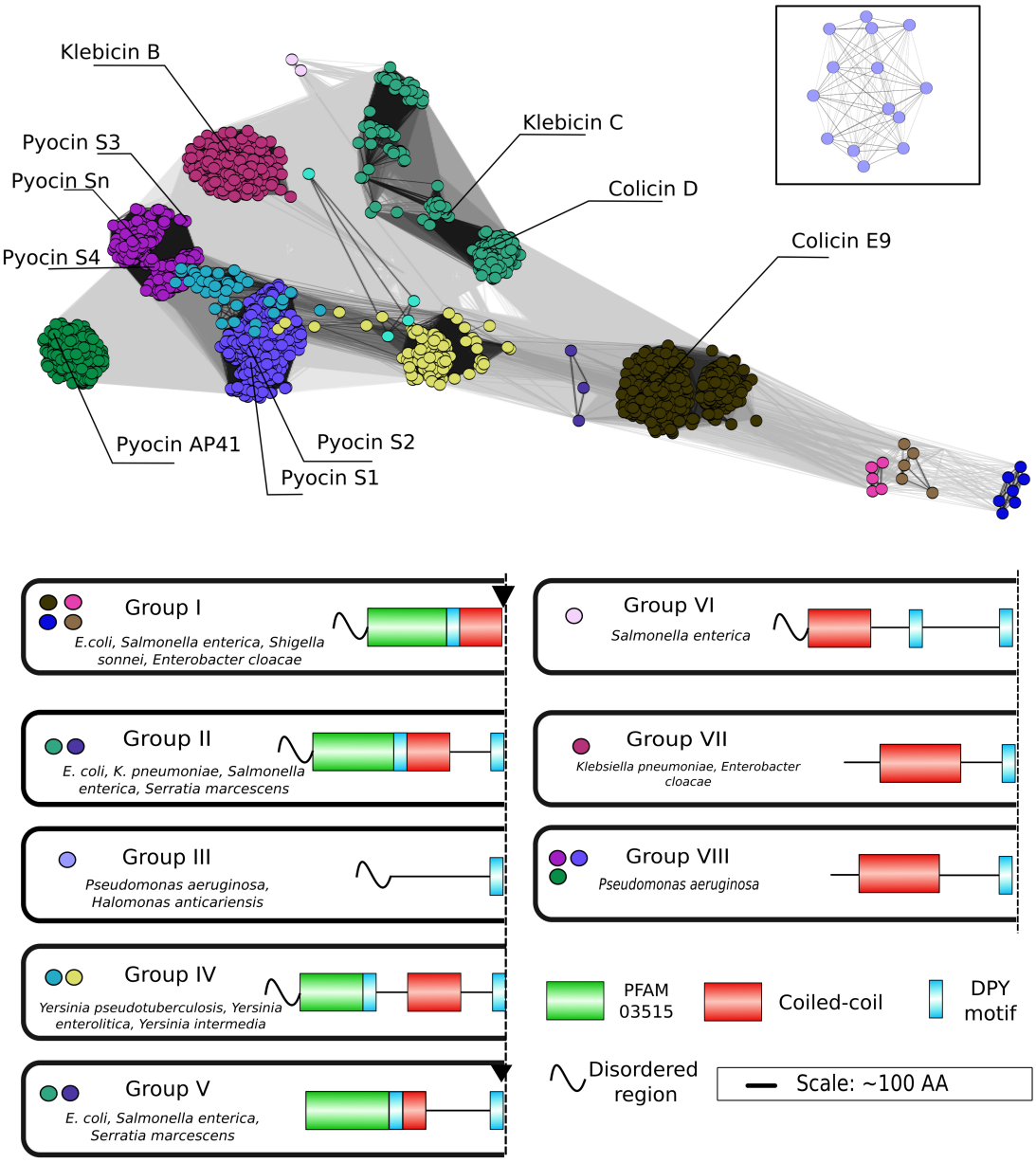


Figure 3.15: **NB gene clustering and domain organisations** *Top panel*: Sequences from *Enterobacteriaceae* and *Pseudomonas aeruginosa* with cytotoxic domains removed (Figure 3.13) clustered by CLANS (Frickey and Lupas, 2004). *Bottom Panel*: Domain organisation predicted for NB sequences using PFAM profiles, coiled-coil regions, disordered regions and the DPY motif. Block arrows indicate known processing sites which have been identified for a few well studied colicins (Chauleau *et al.*, 2011; Mora and de Zamaroczy, 2014). Block inset shows pyocin S9 sequences which are too distantly removed to show at the correct distance. Connections show the High Scoring Segment Pairs (HSP) of $<5x10^{-20}$. Previously identified NBs are indicated as reference points.

E. coli NB which use the Tol system were first identified in the 1970s and are commonly referred to as the 'E-type' colicins (Meyhack *et al.*, 1973; Schaller and Nomura, 1976). They are the most well studied group of NBs (Papadakos *et al.*, 2012) with crystal structures of rRNase ColE3 and DNase NB ColE9 and numerous biophysical studies (Klein *et al.*, 2016). The present analysis places these NBs in group I alongside NBs from *Klebsiella pneumoniae*, *Serratia marcescens*, *Shigella sonnei*, *Enterobacter cloaceae*, *Xenorhabdus*, *Citrobacter* and *Morganella*. The structural features of group I NBs include a T-domain composed of a disordered N-terminus adjoining a folded domain identified within the Pfam database as PF03515 and a central R-domain comprised of a coiled-coil region. E-type colicins use the R-domain to bind the outer membrane B_{12} receptor BtuB while the disordered N-terminal regions are involved in binding OmpF and the periplasmic protein TolB (Housden *et al.*, 2013, 2010). An N-terminal disordered region was identified in five of the eight groups and it is likely that these regions are involved in translocation for all of the NBs which contain them. Once in the periplasm NBs translocate into the cytoplasm using the AAA+ ATPase/protease FtsH. FtsH also proteolytically cleaves the NB releasing the cytotoxic domain. Identified cleavage sites for group I and V NBs are indicated in Figure 3.15 (Chauleau *et al.*, 2011; de Zamaroczy and Buckingham, 2002; Mora and de Zamaroczy, 2014).

K. pneumoniae, is an important human pathogen, a member of the ESKAPE pathogens which covers a group of bacteria which cause nosocomial infections and have a high prevalence of multidrug resistance isolates (Santajit and Indrawattana, 2016). Multiple NBs were identified within *K. pmeumoniae* genomes and NBs from klebicins were found to have domain organisations belonging to groups I, II and VII. The ppHMM pipeline identified, Cloacin DF13 (group I), klebicin B (group VII), klebicin C (group II) and two novel DNases which share <50% similarity to klebicin B or klebicin C at there N-terminus. Group VII NBs have a large domain of coiled-coil similar to group I, however it is found at the N-terminus. Group II NBs differ by containing a much smaller coiled-coil region. Group V NBs also contain a similar small coiled-coil region but differ from group II in the presence of N-terminal disorder and overlap in terms of sequence clustering by CLANS. Group II sequences also contain two copies of the 'DPY' motif whereas group VII NBs only contain one.

Serratia marcescens is a ubiquitous environmental organism that is becoming increasingly resistant to antibiotics and a major cause of healthcare acquired infections (Dessi *et al.*, 2009; Gruber *et al.*, 2015). Currently only one bacteriocin has been identified within *Serratia*, the pore forming bacteriocin N28b (Viejo *et al.*, 1992). The NB pipeline identified 11 families (groups of NBs clustered at 90%) of NBs within 44 *Serratia* genomes, all of which had no significant similarity to N28b. Eight of the eleven clusters were predicted to be chromosomally encoded and have domain organisation which places them within group V. Three of the NBs have higher sequence identity to cloacin DF13 but contain a repeat of the 'DPY' motif which places them within group II. Bacteriocins which utilised a DNase or tRNase cytotoxic domain were interchangeable, suggesting recombination of cytotoxic domains, a trait associated with polymorphic toxins.

The genus *Yersinia* has been shown to produce two bacteriocins as well as a plethora of bactericidal proteins. Pesticin from *Yersinia pestis* targets the periplasm and degrades peptidoglycan using muramidase activity and colicin F_Y is a pore forming bacteriocin with broad specificity, but no NBs have previously been identified within *Yersinia* (Bosák *et al.*, 2012; Vollmer *et al.*, 1997). The NB pipeline identified 18 NB families in 8 species of *Yersinia* with no similarity to either pesticin or colicin F_Y. NBs from Yersinia clustered separately to NBs from other *Enterobacteriaceae* to form group IV. Domain organisation within this group is characterised by a long ~90 residue unstructured region at the N-terminus likely to be a IUTD, a long coiled-coil domains likely to be the receptor domain and a duplicated 'DPY' motif.

Pseudomonas aeruginosa is an opportunistic pathogen, a major pathogen of the cystic fibrosis lung, a leading cause of nosocomial infections worldwide and a member of the ESKAPE group of pathogens. *P. aeruginosa* can produce numerous agents of bacterial competition including the S-type (soluble) pyocins which include NBs. An in-depth *in silico* investigation into the diversity of *Pseudomonas* ribosomally produced toxins has previously been performed (Ghequire and De Mot, 2014). Similarly to this earlier study, the ppHMM identified pyocin S1-9 (excluding the pore former pyocin S5) and pyocins S11-12 (also known as SD2 and SD3). The pipeline excluded a protein designated pyocin S10 due to the presence of an N-terminal PAAR domain which is associated with the

T6SS. The NB pipeline also identified pyocin S13 (or SD1) which contains a colicin D-like tRNase domain associated with the N-terminus of pyocin S1 (McCaughey *et al.*, 2016b), a novel recombination termed pyocin S3C (the rRNase domain of colicin E3 associated with the pyocin S3 N-terminal domain) and a novel pyocin termed Sn which had little sequence identity to pyocin S3 in the first ~200 aa before utilising the central and cytotoxic domains of S3. All pyocins apart from S9 form a single group (group VIII), which is defined as no detectable disorder, a single DPY motif and a large coiled-coil domain at the N-terminus. Recombination of cytotoxic domains appears rife amongst *Pseudomonas aeruginosa*. It has been suggested that the high prevalence of pyocins within *P. aeruginosa* and especially the high prevalence of multiple pyocin genes within a single genome would increase the probability of recombination (Riley, 1998a)

NBs were also detected in 20 *Pseudomonas* species other than *P. aeruginosa*. NBs which contain a rRNase domain were identified in *P. fluorescens*, *P. synxantha*, *P. poae*, *P. brassicacearum* and 10 *Pseudomonas* sp. isolates. 175 non-*aeruginosa Pseudomonas* spp. contain a DNase or rRNase domains (51.5% and 21% respectively) (Figure 3.16). In agreement with previous analysis no tRNase domains, either ColD-like or ColE5-like, were identified in *Pseudomonas* other than *P. aeruginosa* species.

3.2.9 Cytotoxic activity of a predicted NB from *P. aeruginosa*

The ability of the pipeline to accurately identify novel NBs could potentially increase the repertoire of potential antimicrobials for clinically relevant bacteria. To validate the pipeline a novel bacteriocin in combination with its immunity protein, identified within *P. aeruginosa* was cloned, expressed and purified (courtesy of Dr N. Housden). This novel pyocin, termed pyocin Sn from *P. aeruginosa* PA1R (GenBank: AHA26272.1, 832 residues) resides within group VIII, has of no detectable disorder domain, a large coiled coil region and a cytotoxic domain which utilises a non-HNH type DNase domain. Purified Pyocin Sn had concentration dependent cytotoxic activity against three strains of *P. aeruginosa*, therefore validating the pipelines ability to identify novel NBs with antimicrobial activity.

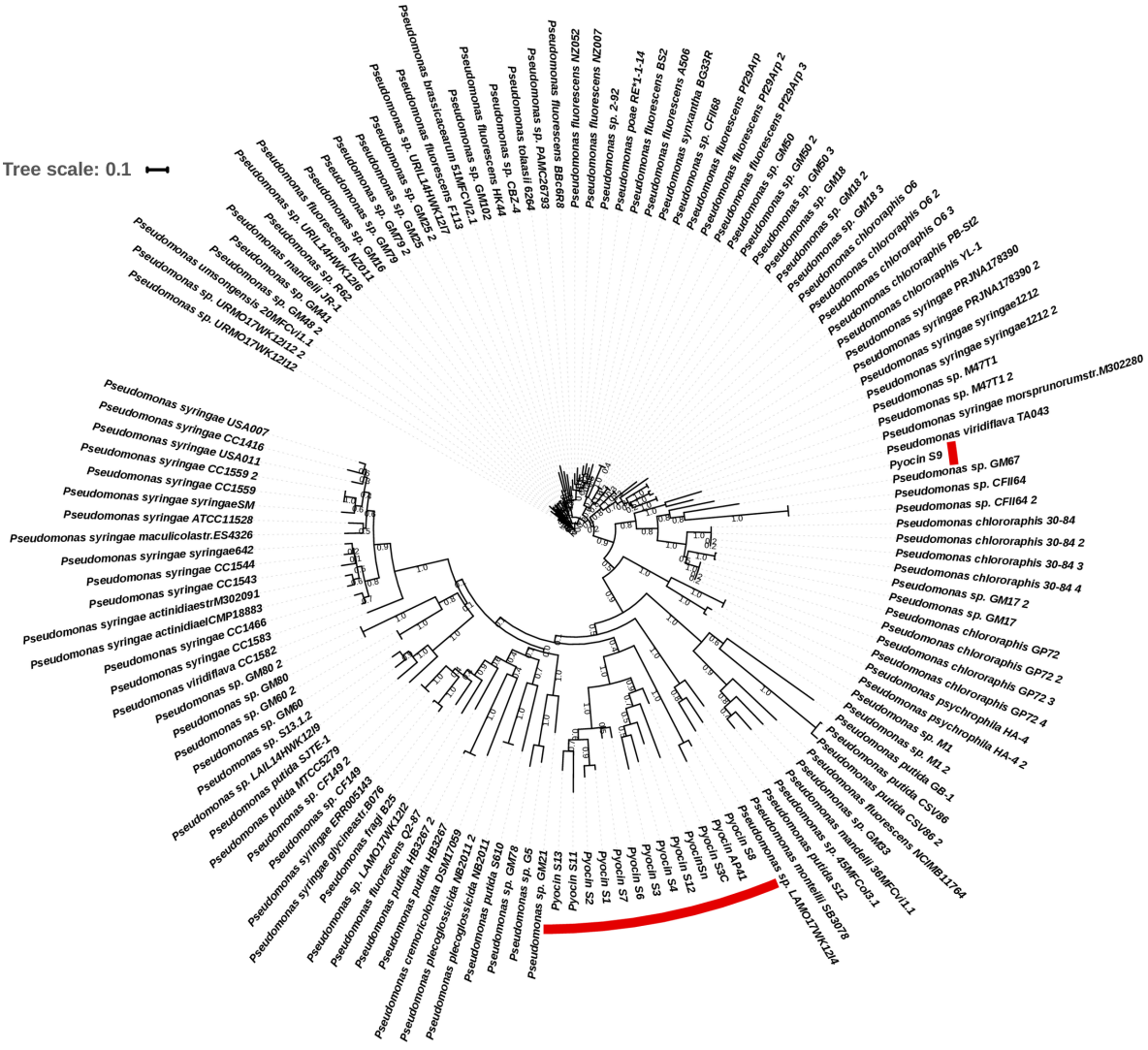


Figure 3.16: **Phylogenetic analysis of NBs identified within *Pseudomonas spp.*** Sequences of TR domains from *Pseudomonas* spp. were aligned using MUSCLE and trees built using a neighbour joining algorithm with 1000 bootstrap replicates, indicated by a fraction (Mega 6.0). Pyocins of *P. aeruginosa* are indicated by a red bar.

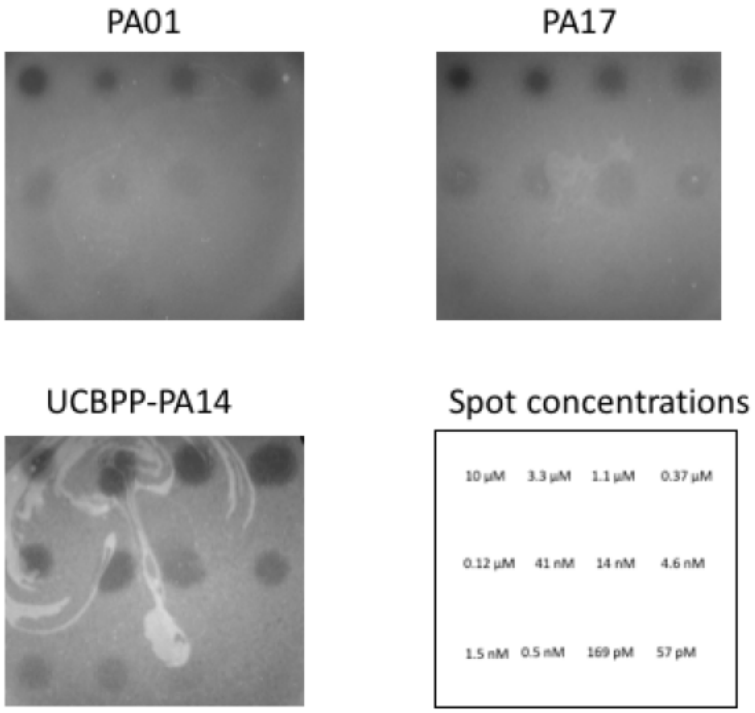


Figure 3.17: **Cytotoxic activity of a novel predicted pyocin** Novel pyocin in combination with its immunity protein, pyocin Sn, was spotted onto lawns of three *P. aeruginosa* strains. Cytotoxicity was measured as a zone of clearance on the lawn (Performed as described in section 2.7).

3.2.10 Pangenome analysis of colicinogenic bacteria

Many studies have focussed on the relation between colicins and virulence factors, primarily using PCR based methods and calculating a simple association (Micenková *et al.*, 2014b; Petkovsek *et al.*, 2012). Whilst this approach has provided some interesting observations, it has certain flaws. A PCR method can only identify a small number of genes (those targeted with primers). Secondly, without knowing the underlying phylogeny of the organisms, relatedness can add a confounding effect to the analysis. Two genes could be linked through a linear ancestry but this association could be specific to a single lineage and not indicative of a true genetic association (Figure 3.18).

To overcome these limitations, a collection of 357 *E. coli* genomes available within the Short Read Archive (SRA) was selected. A pangenome analysis was performed to identify all ORFs and genetic association between those open reading frames and colicins with a HNH-DNase, tRNase or rRNase cytotoxic domain. The underlying phylogeny was also

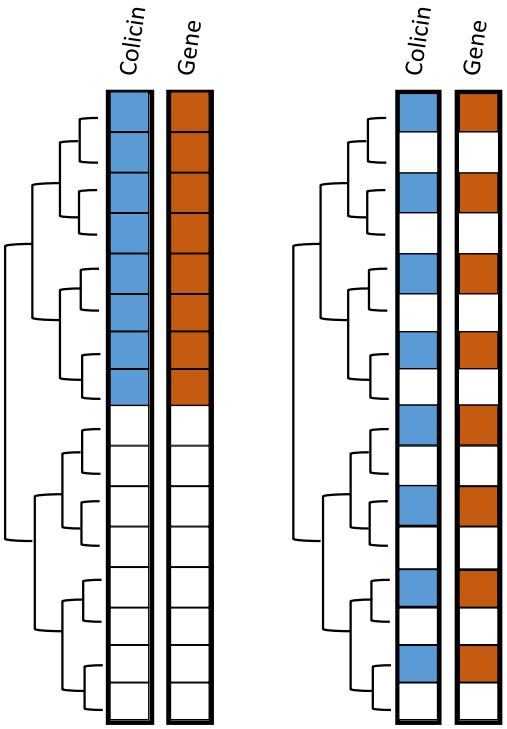


Figure 3.18: **Confounding effects of phylogeny on identifying gene association.** Genetic relatedness can diminish the ability to identify genetic association. *Left*: Whilst the gene is significantly associated with the trait of colicinogenicity (Fishers text), it is due to an ancestor containing both the colicin gene and the gene of interest which is passed linearly. *Right*: Here the same the association is just as significant by the Fishers test, the association has occurred in different lineages and suggests true genetic association

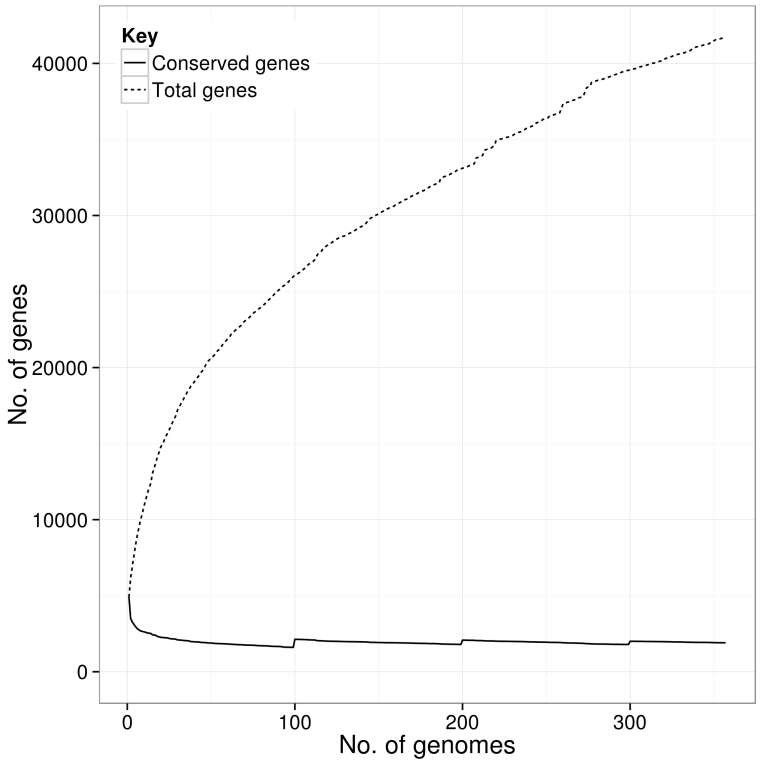


Figure 3.19: **Distribution of total and conserved genes** Increase and decrease of total and conserved genes with increasing genomes (respectively).

taken in account to increase the statistical power to identify true genetic association.

Roary pangenome software identified 41,723 gene clusters in the 357 genomes (using a 95% sequence identity cut off), 1905 core genes (100-99%), 739 soft core genes (99-95%), 3,500 shell genes (95-15%) and 35,579 cloud genes (15-0%). The size of the core genome is similar to other estimates (Kaas *et al.*, 2012). Plotting conserved and total genes against number of genomes gave a U-distribution typical of pangenome analysis (Figure 3.19).

Two controls were used, a strong association between the colicin gene and the immunity protein and no association between colicins and an essential gene, *dnaC*. All cytotoxic motifs when tested had most significant association to both a colicin gene and an immunity protein ($p < 0.00005$). No significant association was observed with *dnaC*.

Analysis identified 28 genes (not including the colicin genes, immunity proteins or plasmid mobilisation genes) which were significantly associated with the trait of colicinogenicity (Table 3.3, Figure 3.20). These genes included virulence factors, toxin/anti-toxin

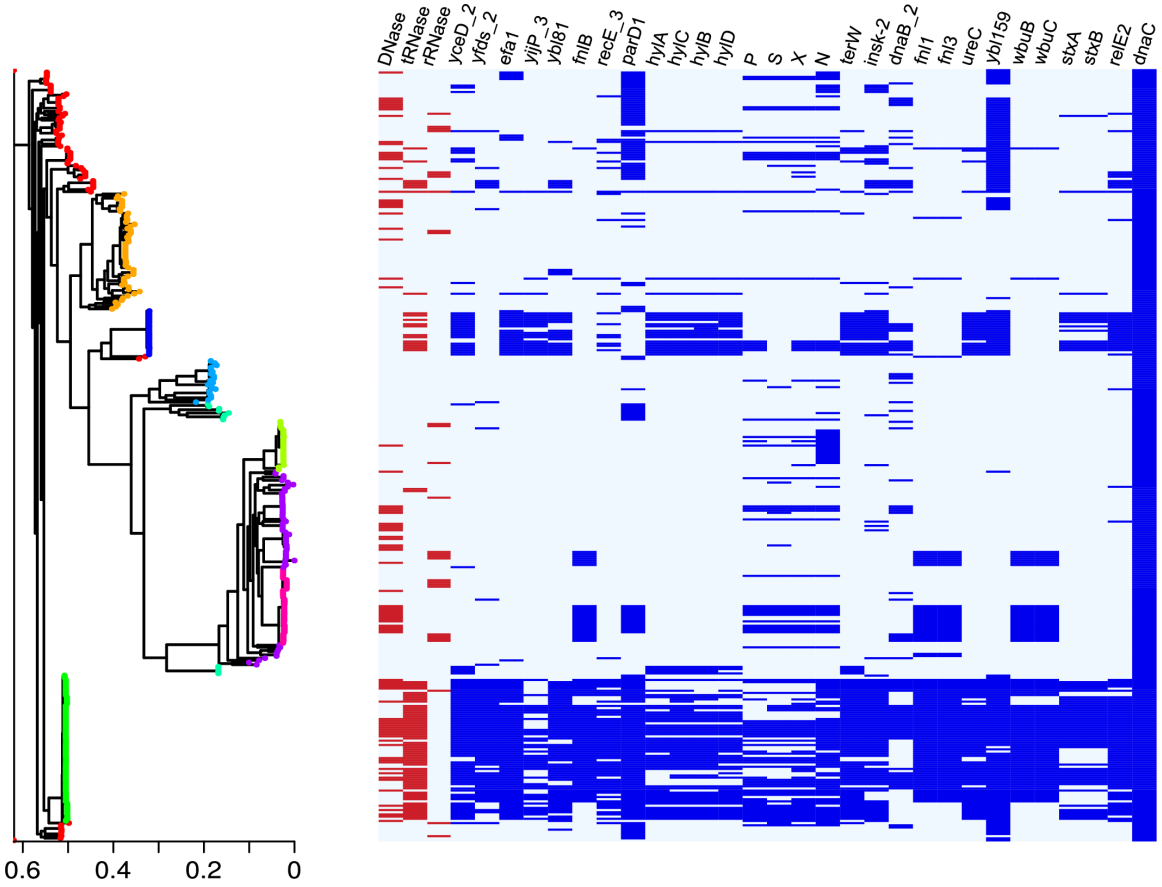


Figure 3.20: **Gene association as revealed by pangenome analysis.** *Left:* Phylogeny of the 357 strains used in the study, coloured nodes indicate the strata as predicted by BAPS. *Right:* Presence of colicin genes, split into three cytotoxic domains, are indicated by red blocks. Genes associated with colicin genes were represented by blue blocks where a gene is present. *dnaC* is a essential gene used as a control.

modules and genes involved in LPS or O-antigen production. Virulence factors were identified, specific to certain *E. coli* pathotypes. Shigella toxin producing *E. coli* are responsible for causing bloody diarrhoea and heamolytic uremic syndrome (Hunt, 2010). The gene yjiP has been shown to aid in the invasion of brain endothelial cells by *E. coli* K1 (Wang *et al.*, 1999). *efa1* encodes the *E. coli* factor for adherence, an adhesin which enhances the adhesion of Enteropathogenic *E. coli* and contributes to colonization of bovine intestines (Deacon *et al.*, 2010). Other virulence factors were not specific to a certain pathotype. Ureases increase the ability of *E. coli* to survive in the host and have a confirmed role in causing disease in many other species (Konieczna *et al.*, 2012). Hemolysin is a major virulence factor of UPEC and is enough to confer pathogenicity to

non pathogenic strains *in vivo* (Elliott *et al.*, 1998).

Table 3.3: **Genes with a significant association to colicins** Gene clusters (as predicted by the pangenome software Roary) which show significant association to colicin genes. Only annotated genes are shown, all 28 genes are shown in Figure 3.20

Cluster	Description	Accession No.
Virulence factors		
yijP3	Uncharacterized protein. Shown to be involved in invasion of brain endothelial cells	WP_000550555.1
stxA/B	Shigella toxin/antitoxin	WP_000691354.1
terW/E	Tellurium resistance protein	WP_001176768.1
ureC	Urease	WP_001301487.1
hylA/B/C/D	Hemolysin operon	WP_000217740.1
efa1	*E. coli* factor for adherence	CTS96436.1
O-antigen biosynthesis		
fnlB	UDP-L-FucNAc biosynthesis pathway	WP_001529290.1
fnl1	UDP-L-FucNAc biosynthesis pathway	WP_000481729.1
fnl3	UDP-L-FucNAc biosynthesis pathway	WP_000734419.1
wbuB	l-FucNAc transferase	WP_000761607.1
Phage		
P	Phage terminase	WP_000613782.1
X	Phage tail	EKV72616.1
S	Phage tail	KME60810.1
N	Phage tail	WP_074468410.1

3.2.11 Testing association between NBs and receptors

CLBs are thought to target closely related bacteria and it has been hypothesised that bacteria which produce a CLB would be sensitive to that CLB, only being saved by producing the immunity protein. To test this hypothesis, analysis was performed to find an association between pyocins and their receptors in *P. aeruginosa*. *P. aeruginosa* offers a model system to test this association. Pyocins S2, S3 and S4 all target outer membrane receptors FpvAI or FpvAII (a third receptor FpvAIII also exists) which act as the receptors for three structurally distinct pyoverdines (de Chial *et al.*, 2003). Pyocins S2 and S4 target FpvAI and S3 binds FpvAII (Baysse *et al.*, 1999; Denayer *et al.*, 2007). Typically, *P. aeruginosa* strains encode a single *fpvA* type (analysis observed only 0.7% of strains which contained more than one *fpvA* type). Pyocins were identified using the NB pipeline and outer membrane receptors identified by blastn (2.28). To simplify analysis, strains which contained multiple pyocins were removed. Analysis identified *fpvAI* in 33% of *P. aeruginosa*, *fpvAII* in 37% and *fpvAIII* in 19% of bacteria. Testing for association observed a significant association between pyocin production and receptors. Pyocin S2 producing strains encoded the target receptor *fpvAI* in 65% of strains, pyocins S3 producing strains encoded its target receptor *fpvAII* in 88% of strains and pyocin S4 producing strains encoded the receptor *fpvAI* in 96% of strains (Figure 3.21). Pyocins which have no identified receptor were also tested for significant association to an *fpvA* or *fpvB* gene but no association was observed.

Testing if this receptor/bacteriocin association was conserved in a different model proved to be difficult as this method requires certain criteria. Firstly the bacteriocin receptor needs to be known, this greatly limits the number of bacteriocins which can be used, especially in species other than *E. coli*. Secondly, the receptor cannot be part of the core genome. Finally, the bacteriocin must be present in significant proportions of the species. Only one other bacteriocin system was identified that fulfilled these requirements, the bacteriocin Cloacin DF13 and its receptor IutA (the receptor for aerobactin) within *E. coli* (*iutA* appears to be an essential gene within *K. pneumoniae*). We found that only 33% of *E. coli* genomes which did not encode *cloDF13* encoded *iutA* however within the sub population of *cloDF13* encoding bacteria, 71% of genomes encoded *iutA*.

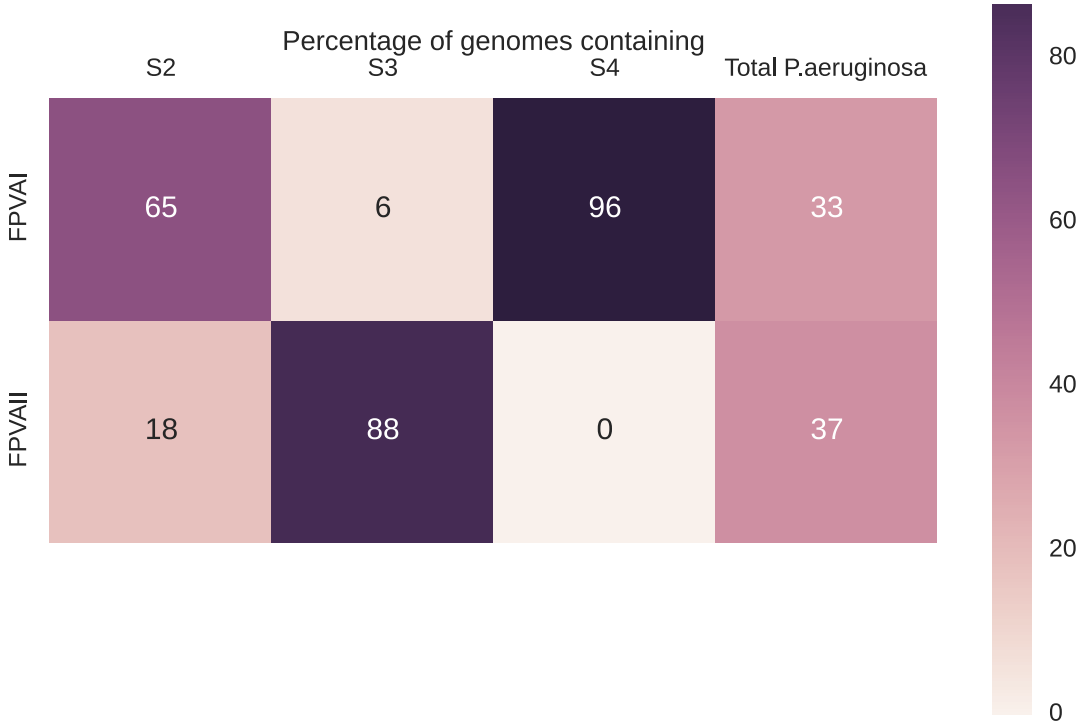
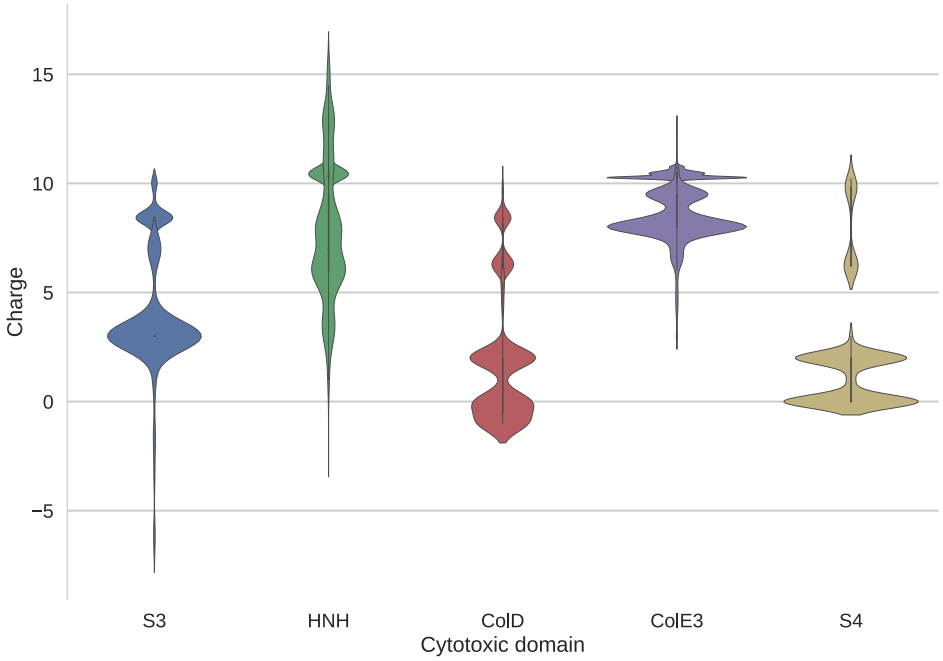


Figure 3.21: **Association between pyocins and their receptors** We observed that *P. aeruginosa* genomes which contained either pyocins S2, S3 or S4 (*x* axis) were significantly more likely to also encode the receptor targeted by the pyocin they encode (*y* axis.).

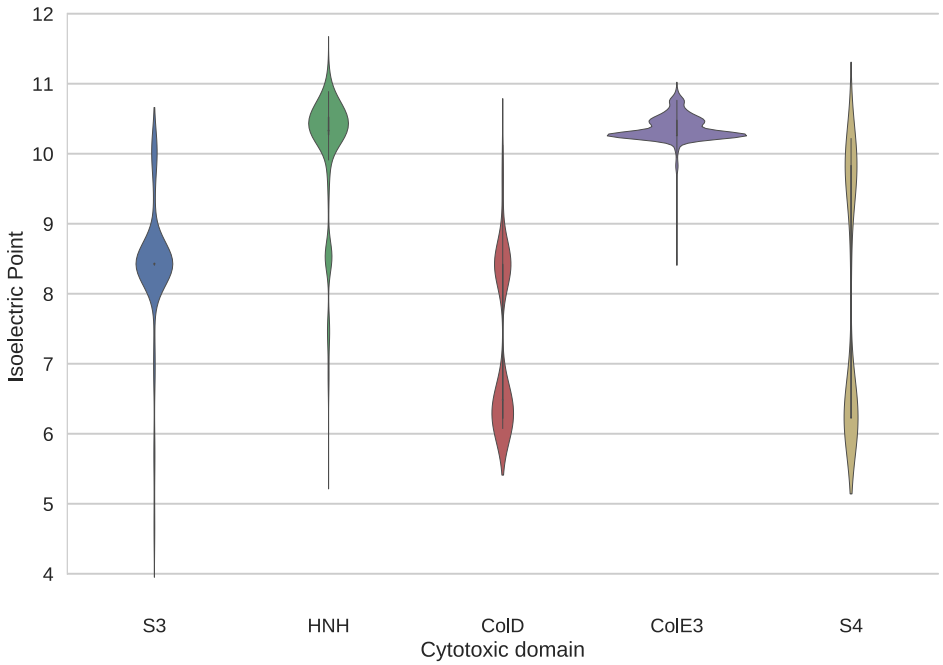
3.2.12 Relation between charge and cytotoxic domain

The electrostatic charge of the cytotoxic domain has been implicated in translocation across the inner membrane (Walker *et al.*, 2007). The rate of cell killing was shown to be proportional to the overall positive charge in HNH DNases. Furthermore, the tRNase domain of colicin D which only has a net charge of +2, requires an additional interaction at the inner membrane, not observed in HNH DNases or E3 rRNases (+11) (Chauleau *et al.*, 2011). Cytotoxic domain sequences were tested to verify if differences in charge were observed across all NBs.

Unsurprisingly we observed that rRNases and both HNH DNases have a high positive charge suggesting that they can interact with the inner membrane through electrostatics (Figure 3.22). Non-HNH DNases have a slightly lower charge. To date there has not been a study which analyses the interaction of this domain with the inner membrane but its charge suggests it may require an additional interaction. Both tRNases showed a lower charge however, both had a highly charged sub-population which suggests that addition interactions with the inner membrane are not always necessary for these cytotoxic domains.



(a) Distribution of overall charges of cytotoxic domains



(b) Distribution of the isoelectric point of cytotoxic domains

Figure 3.22: **Predictions of charge and isoeletric point for cytotoxic domain**

3.2.13 Identification of positive selection in bacterial cytotoxic domains

Previous work has shown by measuring the non-synonymous vs synonymous mutations, that colicins undergo positive selection at the immunity binding domain (Riley, 1993b). Using HNH domains for NBs from *E. coli*, positive selection of the cytotoxic domain was demonstrated to be almost exclusively within residues which interact with the immunity protein. The software OmegaMap was used to predict positive selection at individual residue resolution.

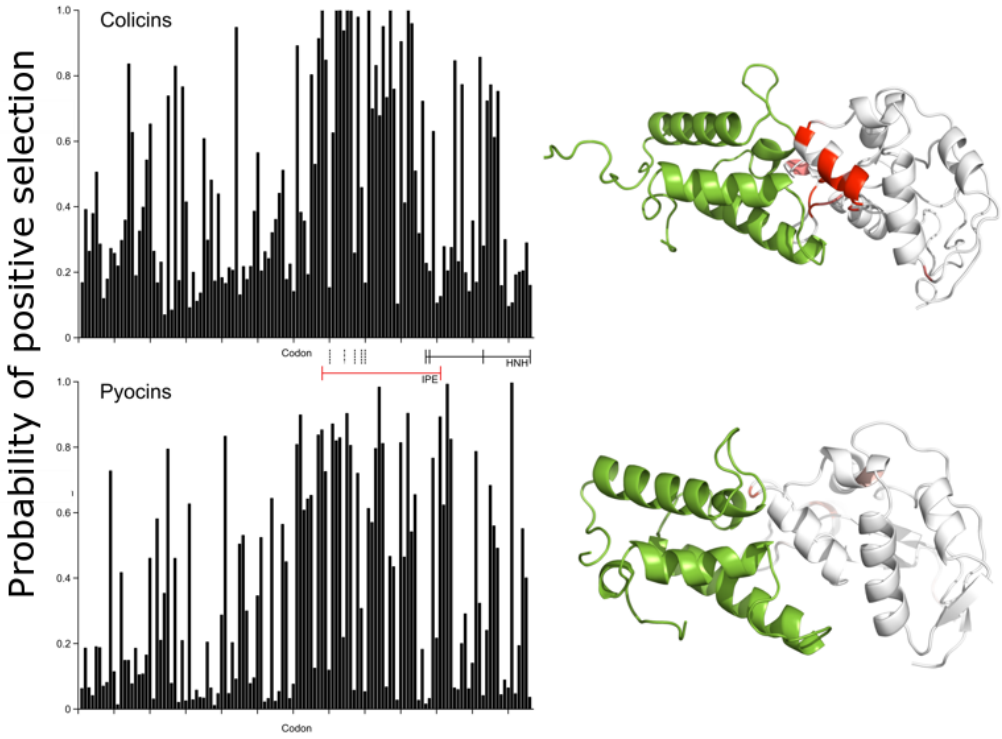


Figure 3.23: **Prediction of positively selected sights for HNH DNase cytotoxic domains** Residues under positive selection (Red) within the DNase domain (white) are mapped onto the crystal structure of colicin E2 bound to its immunity protein (3U43) IPE denotes the immunity protein exosite

Within the colicin DNases, not only were residues under strong positive selection identified, these residues were shown to have specific interactions to the immunity protein. Performing identical analysis on the cytotoxic domain of AP41 HNH DNases however, did not identify the same 'patch' of positively selected residues. This supports the hypothesis of Riley which states that as pyocins are more abundant in populations and more likely to be present with other nucleases within a genome, the driving force behind their evolution would be diversifying recombination and not positive selection (Riley, 1993b).

3.2.14 Identification of pore forming bacteriocins

The ppHMM strategy focusses solely on nuclease bacteriocins as these contain enzymatic motifs that are readily identifiable. Pore forming bacteriocins (PFBs) do not contain any enzymatic motifs but instead insert themselves into the inner membrane. Within colicins, pore forming domains contain two hydrophobic helices surrounded by a shell of amphipahtic helicies. PFB operons do not encode their immunity proteins genes immediately downstream of the toxin. The lack of an identifiable motif or downstream immunity protein limits the usefulness of a ppHMM strategy. PFB only have two recognizable traits which can be used for identification: a C-terminal cytotoxic domain and similarly to NB, a lack of enzymatic domains or signal sequences. To identify the transmembrane helicies of the PFB cytotoxic domain, TMHMM2.0 was used to identify transmembrane helicies within open reading frames. TMHMM2.0 is a sophisticated HMM approach trained on 160 transmembrane proteins and models the 'core' as a 5-25 residue helix. As a control, eight known PFB sequences were analysed to determine the relative positions of the transmembrane helicies. TMHMM2.0 was able to identify transmembrane helicies within PFBs. All TM helicies were identified within the last 3/7th's of the sequence and this was used as a cut-off for testing for PFBs. PFB standard sequences were also analysed for PFAM profiles using Pfam 27a (Figure 3.24).

The ORFs for *Pseudomonas aeruginosa*, *Escherichia coli*, *Klebsiella pneumoniae* and *Yersinia* spp. within PATRIC2 were analysed for pore forming bacteriocins. 211 *Pseudomonas aeruginosa* genomes were downloaded from the PATRIC2 database however using the pipeline only one class of pore forming bacteriocin was identified, pyocin S5. For *K. pneumoniae* and *Yersinia* sp. genomes no pore forming bacteriocins were identified. *E. coli* genomes produced 11 families of pore forming bacteriocins (ColS4, ColN, ColB, ColIa, ColIb, ColU, ColA, ColY, ColK, ColE1, ColS5). It is perhaps unusual to note that *E. coli* genomes encode a plethora of pore forming bacteriocins, whereas they appear to be very few in *Pseudomonadaceae* and are absent *Klebsiella*. NCBI fully assembled genomes were also scanned and, while the size of this database is only a fraction that of the PATRIC database, pore forming bacteriocins were identified in *Pseudomonas fluorescens* and in *Proteus* sp. genomes.

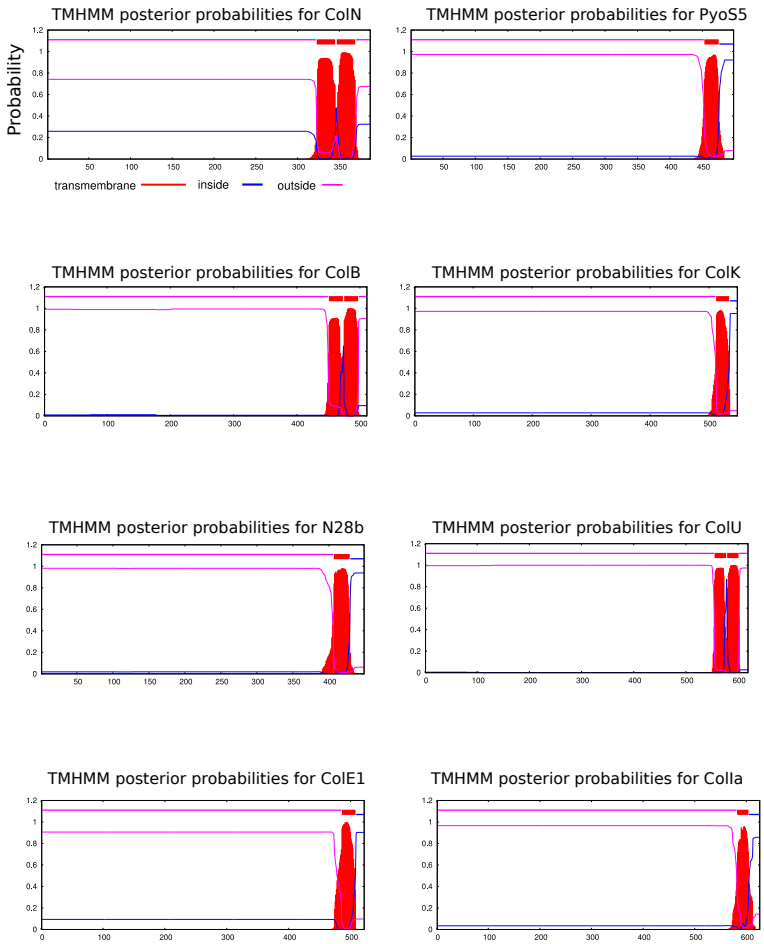


Figure 3.24: **Prediction of TM helicies in pore forming bacteriocins** TMHMM predictions of transmembrane helicies for eight PFBs. Trans membrane helicies are located at the C-terminus of the protein.

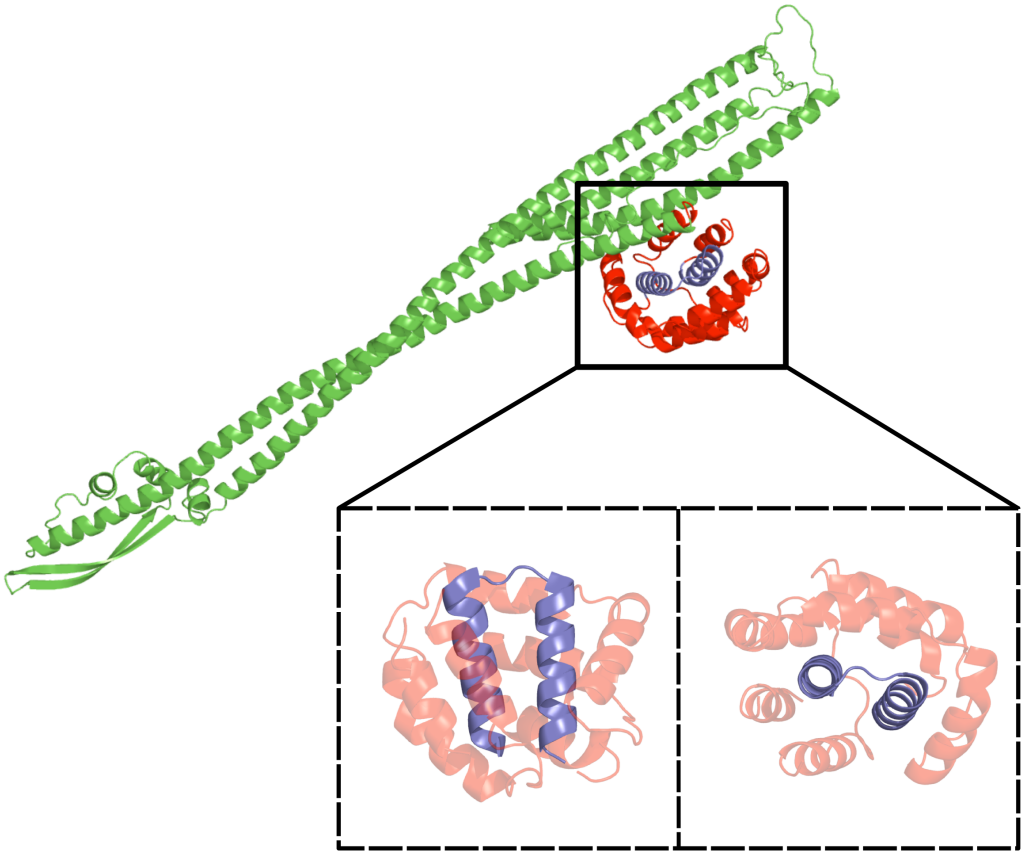


Figure 3.25: **Crystal structure of the pore forming colicin, ColIa** Crystal structure of the pore forming colicin Ia, (1CII). The cytotoxic domain consist of two hydrophobic transmembrane helices (purple) surrounded by a shell of amphipahtic helicies (red).

The computational expense of this pipeline is inhibitory to running it on the full pubMLST database, however it can be used as an effective pipeline to identify pore forming colicins in assembled genomes.

3.2.15 Using the ppHMM pipeline to identify non-NB proteins

Using a combination of profile Hidden Markov Models with defined inter-profile distance metric is a powerful way to identify proteins in large genome databases. In principle this approach could be used to identify any protein or proteins which contain conserved domains which are genetically linked. The paired profile Hidden Markov Model method (ppHMM) was used to identify zinc binding anti-sigma factor (ZAS) proteins and extracytoplasmic function sigma factors (ECF) within bacteria (Hughes and Mathee, 1998; Rajasekar *et al.*, 2016). There has been no previous systematic bioinformatic analysis

of these proteins in bacteria. Anti-sigma factors and sigma factors were identified by zf-HC2 (PF13490.1) or Sigma70_ECF (PF07638.6) respectively before imposing an inter-profile distance metric of 300 bases. NCBI fully assembled genomes database was used to identify >1100 genomes with Zas anti-sigma factors and ECF sigma factor proteins. Zas/ECF systems were widely distributed across multiple phyla. Further to just identifying Zas/ECF domains and underlying their importance across the bacterial kingdom, sequence analysis was used to support the hypothesis that HHCC-type Zas domains require additional domains to act as functional sensors of disulfide or peroxide stress whilst CHCC-type Zas domains do not. Using the >1100 Zas proteins identified a significant association between HHCC-type ZAS domains and cupin-7 vs CHCC-type Zas domain proteins (44/498 vs 1/655) was observed (Rajasekar *et al.*, 2016).

3.3 Discussion

Bacteriocins were one of the first bacterial competition systems identified, however, our understanding of this family of proteins remains limited. With only a handful of known NBs, understanding the role those proteins have within the environment or the range of potential targets and translocation systems is unclear. To date, there has not been a comprehensive *in silico* study to understand the diversity of nuclease bacteriocins across the entire bacterial kingdom. Using the approach of 'paired-profile' Hidden Markov Models a pipeline was develop to identify NB genes and 100,154 genomes from across the bacterial kingdom were analysed. This represents not only the largest attempt to identify NB genes but also one of the largest bioinformatic analysis that has been performed on bacteria. The ppHMM pipeline was able to identify 3094 NB proteins in 53 species. The most surprising outcome of the analysis was the relatively narrow taxonomic distribution of NBs. NBs were identified only within γ-proteobacteria; specifically within the families *Enterobacteriaceae* and *Pseudomonadaceae*. Whilst there had not been any NB identified in other classes it is still surprising to see this trend when the pubMLST database holds genomic data for $>$3000 species. Two potential factors were eliminated in trying to understand the narrow distribution of NBs. Firstly, NBs could be restricted to an environment which only few bacteria could inhabit. Analysing the metadata available

for genomes within the PATRIC database showed that NB genes were identified from bacteria within hundreds of environments and particularly abundant in soil. This supports the hypothesis that NBs would have greater function in heterogeneous structured environments (where multiple microenvironments can arise) with high bacterial competition versus well mixed environments with low bacterial competition. Unfortunately the metadata, even in such a database the size of the PATRIC2 database, was not enough for a statistic analysis of the proportion of NB genes within environments. Secondly, the presence of Ton/Tol operons was tested. As all known NBs require these PMF coupled protein complexes to translocate into the cell, absence of them could limit uptake of NBs. However using a HMM approach Ton/Tol operons were identified across Gram negative bacteria.

To date only one member of the CLB family has been identified in a bacterium outside of the γ-proteobacteria class. Colicin M-like bacteriocins were identified in several *Burkholderia* strains (Ghequire and De Mot, 2015b). Unlike the NB which translocate into the cytoplasm to degrade nucleic acid, colicin M kills via cleaving the lipid II precursor of peptidoglycan within the periplasm (Barreteau *et al.*, 2012). The majority of these proteins were also predicted to be substrates of the Sec or Tat secretion pathways (Ghequire and De Mot, 2015b). It remains unclear why NB genes are exclusively found in γ-proteobacteria.

The cytotoxic domains of NBs share high similarity to a variety of cytotoxic domains used with polymorphic toxin systems. Evolved or specialised effectors are proteins with domains of T6SS structural proteins (Hcp, PAAR domains or Vgr domains) linked to cytotoxic effectors. The extensive use of colicin-like nuclease domains in other polymorphic toxin systems provides more evidence that NBs should be considered polymorphic toxins. Polymorphic toxins are also known to recombine to introduce new effector/immunity protein pairs with N-terminal domains. Within *Pseudomonas aeruginosa*, *Serratia marcescens* and *Yersinia* sp. extensive evidence of recombination was observed, however no recombination was observed in the NBs from *E. coli*. This suggests that the density of NB genes within a population (85% in *P. aeruginosa* vs <5% in *E. coli*) is important in determining the effects of recombination on NBs. Indeed, testing for positive

diversifying selection within colicins vs pyocins, showed a clear effect of positive selection for colicins but none for pyocins suggesting recombination is a major driving force of evolution for pyocins whereas positive selection drives nuclease colicin evolution (Riley, 1998a,b).

Sequence similarity within the TR domains of NBs is low (<20%), which not only limits their detection by alignment methods, it also limits the ability to accurately align sequences and perform a standard analysis of phylogeny. Therefore to perform an accurate and informative analysis of NB diversity, a pairwise similarity method was used. This has a major advantage over alignment and phylogeny methods. It uses unaligned sequences and does not make any inferences regarding deeper phylogeny, both of which could not be considered accurate when dealing with such a large number of sequences with such low diversity. Biophysical studies on NBs have revealed structural features involved in NB translocation; coiled-coil receptor domains and N-terminal disorder (IUTD). Adding a domain identified to be conserved and necessary for translocation (the 'DPY' motif), all of which can be predicted computationally from protein sequence, allowed NBs to be classified based on domain organisation. Eight different domain organisations were observed, three of which are novel and have not been previously characterised in literature (groups III, IV and VI). Studies in translocation have primarily focused on colicins and pyocins which only occupy three of the possible domain organisation groups. It is unknown how these different domain organisations could affect the translocation mechanism of NBs. Studies into the translocation of colicin E9 showed the IUTD was necessary for translocation, however, only five of the eight domains had any detectable disorder.

There have been numerous studies suggesting a link between colicins and virulence, either by an association between virulence factors (Micenková *et al.*, 2014a; Petkovsek *et al.*, 2012), showing colicinogenic bacteria out competing native *E. coli* or colicins helping establish pathogenic bacteria within a population (Holt *et al.*, 2013). Studies using the approach of association between colicins and virulence factors tend to use PCR based approaches only identifying a handful of genes at a time. The pangenome approach used in this study compared 41,723 genes (gene groups at 95% sequence identity) to

colicinogenicity. Not only is the sample size much larger than PCR based methods, it also does not limit the association to virulence factors which could bias statistic analysis. Analysis of the pangenome output revealed a significant association to multiple virulence factors. This supports the hypothesis that colicins could be used to displace native bacteria before causing infection, or used to protect a niche after invasion. Association to protein involved in production of LPS was also observed, though the reason behind this association is currently unknown. A group of phage related genes were also identified as significantly associated, which could be for two reasons. Firstly, phage genes could be involved in the horizontal transfer and evolution of colicins as has been observed for the closely related klebicins (Chavan *et al.*, 2005). Secondly, prophage genes have been proposed as involved in release of group B colicins, where latent phage lysis genes are induced by the SOS response and cause the release of colicin via lysis (Nedialkova *et al.*, 2015).

NBs present an intriguing candidate for the design of novel antimicrobials. The potent activity, modular domain architecture, lack of post translational modifications and high specificity are all desirable traits for a species specific antimicrobial. As NBs are mostly considered to target an organism closely related to the producing strain, it becomes important to know the phylogeny of bacteria which can produce NBs. Whilst NBs were only identified in γ-proteobacteria, they were identified in a number of high important pathogens. *K. pneumonia*, *E. coli*, and *P. aeruginosa* are all included in the 10 most prevalent nosocomial infections and antibiotic resistant Carbapenem-resistant *Enterobacteriaceae* (CRE), Extended spectrum beta-lactamase-producing *Enterobacteriaceae* (ESBLs), multidrug-resistant *Pseudomonas aeruginosa*, drug-resistant nontyphoidal *Salmonella* and drug-resistant *Shigella* are all listed as either serious or urgent threats (CDC, 2013). The NB pipeline identified potential antimicrobials in all of these families/genera/species. Further it was shown (for two species) that bacteria which encode NBs are more likely to encode the receptors which would make them sensitive and that pathogenic bacteria are associated with NBs. Organisms could then be targeted by a clinical dose of the NB that they produce, if the NB is a chimera engineered with a cytotoxic domain from an exogenous species.

The NB pipeline identified >3000 NB genes in 53 species of γ-proteobacteria. The majority of these NBs, such as those from *Yersinia* spp., have never been expressed or characterised. As genome sequencing continues to increase in speed and availability whilst falling in price, a future could be envisioned where pathogens are readily sequenced before antibiotic treatment (Fricke and Rasko, 2014; Köser *et al.*, 2014). Understanding the translocation pathways and the proteins involved in uptake would be the first step in allowing a targeted approach to treating a bacterial infection either by choosing the correct NB to target a pathogen or design an NB to target a pathogen. Further work on large libraries of isolates is needed to prove the hypothesis that bacteria which produce an NB would be sensitive to an external source of that NB with an exogenous cytotoxic domain.

This association between NB genes and receptors not only allows the prediction of proportions of sensitive bacteria, it could also allow a bioinformatic approach to predict an unknown receptor. Using a pangenome approach (similar to that used to identify the virulence factor association) it becomes possible to test all proteins to an association to an NB and then identify associated outer membrane proteins. This could prove to be a rapid approach to identify receptors to novel NBs.

What is arguably the most important future work for this study is the expression of the identified NB and testing for cytotoxic activity against libraries of pathogens. If these proteins, individually or in combination prove effective *in vitro*, a significant challenge will be the design of a delivery system for large protein antibiotics for *in vivo* usage.

Whilst the future work proposed mainly focusses on the use of NBs as antimicrobials for clinical use against human pathogens, they could also be effective against plant pathogens. *Pseudomonas* spp. are significant plant pathogens with 21 different species recognised as phytopathogens, including *P. syringae* (Marcelletti and Scortichini, 2014). Genetically encoded antimicrobials such as S-type pyocins alongside the F- and R- type, could be used in future generation of disease resistant transgenic plants.

In summary, a pipeline for the identification of NB genes within bacterial genomes was developed and validated. Using this pipeline >3000 NB genes were identified, most of

which have never been characterised, and were shown to be widespread in γ-proteobacteria. Identification of structural features allows a novel categorisation of NBs. Using a pangenome approach showed an association between NB genes and virulence factors within *E. coli* supporting the hypothesis that NBs are exploited by invasive bacteria to displace host microflora.

CHAPTER 4

Chapter 4: Non-specific colicin resistance of uropathogenic *E. coli*

4.1 Introduction

Gram-negative bacteria are surrounded by two lipid membranes, the outer membrane and the inner membrane. Between these two membranes lies an aqueous cellular compartment known as the periplasm, which contains a layer of peptidoglycan. The OM acts as a selective permeability barrier, limiting the passage of hydrophobic molecules whilst using porins to allow the passage of nutrients. It also acts to protect the cell from environmental stressors. For antibiotics and bacteriocins, the outer membrane presents a formidable barrier.

4.1.1 The outer membrane

The outer membrane is the defining feature of Gram-negative bacteria. The OM is a lipid bilayer but differs from the IM in its asymmetry. The IM consists of a symmetric phospholipid (PL) bilayer composed of phosphatidylethanolamine (PE) and phosphatidylglycerol (PG). The OM does contain phospholipids, but these are confined to the inner leaflet (the leaflet which is in contact with the periplasm), while the outer leaflet consists of LPS. LPS is the major component of the OM and has roles in permeability, antibiotic resistance, cell adhesion and virulence. LPS has three distinct parts/moieties: lipid A, core oligosaccharides and the O-antigen (Figure 4.1).

Lipid A is a glucosamine-based disaccharide which is acylated at positions 2 and 3 on each monosaccharide and phosphorylated at positions 1 and 4 of the disaccharide. Lipid A has been intensively studied not only for its role in forming the OM but also due to its potent immunogenicity. Lipid A is recognised by TLR-4 and causes stimulation of the immune system. Lipid A can be shed from bacterial OMs via mechanisms such as cell lysis or stripped from the bacterial OM by host produced lipid A-binding protein and presented to the immune system.

Lipid A is linked to the core via 3-deoxy-D-manno-oct-ulosonic acid (Kdo). In *E. coli*, Kdo is further linked to a series of heptose sugars through sequential heptosyltransferase reactions to form the inner core of LPS. Glycosyltransferases then sequentially add hexose

sugar monomers to form the outer core (Figure 4.1). In *E. coli*, four core types have been identified with different sugar patterns in the outer core.

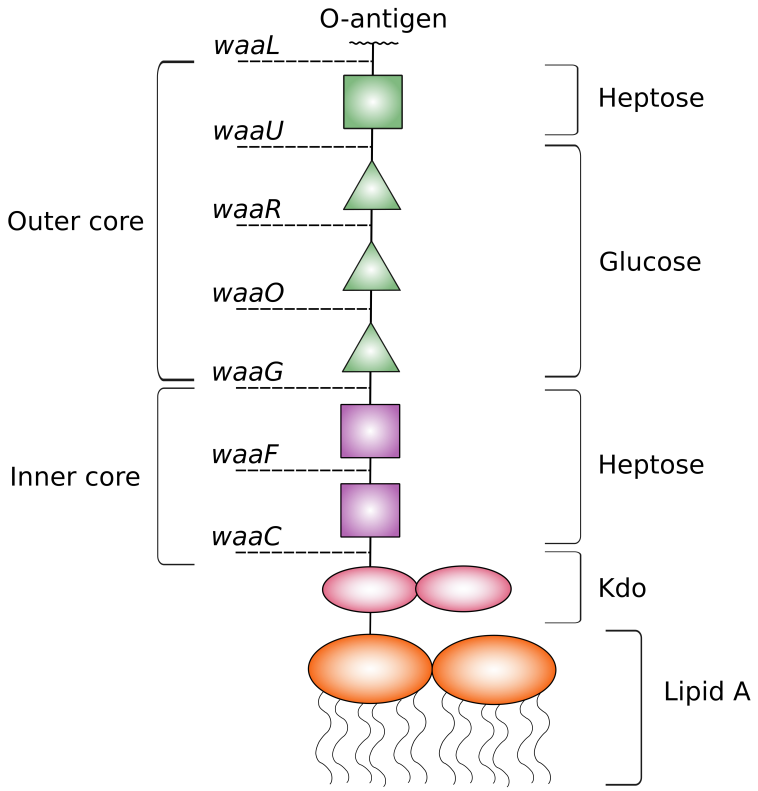


Figure 4.1: **Model structure of core LPS in *E. coli*.** LPS is the major constituent in the OM of Gram-negative bacteria. LPS has a tripartite structure containing lipid A which anchors it in the outer membrane. Followed by a series of sugars which form the inner and outer core. These sugars are sequentially added by heptosyltransferases and glycosytransferases from the *waa* operon. The O-antigen is attached to the outer core by the ligase WaaL.(Adapted from (Qimron *et al.*, 2006)

The O-antigen is not essential and is often lacking in laboratory strains of *E. coli*. Strains that present an O-antigen on their LPS are termed smooth whilst those that lack it are called rough. The O-antigen is comprised of a series of repeating units containing one or more sugar residues. The O-antigen is highly variable. Even just within *E. coli* strains, differences in organisation, combination and modification lead to 185 recognised O-group serotypes (DebRoy *et al.*, 2016). As the O-antigen decorates the outside of the bacteria, it interacts with the immune system and is involved with evading complement dependent killing, decreased internalisation and increased intracellular survival. Like Lipid A, the O-antigen is highly immunogenic and there can be a strong selective pressure on bacteria to lose it, such as occurs with chronic *P. aeruginosa* within the CF lung (Hancock *et al.*,

1983). The length of the O-antigens varies between strains depending on the number and size of repeating units (~2->80). For some E.coli strains, it was measured by AFM to be protruding 37 nm from the cell surface. *E. coli* K-12 is arguably the most well studied organism in biology and has been used as a model organism since the 1940s. During its passages and adaptation to laboratory growth, *E. coli* K-12 acquired the *rfb-50* mutation, an *IS5* insertion in the *wbbL* gene which encodes a rhamnosyltransferase. The *rfb-50* mutation causes a complete loss of O-antigen and the rough phenotype.

The O-antigen has been linked to colicin resistance. Van der Ley used spontaneous O-antigen mutants to show these could engender colicin resistance (van der Ley *et al.*, 1986). This was then further shown to be dependant on length of the O-antigen in a *Shigella* model (Tran *et al.*, 2014). A survey of *E. coli* isolates from multiple culture libraries identified high levels of resistance towards colicins when tested, with >70% of isolates resistant to a single colicin and ~30% of strains resistant to 18 colicins (Feldgarden and Riley, 1998). If this resistance were to be due to mutations in receptor proteins and production of immunity proteins, a single strain would have to encode and express at least 18 different immunity proteins and have evolved a set of mutant receptors. This is highly unlikely suggesting a non specific resistance mechanism such as presence of O-antigen could be prevalent in *E. coli*.

4.1.2 Uropathogenic *E. coli* ST131

Uropathogenic *E. coli* is responsible for approximately 75% of urinary tract infections (UTI) (Flores-Mireles *et al.*, 2015). A clone of *E. coli* (sequence type 131) identified in 2008 as resistant to fluoriquinolones and positive for CTX-M-15 extended spectrum β-lactamases (ESBL), is now recognised as a globally disseminated strain and the major lineage within extra-intestinal *E. coli* (ExPEC) (Nicolas-Chanoine *et al.*, 2014). Two separate studies in Canada (Peirano *et al.*, 2012) and Israel (Karfunkel *et al.*, 2013) found an increase in ESBL+ strains within the *E.coli* population and an increase in the proportion of ST131 isolates amongst ESBL+ *E. coli*. UPEC ST131 is therefore a multidrug resistant bacterium that is globally disseminated and a major cause of UTI, particularly in a health-care setting (Banerjee *et al.*, 2013). UPEC ST131 has a smooth phenotype and the O25b

serotype. The pathogenicity, drug resistance and full O-antigen make UPEC ST131 an ideal strain to test for O-antigen dependant resistance towards colicins.

4.1.3 Transposon-directed insertion site sequencing (TraDIS)

Transposons are genetic elements which can insert themselves into genomes and disrupt genes. They are a commonly used tool to explore gene function by disrupting the gene and examining the resulting phenotype. Transposon libraries are populations of cells where each bacterium contains a single transposon insertion. Combining transposon libraries with next generation sequencing is a powerful approach for genome wide analysis of genotype-phenotype relations (van Opijnen and Camilli, 2013). Transposon-directed insertion site sequencing (TraDIS) is a prominent transposon mutant sequencing technique able to generate highly saturated transposon libraries. In a standard experiment, a transposon library highly saturated with transposons is generated. One population of the library is then exposed to a condition (antibiotic, additive, bacteriocin etc.) whilst another population serves as a control. Cultures are incubated for 24 hours with an optional subculturing. DNA is extracted, cleaved and an adapter added to the DNA ends before a small enrichment step to avoid adding bias. DNA is then sequenced and reads are mapped back to the genome. Differences in the number of reads at each base between the control and the treated population are used to assign the fitness advantage or disadvantage of each gene under that condition. Previously TraDIS was able to generate a *Salmonella typhi* library containing 370,000 unique transposon insertion sites and identified essential genes for both rich and poor media (Langridge *et al.*, 2009). A schematic of the TraDIS method is shown in Figure 4.2. TraDIS can be a powerful tool to examine genes involved in both resistance and sensitivity towards a particular antibiotic. After treatment with an antibiotic at sub-MIC, genes involved with resistance would have a decrease in insertions as the product of that gene provides a fitness benefit and would compete with the population. Genes associated with sensitivity would have an increase in insertions as it would be detrimental to express the product of that gene. Though TraDIS is an exceptionally powerful technique it has weaknesses common to all transposon methods. Firstly, it cannot investigate the role of essential genes, as a trans-

poson insertion would lead to death of the cell. Secondly, an accurate analysis requires the generation of a saturated transposon library, which can be difficult to accomplish.

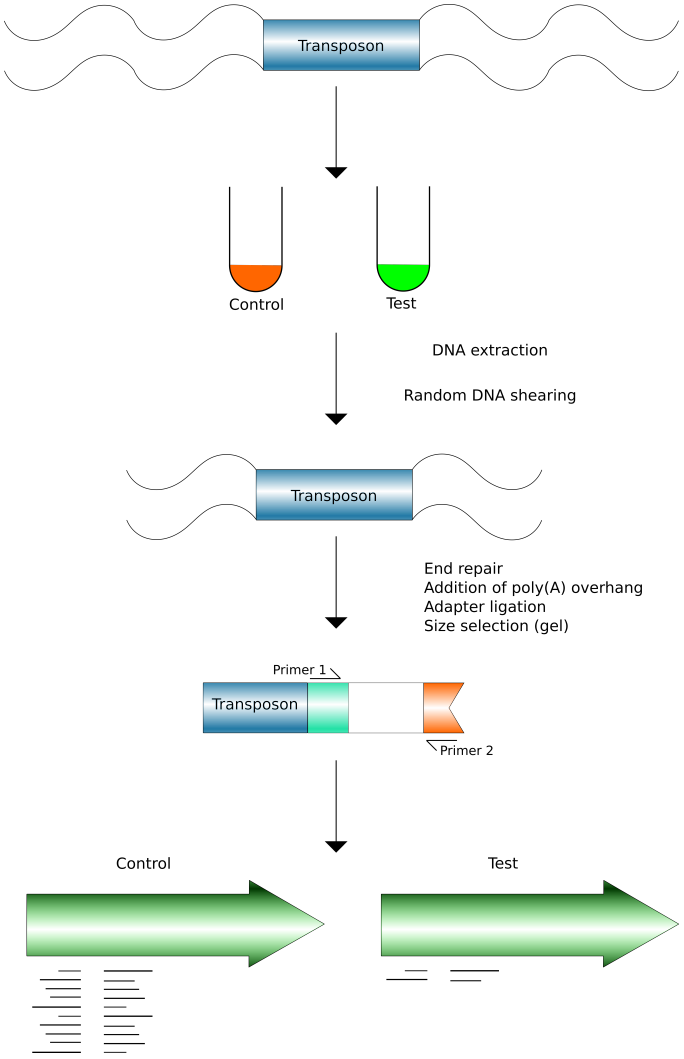


Figure 4.2: **Schematic of a standard TraDIS experiment.** A dense transposon library is created and validated to ensure coverage. Two populations of the library, one as a control and another under the tested condition are incubated for 24 hours before subculturing and incubated for a further 24 hours. DNA is then extracted and randomly sheared. A series of preparation steps are performed and adapters added. DNA undergoes a limited number of PCR reactions before sequencing. Reads are mapped to the reference genome. Differences in the reads between the control and the test are used to evaluate each gene's fitness contribution under the condition. Adapted from (van Opijnen and Camilli, 2013)

4.1.4 Phenomic microarray and BIOLOG

Cultures grown in the lab are usually in ideal conditions, with optimal nutrients, osmolarity and pH. In a natural environment bacteria can be exposed to a variety of harsh and variable conditions. Typically, investigating the phenotype of bacteria in different conditions would involve a laborious time consuming series of multiple growth curves

measuring OD. Phenotypic microarrays can assay 100s of conditions simultaneously. BIOLOG is a popular phenotypic microarray platform which measures the respiration of a culture through the reduction of a tetrazolium dye (Bochner *et al.*, 2001).

4.1.5 Aims

While colicins have potential as next generation antimicrobials, high levels of naturally occurring resistance could limit their impact in the clinic. It therefore becomes necessary to test if a naturally occurring *E. coli* pathogen is capable of O-antigen dependent resistance and if this resistance is an artefact of growth under laboratory conditions or exists in conditions more representative of the bacteria's environment. To accomplish these aims a uropathogenic *E. coli* strain from the highly prevalent pathogenic sequence type 131 (ST131) was used as a model organism. A series of high-throughput genomic and phenomic assays are used to identify the resistance mechanisms and the conditions under which they are relevant. Finally fluorescence microscopy is applied to examine the potential O-antigen dependent blocking of colicin receptors.

4.2 Results

4.2.1 Observed resistance of UPEC ST131

UPEC ST131 exposed to colicins B, N, A and E9 showed little sensitivity or was completely resistant to colicin when grown on M9 minimal media. However, zones of inhibition were observed when media was switched to LB. Solid media killing assays showed a >6000 fold difference in ColE9 MIC of UPEC ST131 compared to the colicin sensitive *E.coli* JM83 (Figure 4.3). The more sensitive liquid killing assays gave a 1000 fold increase of the MIC when performed in LB media and UPEC ST131 was resistant to ColE9 at concentrations up to 250 nM in minimal media.

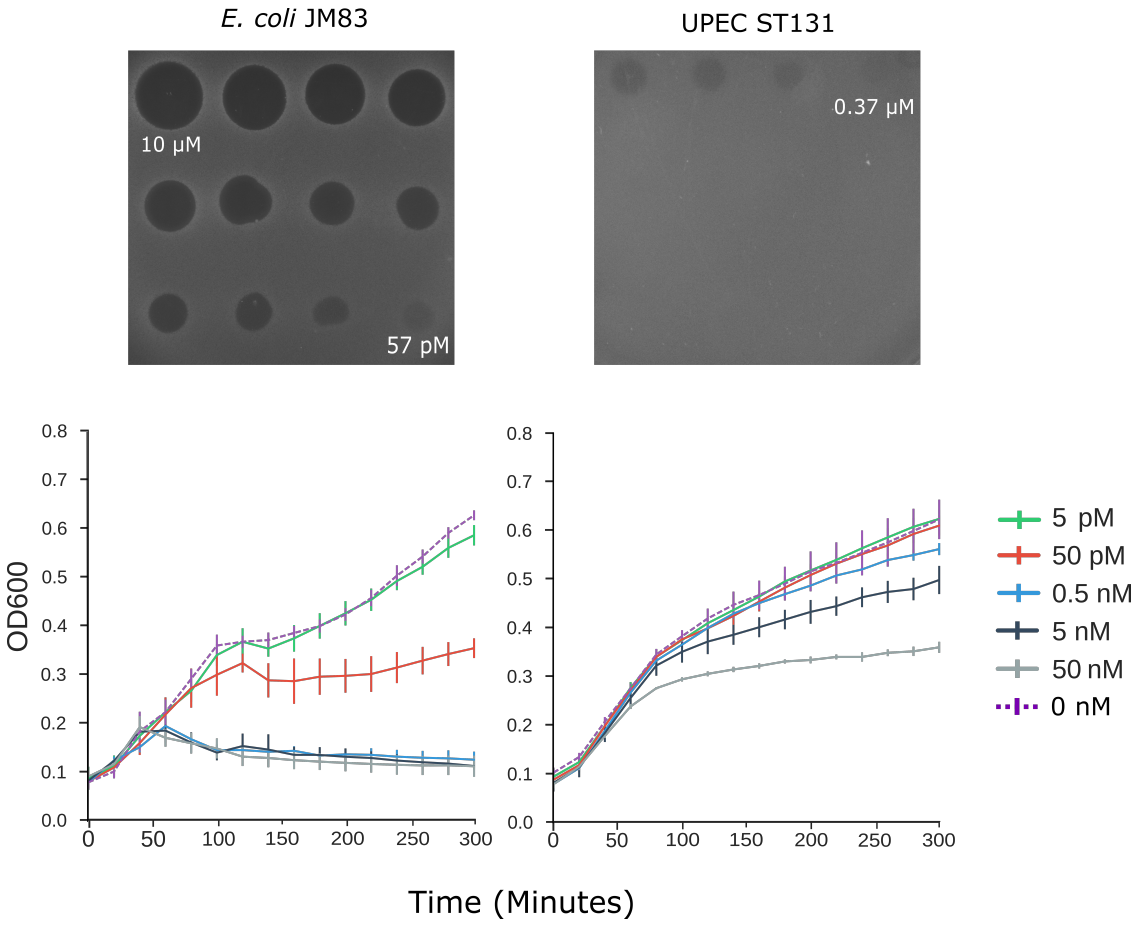


Figure 4.3: **Plate and killing assays of UPEC ST131 and colicin sensitive *E. coli* JM83.** ***Top***: Solid media killing assays on LB agar using serial dilutions of ColE9. UPEC ST131 shows slight sensitivity to the highest concentrations (10μM) of ColE9 whereas *E. coli* JM83 is sensitive at pM concentrations. ***Below***: Liquid media killing assay using LB.

4.2.2 Resistance is independent of immunity proteins or receptor mutations

Resistance to colicins is usually associated with a cognate immunity protein or mutations in receptor/translocator proteins. Two approaches were used to test whether these mechanisms were responsible for the resistance observed in UPEC ST131. Firstly, using a genomic analysis approach, the genome sequence of UPEC ST131 was scanned for presence of colicins and/or immunity proteins using HMMer. Sequences for proteins involved in the translocation of ColE9 were analysed for mutations. Then, UPEC ST131 was exposed to an array of colicins, which utilise a variety of receptor and translocator proteins, as well as the Ton or Tol system.

UPEC ST131 did not contain *im9* or any identifiable colicin immunity genes or mutations in the BtuB or OmpF proteins which have been associated with resistance. Four colicins were used to test for non specific resistance (Table 4.1). UPEC ST131 showed between 10,000 and 100 fold increase in the MIC of tested colicin compared to the sensitive MG1655 strains. As these colicins all target different receptors, translocators, PMF energised complexes and represent cytotoxic domains, and since their immunity proteins are not encoded in the UPEC ST131 genome, the difference in MIC is unlikely to be due mutations in translocation proteins or presence of immunity proteins. This suggests a non-specific mechanism for colicin resistance.

Table 4.1: **Colicins used to test UPEC ST131 utilise different receptor proteins, translocator proteins and PMF coupled complexes.** The four colicins tested use different receptor proteins, translocator proteins, PMF coupled complexes and cytotoxic domains.

Colicin	Group	Cytotoxic domain type	Receptor proteins	Translocation proteins
Ia	B	Pore	Cir	Cir, TonB, ExbBD
D	B	tRNase	FepA	TonB, ExbBD
B	B	Pore	FepA	TonB, ExbBD
E9	A	DNase	BtuB	OmpF, TolABQR

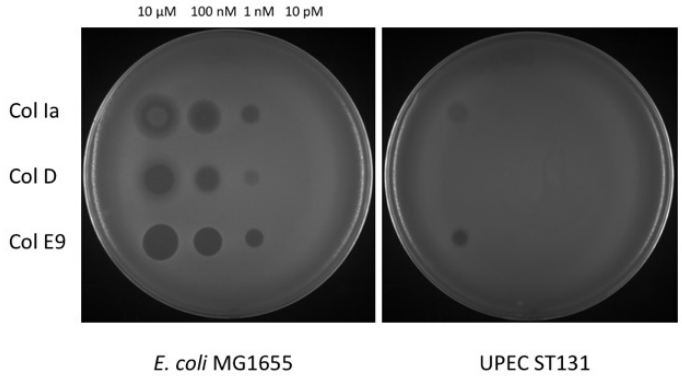


Figure 4.4: **Solid media killing assays using a variety of colicins hijacking different proteins within the target cell**

4.2.3 EDTA significantly lowers the MIC of colicins for UPEC ST131

EDTA is known to destabilise and cause the release of LPS by chelating divalent cations required for its stability (Marvin *et al.*, 1989). To test the role of LPS in the resistance mechanism observed, UPEC ST131 was exposed to three colicins in the presence of increasing concentrations of EDTA (Figure 4.5). In the presence of EDTA, UPEC ST131 showed a 10,000 fold decrease in MIC towards Col Ia, a 100 fold decrease in MIC of ColE9 and sensitivity towards ColD at 10 µM of bacteriocin.

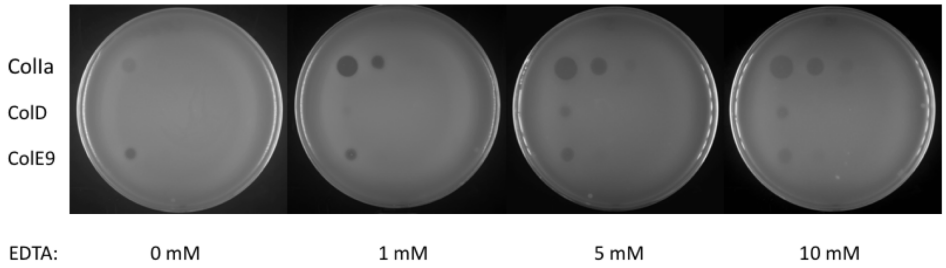


Figure 4.5: **Increasing concentration of EDTA cause an increase in sensitivity towards colicins in UPEC ST131** UPEC ST131 was incubated on LB agar plates with increasing concentrations of EDTA. Colicin Ia, Colicin D and Colicin E9 are spotted onto the plate in a serial dilution (10 µM, 100 nM, 1 nM, 10 pM)

Table 4.2: **MIC testing of UPEC strains.** Ten UPEC strains were tested (in duplicate) with a serial dilution of Colicin E9. Inhibition was determined by a zone of clearance. Numbers indicate the radius of the zone of inhibition (mm). Any inhibition was (green) was considered as sensitive.

ColE9 (nM)	UPEC isolate																					
	2		3		6		8		9		22		34		35		38		50		K-12	
	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
10000	10	8							10	12			6	6			5	6			12	12
3300	9	8							8	10			6	4			4	5			9	10
1100	9	7							7	7			5	3				5			8	9
370	7	5							5	6											8	10
120	5	7							5												5	7
41	5	6																			6	7
14	4	5																			5	6
4.6																					5	7
1.5																					4	5
0.167																					4	5
0.57																						

4.2.4 Resistance of multiple UPEC strains

To determine if resistance towards colicins is widely observed across UPEC strains or only with the UPEC ST131 strain, we tested the MIC of ColE9 for ten UPEC strains (Table 4.2). Six of the 10 strains were not affected by ColE9 at concentrations up to 10 μM. The four more susceptible isolates had an MIC for ColE9 between 30-2000 x the MIC of the sensitive *E. coli* K-12 MG1655.

4.2.5 TraDIS analysis reveals the role of O-antigen in non-specific colicin resistance

To unravel the genetic basis for the colicin resistance observed in UPEC ST131, a TraDIS library had previously been established for UPEC ST131 NCTC 13441 by Dr Christine Boinett (Sanger Institute) using the method described (Langridge *et al.*, 2009). The library had a high insertion density, with 450 000 unique insert sites in the 5.4 Mbp genome giving an average density of 1 insertion per 12 bps. Two experiments were performed, one with EDTA in the media and a second without EDTA. For each experiment two control

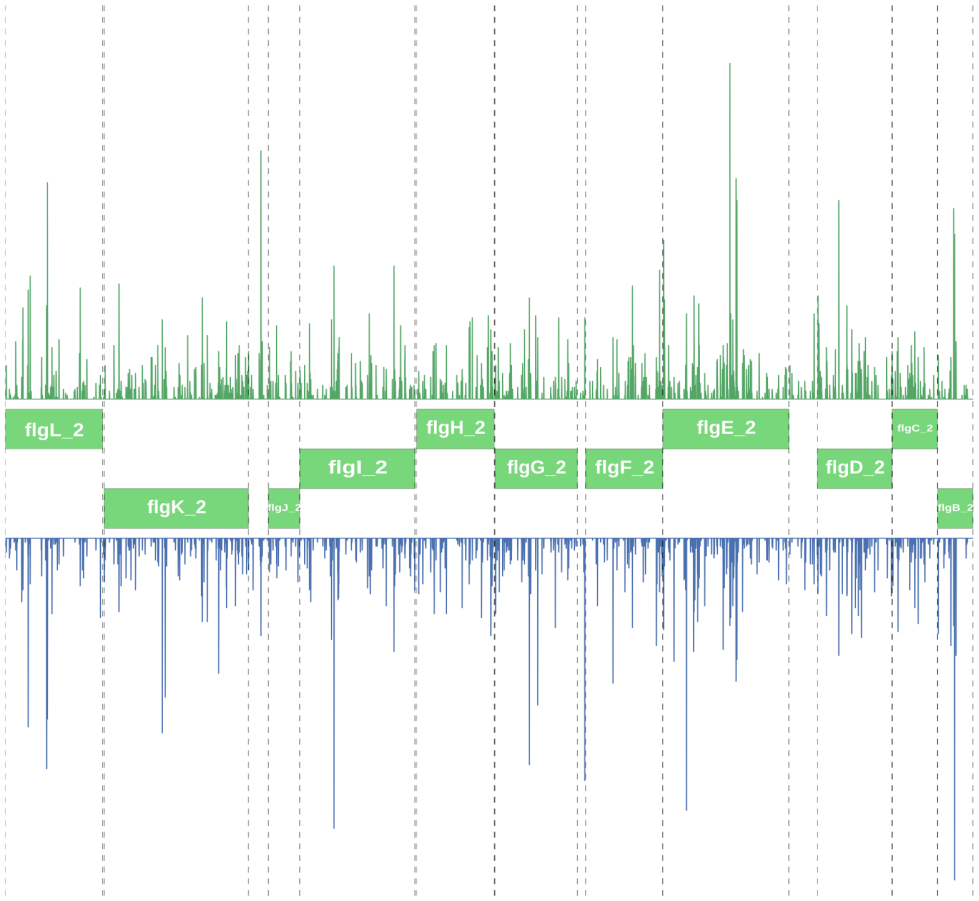
populations were not exposed to colicin and two populations were exposed to a sub-MIC concentration of colicin (0.5 nM ColE9Im9).

In the presence of EDTA, 21 genes were identified which had a significantly decreased number of insertions in the colicin treated sample compared to the non-colicin control, suggesting the products of these genes are necessary for resistance (Table 4.3). No genes were identified with an increased number of insertions in the colicin treated population. These would be genes which make cells sensitive to colicins such as those involved in translocation of ColE9: *btuB*, *ompF* etc.

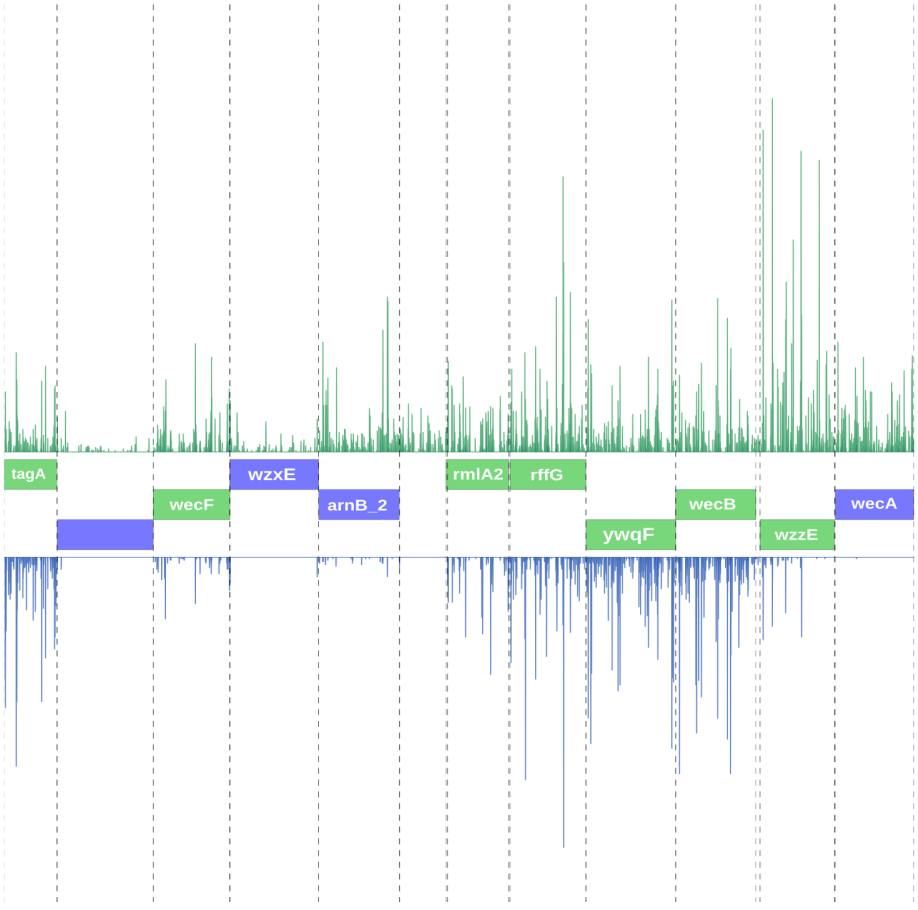
Two genes involved in the modification of lipid A were found to be required for colicin resistance (*arnD* and *arnE*). The *arn* operon is involved in resistance to cationic antimicrobial peptides (CAMPs) by increasing the positive charge of lipid A through modification with a 4-amino-4-deoxy-L-arabinose moiety (L-Ara4N). Increases in L-Ara4N modification have been shown to increase resistance to polymyxins such as colistin in *Salmonella enterica* serovar Typhimurium and *Klebsiella pneumoniae* (Olaitan *et al.*, 2014).

Table 4.3: **Genes highlighted by TraDIS as providing a fitness benefit in the presence of colicins. Results are in the presence of EDTA.**

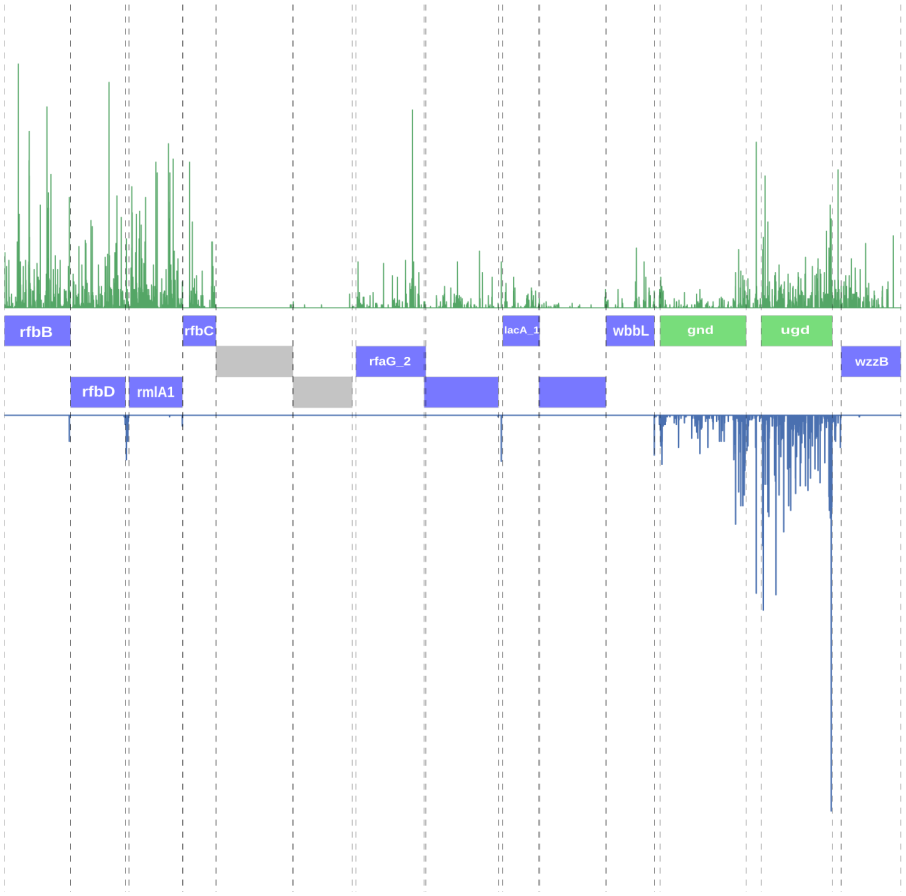
Gene name	Function	logFC	q.value	Acc.
00061	LPS core heptosyltransferase	-12.149	1.06E-235	WP_001236430.1
rfaG_1	LPS core glycosyltransferase	-2.088	4.55E-029	WP_000634259.1
rfaH	Transcription factor involved in LPS biosynthesis	-4.842	1.10E-056	WP_001192400.1
rfaJ_2	LPS core glycosyltransferase	-8.354	0	WP_000376840.1
rfaJ_1	LPS core 3-alpha-galactosyltransferase	-7.005	0	WP_001188027.1
rfaL	O-antigen ligase	-5.297	0	WP_000958045.1
rfbC	rhamnose specific dTDP-4-dehydrorhamnose 3,5-epimerase	-11.967	2.07E-158	WP_001100793.1
lacA_1	WbbJ; catalyzes the transfer of the O-acetyl moiety to the O antigen	-10.953	1.24E-105	WP_001296213.1
rfbD	dTDP-glucose 4,6-dehydratase	-10.919	0	WP_000699407.1
wbbL	glycosyltransferase	-10.651	9.64E-080	WP_000262536.1
rfaG_2	LPS core glycosyl transferase	-9.977	9.71E-214	WP_001296214.1
01679	hypothetical protein	-9.190	1.26E-153	WP_000027654.1
01681	glycosyltransferase	-6.985	2.70E-036	WP_000517629.1
rfbB	rhamnose specific dTDP-glucose 4,6-dehydratase	-3.545	0	WP_000699407.1
rmlA1	rhamnose specific glucose-1-phosphate thymidylyltransferase	-2.980	0	WP_000857525.1
wecA	ECA transferase	-7.915	0	WP_001050960.1
04729	putative common enterobacterial antigen polymerase	-7.360	5.76E-048	WP_000055112.1
wzxE	ECA flippase	-4.547	1.14E-064	WP_000050265.1
arnE	Lipid A modification (CAMP resistance) flippase subunit	-2.413	1.86E-021	WP_000638016.1
arnD	Lipid A modification (CAMP resistance) deformylase	-2.271	1.82E-080	WP_000169713.1
02987	hypothetical protein	-1.960	9.74E-009	WP_001295929.1



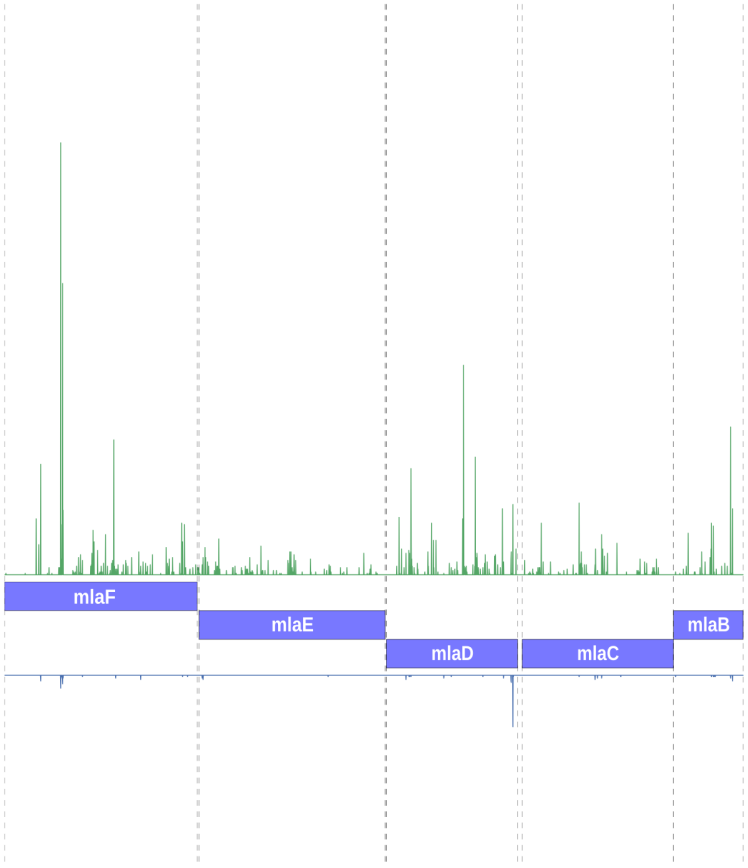
(a) Flagella operon genes which show no significant changes in insertions



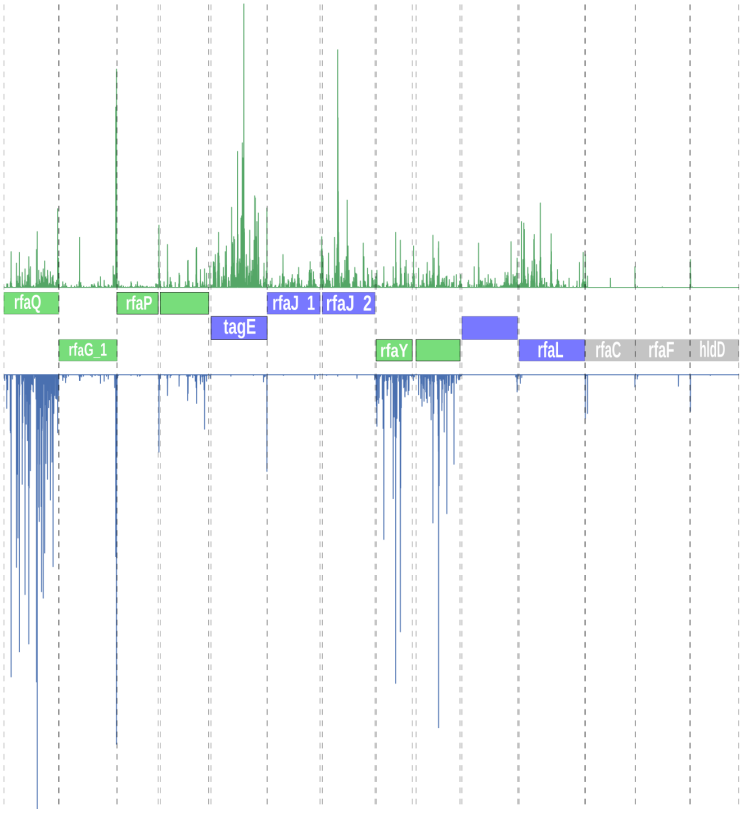
(b) Enterobacterial common antigen operon



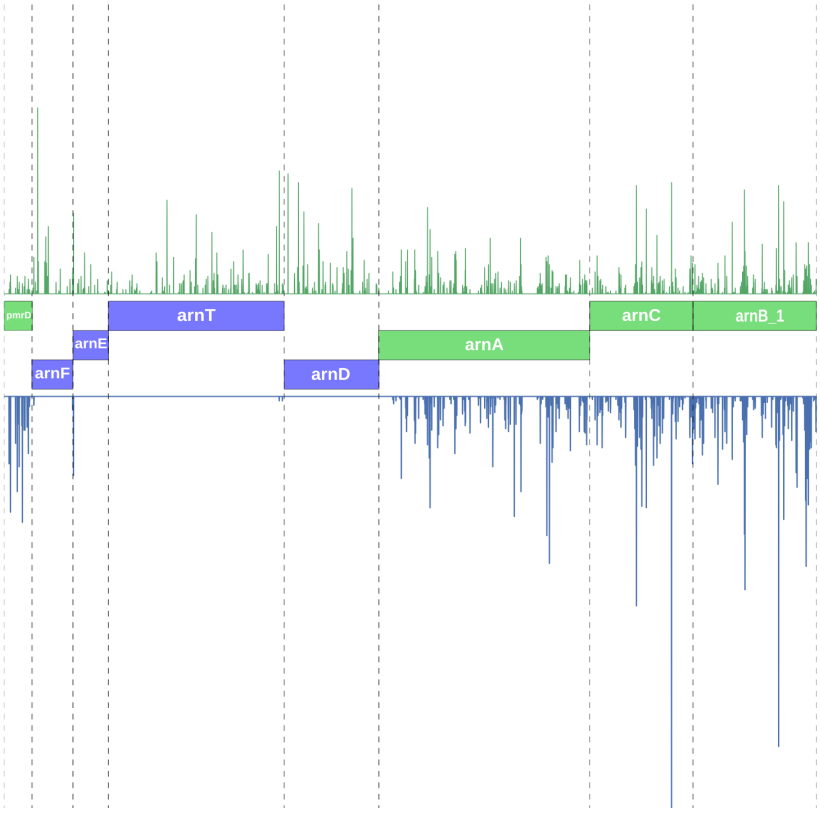
(c) O25b biosynthesis operon



(d) Membrane lipid asymmetry operon



(e) Core LPS operon



(f) arn operon

Figure 4.4: **TraDIS insertion plots.** Insertions are shown for the five operons which had the most significant differences in transposon insertions between the control and colicin treated cultures. Green line represents the insertions in the colicin minus control, blue line is the sample treated with 1/3 MIC ColE9. Genes in blue were found to be associated with resistance, green genes had no significant association and grey genes were essential (no insertions in control or colicin treated sample). Results are for the absence of EDTA.

EDTA affects membrane and LPS stability, which could be masking specific genes involved in colicin resistance. To test this, TraDIS was repeated in identical media (with ColE9Im9) without EDTA. Without addition of EDTA, TraDIS analysis identified 114 genes which had a significant decrease in insertions compared to the non-colicin treated control (involved with resistance) and a further 34 genes with an increase of insertions compared to the control (involved with sensitivity). All 21 of the genes identified in the +EDTA test were again identified in the -EDTA test and additional genes in the *rfa* and *arn* operons were also identified. Five of the genes within the membrane lipid asymmetry (*mla*) operon and *mlaA* were identified as associated with resistance (Figure 4.4). The *mla* operon encodes an ATP transporter responsible for maintaining the asymmetry of lipids in the outer membrane. Mutations in the *mla* operon have been shown to cause sensitivity to SDS and EDTA (Malinverni and Silhavy, 2009). The KEGG Automatic Annotation Server was used to assign the genes highlighted by TraDIS to pathways (Figure 4.6) Three proteins of the ColE9 translocon were identified as genes which make cells sensitive to ColE9 (*ompF*, *tolA* and *btuB*) (Table 4.4).

Table 4.4: **Translocation protein genes identified by TraDIS**

Gene name	Function	logFC	q.value
ompF	Translocator protein	2.030	7.54E-044
btuB	Receptor protein	4.984	6.25E-299
tolA	PMF coupled protein	2.118	6.65E-011

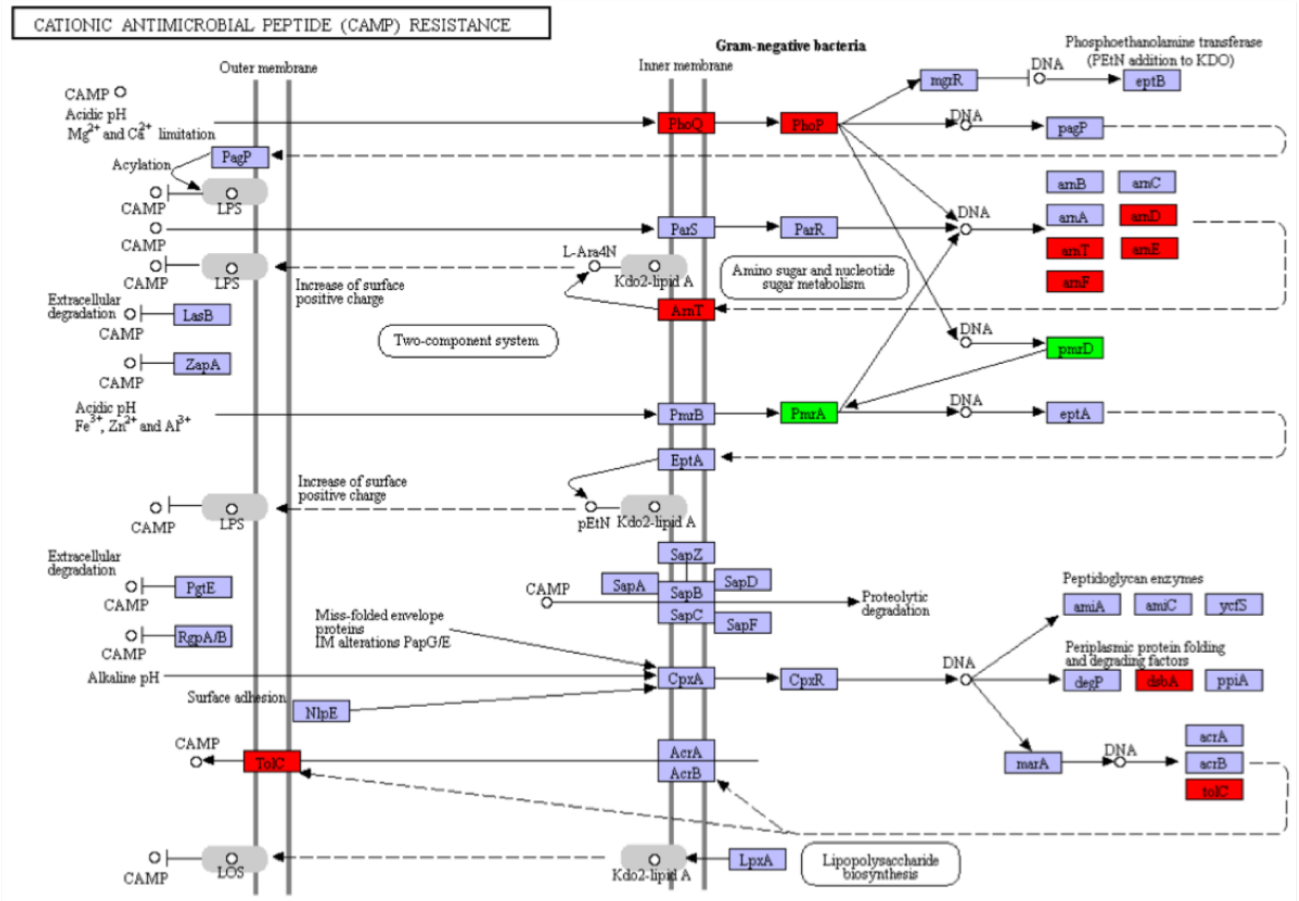


Figure 4.5: **KEGG pathway analysis of genes identified as involved with resistance by TraDIS analysis** Genes involved in resistance (*Red*) or sensitivity (*Green*). Cationic antimicrobial peptide resistance pathway (ko01503). Genes involved in sensing and responding to antimicrobial peptides by modifying lipid A.

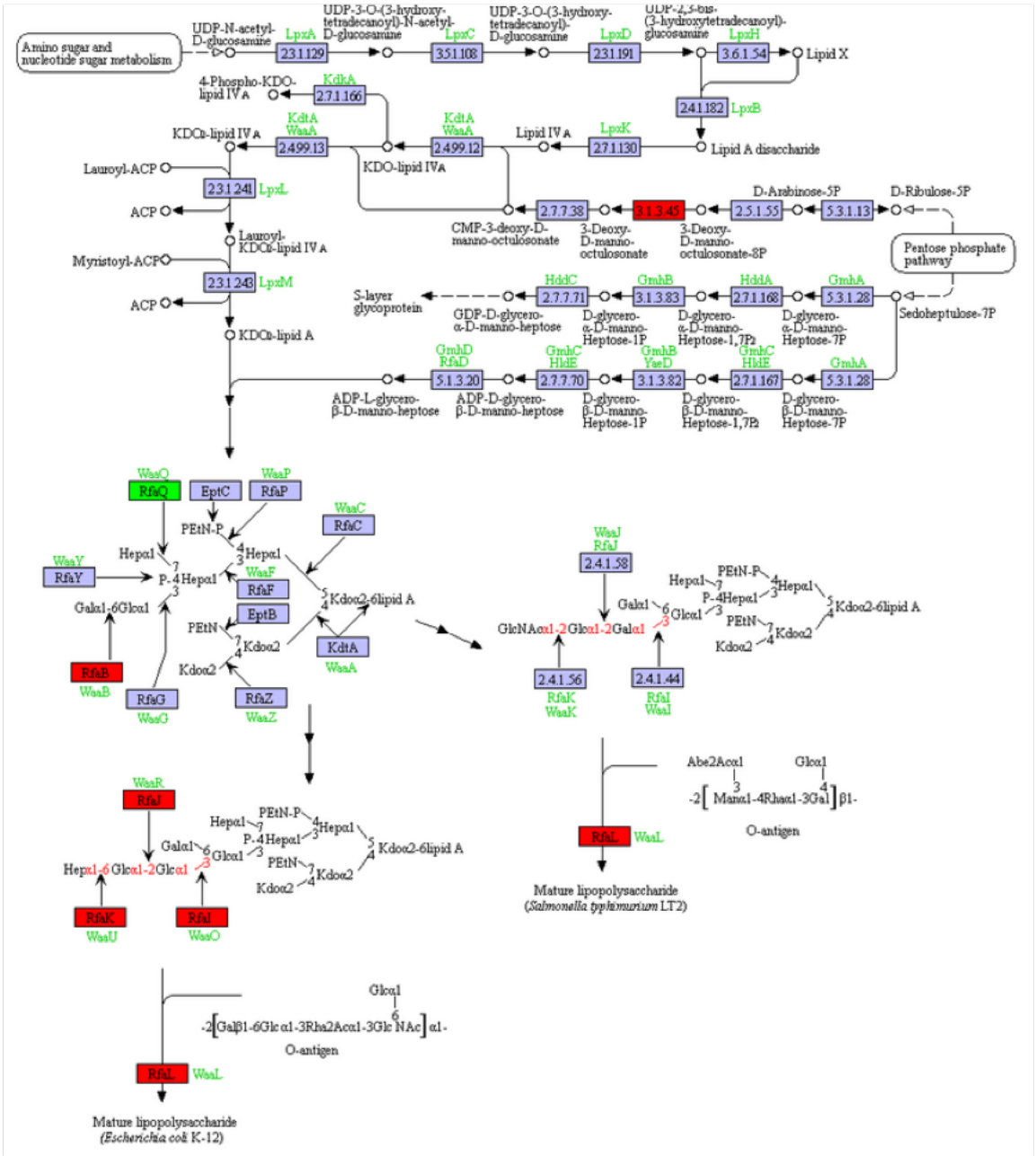


Figure 4.6: **KEGG pathway analysis of genes identified as involved with resistance by TraDIS analysis** Genes involved in resistance (*Red*) or sensitivity (*Green*). Lipopolysaccharide biosynthesis (ko00540). Genes in the biosynthetic pathway of core LPS production and ligation of O-antigen.

4.2.6 Resistance is independent of metabolism

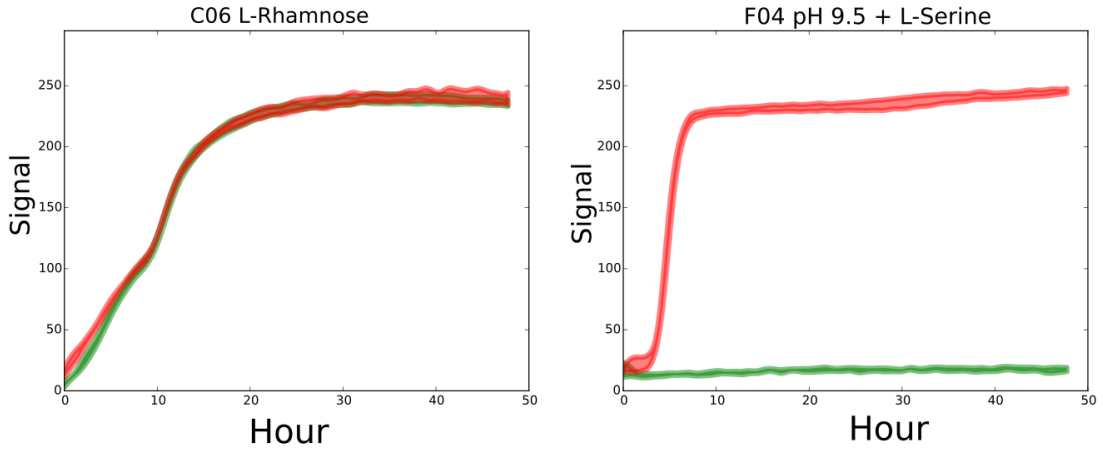


Figure 4.7: **Example of BIOLOG output.** Signal curves of respiration for UPEC ST131 in the presence of WT ColE9Im9His$_6$ (*green*) or the 'null' mutant (*red*). ***Left***: An example condition where UPEC ST131 is resistant to colicin and respiration is not affected. ***Right***: An example condition where UPEC ST131 is sensitive to colicin E9 and shows no significant respiration. Both graphs show results from 3 replicates.

To test the effect of growth conditions on resistance to colicin E9, BIOLOG phenotypic microarrays (PM) were used to study respiration (as a proxy for growth) in the presence and absence of a sub-MIC level of Colicin E9 (Figure 4.7). Four different plates were investigated, two plates which altered carbon source, one plate which explored a range of pH with different additives and a final plate with varying osmolarity.

Carbon source shows no effect on colicin resistance

Two PM plates with various carbon source were tested. No difference was observed between the plates treated with sub MIC colicins (0.37 µM) and the control plates which used an inactive colicin mutant. Across the 191 experimental conditions tested on these plates (2x96 wells minus a control), no significant difference was observed between the inactive and active colicin treatment (Figure 4.8).

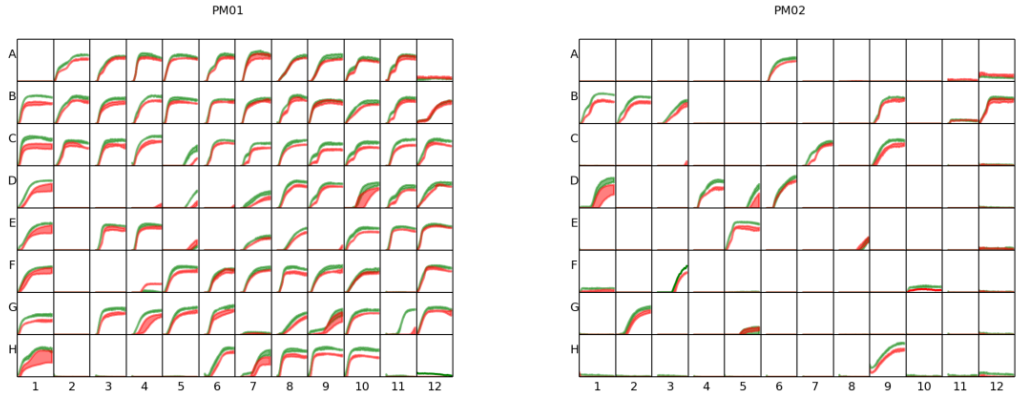


Figure 4.8: **Phenotypic microarray exploring the Colicin E9 sensitivity of UPEC ST131 with varying carbon sources** UPEC ST131 incubated for 50 hours in the presence of an active (green) or inactive (red) colicin. In all conditions, no significant difference in respiration was observed between the active and inactive colicin treatment. Detailed lists of the conditions within PM plates PM1-2 are included in the appendix

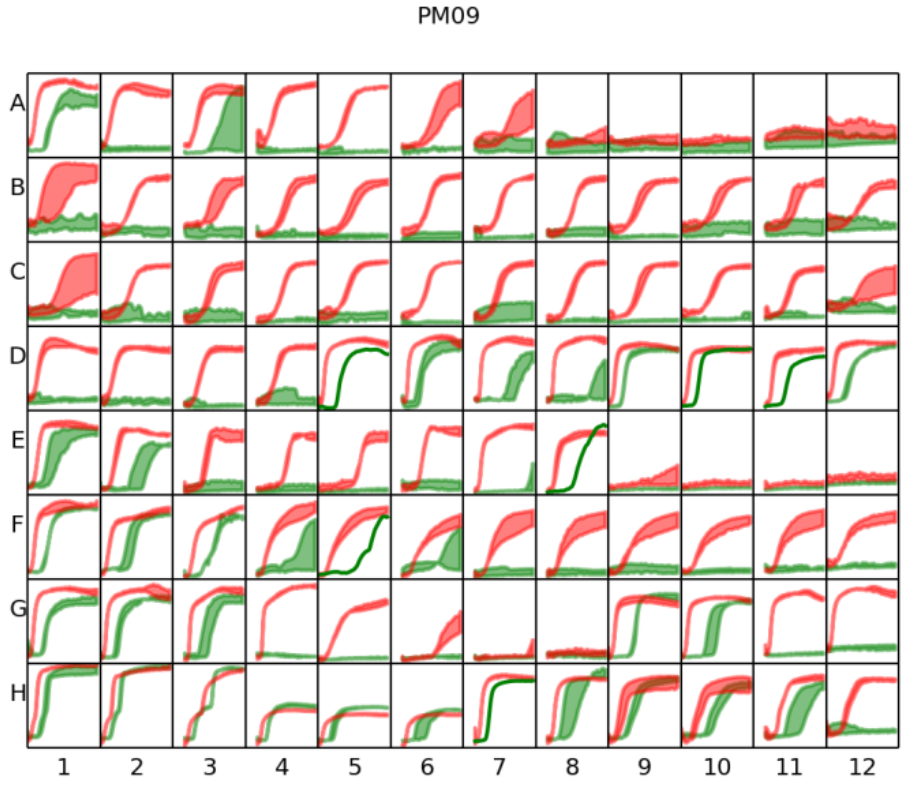
4.2.7 pH and additives can drastically alter sensitivity

Phenotypic microarray plates PM9 and PM10 test a variety of osmolites and pH conditions, respectively. Plate PM9 revealed that resistance is dependent on osmolarity. All conditions with 4-6% NaCl caused sensitivity to 1/3 MIC ColE9 (41 nM). Sodium lactate at 7% and KCl between 3-6% had a similar effect. UPEC ST131 was also sensitive to ColE9 at the highest concentrations of sodium phosphate (pH 7), ammonium sulfate (pH 8) and sodium nitrite. UPEC ST131 was also sensitive to Col E9 in the presence of urea, a compound which would be prevalent in the urinary tract (Figure 4.9).

PM10 plates test varying pH with the addition of amino acids to also evaluate deaminases and decarboxylases. Row A of PM10 tests growth at pH 3.5-10. UPEC ST131 was able to grow at pH 4.5-10. However, resistance was lost with increasing pH. Alkaline pH (pH 8 and above) greatly decreased resistance levels. The resistance at pH 7 agrees with the previous liquid killing assays which were performed at pH 7.2. All conditions at pH 9.5 were sensitive to active colicin suggesting resistance had been lost.

4.2.8 SNP analysis of resistant strains from BIOLOG plates

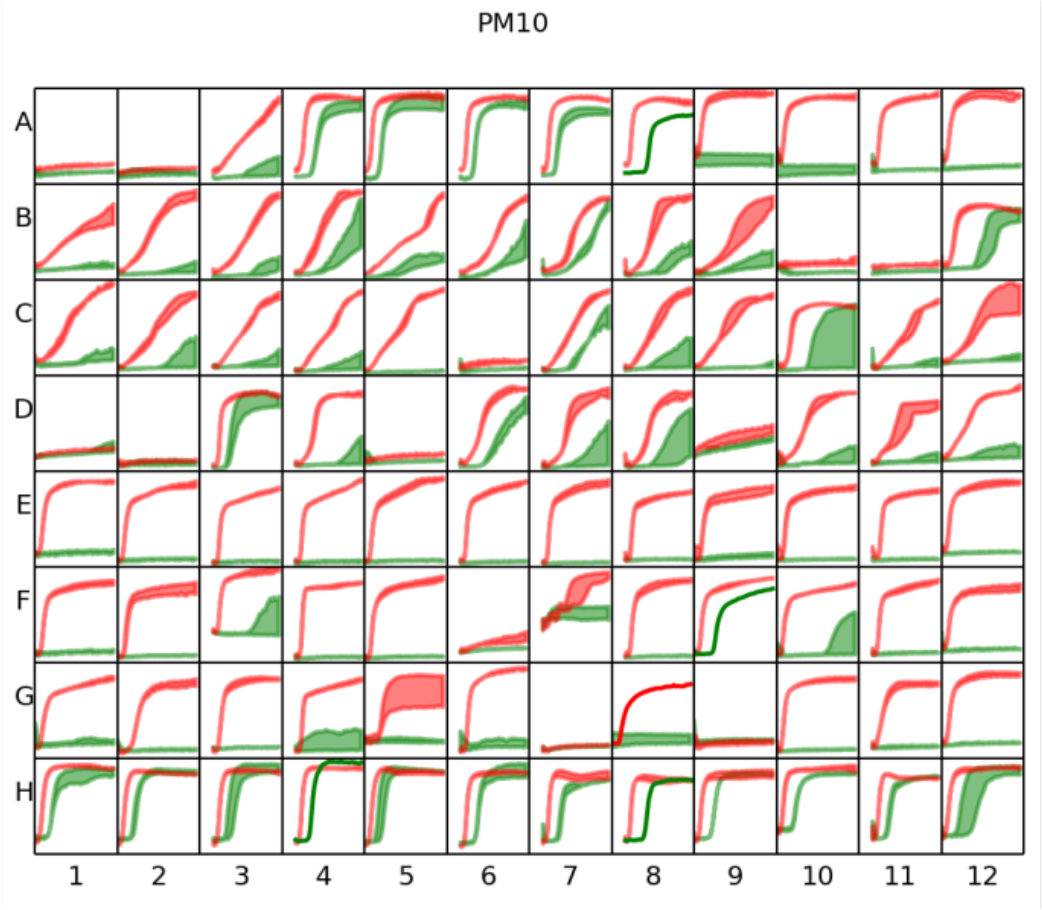
Four BIOLOG conditions generated resistant mutants (one out of three colicin treated replicates were respiring). To determine the mechanism behind this resistance and



PM9 MicroPlate™ Osmolytes

A1 NaCl 1%	A2 NaCl 2%	A3 NaCl 3%	A4 NaCl 4%	A5 NaCl 5%	A6 NaCl 5.5%	A7 NaCl 6%	A8 NaCl 6.5%	A9 NaCl 7%	A10 NaCl 8%	A11 NaCl 9%	A12 NaCl 10%
B1 NaCl 6%	B2 NaCl 6% + Betaine	B3 NaCl 6% + N-N Dimethyl Glycine	B4 NaCl 6% + Sarcosine	B5 NaCl 6% + Dimethyl sulphonyl propionate	B6 NaCl 6% + MOPS	B7 NaCl 6% + Ectoine	B8 NaCl 6% + Choline	B9 NaCl 6% + Phosphoryl Choline	B10 NaCl 6% + Creatine	B11 NaCl 6% + Creatinine	B12 NaCl 6% + L- Carnitine
C1 NaCl 6% + KCl	C2 NaCl 6% + L-Proline	C3 NaCl 6% + N-Acetyl L-Glutamine	C4 NaCl 6% + β-Glutamic Acid	C5 NaCl 6% + γ–Amino -N-Butyric Acid	C6 NaCl 6% + Glutathione	C7 NaCl 6% + Glycerol	C8 NaCl 6% + Trehalose	C9 NaCl 6% + Trimethylamine-N-oxide	C10 NaCl 6% + Trimethylamine	C11 NaCl 6% + Octopine	C12 NaCl 6% + Trigonelline
D1 Potassium chloride 3%	D2 Potassium chloride 4%	D3 Potassium chloride 5%	D4 Potassium chloride 6%	D5 Sodium sulfate 2%	D6 Sodium sulfate 3%	D7 Sodium sulfate 4%	D8 Sodium sulfate 5%	D9 Ethylene glycol 5%	D10 Ethylene glycol 10%	D11 Ethylene glycol 15%	D12 Ethylene glycol 20%
E1 Sodium formate 1%	E2 Sodium formate 2%	E3 Sodium formate 3%	E4 Sodium formate 4%	E5 Sodium formate 5%	E6 Sodium formate 6%	E7 Urea 2%	E8 Urea 3%	E9 Urea 4%	E10 Urea 5%	E11 Urea 6%	E12 Urea 7%
F1 Sodium Lactate 1%	F2 Sodium Lactate 2%	F3 Sodium Lactate 3%	F4 Sodium Lactate 4%	F5 Sodium Lactate 5%	F6 Sodium Lactate 6%	F7 Sodium Lactate 7%	F8 Sodium Lactate 8%	F9 Sodium Lactate 9%	F10 Sodium Lactate 10%	F11 Sodium Lactate 11%	F12 Sodium Lactate 12%
G1 Sodium Phosphate pH 7 20mM	G2 Sodium Phosphate pH 7 50mM	G3 Sodium Phosphate pH 7 100mM	G4 Sodium Phosphate pH 7 200mM	G5 Sodium Benzoate pH 5.2 20mM	G6 Sodium Benzoate pH 5.2 50mM	G7 Sodium Benzoate pH5.2 100mM	G8 Sodium Benzoate pH 5.2 200mM	G9 Ammonium sulfate pH8 10mM	G10 Ammonium sulfate pH 8 20mM	G11 Ammonium sulfate pH 8 50mM	G12 Ammonium sulfate pH8 100mM
H1 Sodium Nitrate 10mM	H2 Sodium Nitrate 20mM	H3 Sodium Nitrate 40mM	H4 Sodium Nitrate 60mM	H5 Sodium Nitrate 80mM	H6 Sodium Nitrate 100mM	H7 Sodium Nitrite 10mM	H8 Sodium Nitrite 20mM	H9 Sodium Nitrite 40mM	H10 Sodium Nitrite 60mM	H11 Sodium Nitrite 80mM	H12 Sodium Nitrite 100mM

Figure 4.9: **Phenotypic microarray plate (PM9) evaluating the effect of osmolarity on resistance to colicin E9.** ***Green***: Respiration UPEC ST131 with 1/3 MIC active ColE9, ***Red:*** Respiration of UPEC with inactive ColE9.



PM10 MicroPlate™ pH

A1 pH 3.5	A2 pH 4	A3 pH 4.5	A4 pH 5	A5 pH 5.5	A6 pH 6	A7 pH 7	A8 pH 8	A9 pH 8.5	A10 pH 9	A11 pH 9.5	A12 pH 10
B1 pH 4.5	B2 pH 4.5 + L-Alanine	B3 pH 4.5 + L-Arginine	B4 pH 4.5 + L-Asparagine	B5 pH 4.5 + L-Aspartic Acid	B6 pH 4.5 + L-Glutamic Acid	B7 pH 4.5 + L-Glutamine	B8 pH 4.5 + Glycine	B9 pH 4.5 + L-Histidine	B10 pH 4.5 + L-Isoleucine	B11 pH 4.5 + L-Leucine	B12 pH 4.5 + L-Lysine
C1 pH 4.5 + L-Methionine	C2 pH 4.5 + L-Phenylalanine	C3 pH 4.5 + L-Proline	C4 pH 4.5 + L-Serine	C5 pH 4.5 + L-Threonine	C6 pH 4.5 + L-Tryptophan	C7 pH 4.5 + L-Citrulline	C8 pH 4.5 + L-Valine	C9 pH 4.5 + Hydroxy-L-Proline	C10 pH 4.5 + L-Ornithine	C11 pH 4.5 + L-Homoarginine	C12 pH 4.5 + L-Homoserine
D-1 pH 4.5 + Anthranilic Acid	D2 pH 4.5 + L-Norleucine	D3 pH 4.5 + L-Norvaline	D4 pH 4.5 + α- Amino-N-Butyric Acid	D5 pH 4.5 + p-Amino-Benzoic Acid	D6 pH 4.5 + L-Cysteic Acid	D7 pH 4.5 + D-Lysine	D8 pH 4.5 + 5-Hydroxy Lysine	D9 pH 4.5 + 5-Hydroxy Tryptophan	D10 pH 4.5 + D,L-Diamino pimelic Acid	D11 pH 4.5 + Trimethylamine-N-oxide	D12 pH 4.5 + Urea
E1 pH 9.5	E2 pH 9.5 + L-Alanine	E3 pH 9.5 + L-Arginine	E4 pH 9.5 + L-Asparagine	E5 pH 9.5 + L-Aspartic Acid	E6 pH 9.5 + L-Glutamic Acid	E7 pH 9.5 + L-Glutamine	E8 pH 9.5 + Glycine	E9 pH 9.5 + L-Histidine	E10 pH 9.5 + L-Isoleucine	E11 pH 9.5 + L-Leucine	E12 pH 9.5 + L-Lysine
F1 pH 9.5 + L-Methionine	F2 pH 9.5 + L-Phenylalanine	F3 pH 9.5 + L-Proline	F4 pH 9.5 + L-Serine	F5 pH 9.5 + L-Threonine	F6 pH 9.5 + L-Tryptophan	F7 pH 9.5 + L-Tyrosine	F8 pH 9.5 + L-Valine	F9 pH 9.5 + Hydroxy-L-Proline	F10 pH 9.5 + L-Ornithine	F11 pH 9.5 + L-Homoarginine	F12 pH 9.5 + L-Homoserine
G1 pH 9.5 + Anthranilic acid	G2 pH 9.5 + L-Norleucine	G3 pH 9.5 + L-Norvaline	G4 pH 9.5 + Agmatine	G5 pH 9.5 + Cadaverine	G6 pH 9.5 + Putrescine	G7 pH 9.5 + Histamine	G8 pH 9.5 + Phenylethylamine	G9 pH 9.5 + Tyramine	G10 pH 9.5 + Creatine	G11 pH 9.5 + Trimethylamine-N-oxide	G12 pH 9.5 + Urea
H1 X-Caprylate	H2 X-α-D-Glucoside	H3 X-β-D-Glucoside	H4 X-α-D-Galactoside	H5 X-β-D-Galactoside	H6 X-α- D-Glucuronide	H7 X-β- D-Glucuronide	H8 X-β-D-Glucosaminide	H9 X-β-D-Galactosaminide	H10 X-α-D-Mannoside	H11 X-PO4	H12 X-SO4

Figure 4.10: **Phenotypic microarray plate (PM10) evaluating the effect of omolarity on resistance to colicin E9.** ***Green***: Respiration UPEC ST131 with 1/3 MIC active ColE9, ***Red:*** Respiration of UPEC with inactive ColE9.

whether it is linked to the LPS, resistant mutants were submitted to whole genome sequencing and analysed for SNPs. Out of the four isolates only one had a significant SNP, an isolate (UPEC ST131_F09) contained a nonsense mutation in *btuB* which causes a stop codon and premature termination of the receptor protein, BtuB.

4.2.9 Non-specific colicin resistance correlates with the presence and density of O-antigen

To directly test the effect of LPS, specifically the O-antigen, on resistance towards colicins, required a model system of near isogenic strains which possess or lack an O-antigen is required. Such a system has previously been created by inserting a second copy of the *wbbl* gene in *E. coli* MG1665 thereby compensating for the rfb-50 mutation and producing an O-antigen (Browning *et al.*, 2013). Two strains were created: L9 with a WT gene order and L5 with an modified gene order and a substantial decrease in O-antigen (Browning *et al.*, 2013). Plate killing assays (solid media) showed a drastic difference in the MIC of the MG1655, L5 and L9 strains. MG1655 and L5 were both inhibited at 57 pM ColE9. The L9 strain, which has fully restored O-antigen, was resistant to 10 μM ColE9 in solid media assays (>100,000 fold increase in MIC) (Figure 4.11) and to 250 μM ColE9 in liquid media killing assays (>5000 fold increase) (Figure 4.12). As the MG1655, L5 and L9 strains are identical apart from the presentation of O-antigen it suggests that the 100,000 fold change in MIC is due to the presence of O-antigen.

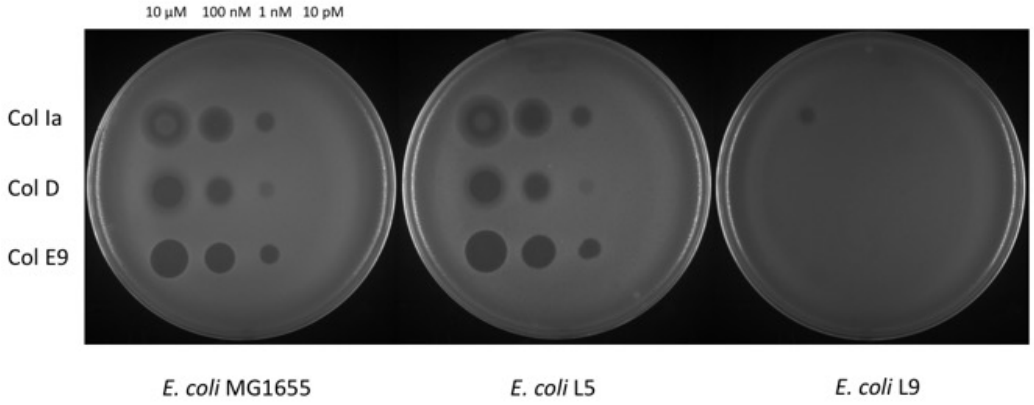


Figure 4.11: **Plate killing assay using a variety of colicins against O-antigen mutants show a non-specific resistance mechanism.** Colicins which hijack different proteins on the target cell were used to highlight a non-specific resistance mechanism. LPS mutants were incubated on LB agar and exposed to a serial dilution of three colicins.

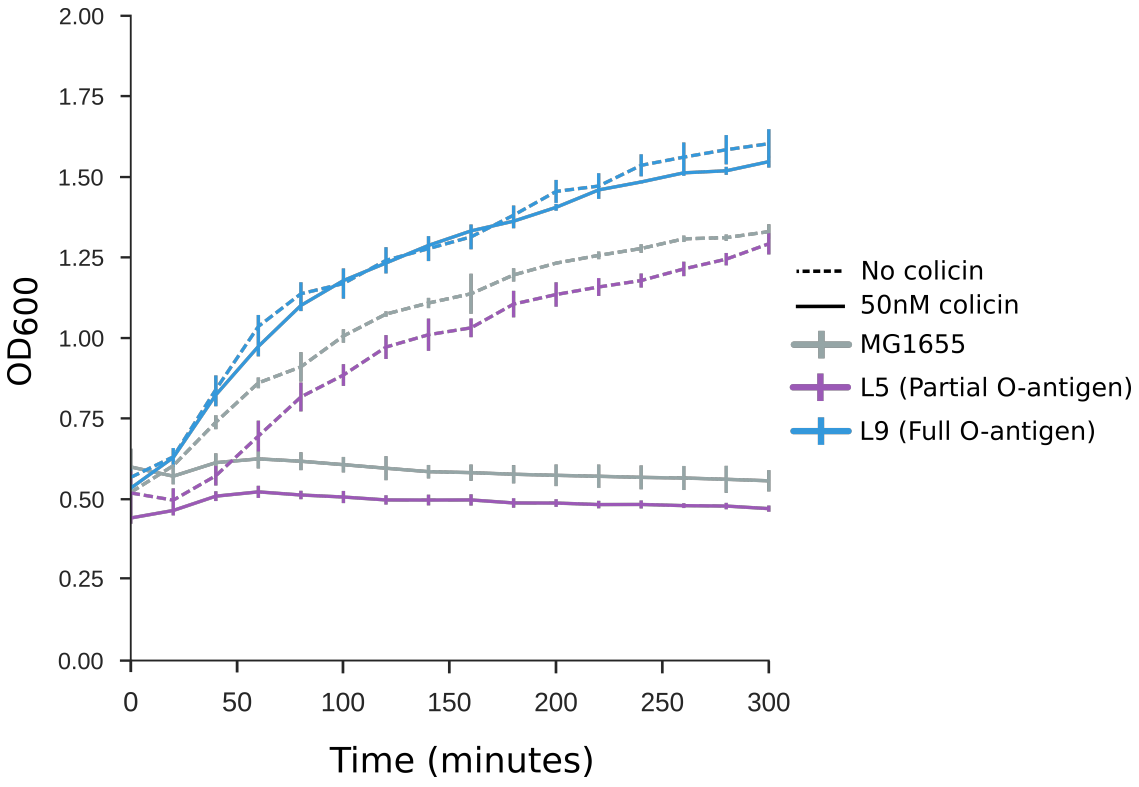


Figure 4.12: **Changes in Colicin E9 susceptibility due to O-antigen.** *E. coli* strains MG1655, L5 and L9 were exposed to 50µM ColE9, the MIC for UPEC ST131. Dashed lines indicate growth in LB media measured using OD_{600}

Four colicins which hijack a variety of different proteins in the target cell were used to ensure that the resistance observed is non-specific. Again, no significant difference was observed between the L5 and WT *E. coli* MG1655 strains on solid media plate killing assays. The L9 strain was resistant to ColD and ColE9 up to 10 μM but showed a slight sensitivity to ColIa (Figure 4.11).

4.2.10 Media additives affect Colicin E9 susceptibility/tolerance of UPEC strains

TraDIS analysis suggested that the resistance mechanisms for UPEC ST131 involves the presence of O-antigen, whilst BIOLOG analysis identified a number of conditions which decrease the MIC of colicin E9 in UPEC ST131. Two conditions: Tris and urea, were chosen for validation. Tris is a commonly used buffer, but it has also been known to trigger LPS release from the OM (Vaara, 1992). Urea was chosen as it is present in the urinary tract as a major constituent of urine. UPEC ST131 was challenged with sub MIC concentrations of ColE9Im9 (33 nM) in the presence of Tris or a physiological concentration of urea (150 mM). UPEC ST131 was sensitive to sub-MIC colicin in the presence of urea, whereas L9 was not (Figure 4.13). The two different O-antigens (L9: O16, UPEC ST131: O25b) could interact differently with urea or, more likely, UPEC ST131 (due to its nature as a uropathogen) could alter its O-antigen in response to urea, sensing that is in the urinary tract. Addition of Tris, which removes LPS from the OM, increased sensitivity of both the L9 and UPEC ST131 strain to sub-MIC colicin.

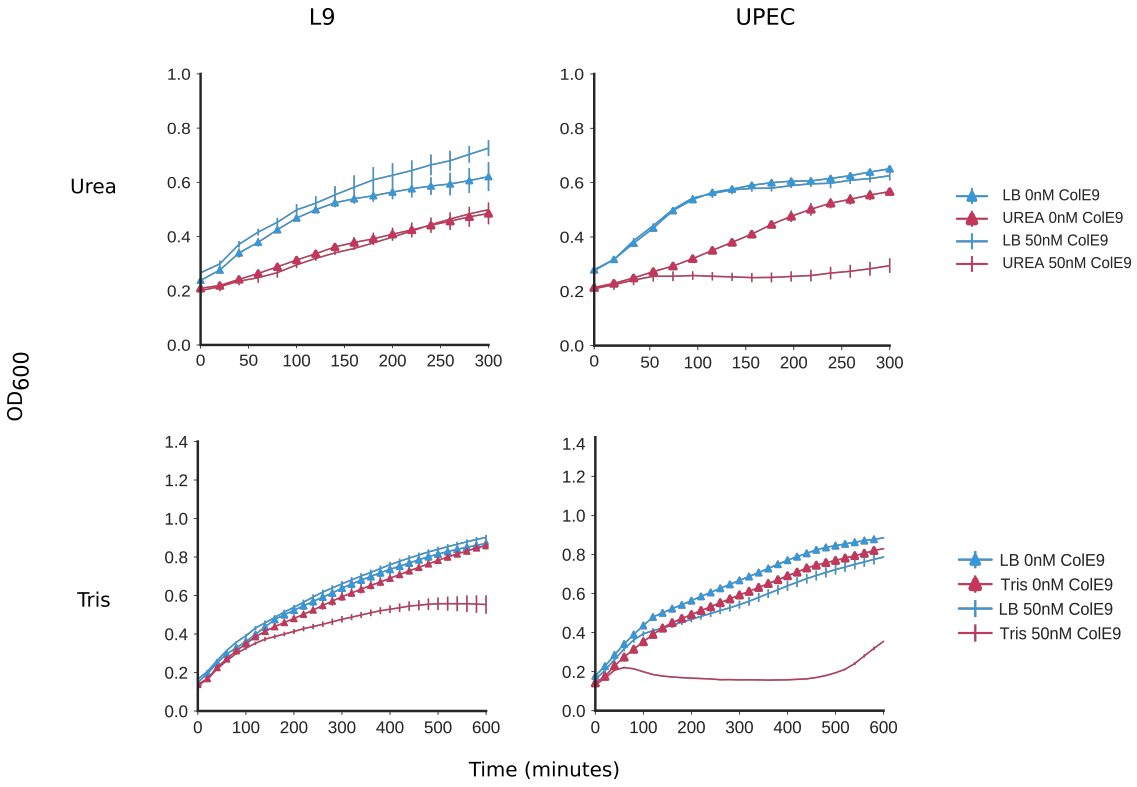


Figure 4.13: **Addition of Tris and urea can increase sensitivity towards Col E9 in strains with an O-antigen.** ***Above***) Addition of physiological concentrations of urea (150 mM) increased the sensitivity of UPEC ST131 but not the L9 strain. ***Below***) Addition of 100 mM Tris was able to decrease the MIC for both UPEC ST131 and *E. coli* L9 strains.

4.2.11 LPS staining revealed differences in concentration of LPS between L9 and L5

To detect differences in the LPS between the L5, L9 and UPEC ST131 strains, LPS was extracted and visualised by SDS-PAGE. UPEC ST131, L5 and L9 showed banding consistant with O-antigen, whereas MG1655 was deficient in O-antigen and only a band corresponding to core LPS was observed. No differences in the presence of length of O-antigen was observed across the different media additives tested. A significant difference was observed between the L5 and L9 strains. L9 with the fully restored O-antigen had an estimated 10-fold higher amount in O-antigen compared to the L5 strain. The length of O-antigen had previously been linked to resistance of a *Shigella* isolate towards colicin E2. The differences in colicin resistance and O-antigen density between the L5 and L9 strains, suggests that the denstity of O-antigen governs non-specific colicin resistance.

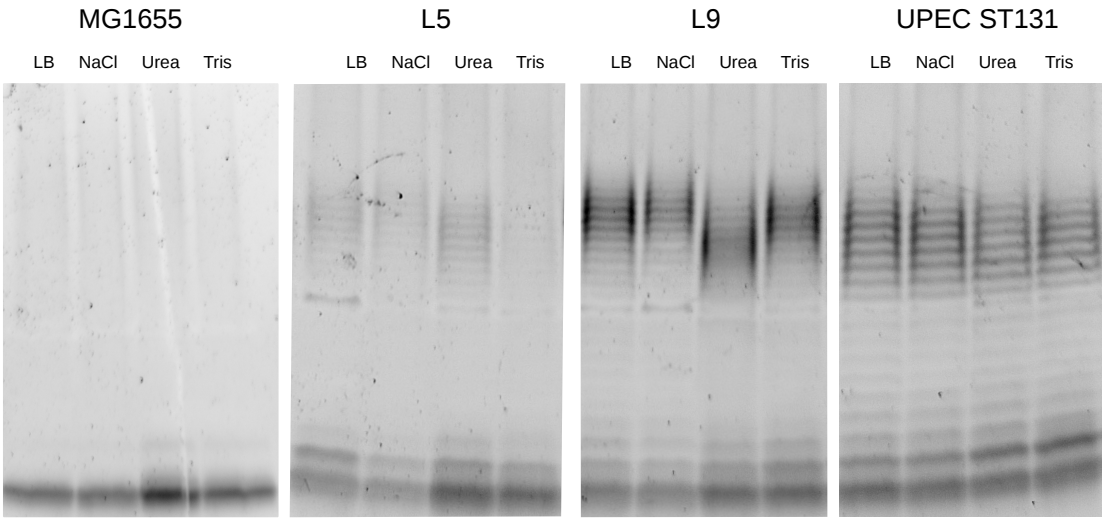


Figure 4.14: **LPS straining shows a distinct difference in the density of O-antigen between L5 and L9** LPS was extracted and visualised using a Pro-Q Emerald stain. The sensitive L5 and the resistant L9 are isogenic apart from a difference in the density of the O-antigen.

4.2.12 Fluorescence microscopy shows O-antigen dependent blocking of the colicin receptor

The O-antigen could provide resistance to colicins by physically blocking access to the receptor proteins in the outer membrane. However, this has never been formally shown. To show a decrease in receptor binding due to occlusion by the O-antigen, fluorescently-labelled cysteine engineered colicin was used to visualise and measure the levels of colicin binding BtuB in the full O-antigen presenting L9 and the partial O-antigen presenting L5 strains (Figure 4.15) (Rassam *et al.*, 2015) (Section 2.12).

A statistically significant difference was observed between the L5 strain, which displayed clear colicin binding and the L9 strain, where no colicin binding was observed (Figure 4.16) (Students t-test, $p<0.01$). Tris, which decreases MIC of colicins for both L9 and UPEC ST131, was tested to evaluate if the decrease in MIC, correlated with an increase in receptor binding. No significant difference in fluorescence intensity was observed for the L5 strain in the presence or absence of Tris suggesting the low density of O-antigen on L5 has no effect on receptor binding. However, addition of Tris caused a significant increase in colicin receptor binding for both the L9 and UPEC ST131 strains. No significant difference in intensity was observed between the L5 strain in LB or Tris and the L9 strain in Tris suggesting 100 mM Tris is enough tos remove any steric hindrance by the O-antigen. Though colicin binding was observed in UPEC ST131, it was significantly increased by the addition of 100mM Tris.

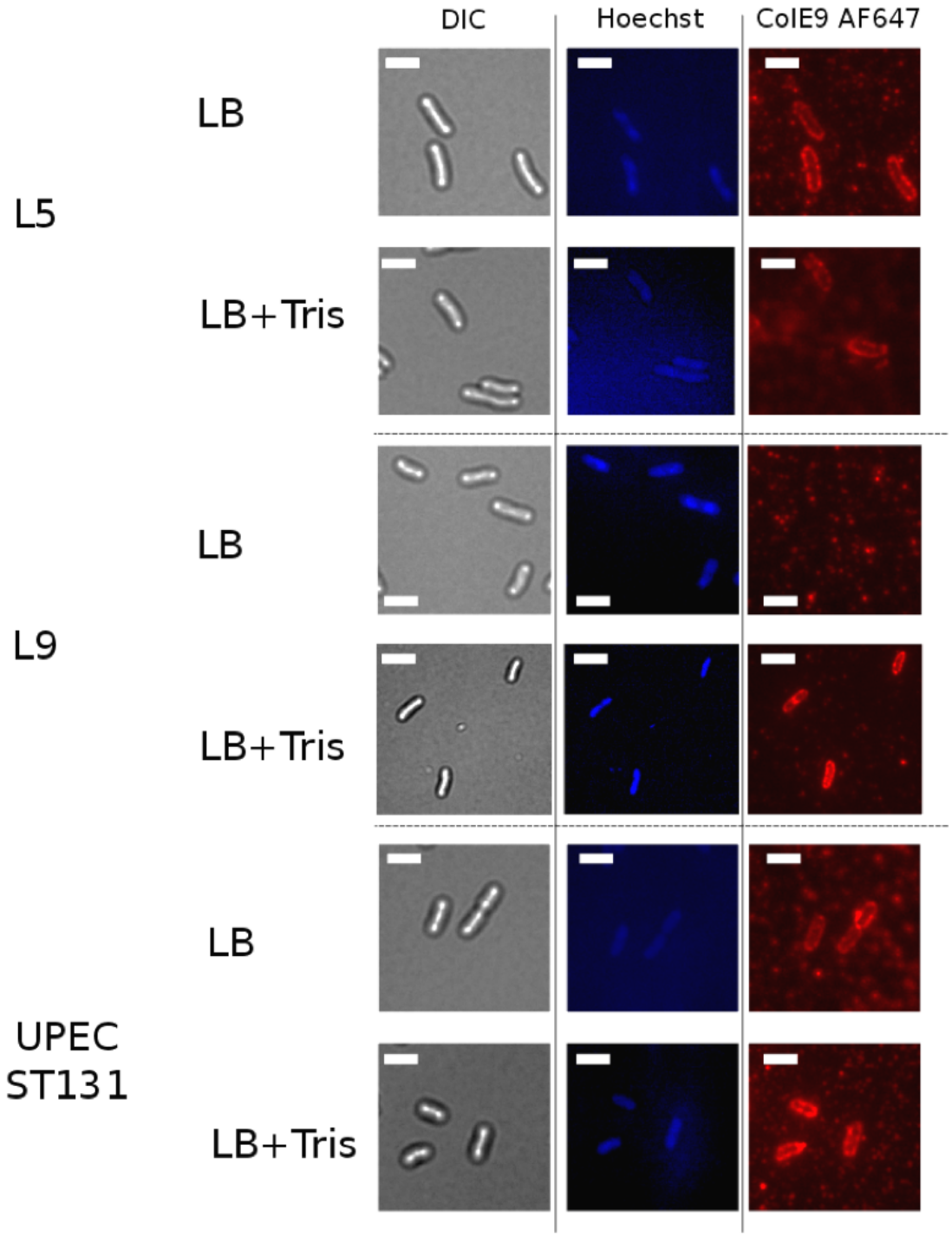


Figure 4.15: **Fluorescently labelled colicin shows difference in receptor binding dependent on the density of O-antigen.** Hoechst stain (blue) was used to stain DNA. Colicin E9 labelled with Alexa Flour 647 (red) was used to visualise the extent of BtuB binding in the outer membrane.

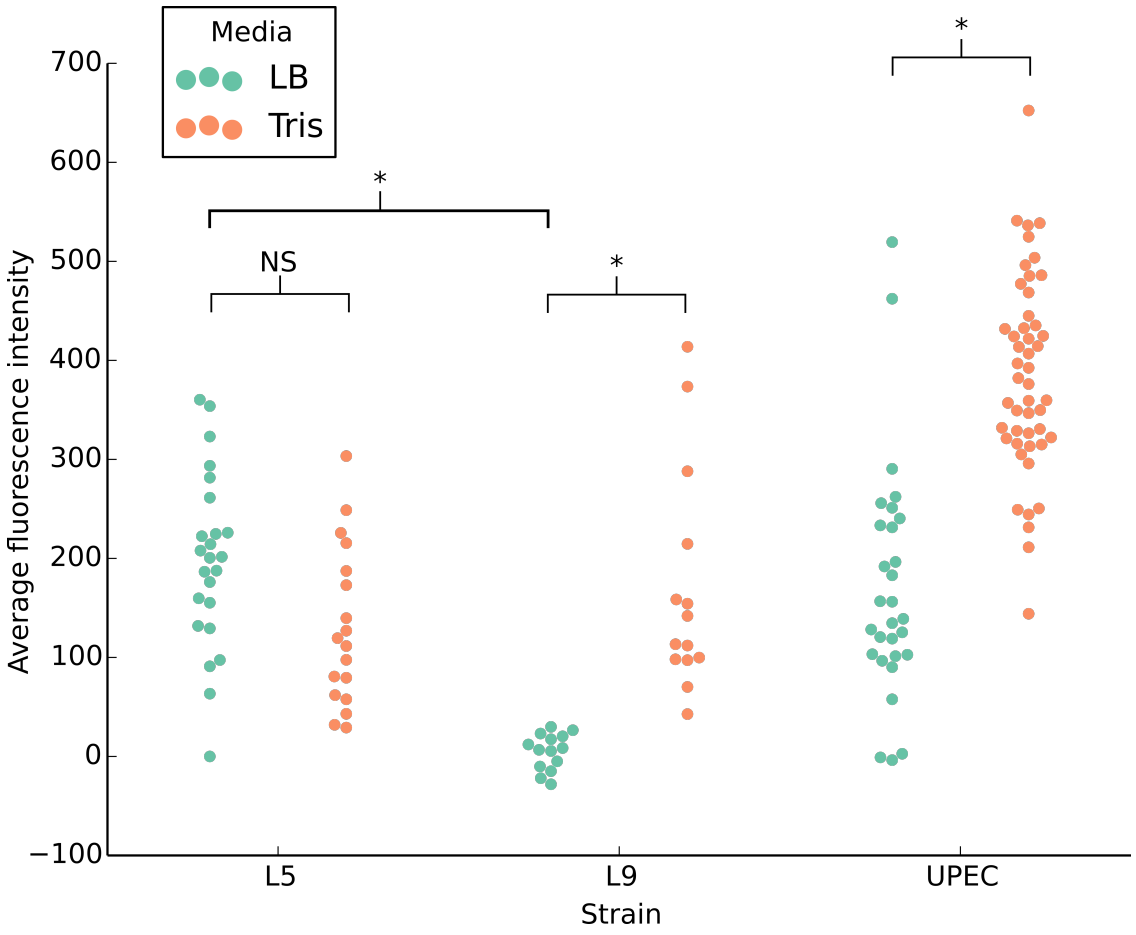


Figure 4.16: **Fluorescence microscopy reveals O-antigen density dependent colicin receptor blocking.** O-antigen producing mutants L5 and L9 were compared to UPEC ST131 by exposure to fluorescently-labelled ColE9 in LB media (*green*) or LB media supplemented with 100mM Tris (*orange*). Each point represents the average fluorescence intensity over a single cell. Statistically significant differences in labelling were observed between the isogenic L5 and L9 strains (t-test). Addition of 100 mM Tris-HCl caused a significant increase in cell labelling for both L9 and UPEC ST131.

4.3 Discussion

Nuclease bacteriocin genes encode small immunity proteins which prevent death of the cell by binding the cytotoxic domain. However, when collections of *E. coli* are tested under laboratory conditions, a high percentage (~30%) of strains showed a non-specific resistance mechanism to multiple colicins (Feldgarden and Riley, 1998). O-antigen displaying (smooth) *E. coli* where shown exhibit colicin resistance while spontaneous mutants which lack an O-antigen (rough) were sensitive to colicins (van der Ley *et al.*, 1986). The length of O-antigen was also shown to be important for resistance in *Shigella* (Tran *et al.*, 2014). In both of these studies, bacteria were cultured under optimal conditions for growth. It is unknown if O-antigen dependent resistance is present in pathogenic bacteria and whether or not this resistance is an consequence of the culture conditions used in the lab.

4.3.1 UPEC ST131 is resistant to multiple colicins

UPEC ST131 is a pathogenic and multi-drug resistant clonal group responsible for a high proportion of urinary tract infections across the globe (Petty *et al.*, 2014). UPEC ST131 NCTC13441 (simply referred to as UPEC ST131) is resistant to multiple colicins but genomic analysis did not reveal any immunity protein genes. The colicins used to test resistance, targeted either the Ton or Tol system and utilised different receptor and translocator proteins suggesting a non-specific resistance mechanism (Figure 4.4). UPEC ST131 has a >6000 fold increased MIC of ColE9 compared to the sensitive *E. coli* JM83 strain (Figure 4.3). Testing a variety of clinical UPEC strains revealed a high proportion (6/10) were resistant to ColE9 (Table 4.2).

4.3.2 UPEC ST131 resistance is dependent on the O-antigen

Addition of low concentrations of EDTA to UPEC ST131 decreased the MIC of three colicins: ColIa, ColD and ColE9 (Figure 4.5). As EDTA has been known to destabilise the LPS by chelating magnesium ions which stablise the interaction between neighbouring

LPS molecules, it suggests that the O-antigen is responsible for the colicin resistance observed in UPEC ST131 (Coughlin *et al.*, 1983).

TraDIS revealed a role of the O-antigen in colicin resistance. TraDIS performed in the presence of EDTA (MMTE media) with sub-MIC ColE9 concentration revealed 21 genes that provide a fitness advantage in the presence of colicin. These genes must remain intact to engender resistance. 19 of these genes were involved in either assembly of the LPS core, lipid A modification, O-antigen assembly or enterobacterial common antigen assembly. Repeating the TraDIS in (MMT) media which lacked EDTA confirmed the role of genes in operons involved in LPS assembly and modification.

4.3.3 Non-specific colicin resistance is heavily influenced by environmental conditions

Bacteria in the environment are likely to be in sub-optimal conditions and face multiple stresses. BIOLOG was used to test the effect of 383 growth conditions on resistance. Carbon sources were tested because altering the carbon source may decrease metabolism and prevent UPEC ST131 from expressing a full O-antigen. However, no changes in resistance were observed. This may be due to the diverse metabolism of UPEC ST131, which is most likely able to utilise numerous carbon sources efficiently. Osmolarity and pH had drastic effects on the resistance observed. It is possible that these two conditions could influence the O-antigen and induce changes by altering its charge for example. A transcriptional change through the envelope stress response/PhoPQ is known to alter LPS composition (Dalebroux and Miller, 2014; Monsieurs *et al.*, 2005). Changes in pH have also been shown to alter the permeability of the outer membrane (Alakomi *et al.*, 2000). LPS is often modified at difference stages of infection and colonisation and these changes can be triggered by changes in environment, such as pH, as bacteria face different compartments of the host (Lerouge and Vanderleyden, 2002).

Many environmental conditions have been shown to affect the permeability of the outer membrane (Vaara, 1992; Weiss *et al.*, 1986). Tris and urea have both been linked to increasing the permeability of the outer membrane and increasing sensitivity to hydropho-

bic molecules. Addition of Tris and urea decreased the MIC of ColE9 for UPEC ST131. An increase in sensitivity in the presence of physiological concentrations of urea presents an intriguing possibility. Uropathogenic *E. coli* are transmitted to a new host by fecal-oral contamination before mechanical transport to the urinary tract from the anus (Lane *et al.*, 2006; Russo *et al.*, 1995). UPEC bacteria could be resistant to colicin in the gut, where it could encounter colicin producers but as it moves to the urinary tract where only uropathogenic *E. coli* can colonise, it may lose its colicin resistance in the presence of urea. It is also possible that colicins could be an effective treatment for UTIs whilst not affecting commensal *E. coli*.

Tris has been shown to cause shedding of LPS which allows the diffusion of small hydrophobic molecules. ColE9 in complex with its immunity protein is a charged ~70kDa protein complex incapable of diffusing through the membranes. ColE9 immunity release is thought to be force dependent and requires the presence of the OM to provide the counter force (Housden *et al.*, 2005). Therefore, addition of Tris is not perforating the membrane to such an extent as to allow the large ColE9 to diffuse through the OM, as ColE9 would be unable to disassociate from its immunity protein and would therefore be inactive.

4.3.4 O-antigen blocks colicin receptor binding

Isogenic O-antigen mutants *E. coli* MG1655, L5 (partially restored O-antigen) and L9 (fully restored O-antigen) demonstrated that density of O-antigen is important for non-specific colicin resistance. L9 had a MIC greater than that of UPEC ST131 for all the colicins tested. Two possible explanations for this could be: (i) the two O-antigens (L9: O16 and UPEC: O25) naturally provide different levels of resistance to colicins or (ii) that the L9 strain, due to its complemented *rfb-50* mutation, might be producing levels of O-antigen greater than wild-type.

O-antigen dependent resistance is predicted to act by blocking binding of receptor proteins, however, this has never been directly shown. Using fluorescently labelled ColE9 to visualise binding to BtuB in the outer membrane, receptor binding was shown to be com-

pletely abolished in the L9 strain but still present in the sensitive and isogenic L5 strain. Addition of Tris, which has been shown to render L9 sensitive to ColE9, was enough to restore receptor binding in L9 and significantly increased receptor binding in UPEC ST131 (Figure 4.16). Unfortunately, addition of urea increased background florescence and made any meaningful statistical analysis impossible.

The difference in colicin sensitivity between L5 and L9, which differ only in O-antigen density, coupled with the observation that Tris, which partially removes LPS, show that subtle differences in the density of O-antigen on the cell surface can block or allow receptor binding and influence resistance. Bacteria in the environment are unlikely to have O-antigen as dense as that in laboratory conditions, meaning the scale of resistance to colicin observed in lab studies could drastically over-represent the real levels of resistance. It also suggests resistance could change dependent on environment, host infection site and stage of infection. The host immune system provides a selective pressure causing many bacteria to alter or entirely lose O-antigen for immunomodulation and immune system evasion. For example, *P. aeruginosa* in a chronic infection lacks an O-antigen while *Burkholderia dolosa* gains the ability to produce O-antigen during infection (Maldonado *et al.*, 2016).

4.3.5 Future work

Colicins are thought to have a role in maintaining a diverse microbial ecology but it is possible that they have an entirely different purpose. In a mixed community of colicin producing and colicin resistant (due to O-antigen) bacteria, the basal level of colicin present in the environment will act as a selective pressure to ensure all bacteria maintain their O-antigen, increasing the fitness of the overall population.

The exploration of O-antigen dependent resistance has generated many hypotheses for future testing. To explore the changes in sensitivity to colicin during infection, an animal model using a strain resistant in the lab (such as UPEC ST131) is required. This model would need to explore different infection sites and chronic versus acute infections.

To date, O-antigen dependent resistance has only been studied using colicins. As the O-

antigen varies greatly between species, so also could its ability to block CLBs binding to OM receptors. It is of particular interest to *P. aeruginosa* and the cystic fibrosis lung and the interplay between the S-type pyocins (which bind outer membrane proteins) and the lectin-like pyocins (which bind O-antigen). It is possible that bacteria sensitive to lectin-like pyocins could be resistant to S-type pyocins and that loss of the O-antigen could cause a reversal in sensitivity. A mixture of lectin-like and S-type pyocins may therefore offer a greater killing spectrum.

In addition to the O-antigen, many bacteria produce a capsule, a thick layer of polysaccharide which could also act to block receptor binding. The thickness of capsule can be up to ~160 nm in *K. pneumoniae* strains and as little as 10 nm in *E. coli* K1 (Amako *et al.*, 1988).

In conclusion, O-antigen dependent colicin resistance was observed in UPEC ST131 a clinical pathogen. Density of O-antigen was found to be important for resistance using isogenic O-antigen mutants. Additives which cause shedding of O-antigen decreased the MIC of colicins for resistant strains, suggesting that in different environments the density of O-antigen is altered and renders strains more sensitive to colicin. Finally, O-antigen dependent blocking of receptor was demonstrated using fluoresently labelled colicin. O-antigen dependent resistance and its sensitivity to environmental conditions has many implications for the use of colicins, and possibly other colicin-like bacteriocins, as future antimicrobials.

CHAPTER 5

Chapter 5: Investigating the role the DPY motif in nuclease bacteriocins

5.1 Introduction

Nuclease bacteriocins (NBs) differ from both pore forming (PFB) and peptidoglycan precursor degrading bacteriocins as their targets are present in the cytoplasm. NBs are capable of delivering relatively large (10-14 kDa) cytotoxic domains through both membranes of Gram-negative bacteria into the cytoplasm. As bacteria have no dedicated protein import system, NBs must hijack outer and inner membrane proteins of the target cell for their import.

In Chapter 3, a paired-profile HMM approach was used to identify large numbers of NBs within Gammaproteobacteria. Within these sequences a conserved motif was identified which was present in all NBs but absent in the majority of pore-forming bacteriocins. This sequence motif had previously been identified as conserved in a small number of colicin and pyocin sequences and deletion of this motif abolished translocation (Sano and Kageyama, 1993). The lack of this motif within Colicin-like Bacteriocins (CLB) which act within the periplasm and IM, and its apparent requirement in nuclease bacteriocins, which act within the cytoplasm, lead to the hypothesis that this motif is required for translocation across the inner membrane.

5.1.1 Translocation across the inner membrane

Colicin DNase and rRNase domains are positively charged and the higher the net positive charge, the higher the cell killing activity (Walker *et al.*, 2007). Significant destabilization of colicin DNases after contact with anionic phospholipids, such as those found in the inner membrane, has also been observed (Mosbahi *et al.*, 2002, 2004). ColE9 also forms spontaneous channels in planar lipid bilayers but addition of a disulphide bond within the colicin domain abolishes channel forming activity and cytotoxicity of the full length ColE9, suggesting a conformational change occurs on interaction with the inner membrane (Mosbahi *et al.*, 2002). The positive charge of the DNase domain correlates with the insertion depth into the inner membrane, suggesting an electrostatically-driven association (Vankemmelbeke *et al.*, 2012).

Proteolytic processing during import has been shown for ColD, ColE3, ColE2 and ColE7 (Chauleau *et al.*, 2011; Liao *et al.*, 2001; Mora and de Zamaroczy, 2014). In these colicins only the cytotoxic domain was identified within the cytoplasm. Cleavage requires the AAA+ protease FtsH, an essential bacterial enzyme that removes misfolded proteins from the inner membrane and transports them to the cytoplasm for degradation (Mora and de Zamaroczy, 2014).

The exact mechanism by which colicin cytotoxic domains pass through the inner membrane has yet to be established. The insertion of DNase domains into the inner membrane may be enough to cause them to be recognised by FtsH as misfolded proteins. FtsH may then cleave the DNase domain and remove it from the membrane. Without interaction with the anionic phospholipids, the cytotoxic domain may then refold to avoid degradation by FtsH. This hypothesis is supported by the observation that cytotoxic domains which lack the necessary positive charge to bury themselves in the membrane require an additional interaction. Colicin D which has a much lower net positive charge (ColD: +2, ColE3: +11) was shown to interact with LepB, an inner membrane signal peptidase (Chauleau *et al.*, 2011; Mora *et al.*, 2015).

5.1.2 Aims

The mechanisms by which Nuclease Bacteriocins (NBs) cross the inner membrane are poorly understood. The presence of a motif associated with bacteriocins which cross the inner membrane presents an intriguing possibility to investigate the translocation mechanism. This chapter identifies a conserved domain present across all identified NBs on both a sequence and structural level. Hidden Markov Models are used to identify additional proteins which contain this motif and use it to identify further nuclease bacteriocins independent of the cytotoxic domain. Finally, mutant NBs are engineered to investigate the essential nature of this motif.

5.2 Results

5.2.1 Identification of a domain using MEME

3094 bacteriocins identified in the analysis described in Chapter 3 were used to identify a domain conserved across all nuclease bacteriocins. Sequences were clustered at 70% sequence identity, giving 114 nuclease bacteriocin clusters. Multiple Expectation maximisation for Motif Elicitation (MEME) revealed a 20 residue motif, which we term the DPY motif, present in 102 of the NB clusters (Figure 5.1). The motif generated was then used to scan the NBs again using MAST (Motif Alignment and Search Tool). This increased the number of sequence clusters which contained the motif to 106. The MAST search also allows for multiple copies of the motif within the same sequence. Using MAST, multiple copies of the DPY motif were identified within certain sequences. MEME is a powerful tool but it cannot identify gapped motifs, where an insertion or deletion is incorporated into the motif. Manual examination of the remaining 8 sequence clusters identified regions corresponding to the DPY motif with an insertion or deletion.

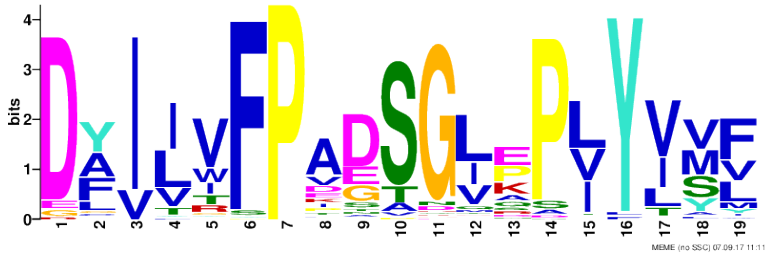


Figure 5.1: **Logo plot of the DPY motif.** Logo plot produced by MEME describing the DPY motif. Heights of each letter indicate the probability of the residue occuring at the position (information measured in bits).

The DPY motif is described as [**D**/E]-X_4-F-**P**-X_2-S-G-X_3-[L/V/I]-**Y**-[L/V/I]. Positions 1, 6 and 7 are extremely conserved even though the analysis covers sequences from 53 species. Position 16, the conserved tyrosine, is invariant. This motif was previously described as conserved in colicins E2 and E3, and pyocins AP41, S1 and S2, and is covered by the Pfam profile pyocin_S (PF06958). The pyocin_S domain is much larger (approximately 140 residues) and contains the highly conserved residues of the DPY motif (Figure 5.2).

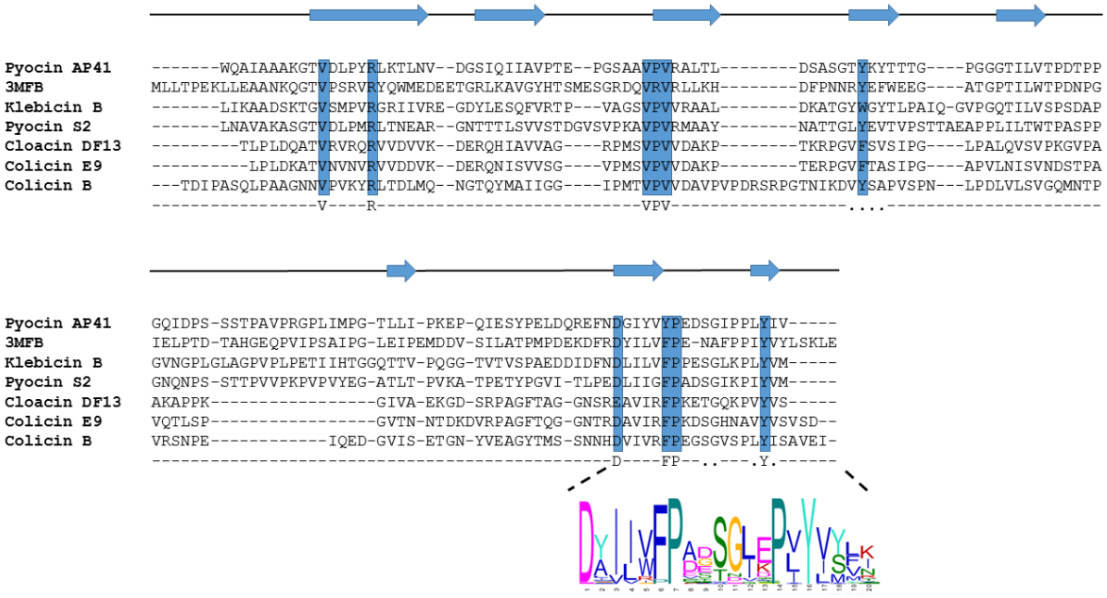


Figure 5.2: **Alignment of the pyocin_S domain from multiple bacteriocins** An alignment of the pyocin_S domain identified in all nuclease bacteriocins revealed a conserved fold in nuclease bacteriocins from many different species. The DPY motif is highlighted at the C-terminal of the pyocin_S domain.

5.2.2 Absence of the DPY in pore forming bacteriocins

The DPY motif, and the larger pyocin_S domain, was identified in all NBs. To test if this domain was present in Colicin-like Bacteriocins (CLBs) which utilize different cytotoxic domains, a selection of pore forming and peptidoglycan degrading bacteriocins were scanned for the DPY motif using MEME, or the entire pyocin_S domain using HMMer3.

Out of the 13 non-nuclease bacteriocins tested, only one example of a non-Nuclease bacteriocin, colicin B, was identified with the DPY motif. The N-terminal domain of ColB shares high sequence identify with the tRNase bacteriocin ColD (~97%). ColD also has a second copy of the DPY motif not present in ColB. It is likely that this DPY motif is an artefact of the shared evolution of colicin B and colicin D. Apart from colicin B, no other pore forming or peptidoglycan precursor degrading bacteriocin contained a DPY motif, suggesting that it is a trait associated with the NBs. Pore-formers and peptidoglycan degrading bacteriocins target the periplasm or the inner membrane suggesting the DPY motif is likely to be associated with translocation across the inner membrane.

5.2.3 Structural alignment

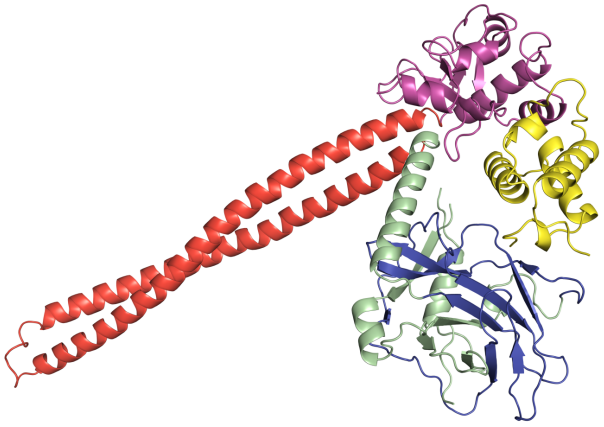


Figure 5.3: **Location of the DPY motif in ColE9** Domain structure of colicin E9 (PDB: 5EW5). Translocation (T-) domain is shown in green with the region covered by the pyocin_s Pfam domain shown in blue. Receptor (R-) domain is shown in red, Cytotoxic domain (C-) domain in purple and the immunity protein in yellow.

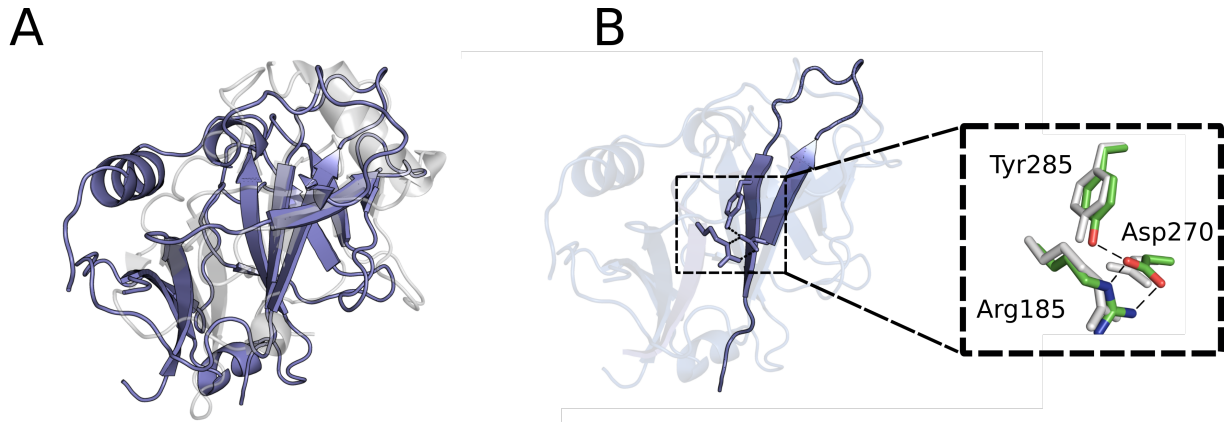


Figure 5.4: **DPY motif is part of a larger structural fold (the pyocin_S domain) conserved in bacteriocin translocation domains A**) alignment of Colicin E9 (blue) and the pyocin_S fragment (PDB: 3mfb, grey). **B**) The ~20 DPY motif forms a β-hairpin. A conserved arginine ~100 residues N-terminal to the DPY motif forms an interacting salt bridge network with the tryosine and aspartate of the DPY motif. Inset shown an alignment of the three interacting residues in ColE9 (coloured) and the pyocin_S structure (grey) (PDB:3mfb)

Few full-length nuclease bacteriocin structures have been solved. Using HHpred, the pdb database was scanned for proteins containing the DPY motif. The DPY motif was identified in structures of full length ColE9 (PDB: 5ew5), ColE3 (PDB: 1jch), ColE7 (PDB: 2axc), ColB (PDB: 1rh1) and a fragment of a pyocin_S domain containing protein from *Pectobacterium carotovora* (PDB: 3mfb). Using the DALI server, the pdb was scanned for proteins which had 3D structural similarity to the pyocin_S domain but no additional structures were identified. Alignment of the pyocin_S domain with the ColE9 translocation domain (RMSD 3.38 Å) showed structural similarity beyond the DPY region (Figure 5.4A). The DPY motif forms a β hairpin at the core of the translocation domain. For ColE9, ColE3 and

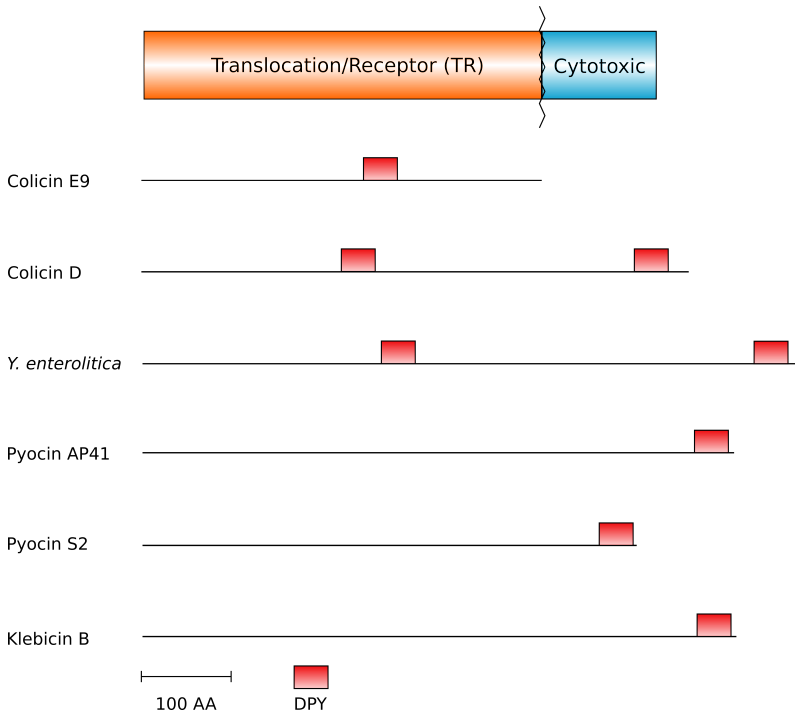


Figure 5.5: **Position of the DPY motif in NBs** The position and number of DPY motifs varies between different NB sequences. TR domains (with cytotoxic domains removed) are shown for 6 different NBs and represented to scale. Positions of the DPY as predicted using MAST are shown as a red box.

ColB, the DPY motif has highly buried residues with the solvent exposed surface area of the Y285 predicted as only 16.6 $Å^2$ (predicted by the POPs server) (Figure 5.4B). Alignment of the pyocin_S domain revealed few conserved residues outside of the DPY motif, however, a single arginine was identified approximately 100 residues N-terminal to the tyrosine of the DPY motif (Figure 5.2). Structurally, this arginine is within distance to form a salt bridge interaction network with the aspartate and tryosine of the DPY motif (Figure 5.2 B inset). This interaction in conserved in all the crystal structures which contain the pyocin_S domain.

5.2.4 Scanning the Multispecies rMLST database for the Pyocin_S domain

The previous bacteriocin identification pipeline required prior knowledge of the cytotoxic/immunity domains to form the paired-profile HMM. The presence of a conserved domain (Pyocin_S) could allow detection of nuclease bacteriocins without prior knowledge of the cytotoxic domain. 154380 genomes of the pubMLST Multispecies rMLST database covering 4030 species were analysed using HMMer3.1 for the presence of

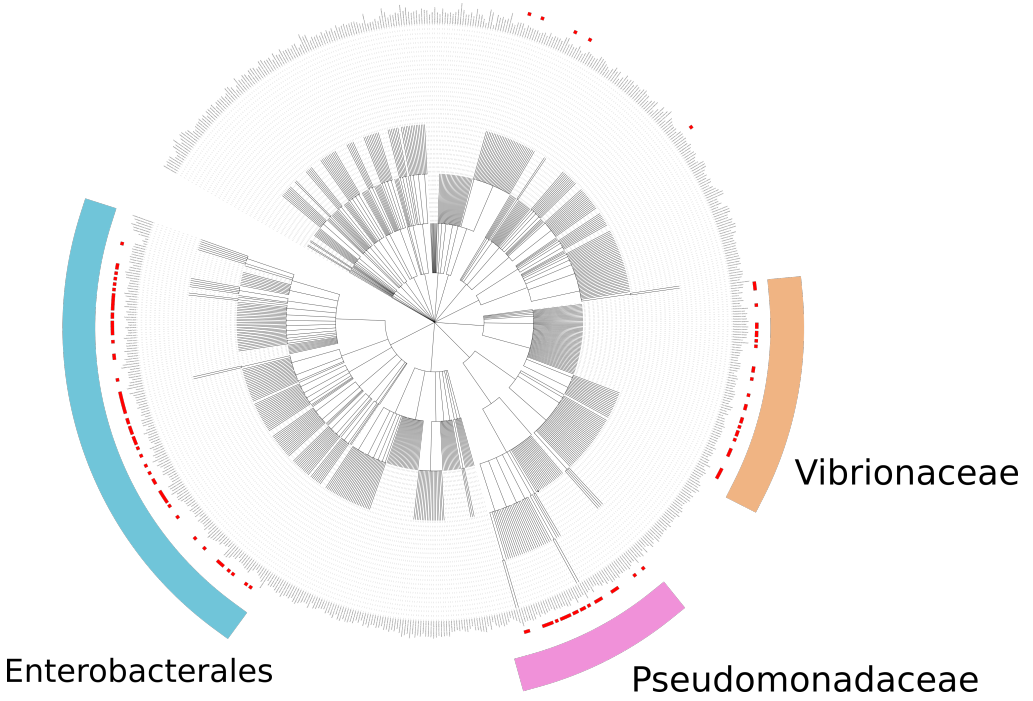


Figure 5.6: **Taxonomic distribution of the DPY motif within *Gammaproteobacteria*.** NCBI common-tree produced taxonomic representation of all the *Gammaproteobacteria* species within the pubMLST. Red squares indicate the species which encoded a protein with contained the pyocin_S domain.

the pyocin_S domain (PF06958). Pyocin_S domain containing proteins were identified in 16875 genomes. The 16923 open reading frames which contained the domain were extracted using Prodigal and analysed by scanning against the Pfam 27 database. The pyocin_S domain was almost exclusively identified within the *Gammaproteobacteria*, primarily within the order *Enterobacterales* and the classes *Pseudomonadaceae* and *Vibrionaceae* (Figure 5.6), with only 4 representatives identified within *Aeromonas* and a single protein in *Halomonas anticariensis*.

5.2.5 Competition system utilising the Pyocin_S domain

Analysing proteins which contained the pyocin_S domain (and the DPY motif contained within that domain) identified a variety of associated Pfam domains (Figure 5.7). Bacteriocin cytotoxic domains were widely associated with pyocin_S, however, the pyocin_S is not specific to NBs. Non-NB proteins, including other competition systems, contained the pyocin_S domain.

The T6SS is a common competition system which uses a molecular 'syringe' to inject

prokaryotic or eukaryotic cells with effector proteins. Proteins with a PAAR (Proline-Alanine-Alanine-aRginine) motif form a trimer spike at the tip of the T6SS needle formed of Hcp proteins. T6SSs can utilize evolved, or specialized effectors, which are effector domains fused with structural elements of the T6SS needle. DUF796 is a Hcp associated protein found in the uropathogenic specific protein (Usp) of *E. coli* and closely related species. It is highly prevalent in the database and present in 2782 genomes. PAAR domains with the pyocin_S domain were identified in 651 proteins across 31 species. Less common T6SS domains, T6SS_Vgr and DUF2345, were also identified with the pyocin_S domain.

Colicin-like nuclease domains were found in 180 PAAR domain proteins (Non-HNH DNase: 85, ColE3 rRNase: 34, HNH-DNase: 61). One protein from *Pectobacterium wasabiae* contained a Toxin52 domain at the C-terminus, a predicted RNase toxin (Zhang *et al.*, 2011). 470 proteins contained no cytotoxic domain/s which could be predicted using Pfam at the C-terminus . Analysis of these proteins identified a conserved domain of ~106 residues (92% sequence identity) at the C-terminus of 457 proteins, 441 of which are within *Yersinia pestis* or *Yersinia pseudotuberculosis*, with low sequence identity to the tRNase domain of colicin D (11% sequence identity). This domain had previously been identified in the genome of *Y. pestis* CO-92 (Parret *et al.*, 2002). The bacterial toxin 17, another predicted RNase toxin (PF15524), was identified as the cytotoxic domain of proteins identified in genomes of *Pectobacterium carotovorum* (Zhang *et al.*, 2012). 98% of the proteins identified with the PAAR and pyocin_S domains contained a cytotoxic domain known to act in the cytoplasm or a predicted RNase domain. 2% of the proteins (12/651) had no identifiable cytotoxic domain. No cytotoxic domains which act within the periplasm or form pores in the inner membrane were identified as associated with T6SS effectors which contained the pyocin_S domain. Usp encodes a HNH-DNase domain and is thought to target mammalian cells during infection. Analysis of proteins containing the DUF796 domain revealed 31 examples of proteins with the DUF796 domain at the N-terminus but an unannotated region at the C-terminus. Scanning the pdb and Pfam database with HHpred predicted similarity to the cytotoxic domain of colicin E3 (1JCH and Pf09000), however the sequence identity is <20%.

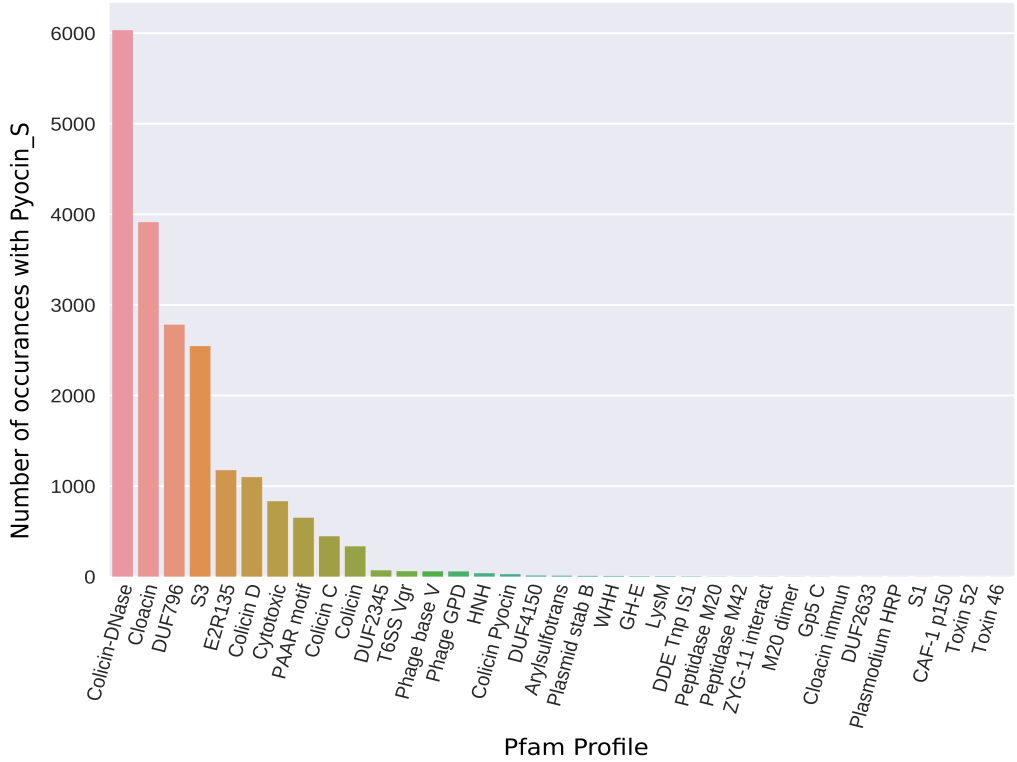


Figure 5.7: **Pfam domains which are associated with the pyocin_S domain.** The pubMLST was scanned for the presence of the pyocin_S domain. Bars indicate the number of times profiles occurred associated with the pyocin_S.

The cytoplasmic final destination of the T6SS effector cytotoxic domains observed in combination with the pyocin_S domain, and the almost complete absence of periplasmic domains, supports the hypothesis that the pyocin_S domain and the DPY motif are associated with translocation across the inner membrane. Multiple studies have suggested that the T6SS only penetrates the OM and injects effectors into the periplasm (Basler *et al.*, 2012; Cotter, 2011). The effectors themselves must therefore pass the inner membrane via an unknown mechanism.

The pore forming colicin domain (PF01024) was identified in 335 pyocin_S domain containing proteins, however, 326 of these proteins had similarity to colicin B which shares N-terminal similarity to colicin. Seven proteins from *Proteus mirabilis* were identified as containing the pyocin_S domain and the pore forming domain. Unlike colicin B, these sequences do not have N-terminal similarity to any NB. Therefore, out of 16923 open reading frames only 335 ORFs were found to contain periplasmic acting cytotoxic domains, of which, all were predicted PFBs and only 7 proteins did not have N-terminal similarity to a recognised NB. The frequent recombination of cytotoxic domains could

account for the small number of proteins which contain both the DPY motif and the pore forming domain.

5.2.6 Identifying NBs using the pyocin_S domain

A pipeline was created to identify NBs using similar criteria to those used in Chapter 1, with one major difference. This pipeline makes no assumption about the cytotoxic domain used by the NBs or the corresponding immunity protein. It instead searches using only the pyocin_S domain and excludes other competition systems by size and N-terminal secretion systems. Identification of an ORF at the 3' of the toxin ORF was also performed, however, the small overlapping ORFs of immunity proteins are often overlooked by ORF-finders and was therefore not used as a validation technique.

6365 potential NBs were identified within the pubMLST database, a significant increase from the previous search due to the increased size of the pubMLST Multispecies rMLST database. NBs were identified with the 5 different cytotoxic motifs used in the profile-paired HMM strategy used in chapter 3; HNH-DNase, rRNase, ColD-tRNase, E5-tRNase and non-HNH DNase. An additional ~143aa cytotoxic domain was identified in 763 predicted NB proteins across 16 species. This domain had high sequence similarity (~60%) to that of carocin D, a predicted nuclease bacteriocin present in *Pectobacterium carotovorum* (formerly *Erwinia carovotora*) (Roh *et al.*, 2010). To avoid confusion this domain is termed NovD, to represent both its novelty (novelty in NBs outside of carocinD) and distant homology to Colicin D. Homology to the pdb structures of colicin D and E3 was detected using HHpred, suggesting this domain functions as a rRNase, however, experimentally carocin D displayed DNase activity (Roh *et al.*, 2010). Previous analysis identified a high number of recombinant pyocins (e.g. S2 and SD2) which give the appearance of cytotoxic domain swapping, a common trait in polymorphic toxin systems. Three pyocins were identified which contained the pyocin S1 TR domains but instead of the HNH-DNase domain, utilised the NovD domain. This protein (Accession: ETV05907.1) and its immunity protein (Accession: ETV05906.1)was cloned, expressed and purified by a similar approach to pyocin Sn courtesy of Dr. Nick Housden (Sharp *et al.*, 2017) and in plate killing assays displayed activity against three strains of *P. aeruginosa*, demonstrat-

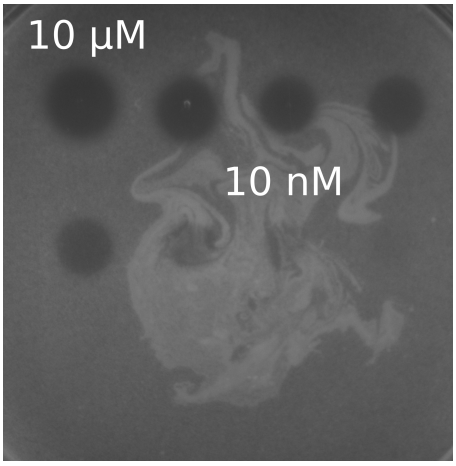


Figure 5.8: **Activity of a novel pyocin cytotoxic domain.** Purified pyocinS1NovD was able to kill all of the 3 strains tested: *P. aeruginosa* PA01, *P. aeruginosa* PA17 and *P. aeruginosa* UCBPP -PA14, at concentrations down to 10 nM. Results are shown for *P. aeruginosa* PA01.

ing that the NovD domain is capable of acting as a cytotoxic domain (Figure 5.8). NBs with the NovD domain at the C-terminus were identified in five species of *Pseudomonas*: *P. aeruginosa*, *P. putida*, *P. savastanoi*, *P. fluorescens* and *Pseudomonas* sp..

Despite the high prevalence of klebicins within the *Klebsiella pneumoniae* population, previous analysis using the paired-profile HMM approach did not reveal cytotoxic domain swapping. 14 klebicins were identified which contained the klebicin B TR domains associated with the NovD domain.

The NovD domain was particularly prevalent within *Serratia marcescens* where it is present in 152 (61%) of the genomes analysed (Figure 5.9). The NovD domain was also present associated with TR domains identified by the ppHMM approach, suggesting cytotoxic domain recombination (Figure 5.10). These TR domains have high similarity with klebicin D and C from *Klebsiella pneumoniae*.

The previous ppHMM strategy of chapter 1 did not identify any NBs within *Y. pestis*. A single 490aa predicted NB, which contained the NovD domain, was identified in 94% of *Y. pestis* genomes analysed. The high prevalence of a single protein is atypical of NBs which are often prevalent in small proportions of the accessory genome. The lack of genetic diversity within *Y. pestis*, which is a relatively recent clonal group of *Y. pseudotuberculosis*, could explain how a competition system becomes fixed in the population (Vogler *et al.*, 2016). The pyocin_S strategy does not identify any other NBs within *Y. pestis*, agreeing

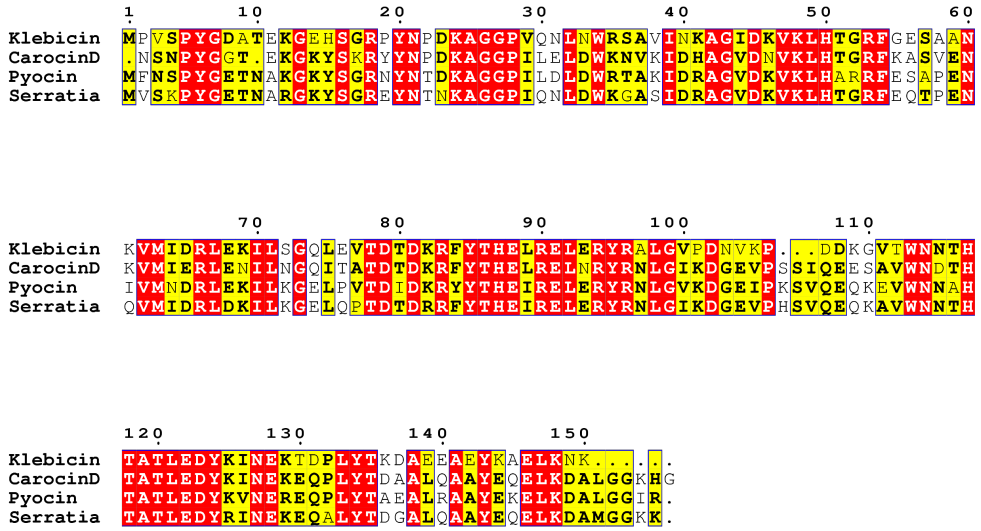


Figure 5.9: **Alignment of a novel cytotoxic domain.** A cytotoxic domain identified in predicted NBs from *K. pneumoniae*, *Serratia* spp. and *Pseudomonas* spp. which has previously only been reported in carocin D, an NB from *Pectobacterium*.

with the results of the ppHMM. It also does not identify the only characterised bacteriocin of *Y. pestis*, the murimidase pesiticin, supporting the hypothesis that the pyocin_S domain is not associated with bacteriocins which do not cross the inner membrane.

5.2.7 Mutation of DPY residues

To test for a role of the three residues which form an interacting network in the pyocin_S domain, RDY (Figure 5.4), a series of single and double alanine mutations were generated in ColE9ImHis$_6$. Mutant proteins were expressed (Figure 5.11) and purified (Figure 5.12). Both crude cell lysate and purified proteins show that the tyrosine at position Y285 is essential for cytotoxic activity against *E. coli* K-12. Any mutation combination which contained the Y285A substitution was unable to kill the sensitive target strain.

An additional mutation, Y285F, was generated to test whether the aromatic property or the hydroxyl group of tyrosine is what makes this residue vital. The ColE9 Y285F mutant was able to kill *E. coli* JM83 and behaved as WT, suggesting that the aromatic side chain is necessary for translocation (Figure 5.15).

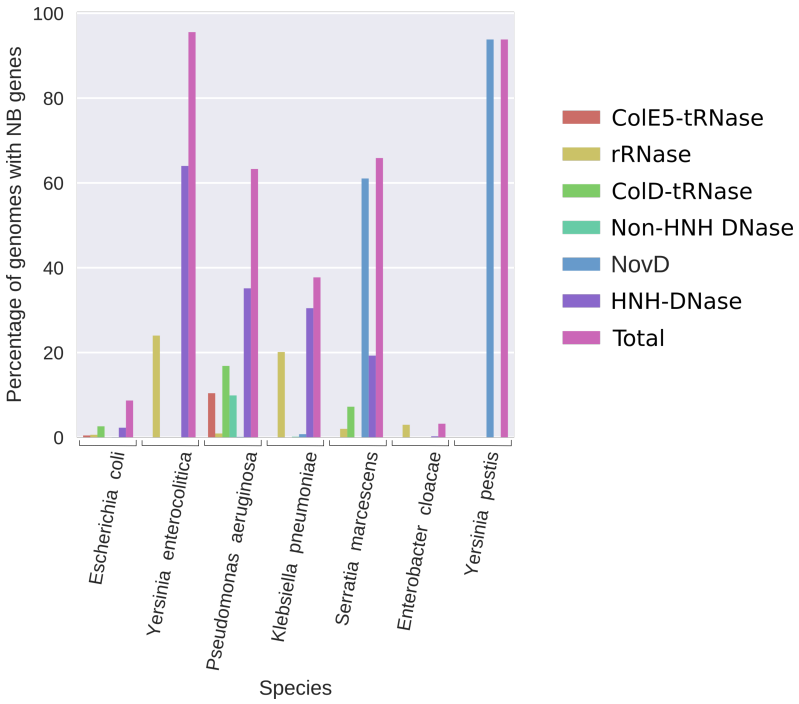


Figure 5.10: **Distribution of predicted NB genes amongst the most abundant species.** Predicted NBs identified by the pipeline separated by species and the predicted cytotoxic domain.

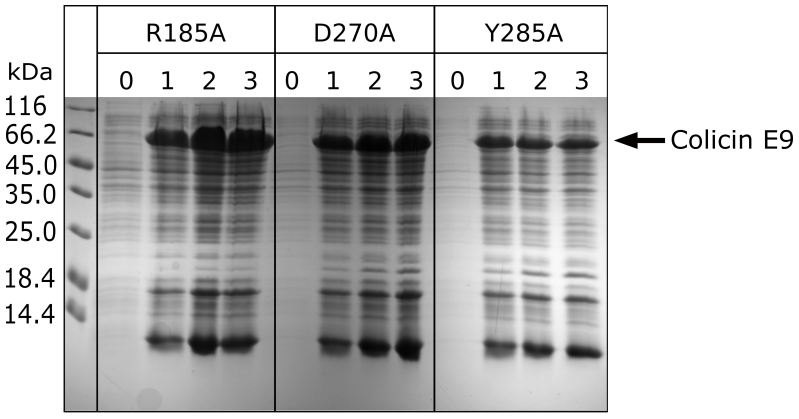


Figure 5.11: **Expression of 3 DPY alanine mutants.** Three pET21a plasmids with individual alanine mutations of the residues which form an interacting network at the core of the pyocin_S domain in ColE9 were induced at 0 hours by addition of 1mM IPTG and samples taken at 1, 2 and 3 hours post induction.

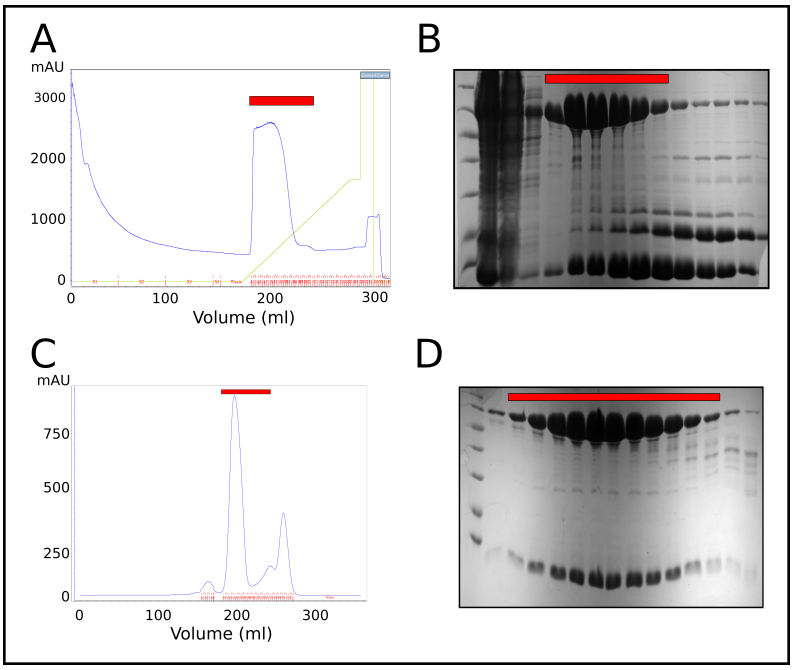


Figure 5.12: **Example purification of a colicin E9Im9His$_6$** A) Chromatogram showing elution of protein from a HisTrap nickel column by measuring absorbance at 280 nm (blue). The imidazole gradient used to elute protein is shown in green. B) SDS-PAGE showing eluted fractions from the HisTrap column. Red bars indicate the fractions that were pooled and passed forward to size exclusion. C) Chromatogram showing protein eluted from the HiLoad 26/60 Superdex S200 column (GE Healthcare) measured using absorbance at 280 nm (blue). D) SDS-PAGE showing eluted fractions from the S200 column. The red bar indicates the fractions pooled and used for further experiments.

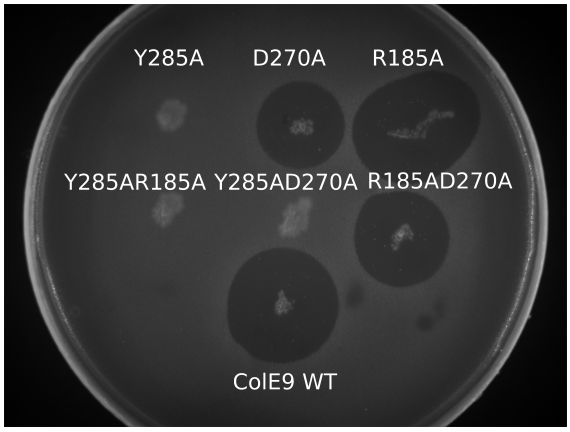


Figure 5.13: **Killing activity assay by chloroform lysis** Cells with Colicin E9 containing plasmids were induced by growth on agar plates with 1mM IPTG. Cells were lysed by exposure to chloroform. Colicin mutants capable of killing produced a zone of clearance in the *E. coli* JM83 overlaid soft agar. All mutant forms containing the Y285A mutation showed no cytotoxicity.

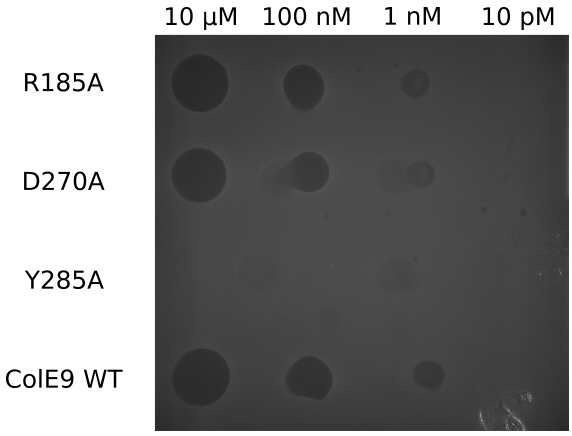


Figure 5.14: **Killing assay using purified colicin E9 DPY mutants** Proteins purified by nickel affinity and size exclusion, were tested for cytotoxicity by spotting a serial dilution onto a soft agar lawn of sensitive *E. coli* JM83 cells.

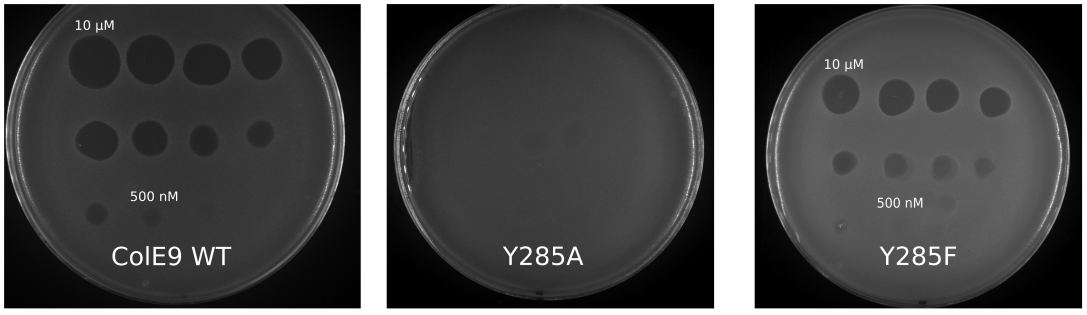


Figure 5.15: **Mutation the Y285 residue within the DPY domain** Killing assays using serial dilutions of purified proftein. The Y285A mutation completely abolishes cytotoxicity of ColE9. The Y285F substitution, which conserves the aromatic property of the side chain, still kills target cells.

Mutation of the DPY motif of dual-DPY proteins

Colicin D contains two pyocin_S domains with DPY motifs, with the conserved tyrosines positioned at Y282 and Y588 (Figure 5.5). Mutation of the tyrosine residue to alanine in ColE9 was sufficient to abolish translocation and cell killing. To test if this substitution can abolish translocation in ColD, both tyrosines, in turn and in combination, were mutated to alanine. Y282A, Y588A and Y282A Y588A, were able to translocate and kill the sensitive *E. coli* JM83 (Figure 5.16).

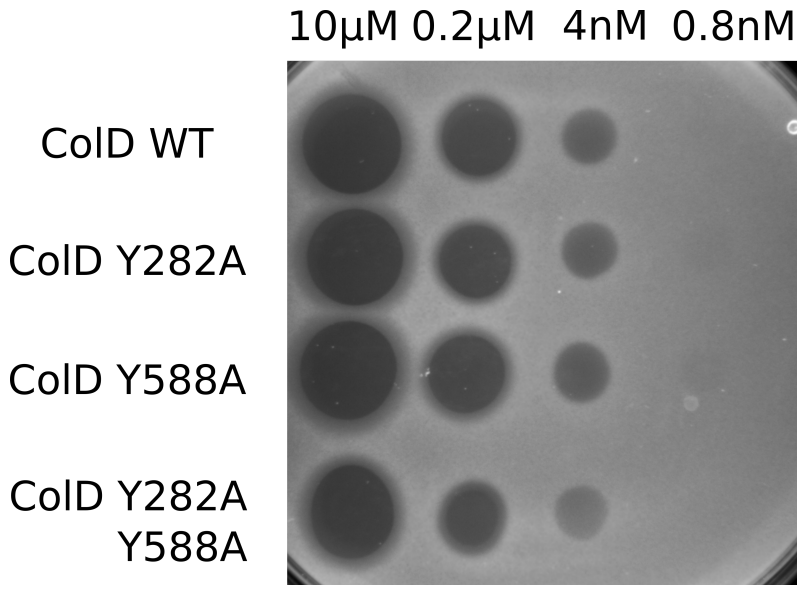


Figure 5.16: **Killing assay for DPY mutants of ColD.** Purified ColD constructs with a mutation of DPY motif tyrosine in one or both of the DPY motifs.

Mutation of the DPY motif in pyocin S2

Pyocin S2 contains a single copy of the DPY motif with the conserved tryosine at position Y552, directly adjacent to the cytotoxic domain, in contrast to ColE9. To test if mutation of the conserved tyrosine would abolish translocation in a manner similar to that in ColE9, whole plasmid mutagenesis was performed to introduce the Y552A mutation. However, test expression of *E. coli* BL21 (DE3) cells containing either WT pyocin S2 or pyocin S2 Y552A encoded in a pET21a plasmid demonstrated that the Y552A mutant is not expressed after induction and incubation for 3 hours (Figure 5.17). Lowering the incubation temperature to 30°C did not increase expression.

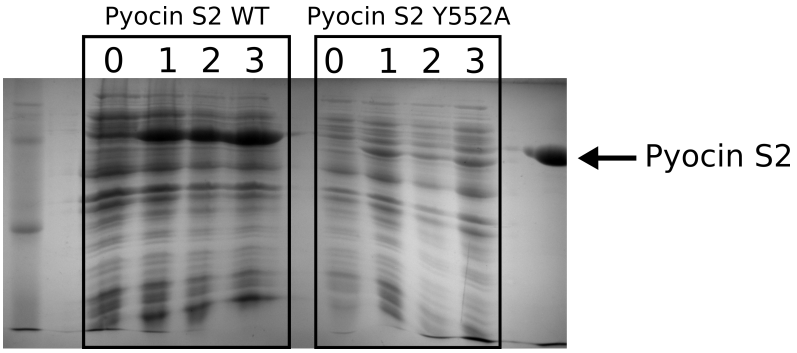


Figure 5.17: **Test expression of pyocin S2ImS2 DPY mutant.** Expression of WT pyocin and an alanine mutation of the conserved DPY tyrosine (Y552A) in pyocin S2. Pyocin S2 Y552A is not expressed 3 hours post induction. Pyocin S2 WT is produced and present in significant concentrations upto 3 hours post induction. *Left*) Ladder, *Right*) purified WT pyocin S2.

5.2.8 Incorporating *p*-benzoylphenylalanine into the DPY motif

p-Benzoylphenylalanine (Bpa) is a unnatural photoreactive amino acid which can be incorporated into proteins (Kauer *et al.*, 1986). After excitation with UV-light (350 nm) Bpa rapidly forms a covalent bond with nearby unactivated C-H bonds (Figure 5.18). Bpa has routinely been incorporated into small peptides by solid-phase peptide synthesis (Kauer *et al.*, 1986) and has been used to identify interactions of steroids, lipids and nucleotide binding peptides (Dorman and Prestwich, 1994; Williams *et al.*, 1986). Expansion of the translation machinery in *E. coli* to allow incorporation of unnatural amino acids, provides a simple method for Bpa incorporation in large proteins and a useful tool for investigating protein-protein interactions (Chin *et al.*, 2002a,b).

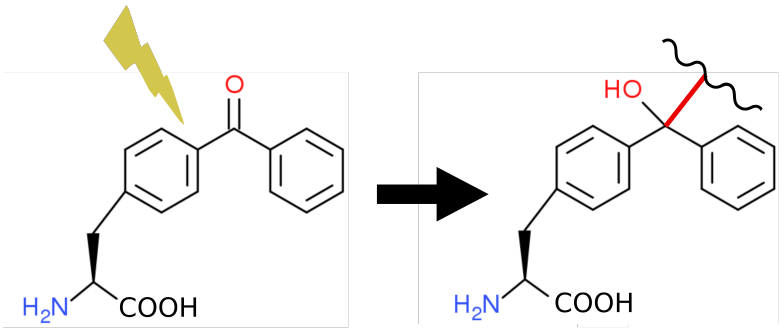


Figure 5.18: **Photo-activatable crosslinking of *p*-benzoylphenylalanine.** Upon excitation with UV-light at 350 nm, Bpa reacts with unactivated C-H bonds forming a stable covalently bonded product.

Using the pEVOL-pBpF plasmid (Chin *et al.*, 2002a,b) Bpa was incorporated into ColE9Im9His$_6$ at positions Y285 and D270. Bpa containing proteins were successfully expressed in the presence of Bpa but became truncated in its absence. Protein constructs are yet to be tested.

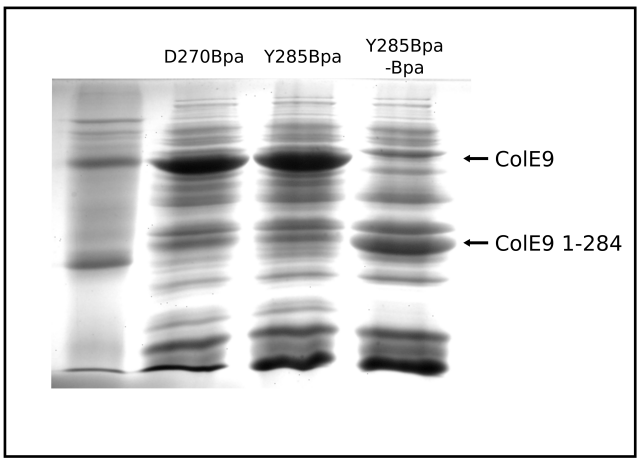


Figure 5.19: **Crude cell extract of Bpa mutant expressing cells.** *E. coli* BL21 strains containing ColE9Im9His$_6$ on an IPTG inducible plasmid and the pEVOL-pBpF plasmid. Cells were induced with addition of 1mM IPTG for 3 hours with or without the addition of Bpa and analysed by SDS-PAGE. In the absence of Bpa truncated proteins are produced.

5.3 Discussion

5.3.1 Characterisation of the pyocin_S domain and the DPY motif

CLBs are a diverse family of proteins targeting a myriad of outer membrane receptor and translocator proteins across over 50 different species. This diversity leads to extremely low sequence similarity across the family. Despite this great diversity, using MEME a single motif, which has previously been identified in a small number of colicin and pyocin sequences (Sano *et al.*, 1993a), was identified within all Nuclease Bacteriocins (NBs) outlined by the ppHMM approach used in Chapter 3. Conserved residues in this motif lead to the name 'the DPY motif'. The ppHMM strategy identified bacteriocins based on conserved residues within the cytotoxic domain. Though a copy of the DPY motif is adjacent to the cytotoxic domain in 7 of the 8 structural domain organisation observed, it is independent of the domain and not included within any of the cytotoxic domain pHMMs. The presence of the pyocin_S domain (which encompasses the DPY motif) was not used as a validation criteria at any point during the ppHMM pipeline, therefore the presence of the 'DPY' motif is an independent observation. Alignment of the pyocin_S domain showed few conserved residues outside of the DPY motif. Alignment of crystal structures however, revealed a conserved fold which at its core contained an interacting network containing the tyrosine and aspartate of the DPY motif interacting with a highly conserved arginine ~100 residues N-terminal to the motif. This interaction was present in all colicin T-domain crystal structures and the crystal structure of the pyocin_S domain of a protein from *Pectobacterium carotovorum*.

5.3.2 The pyocin_S domains is primarily found within Gammaproteobacteria

The Pfam profile for pyocin_S (PF6958) was used to scan the entire pubMLST database, covering >150000 bacterial genomes. The pyocin_S domain was almost exclusively identified within the Gammaproteobacteria (121/127 species) and was particularly prevalent within the order *Enterobacterales* and the classes *Pseudomonadaceae* and *Vibrionaceae*.

This distribution largely matches the distribution of NBs previously identified (apart from the addition of the *Vibrionaceae*). Though the reason for the restricted distribution of NBs was not determined, the DPY motif and whatever function it may serve could be a possible factor. A great deal of further work needs to be performed to firm up this hypothesis.

5.3.3 The DPY motif is present in T6SS effectors

Searching for proteins which contained the pyocin_S domain identified T6SS effectors which contain cytotoxic domains associated with T6SS structural elements: Vgr (Pukatzki *et al.*, 2009) , Hcp (Blondel *et al.*, 2009), DUF4150 (Ma *et al.*, 2014) and PAAR domains (Shneider *et al.*, 2013; Whitney *et al.*, 2014). T6SSs use a molecular syringe to 'inject' effectors directly into the target cell, raising the question of why effectors would require an additional translocation motif? The T6SS is suggested to only penetrate the OM and therefore inject effectors into the periplasm (Basler *et al.*, 2012; Cotter, 2011). The pyocin_S domain was only identified in T6SS effectors encoding cytotoxic domains which act within the cytoplasm: colicin DNase, tRNase, rRNase, Toxin52 and Toxin46. This suggests that the pyocin_S domain could act as a common cytoplasmic targeting motif, able to direct both bacteriocin and T6SS effectors to the cytoplasm.

Searching with the pyocin_S domain did not identify any contact-dependent inhibitors (CDI). While the CDI toxin does cross the inner membrane to target nucleic acids within the cytoplasm, these cytotoxic domains have an additional N-terminal region which is essential for translocation through the inner membrane (Willett *et al.*, 2015).

5.3.4 The pyocin_S domain identify novel nuclease bacteriocins

The ppHMM search strategy used to identify NBs in Chapter 3 suffers from a significant limitation; it can only identify cytotoxic domains where the motif it utilises has previously been identified. Though this approach identified a significant number of NB families and novel TR domains, it failed to identify novel cytotoxic domains. Scanning by the pyocin_S domain identified a domain with similarity to the cytotoxic domain of carocin

D (Roh *et al.*, 2010). This domain is associated with novel TR domains within *Serratia*, and present within the pyocins and klebicins. Expressing a novel pyocin S1NovD protein produced an active bacteriocin with activity against three strains of *P. aeruginosa* (Figure 5.8). Identifying bacteriocins by the pyocin_S domain also has the advantage of reducing computational time as all proteins are identified within a single HMMer search. Using the pyocin_S domain search strategy did not expand the distribution of NBs to species outside of the Gammaproteobacteria. Though more species were identified which encoded NBs, this is due to the increased size of the database.

5.3.5 The role of the DPY domain in translocaton

Previously, NBs were separated into eight different structural organisations. The DPY motif occurs within the translocation domain of E-type colicins which have the T-R-C domain structure. In the NB pyocins however, which are thought to have the R-T-C organisation, the DPY motif occurs proximal to the cytotoxic domain, suggesting that the 'DPY' is also associated with translocation domains. Four of the structural groups contained two copies of the DPY motif, which suggests the distinction between translocation and receptor domain is not as definitive for all NBs. Indeed, colicin D uses a single protein as a receptor and translocator and is described as having a single combined TR domain (Mora *et al.*, 2008).

Mutation of the three residues which interact at the core of the DPY domain in ColE9; R185, D270 and Y285, was performed to determine if these residues were essential for colicin translocation. Only Y285A abolished translocation. The Y285F mutation, which lacks the hydroxyl group necessary to interact with D270, was able to translocate and kill sensitive *E. coli*. The double ColE9 R185A D270A mutant was also able to kill, suggesting the interaction of the three residues is not essential, but the presence of a hydrophobic residue is essential. The reduced yield of the ColE9 Y285A mutant protein and the lack of any protein in the corresponding mutated pyocin S2 Y552A, suggests that the tyrosine has a significant structural role and its mutation prevents correct folding and expression.

Mutation of both DPY tyrosines in ColD was insufficient to block translocation and both

the single and double mutants were able to kill sensitive *E. coli*. ColD is able to interact with LepB at the inner membrane (Mora *et al.*, 2015). This interaction could provide a redundant function to the DPY motif.

Sano *et. al.* previously identified the DPY motif within pyocins and claimed that chimeric pyocins which lacked this motif were unable to kill strains of *P. aeruginosa* (Sano *et al.*, 1993a). However, this approach was not published and it can be presumed that a deletion of the DPY motif, which forms a structured region within the T-domain of colicins, would cause extensive misfolding of the protein.

This chapter has mainly focussed on the possible role in translocating across the inner membrane as this is the major difference between nuclease and pore forming bacteriocins. There is however a second difference which needs to be taken into account. Nuclease bacteriocins are bound to immunity proteins after translation to prevent death of the cell, whereas the pore forming bacteriocin immunity protein is located within the inner membrane. The DPY motif is unlikely to be involved in the rapid and tight binding of the immunity protein. Firstly, the DPY motif is associated with predicted translocation domains, which in E-type colicins are separated from the cytotoxic domain by the ~125aa R-domain. Secondly, nuclease domain/immunity binding occurs when only the nuclease domain is present. If the DPY domain would be required for immunity, leaky expression of DNase/Im constructs would be sufficient to cause death of the producing cell.

5.3.6 Future work

This work has effectively identified a conserved domain across NBs and the evolved effectors of T6SS effectors within Gammaproteobacteria. So far no specific function has been prescribed to this domain.

Investigating the role of a motif which may have dual functions in both translocation and protein folding by mutation will cause confounding factors in data analysis. One possibility is the use of chimeric proteins. If the DPY motif is essential for translocation, swapping the pore forming cytotoxic domain of pyocin S5 with the DNase of S2 is expected produce a protein which cannot translocate. A second chimera which in-

cludes the pyocin_S domain as well as the DNase should then be able to translocate and kill sensitive *P. aeruginosa* strains. A pyocin S5/S2 chimeric protein where the S5 receptor domain is associated with the pyocin S2 translocation and cytotoxic domains has already been shown to be effective (Elfarash *et al.*, 2014). Finding the minimal region required for translocation across the inner membrane will clarify the role of the DPY motif.

To separate the role of the DPY in translocation across the inner and outer membranes, fluorescent microscopy could be applied. Fluorescently labelled colicins, with and without DPY mutants, would be able to distinguish between a colicin bound to the outer membrane receptor or diffusing within the cytoplasm using methods such as trypsin treatment or fluorescent recovery after photobleaching (FRAP).

To test if the DPY is directly responsible for binding a protein in the inner membrane, such as FtsH, the Bpa incorporated constructs will be a valuable tool. This has already been attempted, however, the purification methods, particularly the detergents used for membrane extractions, need to be optimised.

CHAPTER 6

Chapter 6: General Discussion

Bacteriocins are modular protein toxins which have antimicrobial activity again closely related species. Such properties make them ideal candidates as future antimicrobials. However, several obstacles need to be overcome before this family of antimicrobials can be used in the clinic. Bacteriocins target closely related bacteria, therefore without knowing the diversity of the species capable of producing them, the potential range of target bacteria is unknown. Secondly, high rates of resistance have been observed when testing bacteriocins in the lab. Finally, the mechanisms by which these proteins translocate into the target cell are poorly understood. The aims of this study were to use a mixture of bioinformatic, high throughput genomic, and biochemical methods to overcome these obstacles, specially with relation to nuclease bacteriocins. Chapter 3 focuses on the design and implementation of a profile Hidden Markov Model for the identification of NBs, and the subsequent analysis of the sequences identified. Chapter 4 uses a mixture of high throughput genomic and phenomic approaches to determine the resistance mechanism of a strain of uropathogenic *E. coli*. It was shown that resistance was dependent on the presence of O-antigen and that this resistance was heavily dependent on growth conditions. Chapter 5 uses bioinformatics and biochemistry to investigate the possible role of a motif associated with nuclease bacteriocins and other effectors of the T6SS which cross the inner membrane. Attempts to show that this motif is essential for translocation are hindered by its apparent necessity in the correct folding of NBs. Bioinfomatic analysis showed that this motif is conserved across all NBs and is present in T6SS specialised effectors which act within the cytoplasm. Understanding the role of this motif could have implications for our understanding of competition systems and the biochemistry of the bacterial inner membrane.

6.1 Colicin-like NBs are present only within Gammaproteobacteria

Colicins were first identified over 90 years ago, yet the extent to which colicin-like bacteriocins exist has never been elucidated (Gratia, 1925). The pipeline implemented in chapter 3 has significant advantages over previous bacteriocin bioinformatic analysis.

The Hidden Markov Model (HMM) approach identified NBs by the conserved enzymatic motifs found in the cytotoxic domain whilst the addition of an immunity profile adds a powerful tool for false positive removal, allowing a very lenient E-value cut-off and detection of more distant homologues. The pubMLST Multispecies rMLST database itself offers significant advantages. The database size of >100,000 genomes (increasing to >150,000) is a powerful tool for analysing genes which are only present in a species' accessory genome.

Though a great deal of bioinformatic work has been undertaken to analyse nuclease bacteriocins within *Pseudomonadaceae* (Ghequire and De Mot, 2014), to our knowledge this presents the first systematic analysis for the presence of nuclease bacteriocins across the bacterial kingdom. 3094 NB proteins were identified across 53 species. The ppHMM approach used identified NBs for the first time in *Yersinia* and *Serratia*, and novel NBs have been identified in *Pseudomonas aeruginosa* and *Klebsiella pneumoniae*. The 53 species identified using the ppHMM strategy all belong to the class Gammaproteobacteria. Chapter 3 tested two different hypotheses to explain this narrow taxonomic distribution: environmental association or presence of PMF coupled complexes required for import. Neither hypothesis could explain the limitation of NBs to Gammaproteobacteria. The profiles used within the ppHMM strategy, though primarily generated from NBs, managed to identify cytotoxic domains in non-NB proteins from *Burkholderia* spp., *Neiserria* spp. and *Bacillus* spp. (Zhang *et al.*, 2011).

Distributions of other competition systems, the Type Six Secretion System (T6SS) and Contact Dependent Inhibitors (CDI), have also been analysed using bioinformatics. T6SS is a prevalent secretion system present in ~25% of bacteria (Bingle *et al.*, 2008). The T6SS is separated into three distinct subtypes (I,II and III), of which $T6SS^{I}$ is the most common. $T6SS^{I}$ is almost exclusively found within proteobacteria (Abby *et al.*, 2016; Boyer *et al.*, 2009). The two other subtypes ($T6SS^{II}$ and $T6SS^{III}$) show even narrower taxonomic distributions, *Francisella* and Bacteroidetes, respectively (Abby *et al.*, 2016). As for the contact dependent inhibitors, CDI operons are also present throughout the proteobacteria (Aoki *et al.*, 2011). Restriction to proteobacteria appears common for competition systems, however, the restriction to the class Gammaproteobacteria appears unique to

NBs. CLBs which utilize a colicin M-like peptidoglycan degrading or pore forming domain are the only classes of CLBs which have been identified outside of the Gammaproteobacteria (Ghequire and De Mot, 2015b; Ghequire *et al.*, 2017). The reasons behind this apparent expanded phylogeny are still unclear, however, the pyocin_S domain is also almost exclusively found within Gammaproteobacteria. Therefore, whatever function the pyocin_S domain plays, it could be limiting T6SS effectors with the pyocin_S domain and NBs to the Gammaproteobacteria.

6.2 NBs as potential antimicrobials

NBs typically have activity towards closely related species. While broad spectrum antibiotics have been highly sought after in the past, narrow spectrum antibiotics have the potential of targeting a pathogen without producing off-target effects and disturbing the host microflora. *K. pneumoniae*, *E. coli* and *P. aeruginosa* are listed within the ten most prevalent nosocomial infections, and antimicrobial resistant *Enterobacteriaceae* and *Pseudomonadas aeruginosa* are considered serious or urgent threats (CDC, 2013) driving the need for a novel approach for treatment of these species.

As NBs target specific OM proteins of the target cell and the majority of these proteins are still unknown, to understand the possible killing spectrum requires expression of NBs and characterisation of binding partners. Knowing the receptors and their prevalence allows predictions of the activity spectrum for a specific NB or NB cocktail. Receptor identification is typically performed by using large strain collections (Elfarash *et al.*, 2012), spontaneous mutants (Cardelli and Konisky, 1974) or competition with the natural receptor (Wayne *et al.*, 1976). In chapter 3 an association between production of an NB and receptor proteins was observed. This association suggests receptors could be predicted bioinformatically. By calculating the pangenome of a large collection of strains, as performed in Chapter 3, and then finding an association between a particular NB gene and an outer membrane protein gene could produce a list of candidate receptors in a matter of hours. This approach has a number of limitations however. Firstly, a bacteriocin would have to be present in significant number of genomes to make the statistical

analysis meaningful. Secondly, the receptor itself must not be essential. Finally, if the receptor is not encoded by a specific open reading frame e.g. the O-antigen, or if the protein is prone to mutation which abolishes binding, a simple association would not be observed. The increase in popularity of pangenome analysis in bacteria is driving the development of more sensitive statistical analyses for bacterial genome wide association studies (Earle *et al.*, 2016). Transposon based methods, such as TraDIS, could also be a valuable tool for identifying the receptor protein for an unknown NB. Transposon mutagensis has already been applied to identify the receptor for a novel chimeric bacteriocin within *Pseudomonas* (Ghequire *et al.*, 2017).

6.3 Non-specific resistance mechanisms

When challenging libraries of *E. coli* strains with multiple colicins, a high proportion of strains were resistant to a single colicin (>70%) or multiple colicins (~30%) (Feldgarden and Riley, 1998). Such a high rate of naturally occurring resistance would hinder the use of colicins as antimicrobials. The O-antigen has been shown to cause non-specific resistance towards colicins (Rendueles *et al.*, 2014; Tran *et al.*, 2014; van der Ley *et al.*, 1986). In Chapter 4, the O-antigen is shown to provide *E. coli* strains with resistance to multiple colicins in a density dependent manner. The development of reistance to colicin E9 due to O-antigen restoration was not only shown for *E. coli* K-12 but also for a clinically relevant uropathogenic *E. coli* UPEC ST131. O-antigen has also been implicated in the resistance to polymyxin B (Allen *et al.*, 1998; Loutet *et al.*, 2006). How prevalent this resistance is across Gammaproteobacteria is still unknown. As well as the O-antigen and LPS, Gram-negative bacteria can produce a thick layer of polysaccharide called the capsule (10-160nm (Amako *et al.*, 1988)). Like the O-antigen, the capsule has been shown to increase resistance towards polymyxin B. The capsule is particularly pronounced in *Klebsiella pneumoniae* where it is also involved in virulence (Struve and Krogfelt, 2003). The capsule is likely to provide a significant barrier to CLBs. Chapter 4 shows the changes to O-antigen in response to environmental stresses can effect resistance towards colicins. The changes in concentration of ions, osmolarity, pH and the influence of the immune system within a host will affect density and length of the O-antigen (Lerouge

and Vanderleyden, 2002). Screening libraries under lab conditions is likely to drastically over-estimate resistance within a population. *In vivo* models of infection are needed to examine if O-antigen dependent resistance is relevant within the host. The effect of different infections sites or chronic versus acute infection are of particular interest as the O-antigen in known to be lost during some chronic infections (Maldonado *et al.*, 2016).

6.4 Roles of NBs within the environment

The prevalence of NBs suggests they play a role throughout *Enterobacteriaceae* and *Pseudomonadaecae*. The proportions of NB producing strains across species varies, between <5% to >80% for *E. coli* and *P. aeruginosa*, respectively. The differences in prevalence within a strain have remarkable differences on the evolution of sequences, causing a switch from diversifying selection to diversifying recombination (Riley, 1993a). The differences in prevalence could also drive selection of strains with orphan immunity proteins, as more than one NB is likely to exist within the environment.

The association observed between colicins and virulence factors suggests colicins could act to displace native bacteria and help become fixed in the environment (Holt *et al.*, 2013). The presence of virulence factors is not a definitive measurement of a strains' pathogenicity. Future work is needed to test colicin deletions for defects in virulence to prove this hypothesis. The O-antigen dependent resistance described in Chapter 4 suggests colicins could have another role in maintaining O-antigen within a population as bacteria which lose the ability to produce O-antigen will become sensitive to colicins.

6.5 Translocation of Nuclease Bacteriocins

Chapter 3 describes a novel grouping system for NBs based on structural elements. Elements such as coiled-coil and disordered domains have significant impact on the predicted translocation mechanism. For example, all known NBs within group I require a receptor, separate translocator protein and the Tol complex. Group V NBs require a single translocation/receptor protein and utilise the Ton system (Mora *et al.*, 2005).

Chapter 5 deals with a domain associated with bacteriocins and T6SS effectors which cross the inner membrane but is absent in pore forming and peptidoglycan degrading bacteriocins. Though this would suggest a role in translocation across the inner membrane, there are still many questions left to answer. If the pyocin_S domain (which contains the DPY motif) indicates the translocation domain as predicted (Michel-Briand and Baysse, 2002; Sano *et al.*, 1993a), then why do four of the eight structural organisations contain two copies? Secondly, in group I bacteriocins (such as ColE9), the pyocin_S domain is located 135 residues N-terminal to the cytotoxic domain. If the domain is involved with helping the cytotoxic domain cross the inner membrane, how does this interaction occur? Understanding the role that the pyocin_S domain plays in translocation has ramifications beyond NBs. If the domain is involved with crossing the inner membrane, it also implies that the T6SS injects effectors into the periplasm, which is a contentious unanswered question for T6SS. Availability of a universal translocation domain across the inner membrane for peptides could also be a valuable biotechnological tool.

Investigating the role of a motif which may have dual functions in translocation and protein folding is challenging. Future work may involve creating chimeric proteins of pore forming translocation domain and a nuclease cytotoxic domain with or without the presence of the DPY motif. To separate defects in translocation across the OM and IM, florescently labelled DPY mutants and spheroplasts will distinguish between a periplasmic and cytotoxic location. Finally, if the DPY motif directly interacts with a binding partner, the already generated constructs for photo-inducible crosslinking could be used to trap the interacting partners and mass spectrometry used to identify the binding partner.

In conclusion, the work presented in this study has used a mixture of bioinformatic, genomic and biochemical approaches to investigate several of the obstacles which could limit the use of NBs as future antimicrobials. Though questions still remain, particularly involving the role of the DPY motif, the results obtained can be used to direct future *in vivo* testing of NBs as novel species specific antimicrobials.

Bibliography

Centers for Disease Control and Prevention, Office of Infectious Disease Antibiotic resistance threats in the United States. Journal article, Centers for Disease Control and Prevention, Office of Infectious Disease, 2013.

ABBY, S. S., CURY, J., GUGLIELMINI, J., NÉRON, B., TOUCHON, M., AND ROCHA, E. P. C. Identification of protein secretion systems in bacterial genomes. *Scientific Reports*, **6**:23080, 2016. doi: 10.1038/srep23080.

ALAKOMI, H.-L., SKYTTÄ, E., SAARELA, M., MATTILA-SANDHOLM, T., LATVA-KALA, K., AND HELANDER, I. M. Lactic Acid Permeabilizes Gram-Negative Bacteria by Disrupting the Outer Membrane. *Applied and Environmental Microbiology*, **66**(5):2001–2005, 2000.

ALCOFORADO DINIZ, J. AND COULTHURST, S. J. Intraspecies Competition in *Serratia marcescen* Is Mediated by Type VI-Secreted Rhs Effectors and a Conserved Effector-Associated Accessory Protein. *Journal of Bacteriology*, **197**(14):2350–2360, 2015. doi: 10.1128/JB.00199-15.

ALCOFORADO DINIZ, J., LIU, Y., AND COULTHURST, S. J. Molecular weaponry: diverse effectors delivered by the Type VI secretion system. *Cellular Microbiology*, **17**(12):1742–1751, 2015. doi: 10.1111/cmi.12532.

ALLEN, C. A., ADAMS, L. G., AND FICHT, T. A. Transposon-derived *Brucella abortus* rough mutants are attenuated and exhibit reduced intracellular survival. *Infection and immunity*, **66**(3):1008–1016, 1998.

AMAKO, K., MENO, Y., AND TAKADE, A. Fine structures of the capsules of *Klebsiella pneumoniae* and *Escherichia coli* K1. *Journal of Bacteriology*, **170**(10):4960–4962, 1988.

ANDERSON, M. S., GARCIA, E. C., AND COTTER, P. A. The Burkholderia *bcpAIOB* genes define unique classes of two-partner secretion and contact dependent growth inhibition systems. *PLoS Genet*, **8**(8):e1002877, 2012. doi: 10.1371/journal.pgen.1002877.

AOKI, S. K., DINER, E. J., DE ROODENBEKE, C. T., BURGESS, B. R., POOLE, S. J., BRAATEN, B. A., JONES, A. M., WEBB, J. S., HAYES, C. S., COTTER, P. A., AND LOW, D. A. A widespread family of polymorphic contact-dependent toxin delivery systems in bacteria. *Nature*, **468**(7322):439–442, 2010. doi: 10.1038/nature09490.

AOKI, S. K., MALINVERNI, J. C., JACOBY, K., THOMAS, B., PAMMA, R., TRINH, B. N., REMERS, S., WEBB, J., BRAATEN, B. A., SILHAVY, T. J., AND LOW, D. A. Contact-dependent growth inhibition requires the essential outer membrane protein BamA (YaeT) as the receptor and the inner membrane transport protein AcrB. *Mol Microbiol*, **70**(2):323–340, 2008. doi: 10.1111/j.1365-2958.2008.06404.x.

AOKI, S. K., PAMMA, R., HERNDAY, A. D., BICKHAM, J. E., BRAATEN, B. A., AND LOW, D. A. Contact-dependent inhibition of growth in *Escherichia coli*. *Science*, **309**(5738):1245–1248, 2005. doi: 10.1126/science.1115109.

AOKI, S. K., POOLE, S. J., HAYES, C. S., AND LOW, D. A. Toxin on a stick: Modular CDI toxin delivery systems play roles in bacterial competition. *Virulence*, **2**(4):356–359, 2011. doi: 10.4161/viru.2.4.16463.

ARNISON, P. G., BIBB, M. J., BIERBAUM, G., BOWERS, A. A., BUGNI, T. S., BULAJ, G., CAMARERO, J. A., CAMPOPIANO, D. J., CHALLIS, G. L., CLARDY, J., COTTER, P. D., CRAIK, D. J., DAWSON, M., DITTMANN, E., DONADIO, S., DORRESTEIN, P. C., ENTIAN, K.-D., FISCHBACH, M. A., GARAVELLI, J. S., GORANSSON, U., GRUBER, C. W., HAFT, D. H., HEMSCHEIDT, T. K., HERTWECK, C., HILL, C., HORSWILL, A. R., JASPARS, M., KELLY, W. L., KLINMAN, J. P., KUIPERS, O. P., LINK, A. J., LIU, W., MARAHIEL, M. A., MITCHELL, D. A., MOLL, G. N., MOORE, B. S., MULLER, R., NAIR, S. K., NES, I. F., NORRIS, G. E., OLIVERA, B. M., ONAKA, H., PATCHETT, M. L., PIEL, J., REANEY, M. J. T., REBUFFAT, S., ROSS, R. P., SAHL, H.-G., SCHMIDT, E. W., SELSTED, M. E., SEVERINOV, K., SHEN, B., SIVONEN, K., SMITH, L., STEIN, T., SUSSMUTH, R. D., TAGG, J. R., TANG, G.-L., TRUMAN, A. W., VEDERAS, J. C., WALSH, C. T., WALTON, J. D., WENZEL, S. C., WILLEY, J. M., AND VAN DER DONK, W. A. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.*, **30**:108–160, 2013. doi: 10.1039/C2NP20085F. URL `http://dx.doi.org/10.1039/C2NP20085F`.

ARNOLD, T., ZETH, K., AND LINKE, D. Structure and function of colicin S4, a colicin with a duplicated receptor-binding domain. *J Biol Chem*, **284**(10):6403–6413, 2009.

BAHAROGLU, Z. AND MAZEL, D. SOS, the formidable strategy of bacteria against aggressions. *FEMS microbiology reviews*, pages 1–20, 2014. doi: 10.1111/1574-6976.12077.

BAILEY, T. L., BODEN, M., BUSKE, F. A., FRITH, M., GRANT, C. E., CLEMENTI, L., REN, J., LI, W. W., AND NOBLE, W. S. MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Research*, **37**(SUPPL. 2):202–208, 2009. doi: 10.1093/nar/gkp335.

BALLISTER, E. R., LAI, A. H., ZUCKERMANN, R. N., CHENG, Y., AND MOUGOUS, J. D. In vitro self-assembly of tailorable nanotubes from a simple protein building block. *Proc Natl Acad Sci USA*, **105**(10):3733–3738, 2008. doi: 10.1073/pnas.0712247105.

BANERJEE, R., JOHNSTON, B., LOHSE, C., PORTER, S. B., CLABOTS, C., AND JOHNSON, J. R. *Escherichia coli* Sequence Type 131 Is a Dominant, Antimicrobial-Resistant Clonal Group Associated with Healthcare and Elderly Hosts. *Infection control and hospital epidemiology*, **34**(4):361–369, 2013. doi: 10.1086/669865.

BAQUERO, F., BOUANCHAUD, D., MARTINEZ-PEREZ, M. C., AND FERNANDEZ, C. Microcin plasmids: a group of extrachromosomal elements coding for low-molecular-weight antibiotics in *Escherichia coli*. *Journal of Bacteriology*, **135**(2):342–347, 1978.

BARQUIST, L., MAYHO, M., CUMMINS, C., CAIN, A. K., BOINETT, C. J., PAGE, A. J., LANGRIDGE, G. C., QUAIL, M. A., KEANE, J. A., AND PARKHILL, J. The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries. *Bioinformatics*, **32**(7):1109–1111, 2016. doi: 10.1093/bioinformatics/btw022.

BARRET, M., EGAN, F., FARGIER, E., MORRISSEY, J. P., AND O'GARA, F. Genomic analysis of the type VI secretion systems in *Pseudomonas* spp.: novel clusters and putative effectors uncovered. *Microbiology*, **157**(Pt 6):1726–1739, 2011. doi: 10.1099/mic.0.048645-0.

BARRETEAU, H., TIOUAJNI, M., GRAILLE, M., JOSSEAUME, N., BOUHSS, A., PATIN, D., BLANOT, D., FOURGEAUD, M., MAINARDI, J. L., ARTHUR, M., VAN TILBEURGH, H., MENGIN-LECREULX, D., AND TOUZE, T. Functional and structural characterization of PaeM, a colicin M-like bacteriocin produced by *Pseudomonas aeruginosa*. *J Biol Chem*, **287**(44):37395–37405, 2012. doi: 10.1074/jbc.M112.406439.

BASLER, M. AND MEKALANOS, J. J. Type 6 secretion dynamics within and between bacterial cells. *Science (New York, N.Y.)*, **337**(6096):815, 2012. doi: 10.1126/science.1222901.

BASLER, M., PILHOFER, M., HENDERSON, G. P., JENSEN, G. J., AND MEKALANOS, J. J. Type VI secretion requires a dynamic contractile phage tail-like structure. *Nature*, **483**(7388):182–186, 2012. doi: 10.1038/nature10846.

BAYRAMOGLU, B., TOUBIANA, D., VAN VLIET, S., INGLIS, R. F., SHNERB, N., AND GILLOR, O. Bet-hedging in bacteriocin producing *Escherichia coli* populations: the single cell perspective. *Scientific Reports*, **7**:42068, 2017.

BAYSSE, C., MEYER, J. M., PLESIAT, P., GEOFFROY, V., MICHEL-BRIAND, Y., AND CORNELIS, P. Uptake of pyocin S3 occurs through the outer membrane ferripyoverdine type II receptor of *Pseudomonas aeruginosa*. *Journal of Bacteriology*, **181**(12):3849–3851, 1999.

BECK, C. M., WILLETT, J. L., AND CUNNINGHAM, D. A. CdiA Effectors from Uropathogenic *Escherichia coli* Use Heterotrimeric Osmoporins as Receptors to Recognize Target Bacteria. *PLoS pathogens*, **12**(10):e1005925, 2016. doi: 10.1371/journal.ppat.1005925.

BEHRENS, H. M., SIX, A., WALKER, D., AND KLEANTHOUS, C. The therapeutic potential of bacteriocins as protein antibiotics. *Emerging Topics in Life Sciences*, **1**(1):65–74, 2017.

BINGLE, L. E. H., BAILEY, C. M., AND PALLEN, M. J. Type VI secretion: a beginner's guide. *Current Opinion in Microbiology*, **11**(1):3–8, 2008. doi: http://dx.doi.org/10.1016/j.mib.2008.01.006.

BLONDEL, C. J., JIMÉNEZ, J. C., CONTRERAS, I., AND SANTIVIAGO, C. A. Comparative genomic analysis uncovers 3 novel loci encoding type six secretion systems differentially distributed in *Salmonella* serotypes. *BMC Genomics*, **10**(1):354, 2009. doi: 10.1186/1471-2164-10-354.

BOCHNER, B. R., GADZINSKI, P., AND PANOMITROS, E. Phenotype MicroArrays for High-Throughput Phenotypic Testing and Assay of Gene Function. *Genome Research*, **11**(7):1246–1255, 2001. doi: 10.1101/gr.186501.

BOSÁK, J., LAIBLOVÁ, P., ŠMARDA, J., DĚDIČOVÁ, D., AND ŠMAJS, D. Novel colicin FY of *Yersinia frederiksenii* inhibits pathogenic *Yersinia* strains via YiuR-mediated reception, TonB import, and cell membrane pore formation. *Journal of Bacteriology*, **194**(8):1950–1959, 2012. doi: 10.1128/JB.05885-11.

BOWMAN, C. M., DAHLBERG, J. E., IKEMURA, T., KONISKY, J., AND NOMURA, M. Specific inactivation of 16S ribosomal RNA induced by colicin E3 in vivo. *Proc. Natl. Acad. Sci. USA*, **68**(5):964, 1971.

BOYER, F., FICHANT, G., BERTHOD, J., VANDENBROUCK, Y., AND ATTREE, I. Dissecting the bacterial type VI secretion system by a genome wide in silico analysis: what can be learned from available microbial genomic resources? *BMC Genomics*, **10**:104, 2009. doi: 10.1186/1471-2164-10-104.

BROOKS, T. M., UNTERWEGER, D., BACHMANN, V., KOSTIUK, B., AND PUKATZKI, S. Lytic Activity of the *Vibrio cholerae* Type VI Secretion Toxin VgrG-3 Is Inhibited by the Antitoxin TsaB. *The Journal of Biological Chemistry*, **288**(11):7618–7625, 2013. doi: 10.1074/jbc.M112.436725.

BROWNING, D. F., WELLS, T. J., FRANCA, F. L., MORRIS, F. C., SEVASTSYANOVICH, Y. R., BRYANT, J. A., JOHNSON, M. D., LUND, P. A., CUNNINGHAM, A. F., HOBMAN, J. L., MAY, R. C., WEBBER, M. A., HENDERSON, I. R., FRANÇA, F. L. S., MORRIS, F. C., SEVASTSYANOVICH, Y. R., BRYANT, J. A., JOHNSON, M. D., LUND, P. A., CUNNINGHAM, A. F., HOBMAN, J. L., MAY, R. C., WEBBER, M. A., AND HENDERSON, I. R. Laboratory adapted *Escherichia coli* K-12 becomes a pathogen of *Caenorhabditis elegans* upon restoration of O antigen biosynthesis. *Molecular Microbiology*, **87**(5):939–50, 2013. doi: 10.1111/mmi.12144.

BRUNET, Y. R., ZOUED, A., BOYER, F., DOUZI, B., AND CASCALES, E. The Type VI Secretion TssEFGK-VgrG Phage-Like Baseplate Is Recruited to the TssJLM Membrane Complex via Multiple Contacts and Serves As Assembly Platform for Tail Tube/Sheath Polymerization. *PLoS Genetics*, **11**(10):e1005545, 2015. doi: 10.1371/journal.pgen.1005545.

CARATTOLI, A., ZANKARI, E., GARCÍA-FERNÁNDEZ, A., VOLDBY LARSEN, M., LUND, O., VILLA, L., MØLLER AARESTRUP, F., AND HASMAN, H. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrobial agents and chemotherapy*, **58**(7):3895–3903, 2014. doi: 10.1128/AAC.02412-14.

CARDELLI, J. AND KONISKY, J. Isolation and characterization of an *Escherichia coli* mutant tolerant to colicins Ia and Ib. *Journal of bacteriology*, **119**(2):379–385, 1974.

CASCALES, E., BUCHANAN, S. K., DUCHÉ, D., KLEANTHOUS, C., LLOUBÈS, R., POSTLE, K., RILEY, M., SLATIN, S., CAVARD, D., DUCHE, D., KLEANTHOUS, C., LLOUBES, R., POSTLE, K., RILEY, M., SLATIN, S., AND CAVARD, D. Colicin biology. *Microbiology and molecular biology reviews : MMBR*, **71**(1):158–229, 2007. doi: 10.1128/MMBR.00036-06.

CAVARD, D. Colicin A and colicin E1 lysis proteins differ in their dependence on secA and secY gene products. *FEBS Lett*, **298**(1):84–88, 1992. doi: 0014-5793(92)80027-E[pii].

CAVARD, D. AND LAZDUNSKI, C. Colicin cleavage by OmpT protease during both entry into and release from *Escherichia coli* cells. *Journal of Bacteriology*, **172**(2):648–652, 1990.

CHAK, K. F. AND JAMES, R. Analysis of the promoters for the two immunity genes present in the ColE3-CA38 plasmid using two new promoter probe vectors. *Nucleic Acids Research*, **13**(7):2519–2531, 1985.

CHANG, Y. W., RETTBERG, L. A., ORTEGA, D. R., AND JENSEN, G. J. In vivo structures of an intact type VI secretion system revealed by electron cryotomography. *EMBO Rep*, 2017. doi: 10.15252/embr.201744072.

CHAO, L. AND LEVIN, B. R. Structured habitats and the evolution of anticompetitor toxins in bacteria. *Proc Natl Acad Sci USA*, **78**(10):6324–6328, 1981.

CHATTERJEE, C., PAUL, M., XIE, L., AND VAN DER DONK, W. A. Biosynthesis and mode of action of lantibiotics. *Chem Rev*, **105**(2):633–684, 2005. doi: 10.1021/cr030105v.

CHAULEAU, M., MORA, L., SERBA, J., AND DE ZAMAROCZY, M. FtsH-dependent processing of RNase colicins D and E3 means that only the cytotoxic domains are imported into the cytoplasm. *J Biol Chem*, **286**(33):29397–29407, 2011. doi: 10.1074/jbc.M111.242354.

CHAVAN, M., RAFI, H., WERTZ, J., GOLDSTONE, C., AND RILEY, M. A. Phage Associated Bacteriocins Reveal a Novel Mechanism for Bacteriocin Diversification in *Klebsiella*. *Journal of Molecular Evolution*, **60**(4):546–556, 2005. doi: 10.1007/s00239-004-0263-9.

CHEN, Y., SHAPIRA, R., EISENSTEIN, M., AND MONTVILLE, T. J. Functional characterization of pediocin PA-1 binding to liposomes in the absence of a protein receptor and its relationship to a predicted tertiary structure. *Applied and Environmental Microbiology*, **63**(2):524–531, 1997.

CHIN, J. W., MARTIN, A. B., KING, D. S., WANG, L., AND SCHULTZ, P. G. Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proc Natl Acad Sci USA*, **99**(17):11020–11024, 2002a. doi: 10.1073/pnas.172226299.

CHIN, J. W., SANTORO, S. W., MARTIN, A. B., KING, D. S., WANG, L., AND SCHULTZ, P. G. Addition of p-azido-L-phenylalanine to the genetic code of *Escherichia coli*. *J Am Chem Soc*, **124**(31):9026–9027, 2002b.

CHOW, J. AND MAZMANIAN, S. K. A Pathobiont of the Microbiota Balances Host Colonization and Intestinal Inflammation. *Cell host & microbe*, **7**(4):265–276, 2010. doi: 10.1016/j.chom.2010.03.004.

CHRISTENSON, J. K. AND GORDON, D. M. Evolution of colicin BM plasmids: the loss of the colicin B activity gene. *Microbiology*, **155**(Pt 5):1645–1655, 2009. doi: 10.1099/mic.0.026666-0.

COLE, S. T., SAINT-JOANIS, B., AND PUGSLEY, A. P. Molecular characterisation of the colicin E2 operon and identification of its products. *Mol Gen Genet*, **198**(3):465–472, 1985.

CORANDER, J., WALDMANN, P., MARTTINEN, P., AND SILLANPÄÄ, M. J. BAPS 2: enhanced possibilities for the analysis of genetic population structure. *Bioinformatics (Oxford, England)*, **20**(15):2363–2369, 2004. doi: 10.1093/bioinformatics/bth250.

COTTER, P. Microbiology: Molecular syringes scratch the surface. *Nature*, **475**(7356):301–303, 2011.

COTTER, P. D., HILL, C., AND ROSS, R. P. Bacteriocins: developing innate immunity for food. *Nat Rev Microbiol*, **3**(10):777–788, 2005. doi: nrmicro1273[pii]10.1038/nrmicro1273.

COTTER, P. D., ROSS, R. P., AND HILL, C. Bacteriocins - a viable alternative to antibiotics? *Nat Rev Microbiol*, **11**(2):95–105, 2013. doi: 10.1038/nrmicro2937.

COUGHLIN, R. T., HAUG, A., AND MCGROARTY, E. J. Physical properties of defined lipopolysaccharide salts. *Biochemistry*, **22**(8):2007–2013, 1983. doi: 10.1021/bi00277a042.

DALEBROUX, Z. AND MILLER, S. *Salmonella* PhoPQ regulation of the outer membrane to resist innate immunity. *Current Opinion in Microbiology*, **17**(1):106–113, 2014. ISSN 1369-5274. doi: 10.1016/j.mib.2013.12.005.

DANKA, E. S., GARCIA, E. C., AND COTTER, P. A. Are CDI Systems Multicolored, Facultative, Helping Greenbeards? *Trends Microbiol*, **25**(5):391–401, 2017. doi: 10.1016/j.tim.2017.02.008.

DAVIES, J. K. AND REEVES, P. Colicin tolerance and map location of conjugation-deficient mutants. *Journal of Bacteriology*, **123**(1):372–373, 1975a.

DAVIES, J. K. AND REEVES, P. Genetics of resistance to colicins in *Escherichia coli* K-12: cross-resistance among colicins of group B. *Journal of Bacteriology*, **123**(1):96–101, 1975b.

DAWKINS, R. The Selfish Gene. Oxford Univ. Press, Oxford, 1976.

DE CHIAL, M., GHYSELS, B., BEATSON, S. A., GEOFFROY, V., MEYER, J. M., PATTERY, T., BAYSSE, C., CHABLAIN, P., PARSONS, Y. N., WINSTANLEY, C., CORDWELL, S. J., AND CORNELIS, P. Identification of type II and type III pyoverdine receptors from *Pseudomonas aeruginosa*. *Microbiology*, **149**(Pt 4):821–831, 2003. doi: 10.1099/mic.0.26136-0.

DE ZAMAROCZY, M. AND BUCKINGHAM, R. H. Importation of nuclease colicins into *E. coli cells*: endoproteolytic cleavage and its prevention by the immunity protein. *Biochimie*, **84**(5-6):423–432, 2002.

DE ZAMAROCZY, M., MORA, L., LECUYER, A., GELI, V., AND BUCKINGHAM, R. H. Cleavage of colicin D is necessary for cell killing and requires the inner membrane peptidase LepB. *Mol Cell*, **8**(1):159–168, 2001.

DEACON, V., DZIVA, F., VAN DIEMEN, P. M., FRANKEL, G., AND STEVENS, M. P. Efa-1/LifA mediates intestinal colonization of calves by enterohaemorrhagic *Escherichia coli* O26 in a manner independent of glycosyltransferase and cysteine protease motifs or effects on type III secretion. *Microbiology*, **156**(8):2527–2536, 2010.

DEBROY, C., FRATAMICO, P. M., YAN, X., BARANZONI, G., LIU, Y., NEEDLEMAN, D. S., TEBBS, R., O'CONNELL, C. D., ALLRED, A., SWIMLEY, M., MWANGI, M., KAPUR, V., RAYGOZA GARAY, J. A., ROBERTS, E. L., AND KATANI, R. Comparison of O-Antigen Gene Clusters of All O-Serogroups of *Escherichia coli and Proposal* for Adopting a New Nomenclature for O-Typing. *PLoS ONE*, **11**(1):e0147434, 2016. doi: 10.1371/journal.pone.0147434.

DELVES-BROUGHTON, J., BLACKBURN, P., EVANS, R. J., AND HUGENHOLTZ, J. Applications of the bacteriocin, nisin. *Antonie Van Leeuwenhoek*, **69**(2):193–202, 1996.

DENAYER, S., MATTHIJS, S., AND CORNELIS, P. Pyocin S2 (Sa) kills *Pseudomonas aeruginosa* strains via the FpvA type I ferripyoverdine receptor. *Journal of Bacteriology*, **189**(21):7663–7668, 2007. doi: 10.1128/JB.00992-07.

DESSI, A., PUDDU, M., TESTA, M., MARCIALIS, M. A., PINTUS, M. C., AND FANOS, V. *Serratia marcescens* infections and outbreaks in neonatal intensive care units. *Journal of chemotherapy (Florence, Italy)*, **21**(5):493–499, 2009. doi: 10.1179/joc.2009.21.5.493.

DEVANATHAN, S. AND POSTLE, K. Studies on colicin B translocation: FepA is gated by TonB. *Mol Microbiol*, **65**(2):441–453, 2007. doi: 10.1111/j.1365-2958.2007.05808.x.

DIGGLE, S. P., GRIFFIN, A. S., CAMPBELL, G. S., AND WEST, S. A. Cooperation and conflict in quorum-sensing bacterial populations. *Nature*, **450**(7168):411–414, 2007. doi: 10.1038/nature06279.

DIMOV, S. G., IVANOVA, P. M., HARIZANOVA, N. T., AND IVANOVA, I. V. Bioactive Peptides used by Bacteria in the Concur-Rence for the Ecological Niche: General Classification and Mode of Action (Overview). *Biotechnology & Biotechnological Equipment*, **19**(2):3–22, 2005. doi: 10.1080/13102818.2005.10817185.

DINER, E. J., BECK, C. M., WEBB, J. S., LOW, D. A., AND HAYES, C. S. Identification of a target cell permissive factor required for contact-dependent growth inhibition (CDI). *Genes Dev*, **26**(5):515–525, 2012. doi: 10.1101/gad.182345.111.

DORMAN, G. AND PRESTWICH, G. D. Benzophenone photophores in biochemistry. *Biochemistry*, **33**(19):5661–5673, 1994.

DOSZTANYI, Z., CSIZMOK, V., TOMPA, P., AND SIMON, I. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *Journal of Molecular Biology*, **347**(4):827–839, 2005. doi: 10.1016/j.jmb.2005.01.071.

DUPORT, C., BAYSSE, C., MICHELBRIAND, Y., AND MICHEL-BRIAND, Y. Molecular characterization of pyocin S3, a novel S-type from *Pseudomonas aeruginosa*. *Journal of Biological Chemistry*, **270**(15):8920–8927, 1995.

DUQUESNE, S., DESTOUMIEUX-GARZON, D., PEDUZZI, J., AND REBUFFAT, S. Microcins, gene-encoded antibacterial peptides from enterobacteria. *Nat Prod Rep*, **24**(4):708–734, 2007. doi: 10.1039/b516237h.

DURAND, E., NGUYEN, V. S., ZOUED, A., LOGGER, L., PEHAU-ARNAUDET, G., ASCHTGEN, M. S., SPINELLI, S., DESMYTER, A., BARDIAUX, B., DUJEANCOURT, A., ROUSSEL, A., CAMBILLAU, C., CASCALES, E., AND FRONZES, R. Biogenesis and structure of a type VI secretion membrane core complex. *Nature*, **523**(7562):555–560, 2015. doi: 10.1038/nature14667.

EARLE, S. G., WU, C.-H., CHARLESWORTH, J., STOESSER, N., GORDON, N. C., WALKER, T. M., SPENCER, C. C. A., IQBAL, Z., CLIFTON, D. A., HOPKINS, K. L., WOODFORD, N., SMITH, E. G., ISMAIL, N., LLEWELYN, M. J., PETO, T. E., CROOK, D. W., MCVEAN, G., WALKER, A. S., AND WILSON, D. J. Identifying lineage effects when controlling for population structure improves power in bacterial association studies. *Nature microbiology*, **1**:16041, 2016. doi: 10.1038/nmicrobiol.2016.41.

EDDY, S. R. Accelerated Profile HMM Searches. *PLoS computational biology*, **7**(10):e1002195, 2011. doi: 10.1371/journal.pcbi.1002195.

EIJSINK, V. G. H., SKEIE, M., MIDDELHOVEN, P. H., BRURBERG, M. B., AND NES, I. F. Comparative Studies of Class IIa Bacteriocins of Lactic Acid Bacteria. *Applied and Environmental Microbiology*, **64**(9):3275–3281, 1998.

ELFARASH, A., DINGEMANS, J., YE, L., HASSAN, A. A., CRAGGS, M., REIMMANN, C., THOMAS, M. S., AND CORNELIS, P. Pore-forming pyocin S5 utilizes the FptA ferripyochelin receptor to kill *Pseudomonas aeruginosa*. *Microbiology-Sgm*, **160**(Pt 2):261–269, 2014. doi: 10.1099/mic.0.070672-0.

ELFARASH, A., WEI, Q., AND CORNELIS, P. The soluble pyocins S2 and S4 from *Pseudomonas aeruginosa* bind to the same FpvAI receptor. *Microbiologyopen*, **1**(3):268–275, 2012. doi: 10.1002/mbo3.27.

ELLIOTT, S. J., SRINIVAS, S., ALBERT, M. J., ALAM, K., ROBINS-BROWNE, R. M., GUNZBURG, S. T., MEE, B. J., AND CHANG, B. J. Characterization of the Roles of Hemolysin and Other Toxins in Enteropathy Caused by Alpha-Hemolytic *Escherichia coli* Linked to Human Diarrhea. *Infection and Immunity*, **66**(5):2040–2051, 1998.

FELDGARDEN, M. AND RILEY, M. A. High levels of colicin resistance in *Escherichia coli*. *Evolution*, **52**(5):1270–1276, 1998. doi: 10.1111/j.1558-5646.1998.tb02008.x.

FINN, R. D., MISTRY, J., TATE, J., COGGILL, P., HEGER, A., POLLINGTON, J. E., GAVIN, O. L., GUNASEKARAN, P., CERIC, G., FORSLUND, K., HOLM, L., SONNHAMMER, E. L. L., EDDY, S. R., AND BATEMAN, A. The Pfam protein families database. *Nucleic Acids Research*, **38**(Database issue):D211–D222, 2010. doi: 10.1093/nar/gkp985.

FLORES-MIRELES, A. L., WALKER, J. N., CAPARON, M., AND HULTGREN, S. J. Urinary tract infections: epidemiology, mechanisms of infection and treatment options. *Nature reviews. Microbiology*, **13**(5):269–284, 2015. doi: 10.1038/nrmicro3432.

FRANCINO, M. P. Antibiotics and the Human Gut Microbiome: Dysbioses and Accumulation of Resistances. *Frontiers in Microbiology*, **6**:1543, 2015. doi: 10.3389/fmicb.2015.01543.

FRICKE, W. F. AND RASKO, D. A. Bacterial genome sequencing in the clinic: bioinformatic challenges and solutions. *Nat Rev Genet*, **15**(1):49–55, 2014.

FRICKEY, T. AND LUPAS, A. CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics*, **20**(18):3702–3704, 2004. doi: 10.1093/bioinformatics/bth444.

GALARDINI, M., MENGONI, A., BIONDI, E. G., SEMERARO, R., FLORIO, A., BAZZICALUPO, M., BENEDETTI, A., AND MOCALI, S. DuctApe: A suite for the analysis and correlation of genomic and OmniLog Phenotype Microarray data. *Genomics*, **103**(1):1–10, 2014. doi: https://doi.org/10.1016/j.ygeno.2013.11.005.

GARCIA, E. C., ANDERSON, M. S., HAGAR, J. A., AND COTTER, P. A. Burkholderia BcpA mediates biofilm formation independently of interbacterial contact dependent growth inhibition. *Molecular microbiology*, **89**(6):1213–1225, 2013. doi: 10.1111/mmi.12339.

GARCIA-BUSTOS, J. F., PEZZI, N., AND MENDEZ, E. Structure and mode of action of microcin 7, an antibacterial peptide produced by *Escherichia coli*. *Antimicrobial Agents and Chemotherapy*, **27**(5):791–797, 1985.

GARNEAU, S., MARTIN, N. I., AND VEDERAS, J. C. Two-peptide bacteriocins produced by lactic acid bacteria. *Biochimie*, **84**(5-6):577–592, 2002.

GE, P., SCHOLL, D., AND LEIMAN, P. G. Atomic structures of a bactericidal contractile nanotube in its pre- and postcontraction states. **22**(5):377–382, 2015. doi: 10.1038/nsmb.2995.

GENILLOUD, O., MORENO, F., AND KOLTER, R. DNA sequence, products, and transcriptional pattern of the genes involved in production of the DNA replication inhibitor microcin B17. *Journal of Bacteriology*, **171**(2):1126–1135, 1989.

GERARD, F., PRADEL, N., WU, L.-F. F., GÉRARD, F., PRADEL, N., AND WU, L.-F. F. Bactericidal activity of colicin V is mediated by an inner membrane protein, SdaC, of *Escherichia coli*. *Journal of Bacteriology*, **187**(6):1945–1950, 2005. doi: 187/6/1945[pii]10.1128/JB.187.6.1945-1950.2005.

GHEQUIRE, M., LORIS, R., AND DE MOT, R. MMBL proteins: from lectin to bacteriocin. *Biochemical Society Transactions*, **40**(6):1553–1559, 2012. doi: 10.1042/BST20120170.

GHEQUIRE, M. G. AND DE MOT, R. The Tailocin Tale: Peeling off Phage Tails. *Trends Microbiol*, **23**(10):587–590, 2015a. doi: 10.1016/j.tim.2015.07.011.

GHEQUIRE, M. G. K. AND DE MOT, R. Ribosomally encoded antibacterial proteins and peptides from Pseudomonas. *FEMS Microbiology Reviews*, **38**(4):523–568, 2014. doi: 10.1111/1574-6976.12079.

GHEQUIRE, M. G. K. AND DE MOT, R. Distinct colicin M-like bacteriocin-immunity pairs in Burkholderia. *Scientific reports*, **5**:17368, 2015b. doi: 10.1038/srep17368.

GHEQUIRE, M. G. K., KEMLAND, L., ANOZ-CARBONELL, E., BUCHANAN, S. K., AND DE MOT, R. A Natural Chimeric Pseudomonas Bacteriocin with Novel Pore-Forming Activity Parasitizes the Ferrichrome Transporter. *MBio*, **8**(1), 2017. doi: 10.1128/mBio.01961-16.

GILLOR, O., VRIEZEN, J. A., AND RILEY, M. A. The role of SOS boxes in enteric bacteriocin regulation. *Microbiology*, **154**(Pt 6):1783–1792, 2008. doi: 154/6/1783[pii]10.1099/mic.0.2007/016139-0.

GILSON, L., MAHANTY, H. K., AND KOLTER, R. Genetic analysis of an MDR-like export system: the secretion of colicin V. *EMBO J*, **9**(12):3875–3884, 1990.

GORDON, D. M. AND O'BRIEN, C. L. Bacteriocin diversity and the frequency of multiple bacteriocin production in *Escherichia coli*. *Microbiology (Reading, England)*, **152**(Pt 11):3239–44, 2006. doi: 10.1099/mic.0.28690-0.

GRAFEN, A. A geometric view of relatedness. *Oxf. Surv. Evol. Biol*, **2**:28–89, 1985.

GRAILLE, M., MORA, L., BUCKINGHAM, R. H., VAN TILBEURGH, H., AND DE ZAMAROCZY, M. Structural inhibition of the colicin D tRNase by the tRNA-mimicking immunity protein. *EMBO J*, **23**(7):1474–1482, 2004.

GRATIA, A. Sur un remarquable exemple d'antagonisme entre deux souches de colibacille. *C.R. Soc.Biol. (Paris)*, (93):1040–1041, 1925.

GRIFFIN, J. E., GAWRONSKI, J. D., DEJESUS, M. A., IOERGER, T. R., AKERLEY, B. J., AND SASSETTI, C. M. High-Resolution Phenotypic Profiling Defines Genes Essential for Mycobacterial Growth and Cholesterol Catabolism. *PLoS Pathogens*, **7**(9):1–9, 2011. doi: 10.1371/journal.ppat.1002251.

GRINTER, R., JOSTS, I., ZETH, K., ROSZAK, A. W., MCCAUGHEY, L. C., COGDELL, R. J., MILNER, J. J., KELLY, S. M., BYRON, O., AND WALKER, D. Structure of the atypical bacteriocin pectocin M2 implies a novel mechanism of protein uptake. *Mol Microbiol*, **93**(2):234–246, 2014. doi: 10.1111/mmi.12655.

GRINTER, R., ROSZAK, A. W., COGDELL, R. J., MILNER, J. J., AND WALKER, D. The crystal structure of the lipid II-degrading bacteriocin syringacin M suggests unexpected evolutionary relationships between colicin M-like bacteriocins. *J Biol Chem*, **287**(46):38876–38888, 2012. doi: 10.1074/jbc.M112.400150.

GRUBER, T. M., GOTTIG, S., MARK, L., CHRIST, S., KEMPF, V. A. J., WICHELHAUS, T. A., AND HAMPRECHT, A. Pathogenicity of pan-drug-resistant *Serratia marcescens* harbouring blaNDM-1. *The Journal of antimicrobial chemotherapy*, **70**(4):1026–1030, 2015. doi: 10.1093/jac/dku482.

GUASCH, J. F., ENFEDAQUE, J., FERRER, S., GARGALLO, D., AND REGUE, M. Bacteriocin 28b, a chromosomally encoded bacteriocin produced by most *Serratia marcescens* biotypes. *Res Microbiol*, **146**(6):477–483, 1995. doi: 0923250896802932[pii].

HACHANI, A., LOSSI, N. S., HAMILTON, A., JONES, C., BLEVES, S., ALBESA-JOVÉ, D., AND FILLOUX, A. Type VI Secretion System in *Pseudomonas aeruginosa*: secretion and multimerization of VgrG proteins. *The Journal of Biological Chemistry*, **286**(14):12317–12327, 2011. doi: 10.1074/jbc.M110.193045.

HACHANI, A., WOOD, T. E., AND FILLOUX, A. Type VI secretion and anti-host effectors. *Curr Opin Microbiol*, **29**:81–93, 2016. doi: 10.1016/j.mib.2015.11.006.

HAMILTON, W. D. The genetical evolution of social behaviour. I. *J Theor Biol*, **7**(1):1–16, 1964.

HAMILTON, W. D. Selfish and spiteful behaviour in an evolutionary model. *Nature*, **228**(5277):1218–1220, 1970.

HAMILTON, W. D. in Biosocial Anthropology. R. Fox„ 1975.

HANCOCK, R. E., MUTHARIA, L. M., CHAN, L., DARVEAU, R. P., SPEERT, D. P., AND PIER, G. B. *Pseudomonas aeruginosa* isolates from patients with cystic fibrosis: a class of serum-sensitive, nontypable strains deficient in lipopolysaccharide O side chains. *Infection and Immunity*, **42**(1):170–177, 1983.

HARDY, K. G., MEYNELL, G. G., DOWMAN, J. E., AND SPRATT, B. G. Two major groups of colicin factors: their evolutionary significance. *Mol Gen Genet*, **125**(3):217–230, 1973.

HECHARD, Y. AND SAHL, H. G. Mode of action of modified and unmodified bacteriocins from Gram-positive bacteria. *Biochimie*, **84**(5-6):545–557, 2002. doi: S0300908402014177[pii].

HEDDLE, J. G., BLANCE, S. J., ZAMBLE, D. B., HOLLFELDER, F., MILLER, D. A., WENTZELL, L. M., WALSH, C. T., AND MAXWELL, A. The antibiotic microcin B17 is a DNA gyrase poison: characterisation of the mode of inhibition. *Journal of Molecular Biology*, **307**(5):1223–1234, 2001. doi: 10.1006/jmbi.2001.4562.

HERSCHMAN, H. R. AND HELINSKI, D. R. Comparative study of the events associated with colicin induction. *Journal of Bacteriology*, **94**(3):691–699, 1967.

HILSENBECK, J. L., PARK, H., CHEN, G., YOUN, B., POSTLE, K., AND KANG, C. Crystal structure of the cytotoxic bacterial protein colicin B at 2.5 Åresolution. *Mol.Microbiol.*, **51**(3):711–720, 2004.

HO, B. T., DONG, T. G., AND MEKALANOS, J. J. A view to a kill: the bacterial type 6 secretion system. *Cell host & microbe*, **15**(1):9–21, 2014. doi: 10.1016/j.chom.2013.11.008.

HOL, F. J. H., VOGES, M. J., DEKKER, C., AND KEYMER, J. E. Nutrient-responsive regulation determines biodiversity in a colicin-mediated bacterial community. *BMC biology*, **12**:68, 2014. doi: 10.1186/s12915-014-0068-2.

HOLT, K., NGA, T., AND THANH, D. Tracking the establishment of local endemic populations of an emergent enteric pathogen. *Proc Natl Acad Sci USA*, **110**(43), 2013. doi: 10.1073/pnas.1308632110/-/DCSupplemental. www.pnas.org/cgi/doi/10.1073/pnas.1308632110.

HOOD, R. D., SINGH, P., HSU, F., GÜVENER, T., CARL, M. A., TRINIDAD, R. R. S., SILVERMAN, J. M., OHLSON, B. B., HICKS, K. G., PLEMEL, R. L., LI, M., SCHWARZ, S., WANG, W. Y., MERZ, A. J., GOODLETT, D. R., AND MOUGOUS, J. D. A Type VI Secretion System of *Pseudomonas aeruginosa* Targets a Toxin to Bacteria. *Cell host & microbe*, **7**(1):25–37, 2010. doi: 10.1016/j.chom.2009.12.007.

HOUSDEN, N. G., HOPPER, J. T., LUKOYANOVA, N., RODRIGUEZ-LARREA, D., WOJDYLA, J. A., KLEIN, A., KAMINSKA, R., BAYLEY, H., SAIBIL, H. R., ROBINSON, C. V., AND KLEANTHOUS, C. Intrinsically disordered protein threads through the bacterial outer-membrane porin OmpF. *Science*, **340**(6140):1570–1574, 2013. doi: 10.1126/science.1237864.

HOUSDEN, N. G., LOFTUS, S. R., MOORE, G. R., JAMES, R., AND KLEANTHOUS, C. Cell entry mechanism of enzymatic bacterial colicins: porin recruitment and the thermodynamics of receptor binding. *Proc Natl Acad Sci USA*, **102**(39):13849–13854, 2005.

HOUSDEN, N. G., WOJDYLA, J. A., KORCZYNSKA, J., GRISHKOVSKAYA, I., KIRKPATRICK, N., BRZOZOWSKI, A. M., AND KLEANTHOUS, C. Directed epitope delivery across the *Escherichia coli* outer membrane through the porin OmpF. *Proc Natl Acad Sci USA*, **107**(50):21412–21417, 2010. doi: 10.1073/pnas.1010780107.

HSIA, K. C., CHAK, K. F., LIANG, P. H., CHENG, Y. S., KU, W. Y., AND YUAN, H. S. DNA binding and degradation by the HNH protein ColE7. *Structure*, **12**(2):205–214, 2004.

HUGHES, K. T. AND MATHEE, K. The anti-sigma factors. *Annu Rev Microbiol*, **52**:231–286, 1998. doi: 10.1146/annurev.micro.52.1.231.

HUNT, J. M. Shiga Toxin-Producing *Escherichia coli* (STEC). *Clinics in Laboratory Medicine*, **30**(1):21–45, 2010. doi: http://dx.doi.org/10.1016/j.cll.2009.11.001.

JAKES, K. S. AND MODEL, P. Mechanism of export of colicin E1 and colicin E3. *Journal of Bacteriology*, **138**(3):770–778, 1979.

JAMES, R., SCHNEIDER, J., AND COOPER, P. C. Characterization of three group A klebicin plasmids: localization of their E colicin immunity genes. *Journal of general microbiology*, **133**(8):2253–2262, 1987. doi: 10.1099/00221287-133-8-2253.

JAMET, A., JOUSSET, A. B., EUPHRASIE, D., MUKORAKO, P., BOUCHARLAT, A., DUCOUSSO, A., CHARBIT, A., AND NASSIF, X. A new family of secreted toxins in pathogenic Neisseria species. *PLoS Pathog*, **11**(1):e1004592, 2015. doi: 10.1371/journal.ppat.1004592.

JAMET, A. AND NASSIF, X. New Players in the Toxin Field : Polymorphic Toxin Systems. *MBio*, **6**(3):1–8, 2015. doi: 10.1128/mBio.00285-15.Copyright.

JIANG, W., BIKARD, D., COX, D., ZHANG, F., AND MARRAFFINI, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol*, **31**(3):233–239, 2013. doi: 10.1038/nbt.2508.

JOHNSON, P. M., BECK, C. M., MORSE, R. P., GARZA-SANCHEZ, F., LOW, D. A., HAYES, C. S., AND GOULDING, C. W. Unraveling the essential role of CysK in CDI toxin activation. *Proc Natl Acad Sci USA*, **113**(35):9792–9797, 2016. doi: 10.1073/pnas.1607112113.

JOLLEY, K. A., BLISS, C. M., BENNETT, J. S., BRATCHER, H. B., BREHONY, C., COLLES, F. M., WIMALARATHNA, H., HARRISON, O. B., SHEPPARD, S. K., CODY, A. J., AND MAIDEN, M. C. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology*, **158**(Pt 4):1005–1015, 2012. doi: 10.1099/mic.0.055459-0.

JOSHI, A., GRINTER, R., JOSTS, I., CHEN, S., WOJDYLA, J. A., LOWE, E. D., KAMINSKA, R., SHARP, C., MCCAUGHEY, L., ROSZAK, A. W., COGDELL, R. J., BYRON, O., WALKER, D., AND KLEANTHOUS, C. Structures of the Ultra-High-Affinity Protein-Protein Complexes of Pyocins S2 and AP41 and Their Cognate Immunity Proteins from *Pseudomonas aeruginosa*. *Journal of Molecular Biology*, **427**(17):2852–2866, 2015. doi: 10.1016/j.jmb.2015.07.014.

KAAS, R. S., FRIIS, C., USSERY, D. W., AND AARESTRUP, F. M. Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse *Escherichia coli* genomes. *BMC Genomics*, **13**:577, 2012. doi: 10.1186/1471-2164-13-577.

KAGEYAMA, M., KOBAYASHI, M., SANO, Y., AND MASAKI, H. Construction and characterization of pyocin-colicin chimeric proteins. *Journal of Bacteriology*, **178**(1):103–110, 1996.

KAMENŠEK, S., PODLESEK, Z., GILLOR, O., AND ŽGUR-BERTOK, D. Genes regulated by the *Escherichia coli* SOS repressor LexA exhibit heterogenous expression. *BMC Microbiology*, **10**:283, 2010. doi: 10.1186/1471-2180-10-283.

KARFUNKEL, D., CARMELI, Y., CHMELNITSKY, I., KOTLOVSKY, T., AND NAVON-VENEZIA, S. The emergence and dissemination of CTX-M-producing *Escherichia coli* sequence type 131 causing community-onset bacteremia in Israel. *European Journal of Clinical Microbiology & Infectious Diseases*, **32**(4):513–521, 2013. doi: 10.1007/s10096-012-1765-9.

KAUER, J. C., ERICKSON-VIITANEN, S., WOLFE, H. R., AND DEGRADO, W. F. p-Benzoyl-L-phenylalanine, a new photoreactive amino acid. Photolabeling of calmodulin with a synthetic calmodulin-binding peptide. *Journal of Biological Chemistry*, **261**(23):10695–10700, 1986.

KAUNDAL, S., UTTAM, M., AND THAKUR, K. G. Dual Role of a Biosynthetic Enzyme, CysK, in Contact Dependent Growth Inhibition in Bacteria. *PLoS One*, **11**(7):e0159844, 2016. doi: 10.1371/journal.pone.0159844.

KAWAI, Y., ISHII, Y., ARAKAWA, K., UEMURA, K., SAITOH, B., NISHIMURA, J., KITAZAWA, H., YAMAZAKI, Y., TATENO, Y., ITOH, T., AND SAITO, T. Structural and Functional Differences in Two Cyclic Bacteriocins with the Same Sequences Produced by Lactobacilli. *Applied and Environmental Microbiology*, **70**(5):2906–2911, 2004. doi: 10.1128/AEM.70.5.2906-2911.2004.

KEEBLE, A. H. AND KLEANTHOUS, C. The kinetic basis for dual recognition in colicin endonuclease-immunity protein complexes. *Journal of Molecular Biology*, **352**(3):656–671, 2005. doi: S0022-2836(05)00823-5[pii] 10.1016/j.jmb.2005.07.035.

KERR, B., RILEY, M. A., FELDMAN, M. W., AND BOHANNAN, B. J. M. Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. *Nature*, **418**(6894):171–4, 2002. doi: 10.1038/nature00823.

KLEANTHOUS, C. Swimming against the tide: progress and challenges in our understanding of colicin translocation. *Nat Rev Microbiol*, **8**(12):843–848, 2010a. doi: nrmicro2454[pii]10.1038/nrmicro2454.

KLEANTHOUS, C. Translocator hunt comes full Cir-Col. *Mol Microbiol*, **75**(3):529–533, 2010b. doi: MMI6967[pii] 10.1111/j.1365-2958.2009.06967.x.

KLEANTHOUS, C., HEMMINGS, A. M., MOORE, G. R., AND JAMES, R. Immunity proteins and their specificity for endonuclease colicins: telling right from wrong in protein-protein recognition. *Mol Microbiol*, **28**(2):227–233, 1998.

KLEANTHOUS, C. AND WALKER, D. Immunity proteins: enzyme inhibitors that avoid the active site. *Trends Biochem Sci*, **26**(10):624–631, 2001.

KLEIN, A., WOJDYLA, J. A., JOSHI, A., JOSTS, I., MCCAUGHEY, L. C., HOUSDEN, N. G., KAMINSKA, R., BYRON, O., WALKER, D., AND KLEANTHOUS, C. Structural and biophysical analysis of nuclease protein antibiotics. *The Biochemical journal*, **473**(18):2799–2812, 2016. doi: 10.1042/BCJ20160544.

KONIECZNA, I., ŻARNOWIEC, P., KWINKOWSKI, M., KOLESIŃSKA, B., FRĄCZYK, J., KAMIŃSKI, Z., AND KACA, W. Bacterial Urease and its Role in Long-Lasting Human Diseases. *Current Protein & Peptide Science*, **13**(8):789–806, 2012. doi: 10.2174/138920312804871094.

KÖSER, C. U., ELLINGTON, M. J., AND PEACOCK, S. J. Whole-genome sequencing to control antimicrobial resistance. *Trends in Genetics*, **30**(9):401–407, 2014. doi: 10.1016/j.tig.2014.07.003.

KROGH, A., LARSSON, B., VON HEIJNE, G., AND SONNHAMMER, E. L. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology*, **305**(3):567–580, 2001. doi: http://dx.doi.org/10.1006/jmbi.2000.4315.

KUBE, S., KAPITEIN, N., ZIMNIAK, T., HERZOG, F., MOGK, A., AND WENDLER, P. Structure of the VipA/B type VI secretion complex suggests a contraction-state-specific recycling mechanism. *Cell Rep*, **8**(1):20–30, 2014. doi: 10.1016/j.celrep.2014.05.034.

KUHLMANN, U. C., POMMER, A. J., MOORE, G. R., JAMES, R., AND KLEANTHOUS, C. Specificity in protein-protein interactions: the structural basis for dual recognition in endonuclease colicin-immunity protein complexes. *Journal of Molecular Biology.*, **301**(5):1163, 2000.

KURISU, G., ZAKHAROV, S. D., ZHALNINA, M. V., BANO, S., EROUKOVA, V. Y., ROKITSKAYA, T. I., ANTONENKO, Y. N., WIENER, M. C., AND CRAMER, W. A. The structure of BtuB with bound colicin E3 R-domain implies a translocon. *Nat. Struct. Biol.*, **10**(11):948–954, 2003. doi: 10.1038/nsb997.

LAGOS, R., BAEZA, M., CORSINI, G., HETZ, C., STRAHSBURGER, E., CASTILLO, J. A., VERGARA, C., AND MONASTERIO, O. Structure, organization and characterization of the gene cluster involved in the production of microcin E492, a channel-forming bacteriocin. *Mol Microbiol*, **42**(1):229–243, 2001.

LANE, M. C., LLOYD, A. L., MARKYVECH, T. A., HAGAN, E. C., AND MOBLEY, H. L. T. Uropathogenic *Escherichia coli* Strains Generally Lack Functional Trg and Tap Chemoreceptors Found in the Majority of *E. coli* Strains Strictly Residing in the Gut. *Journal of Bacteriology*, **188**(15):5618–5625, 2006. doi: 10.1128/JB.00449-06.

LANGRIDGE, G. C., PHAN, M. D., TURNER, D. J., PERKINS, T. T., PARTS, L., HAASE, J., CHARLES, I., MASKELL, D. J., PETERS, S. E., DOUGAN, G., WAIN, J., PARKHILL, J., AND TURNER, A. K. Simultaneous assay of every *Salmonella Typhi* gene using one million transposon mutants. *Genome Res*, **19**(12):2308–2316, 2009. doi: 10.1101/gr.097097.109.

LAVINA, M., PUGSLEY, A. P., AND MORENO, F. Identification, mapping, cloning and characterization of a gene (sbmA) required for microcin B17 action on *Escherichia coli* K12. *J Gen Microbiol*, **132**(6):1685–1693, 1986. doi: 10.1099/00221287-132-6-1685.

LEHMANN, L., BARGUM, K., AND REUTER, M. An evolutionary analysis of the relationship between spite and altruism. *J Evol Biol*, **19**(5):1507–1516, 2006. doi: 10.1111/j.1420-9101.2006.01128.x.

LEIMAN, P. G., BASLER, M., RAMAGOPAL, U. A., BONANNO, J. B., SAUDER, J. M., PUKATZKI, S., BURLEY, S. K., ALMO, S. C., AND MEKALANOS, J. J. Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proc Natl Acad Sci USA*, **106**(11):4154–4159, 2009. doi: 10.1073/pnas.0813360106.

LEROUGE, I. AND VANDERLEYDEN, J. O-antigen structural variation: mechanisms and possible roles in animal/plant-microbe interactions. *FEMS Microbiology Reviews*, **26**(1):17–47, 2002. doi: 10.1111/j.1574-6976.2002.tb00597.x.

LESIC, B., STARKEY, M., HE, J., HAZAN, R., AND RAHME, L. G. Quorum sensing differentially regulates *Pseudomonas aeruginosa* type VI secretion locus I and homologous loci II and III, which are required for pathogenesis. *Microbiology*, **155**(Pt 9):2845–2855, 2009. doi: 10.1099/mic.0.029082-0.

LI, W., KEEBLE, A. H., GIFFARD, C., JAMES, R., MOORE, G. R., AND KLEANTHOUS, C. Highly discriminating protein-protein interaction specificities in the context of a conserved binding energy hotspot. *Journal of Molecular Biology*, **337**(3):743–759, 2004. doi: 10.1016/j.jmb.2004.02.005S0022283604001512[pii].

LIANG, X., MOORE, R., WILTON, M., WONG, M. J. Q., LAM, L., AND DONG, T. G. Identification of divergent type VI secretion effectors using a conserved chaperone domain. *Proc Natl Acad Sci USA*, **112**(29):9106–9111, 2015. doi: 10.1073/pnas.1505317112.

LIAO, C. C., HSIAO, K. C., LIU, Y. W., LENG, P. H., YUEN, H. S., AND CHAK, K. F. Processing of DNase domain during translocation of colicin E7 across the membrane of *Escherichia coli*. *Biochem Biophys Res Commun*, **284**(3):556–562, 2001.

LIN, L. J. R., LIAO, C. C., CHEN, Y. R., AND CHAK, K. F. Induction of membrane permeability in *Escherichia coli* mediated by lysis protein of the ColE7 operon. *FEMS Microbiology Letters*, **298**(1):85–92, 2009. doi: 10.1111/j.1574-6968.2009.01705.x.

LIN, Y. L., ELIAS, Y., AND HUANG, R. H. Structural and mutational studies of the catalytic domain of colicin E5: a tRNA-specific ribonuclease. *Biochemistry*, **44**(31):10494–10500, 2005.

LLOUBES, R., BATY, D., AND LAZDUNSKI, C. The promoters of the genes for colicin production, release and immunity in the ColA plasmid: effects of convergent transcription and Lex A protein. *Nucleic Acids Res*, **14**(6):2621–2636, 1986.

LOFTUS, S. R., WALKER, D., MATE, M. J., BONSOR, D. A., JAMES, R., MOORE, G. R., AND KLEANTHOUS, C. Competitive recruitment of the periplasmic translocation portal TolB by a natively disordered domain of colicin E9. *Proc. Natl. Acad. Sci. USA*, **103**(33):12353–12358, 2006.

LOUTET, S. A., FLANNAGAN, R. S., KOOI, C., SOKOL, P. A., AND VALVANO, M. A. A complete lipopolysaccharide inner core oligosaccharide is required for resistance of *Burkholderia cenocepacia* to antimicrobial peptides and bacterial survival in vivo. *Journal of bacteriology*, **188**(6):2073–2080, 2006. doi: 10.1128/JB.188.6.2073-2080.2006.

LUNA-CHAVEZ, C., LIN, Y. L., AND HUANG, R. H. Molecular basis of inhibition of the ribonuclease activity in colicin E5 by its cognate immunity protein. *Journal of Molecular Biology*, **358**(2):571–579, 2006.

LUPAS, A., VAN DYKE, M., AND STOCK, J. Predicting coiled coils from protein sequences. *Science (New York, N.Y.)*, **252**(5010):1162–4, 1991. doi: 10.1126/science.252.5009.1162.

MA, A. T., MCAULEY, S., PUKATZKI, S., AND MEKALANOS, J. J. Translocation of a *Vibrio cholerae* Type VI Secretion Effector Requires Bacterial Endocytosis by Host Cells. *Cell host & microbe*, **5**(3):234–243, 2009. doi: 10.1016/j.chom.2009.02.005.

MA, J., PAN, Z., HUANG, J., SUN, M., LU, C., AND YAO, H. The Hcp proteins fused with diverse extended-toxin domains represent a novel pattern of antibacterial effectors in type VI secretion systems. *Virulence*, pages 1–14, 2017. doi: 10.1080/21505594.2017.1279374.

MA, L.-S., HACHANI, A., LIN, J.-S., FILLOUX, A., AND LAI, E.-M. *Agrobacterium tumefaciens* Deploys a Superfamily of Type VI Secretion DNase Effectors as Weapons for Interbacterial Competition In Planta. *Cell Host & Microbe*, **16**(1):94–104, 2014. doi: https://doi.org/10.1016/j.chom.2014.06.002.

MACINTYRE, D. L., MIYATA, S. T., KITAOKA, M., AND PUKATZKI, S. The *Vibrio cholerae* type VI secretion system displays antimicrobial properties. *Proc Natl Acad Sci USA*, **107**(45):19520–19524, 2010. doi: 10.1073/pnas.1012931107.

MAJEED, H., GHAZARYAN, L., HERZBERG, M., AND GILLOR, O. Bacteriocin expression in sessile and planktonic populations of *Escherichia coli*. *J Antibiot*, **68**(1):52–55, 2015.

MAJEED, H., GILLOR, O., KERR, B., AND RILEY, M. A. Competitive interactions in *Escherichia coli*populations: the role of bacteriocins. *The ISME journal*, **5**(1):71–81, 2011. doi: 10.1038/ismej.2010.90.

MALDONADO, R. F., SÁ-CORREIA, I., AND VALVANO, M. A. Lipopolysaccharide modification in Gram-negative bacteria during chronic infection. *FEMS Microbiology Reviews*, **40**(4):480–493, 2016. doi: 10.1093/femsre/fuw007.

MALINVERNI, J. C. AND SILHAVY, T. J. An ABC transport system that maintains lipid asymmetry in the Gram-negative outer membrane. *Proc Natl Acad Sci USA*, **106**(19):8009–8014, 2009. doi: 10.1073/pnas.0903229106.

MAQUEDA, M., SANCHEZ-HIDALGO, M., FERNANDEZ, M., MONTALBAN-LOPEZ, M., VALDIVIA, E., AND MARTINEZ-BUENO, M. Genetic features of circular bacteriocins produced by Gram-positive bacteria. *Fems Microbiology Reviews*, **32**(1):2–22, 2008. doi: 10.1111/j.1574-6976.2007.00087.x.

MARCELLETTI, S. AND SCORTICHINI, M. Definition of Plant-Pathogenic Pseudomonas Genomospecies of the *Pseudomonas syringae* Complex Through Multiple Comparative Approaches. *Phytopathology*, **104**(12):1274–1282, 2014. doi: 10.1094/PHYTO-12-13-0344-R.

MARSH, A. J., O'SULLIVAN, O., ROSS, R. P., COTTER, P. D., AND HILL, C. In silico analysis highlights the frequency and diversity of type 1 lantibiotic gene clusters in genome sequenced bacteria. *BMC Genomics*, **11**:679, 2010. doi: 10.1186/1471-2164-11-679.

MARVIN, H. J., TER BEEST, M. B., AND WITHOLT, B. Release of outer membrane fragments from wild-type *Escherichia coli* and from several *E. coli* lipopolysaccharide mutants by EDTA and heat shock treatments. *Journal of Bacteriology*, **171**(10):5262–5267, 1989.

MATSUI, H., SANO, Y., ISHIHARA, H., AND SHINOMIYA, T. Regulation of pyocin genes in *Pseudomonas aeruginosa* by positive (prtN) and negative (prtR) regulatory genes. *Journal of Bacteriology*, **175**(5):1257–1263, 1993.

MCCAUGHEY, L. C., GRINTER, R., JOSTS, I., ROSZAK, A. W., WALOEN, K. I., COGDELL, R. J., MILNER, J., EVANS, T., KELLY, S., TUCKER, N. P., BYRON, O., SMITH, B., AND WALKER, D. Lectin-Like Bacteriocins from *Pseudomonas* spp. Utilise D-Rhamnose Containing Lipopolysaccharide as a Cellular Receptor. *PLoS Pathog*, **10**(2):e1003898, 2014. doi: 10.1371/journal.ppat.1003898.

MCCAUGHEY, L. C., RITCHIE, N. D., DOUCE, G. R., EVANS, T. J., AND WALKER, D. Efficacy of species-specific protein antibiotics in a murine model of acute *Pseudomonas aeruginosa* lung infection. *Scientific Reports*, **6**:30201, 2016a. doi: 10.1038/srep30201.

McCaughey, L. C. C., Josts, I., Grinter, R., White, P., Byron, O., Tucker, N. P. P., Matthews, J. M. M., Kleanthous, C., Whitchurch, C. B. B., and Walker, D. Discovery, characterization and in vivo activity of pyocin SD2, a protein antibiotic from *Pseudomonas aeruginosa*. *The Biochemical journal*, **473**(15):2345–2358, 2016b. doi: 10.1042/bcj20160470.

McNally, L., Bernardy, E., Thomas, J., Kalziqi, A., Pentz, J., Brown, S. P., Hammer, B. K., Yunker, P. J., and Ratcliff, W. C. Killing by Type VI secretion drives genetic phase separation and correlates with increased cooperation. *Nature Communications*, **8**:14371, 2017. doi: 10.1038/ncomms14371.

Meenan, N. A., Sharma, A., Fleishman, S. J., Macdonald, C. J., Morel, B., Boetzel, R., Moore, G. R., Baker, D., and Kleanthous, C. The structural and energetic basis for high selectivity in a high-affinity protein-protein interaction. *Proc Natl Acad Sci USA*, **107**(22):10080–10085, 2010. doi: 0910756107[pii]10.1073/pnas.0910756107.

Menozzi, F. D., Mutombo, R., Renauld, G., Gantiez, C., Hannah, J. H., Leininger, E., Brennan, M. J., and Locht, C. Heparin-inhibitable lectin activity of the filamentous hemagglutinin adhesin of *Bordetella pertussis*. *Infect Immun*, **62**(3):769–778, 1994.

Meyhack, B., Meyhack, I., and Apirion, D. Colicin E3: a unique endoribonuclease. *Proc Natl Acad Sci USA*, **70**(1):156–160, 1973.

Micenková, L., Staudová, B., Bosák, J., Mikalová, L., Littnerová, S., Vrba, M., Sevčíková, A., Woznicová, V., and Smajs, D. Bacteriocin-encoding genes and ExPEC virulence determinants are associated in human fecal *Escherichia coli* strains. *BMC microbiology*, **14**(1):109, 2014a. doi: 10.1186/1471-2180-14-109.

Micenková, L., Staudová, B., Bosák, J., Mikalová, L., Littnerová, S., Vrba, M., Sevčíková, A., Woznicová, V., Smajs, D., Micenkova, L., Staudova, B., Bosak, J., Mikalova, L., Littnerova, S., Vrba, M., Sevcikova, A., Woznicova, V., Smajs, D., Micenková, L., Staudová, B., Bosák, J., Mikalová, L., Littnerová, S., Vrba, M., Sevčíková, A., Woznicová, V., and Smajs, D. Bacteriocin-encoding genes and ExPEC virulence determinants are associated in human fecal *Escherichia coli* strains. *BMC Microbiol*, **14**:109, 2014b. doi: 10.1186/1471-2180-14-109.

Michel-Briand, Y. and Baysse, C. The pyocins of *Pseudomonas aeruginosa*. *Biochimie*, **84**(5-6):499–510, 2002. doi: S0300908402014220[pii].

Monsieurs, P., De Keersmaecker, S., Navarre, W. W., Bader, M. W., De Smet, F., McClelland, M., Fang, F. C., De Moor, B., Vanderleyden, J., and Marchal, K. Comparison of the PhoPQ Regulon in *Escherichia coli* and *Salmonella typhimurium*. *Journal of Molecular Evolution*, **60**(4):462–474, 2005. doi: 10.1007/s00239-004-0212-7.

Mora, L. and de Zamaroczy, M. In vivo processing of DNase colicins E2 and E7 is required for their import into the cytoplasm of target cells. *PLoS One*, **9**(5):e96549, 2014. doi: 10.1371/journal.pone.0096549.

MORA, L., DIAZ, N., BUCKINGHAM, R. H., AND DE ZAMAROCZY, M. Import of the transfer RNase colicin D requires site-specific interaction with the energy-transducing protein TonB. *Journal of Bacteriology*, **187**(8):2693–2697, 2005. doi: 187/8/2693[pii]10.1128/JB.187.8.2693-2697.2005.

MORA, L., KLEPSCH, M., BUCKINGHAM, R. H., HEURGUÉ-HAMARD, V., KERVESTIN, S., AND DE ZAMAROCZY, M. Dual roles of the central domain of colicin D tRNase in TonB-mediated import and in immunity. *The Journal of biological chemistry*, **283**(8):4993–5003, 2008. doi: 10.1074/jbc.M706846200.

MORA, L., MONCOQ, K., ENGLAND, P., OBERTO, J., AND DE ZAMAROCZY, M. The Stable Interaction Between Signal Peptidase LepB of *Escherichia coli* and Nuclease Bacteriocins Promotes Toxin Entry into the Cytoplasm. *The Journal of biological chemistry*, **290**(52):30783–96, 2015. doi: 10.1074/jbc.M115.691907.

MOSBAHI, K., LEMAITRE, C., KEEBLE, A. H., MOBASHERI, H., MOREL, B., JAMES, R., MOORE, G. R., LEA, E. J., AND KLEANTHOUS, C. The cytotoxic domain of colicin E9 is a channel-forming endonuclease. *Nat Struct Biol*, **9**(6):476–484, 2002.

MOSBAHI, K., WALKER, D., LEA, E., MOORE, G. R., JAMES, R., AND KLEANTHOUS, C. Destabilization of the colicin E9 Endonuclease domain by interaction with negatively charged phospholipids: implications for colicin translocation into bacteria. *J Biol Chem*, **279**(21):22145–22151, 2004. doi: 10.1074/jbc.M400402200M400402200[pii].

MOUGOUS, J. D., CUFF, M. E., RAUNSER, S., SHEN, A., ZHOU, M., GIFFORD, C. A., GOODMAN, A. L., JOACHIMIAK, G., ORDOÑEZ, C. L., LORY, S., WALZ, T., JOACHIMIAK, A., AND MEKALANOS, J. J. A Virulence Locus of *Pseudomonas aeruginosa* Encodes a Protein Secretion Apparatus. *Science (New York, N.Y.)*, **312**(5779):1526–1530, 2006. doi: 10.1126/science.1128393.

MURDOCH, S. L., TRUNK, K., ENGLISH, G., FRITSCH, M. J., POURKARIMI, E., AND COULTHURST, S. J. The Opportunistic Pathogen *Serratia marcescens* Utilizes Type VI Secretion To Target Bacterial Competitors . *Journal of Bacteriology*, **193**(21):6057–6069, 2011. doi: 10.1128/JB.05671-11.

NAKAYAMA, K., TAKASHIMA, K., ISHIHARA, H., SHINOMIYA, T., KAGEYAMA, M., KANAYA, S., OHNISHI, M., MURATA, T., MORI, H., AND HAYASHI, T. The R-type pyocin of *Pseudomonas aeruginosa* is related to P2 phage, and the F-type is related to lambda phage. *Mol Microbiol*, **38**(2):213–231, 2000. doi: mmi2135[pii].

NEDIALKOVA, L. P., DENZLER, R., KOEPPEL, M. B., DIEHL, M., RING, D., WILLE, T., GERLACH, R. G., AND STECHER, B. Inflammation fuels colicin Ib-dependent competition of Salmonella serovar Typhimurium and E. coli in enterobacterial blooms. *PLoS pathogens*, **10**(1):e1003844, 2014. doi: 10.1371/journal.ppat.1003844.

NEDIALKOVA, L. P., SIDSTEDT, M., KOEPPEL, M. B., SPRIEWALD, S., RING, D., GERLACH, R. G., BOSSI, L., AND STECHER, B. Temperate phages promote colicin-dependent fitness of Salmonella enterica serovar Typhimurium. *Environmental microbiology*, 2015. doi: 10.1111/1462-2920.13077.

NG, C. L., LANG, K., MEENAN, N. A., SHARMA, A., KELLEY, A. C., KLEANTHOUS, C., AND RAMAKRISHNAN, V. Structural basis for 16S ribosomal RNA cleavage by the cytotoxic domain of colicin E3. *Nat Struct Mol Biol*, **17**(10):1241–1246, 2010. doi: nsmb.1896[pii]10.1038/nsmb.1896.

NICOLAS-CHANOINE, M.-H., BERTRAND, X., AND MADEC, J.-Y. *Escherichia coli* ST131, an Intriguing Clonal Group. *Clinical Microbiology Reviews*, **27**(3):543–574, 2014. doi: 10.1128/CMR.00125-13.

NIPIČ, D., PODLESEK, Z., BUDIČ, M., ČRNIGOJ, M., AND ŽGUR-BERTOK, D. *Escherichia coli* uropathogenic-specific protein, USP, is a bacteriocin-like genotoxin. *Journal of Infectious Diseases*, **208**(10):1545–1552, 2013. doi: 10.1093/infdis/jit480.

NOINAJ, N., KUSZAK, A. J., GUMBART, J. C., LUKACIK, P., CHANG, H., EASLEY, N. C., LITHGOW, T., AND BUCHANAN, S. K. Structural insight into the biogenesis of beta-barrel membrane proteins. *Nature*, **501**(7467):385–390, 2013. doi: 10.1038/nature12521.

NOVOA, M. A., DIAZ-GUERRA, L., SAN MILLAN, J. L., AND MORENO, F. Cloning and mapping of the genetic determinants for microcin C7 production and immunity. *Journal of Bacteriology*, **168**(3):1384–1391, 1986.

OGAWA, T., INOUE, S., YAJIMA, S., HIDAKA, M., AND MASAKI, H. Sequence-specific recognition of colicin E5, a tRNA-targeting ribonuclease. *Nucleic Acids Research*, **34**(21):6065–6073, 2006. doi: 10.1093/nar/gkl629.

OHKAWA, I., SHIGA, S., AND KAGEYAMA, M. Effect of Iron Concentration in the Growth Medium on the Sensitivity of *Pseudomonas aeruginosa* to Pyocin S2. *Journal of Biochemistry*, **87**(1):323–331, 1980.

OLAITAN, A. O., MORAND, S., AND ROLAIN, J.-M. Mechanisms of polymyxin resistance: acquired and intrinsic resistance in bacteria. *Frontiers in Microbiology*, **5**:643, 2014. doi: 10.3389/fmicb.2014.00643.

OSIPIUK, J., XU, X., CUI, H., SAVCHENKO, A., EDWARDS, A., AND JOACHIMIAK, A. Crystal structure of secretory protein Hcp3 from *Pseudomonas aeruginosa*. *Journal of Structural and Functional Genomics*, **12**(1):21–26, 2011. doi: 10.1007/s10969-011-9107-1.

OUDEGA, B., MOL, O., VAN ULSEN, P., STEGEHUIS, F., VAN DER WAL, F. J., AND LUIRINK, J. *Escherichia coli* SecB, SecA, and SecY proteins are required for expression and membrane insertion of the bacteriocin release protein, a small lipoprotein. *Journal of Bacteriology*, **175**(5):1543–1547, 1993.

PACE, C. N., VAJDOS, F., FEE, L., GRIMSLEY, G., AND GRAY, T. How to measure and predict the molar absorption coefficient of a protein. *Protein Science*, **4**(11):2411–2423, 1995. doi: 10.1002/pro.5560041120.

PAGE, A. J., CUMMINS, C. A., HUNT, M., WONG, V. K., REUTER, S., HOLDEN, M. T. G., FOOKES, M., FALUSH, D., KEANE, J. A., AND PARKHILL, J. Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, **31**(July):btv421, 2015. doi: 10.1093/bioinformatics/btv421.

PAGE, A. J., DE SILVA, N., HUNT, M., QUAIL, M. A., PARKHILL, J., HARRIS, S. R., OTTO, T. D., AND KEANE, J. A. Robust high-throughput prokaryote de novo assembly and improvement pipeline for Illumina data. *Microbial Genomics*, **2**(8):e000083, 2016. doi: 10.1099/mgen.0.000083.

PAPADAKOS, G., WOJDYLA, J. A., AND KLEANTHOUS, C. Nuclease colicins and their immunity proteins., volume 45. 2011/11/17 edition, 2012. ISBN 4419043288. doi: 10.1017/S0033583511000114.

PARRET, A. H. AND DE MOT, R. Bacteria killing their own kind: novel bacteriocins of Pseudomonas and other gamma-proteobacteria. *Trends Microbiol*, **10**(3):107–112, 2002.

PARRET, A. H., DE MOT, R., AND MOT, R. D. Bacteria killing their own kind: novel bacteriocins of Pseudomonas and other γ-proteobacteria. *Trends Microbiol*, **10**(3):107–112, 2002.

PARRET, A. H. A., SCHOOFS, G., PROOST, P., AND DE MOT, R. Plant lectin-like bacteriocin from a rhizosphere-colonizing Pseudomonas isolate. *Journal of bacteriology*, **185**(3):897–908, 2003.

PEIRANO, G., VAN DER BIJ, A. K., GREGSON, D. B., AND PITOUT, J. D. D. Molecular Epidemiology over an 11-Year Period (2000 to 2010) of Extended-Spectrum β-Lactamase-Producing *Escherichia coli* Causing Bacteremia in a Centralized Canadian Region. *Journal of Clinical Microbiology*, **50**(2):294–299, 2012. doi: $10.1128/\mathrm{JCM.06025\text{-}11}$.

PELL, L. G., KANELIS, V., DONALDSON, L. W., LYNNE HOWELL, P., AND DAVIDSON, A. R. The phage λ major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system. *Proc Natl Acad Sci USA*, **106**(11):4160–4165, 2009. doi: $10.1073/\mathrm{pnas.0900044106}$.

PETKOVSEK, Z., ZGUR-BERTOK, D., AND STARCIC ERJAVEC, M. Colicin insensitivity correlates with a higher prevalence of extraintestinal virulence factors among *Escherichia coli* isolates from skin and soft-tissue infections. *Journal of medical microbiology*, **61**(Pt 6):762–5, 2012. doi: $10.1099/\mathrm{jmm.0.037234\text{-}0}$.

PETTY, N. K., BEN ZAKOUR, N. L., STANTON-COOK, M., SKIPPINGTON, E., TOTSIKA, M., FORDE, B. M., PHAN, M.-D., GOMES MORIEL, D., PETERS, K. M., DAVIES, M., ROGERS, B. A., DOUGAN, G., RODRIGUEZ-BAÑO, J., PASCUAL, A., PITOUT, J. D. D., UPTON, M., PATERSON, D. L., WALSH, T. R., SCHEMBRI, M. A., AND BEATSON, S. A. Global dissemination of a multidrug resistant *Escherichia coli* clone. *Proc Natl Acad Sci USA*, **111**(15):5694–5699, 2014. doi: $10.1073/\mathrm{pnas.1322678111}$.

POMMER, A. J., CAL, S., KEEBLE, A. H., WALKER, D., EVANS, S. J., KUHLMANN, U. C., COOPER, A., CONNOLLY, B. A., HEMMINGS, A. M., MOORE, G. R., JAMES, R., AND KLEANTHOUS, C. Mechanism and cleavage specificity of the H-N-H endonuclease colicin E9. *Journal of Molecular Biology.*, **314**(4):735, 2001.

PUGSLEY, A. P. The immunity and lysis genes of ColN plasmid pCHAP4. *Mol Gen Genet*, **211**(2):335–341, 1988.

PUGSLEY, A. P. AND SCHWARTZ, M. Colicin E2 release: lysis, leakage or secretion? Possible role of a phospholipase. *EMBO J*, **3**(10):2393–2397, 1984.

PUKATZKI, S., MA, A. T., STURTEVANT, D., KRASTINS, B., SARRACINO, D., NELSON, W. C., HEIDELBERG, J. F., AND MEKALANOS, J. J. Identification of a conserved bacterial protein secretion system in Vibrio cholerae using the Dictyostelium host model system. *Proc Natl Acad Sci USA*, **103**(5):1528–1533, 2006. doi: $10.1073/\mathrm{pnas.0510322103}$.

PUKATZKI, S., MCAULEY, S. B., AND MIYATA, S. T. The type VI secretion system: translocation of effectors and effector-domains. *Current Opinion in Microbiology*, **12**(1):11–17, 2009. doi: $\mathrm{https://doi.org/10.1016/j.mib.2008.11.010}$.

QIMRON, U., MARINTCHEVA, B., TABOR, S., AND RICHARDSON, C. C. Genomewide screens for *Escherichia coli* genes affecting growth of T7 bacteriophage. *Proc Natl Acad Sci USA*, **103**(50):19039–19044, 2006. doi: 10.1073/pnas.0609428103.

RAJASEKAR, K. V., ZDANOWSKI, K., YAN, J., HOPPER, J. T., FRANCIS, M. L., SEEPERSAD, C., SHARP, C., PECQUEUR, L., WERNER, J. M., ROBINSON, C. V., MOHAMMED, S., POTTS, J. R., AND KLEANTHOUS, C. The anti-sigma factor RsrA responds to oxidative stress by reburying its hydrophobic core. *Nat Commun*, **7**:12194, 2016. doi: 10.1038/ncomms12194.

RAKIN, A., BOOLGAKOWA, E., AND HEESEMANN, J. Structural and functional organization of the Yersinia pestis bacteriocin pesticin gene cluster. *Microbiology*, **142**(12):3415–3424, 1996. doi: 10.1099/13500872-142-12-3415.

RASSAM, P., COPELAND, N. A., BIRKHOLZ, O., TOTH, C., CHAVENT, M., DUNCAN, A. L., CROSS, S. J., HOUSDEN, N. G., KAMINSKA, R., SEGER, U., QUINN, D. M., GARROD, T. J., SANSOM, M. S. P., PIEHLER, J., BAUMANN, C. G., AND KLEANTHOUS, C. Supramolecular assemblies underpin turnover of outer membrane proteins in bacteria. *Nature*, **523**(7560):333–336, 2015. doi: 10.1038/nature14461http://www.nature.com/nature/journal/v523/n7560/abs/nature14461.html#supplementary-information.

RENDUELES, O., BELOIN, C., LATOUR-LAMBERT, P., AND GHIGO, J.-M. A new biofilm-associated colicin with increased efficiency against biofilm bacteria. *ISME J*, **8**(6):1275–1288, 2014.

RICE, P., LONGDEN, I., AND BLEASBY, A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics*, **16**(6):276–277, 2000. doi: http://dx.doi.org/10.1016/S0168-9525(00)02024-2.

RIJAVEC, M., BUDIC, M., MRAK, P., MÜLLER-PREMRU, M., PODLESEK, Z., AND ZGUR-BERTOK, D. Prevalence of ColE1-like plasmids and colicin K production among uropathogenic *Escherichia coli* strains and quantification of inhibitory activity of colicin K. *Applied and environmental microbiology*, **73**(3):1029–32, 2007. doi: 10.1128/AEM.01780-06.

RILEY, M. A. Molecular mechanisms of colicin evolution. *Mol Biol Evol*, **10**(6):1380–1395, 1993a.

RILEY, M. A. Molecular mechanisms of bacteriocin evolution. *Annu Rev Genet*, **32**:255–278, 1998a.

RILEY, M. A. MOLECULAR MECHANISMS OF Producing antimicrobial compounds seems to be a generic phenomenon for. *Annu. Rev. Genet.*, **32**:255–278, 1998b.

RILEY, M. A. AND GORDON, D. M. A survey of Col plasmids in natural isolates of *Escherichia coli* and an investigation into the stability of Col-plasmid lineages. *J Gen Microbiol*, **138**(7):1345–1352, 1992.

RILEY, M. A. AND GORDON, D. M. The ecology and evolution of bacteriocins. *Journal of Industrial Microbiology & Biotechnology*, **17**(3-4):151–158, 1996. doi: 10.1007/BF01574688.

RILEY, M. A., PINOU, T., WERTZ, J. E., TAN, Y., AND VALLETTA, C. M. Molecular characterization of the klebicin B plasmid of *Klebsiella pneumoniae*. *Plasmid*, **45**(3):209–221, 2001. doi: 10.1006/plas.2001.1519S0147-619X(01)91519-7[pii].

RILEY, M. A. AND WERTZ, J. E. Bacteriocin diversity: ecological and evolutionary perspectives. *Biochimie*, **84**(5-6):357–64, 2002. doi: S0300908402014219[pii].

RILEY, M. A. A. Positive selection for colicin diversity in bacteria. *Mol Biol Evol*, **10**(5):1048–1059, 1993b.

ROH, E., PARK, T. H., KIM, M. I., LEE, S., RYU, S., OH, C. S., RHEE, S., KIM, D. H., PARK, B. S., AND HEU, S. Characterization of a new bacteriocin, Carocin D, from Pectobacterium carotovorum subsp. carotovorum Pcc21. *Appl Environ Microbiol*, **76**(22):7541–7549, 2010. doi: AEM.03103-09[pii]10.1128/AEM.03103-09.

ROJAS, C. M., HAM, J. H., DENG, W. L., DOYLE, J. J., AND COLLMER, A. HecA, a member of a class of adhesins produced by diverse pathogenic bacteria, contributes to the attachment, aggregation, epidermal cell killing, and virulence phenotypes of *Erwinia chrysanthemi* EC16 on *Nicotiana clevelandii* seedlings. *Proc Natl Acad Sci USA*, **99**(20):13142–13147, 2002.

RUHE, Z. C., NGUYEN, J. Y., BECK, C. M., LOW, D. A., AND HAYES, C. S. The proton-motive force is required for translocation of CDI toxins across the inner membrane of target bacteria. *Mol Microbiol*, **94**(2):466–481, 2014. doi: 10.1111/mmi.12779.

RUHE, Z. C., NGUYEN, J. Y., CHEN, A. J., LEUNG, N. Y., HAYES, C. S., AND LOW, D. A. CDI Systems Are Stably Maintained by a Cell-Contact Mediated Surveillance Mechanism. *PLoS Genetics*, **12**(6):e1006145, 2016. doi: 10.1371/journal.pgen.1006145.

RUHE, Z. C., NGUYEN, J. Y., XIONG, J., KOSKINIEMI, S., BECK, C. M., PERKINS, B. R., LOW, D. A., AND HAYES, C. S. CdiA Effectors Use Modular Receptor-Binding Domains To Recognize Target Bacteria. *MBio*, **8**(2), 2017. doi: 10.1128/mBio.00290-17.

RUHE, Z. C., TOWNSLEY, L., WALLACE, A. B., KING, A., DER WOUDE, M. W. V., LOW, D. A., YILDIZ, F. H., AND HAYES, C. S. CdiA promotes receptor-independent intercellular adhesion. *Molecular microbiology*, **98**(1):175–192, 2015. doi: 10.1111/mmi.13114.

RUHE, Z. C., WALLACE, A. B., LOW, D. A., AND HAYES, C. S. Receptor polymorphism restricts contact-dependent growth inhibition to members of the same species. *MBio*, **4**(4), 2013. doi: 10.1128/mBio.00480-13.

RUSSELL, A. B., LEROUX, M., HATHAZI, K., AGNELLO, D. M., ISHIKAWA, T., WIGGINS, P. A., WAI, S. N., AND MOUGOUS, J. D. Diverse type VI secretion phospholipases are functionally plastic antibacterial effectors. *Nature*, **496**(7446):508–512, 2013. doi: 10.1038/nature12074.

RUSSELL, A. B., PETERSON, S. B., AND MOUGOUS, J. D. Type VI secretion system effectors: poisons with a purpose. *Nat Rev Microbiol*, **12**(2):137–148, 2014. doi: 10.1038/nrmicro3185.

RUSSO, T. A., STAPLETON, A., WENDEROTH, S., HOOTON, T. M., AND STAMM, W. E. Chromosomal restriction fragment length polymorphism analysis of *Escherichia coli* strains causing recurrent urinary tract infections in young women. *The Journal of infectious diseases*, **172**(2):440–445, 1995.

SALLES, B., WEISEMANN, J. M., AND WEINSTOCK, G. M. Temporal control of colicin E1 induction. *Journal of Bacteriology*, **169**(11):5028–5034, 1987.

SALOMON, D., KINCH, L. N., TRUDGIAN, D. C., GUO, X., KLIMKO, J. A., GRISHIN, N. V., MIRZAEI, H., AND ORTH, K. Marker for type VI secretion system effectors. *Proc Natl Acad Sci USA*, **111**(25):1406110111–, 2014. doi: $10.1073/\mathrm{pnas}.1406110111$.

SANKARAN, K. AND WU, H. C. Lipid modification of bacterial prolipoprotein. Transfer of diacylglyceryl moiety from phosphatidylglycerol. *Journal of Biological Chemistry*, **269**(31):19701–19706, 1994.

SANO, Y. AND KAGEYAMA, M. A novel transposon-like structure carries the genes for pyocin AP41, a *Pseudomonas aeruginosa* bacteriocin with a DNase domain homology to E2 group colicins. *Mol Gen Genet*, **237**(1-2):161–170, 1993.

SANO, Y., KOBAYASHI, M., AND KAGEYAMA, M. Functional domains of S-type pyocins deduced from chimeric molecules. *Journal of Bacteriology*, **175**(19):6179–6185, 1993a.

SANO, Y., MATSUI, H., KOBAYASHI, M., AND KAGEYAMA, M. Molecular structures and functions of pyocins S1 and S2 in *Pseudomonas aeruginosa*. *Journal of Bacteriology*, **175**(10):2907–2916, 1993b.

SANTAJIT, S. AND INDRAWATTANA, N. Mechanisms of Antimicrobial Resistance in ESKAPE Pathogens. *BioMed Research International*, **2016**:2475067, 2016. doi: $10.1155/2016/2475067$.

SCHALLER, K. AND NOMURA, M. Colicin E2 is DNA endonuclease. *Proc Natl Acad Sci USA*, **73**(11):3989–3993, 1976.

SCHINDELIN, J., ARGANDA-CARRERAS, I., FRISE, E., KAYNIG, V., LONGAIR, M., PIETZSCH, T., PREIBISCH, S., RUEDEN, C., SAALFELD, S., SCHMID, B., TINEVEZ, J.-Y., WHITE, D. J., HARTENSTEIN, V., ELICEIRI, K., TOMANCAK, P., AND CARDONA, A. Fiji - an Open Source platform for biological image analysis. *Nature methods*, **9**(7):10.1038/nmeth.2019, 2012. doi: $10.1038/\mathrm{nmeth}.2019$.

SCHWARZ, S., WEST, T. E., BOYER, F., CHIANG, W.-C., CARL, M. A., HOOD, R. D., ROHMER, L., TOLKER-NIELSEN, T., SKERRETT, S. J., AND MOUGOUS, J. D. Burkholderia Type VI Secretion Systems Have Distinct Roles in Eukaryotic and Bacterial Cell Interactions. *PLoS Pathogens*, **6**(8):e1001068, 2010.

SHANNON, R. AND HEDGES, A. J. Kinetics of Lethal Adsorption of Colicin E2 by *Escherichia coli*. *Journal of Bacteriology*, **93**(4):1353–1359, 1967.

SHARMA, O. AND CRAMER, W. A. Minimum length requirement of the flexible N-terminal translocation subdomain of colicin E3. *Journal of Bacteriology*, **189**(2):363–368, 2007. doi: $\mathrm{JB}.01344\text{-}06[\mathrm{pii}]10.1128/\mathrm{JB}.01344\text{-}06$.

SHARP, C., BRAY, J., HOUSDEN, N. G., MAIDEN, M. C. J., AND KLEANTHOUS, C. Diversity and distribution of nuclease bacteriocins in bacterial genomes revealed using Hidden Markov Models. *PLoS Computational Biology*, **13**(7):e1005652, 2017.

SHEN, B. W., LANDTHALER, M., SHUB, D. A., AND STODDARD, B. L. DNA binding and cleavage by the HNH homing endonuclease I-HmuI. *Journal of Molecular Biology*, **342**(1):43–56, 2004.

SHIN, J. M., GWAK, J. W., KAMARAJAN, P., FENNO, J. C., RICKARD, A. H., AND KAPILA, Y. L. Biomedical Applications of Nisin. *Journal of applied microbiology*, **120**(6):1449–1465, 2016. doi: 10.1111/jam.13033.

SHINOMIYA, T., SHIGA, S., AND KAGEYAMA, M. Genetic determinant of pyocin R2 in *Pseudomonas aeruginosa* PAO. I. Localization of the pyocin R2 gene cluster between the trpCD and trpE genes. *Mol Gen Genet*, **189**(3):375–381, 1983.

SHNEIDER, M. M., BUTH, S. A., HO, B. T., BASLER, M., MEKALANOS, J. J., AND LEIMAN, P. G. PAAR-repeat proteins sharpen and diversify the Type VI secretion system spike. *Nature*, **500**(7462):350–353, 2013. doi: 10.1038/nature12453.

SILVERMAN, J. M., AGNELLO, D. M., ZHENG, H., ANDREWS, B. T., LI, M., CATALANO, C. E., GONEN, T., AND MOUGOUS, J. D. Haemolysin Co-regulated Protein is an Exported Receptor and Chaperone of Type VI Secretion Substrates. *Molecular cell*, **51**(5):10.1016/j.molcel.2013.07.025, 2013. doi: 10.1016/j.molcel.2013.07.025.

SIMMONDS, R. S., SIMPSON, W. J., AND TAGG, J. R. Cloning and sequence analysis of zooA, a Streptococcus zooepidemicus gene encoding a bacteriocin-like inhibitory substance having a domain structure similar to that of lysostaphin. *Gene*, **189**(2):255–261, 1997. doi: https://doi.org/10.1016/S0378-1119(96)00859-1.

SOELAIMAN, S., JAKES, K., WU, N., LI, C. M., AND SHOHAM, M. Crystal structure of colicin E3: implications for cell entry and ribosome inactivation. *Mol.Cell*, **8**(5):1053–1062, 2001. doi: 10.1016/s1097-2765(01)00396-3.

SOONG, B. W., HSIEH, S. Y., AND CHAK, K. F. Mapping of transcriptional start sites of the cea and cei genes of the ColE7 operon. *Mol Gen Genet*, **243**(4):477–481, 1994.

SPAIN, A. M., KRUMHOLZ, L. R., AND ELSHAHED, M. S. Abundance, composition, diversity and novelty of soil Proteobacteria. *ISME J*, **3**(8):992–1000, 2009. doi: 10.1038/ismej.2009.43.

STAMATAKIS, A., LUDWIG, T., AND MEIER, H. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics (Oxford, England)*, **21**(4):456–463, 2005. doi: 10.1093/bioinformatics/bti191.

STRUVE, C. AND KROGFELT, K. A. Role of capsule in *Klebsiella pneumoniae* virulence: lack of correlation between in vitro and in vivo studies. *FEMS Microbiology Letters*, **218**(1):149–154, 2003. doi: 10.1111/j.1574-6968.2003.tb11511.x.

THOMAS, X., DESTOUMIEUX-GARZÓN, D., PEDUZZI, J., AFONSO, C., BLOND, A., BIRLIRAKIS, N., GOULARD, C., DUBOST, L., THAI, R., TABET, J.-C., AND REBUFFAT, S. Siderophore Peptide, a New Type of Post-translationally Modified Antibacterial Peptide with Potent Activity. *Journal of Biological Chemistry*, **279**(27):28233–28242, 2004. doi: 10.1074/jbc.M400228200.

TOMITA, K., OGAWA, T., UOZUMI, T., WATANABE, K., AND MASAKI, H. A cytotoxic ribonuclease which specifically cleaves four isoaccepting arginine tRNAs at their anticodon loops. *Proc Natl Acad Sci USA*, **97**(15):8278–8283, 2000.

TRAN, E. N. H., PAPADOPOULOS, M., AND MORONA, R. Relationship between O-antigen chain length and resistance to colicin E2 in Shigella flexneri. *Microbiology (Reading, England)*, **160**(Pt 3):589–601, 2014. doi: 10.1099/mic.0.074955-0.

TRAUTNER, B. W., HULL, R. A., AND DAROUICHE, R. O. Colicins prevent colonization of urinary catheters. *The Journal of antimicrobial chemotherapy*, **56**(2):413–415, 2005. doi: 10.1093/jac/dki228.

URATANI, Y. AND HOSHINO, T. Pyocin R1 inhibits active transport in *Pseudomonas aeruginosa* and depolarizes membrane potential. *Journal of Bacteriology*, **157**(2):632–636, 1984.

VAARA, M. Agents that increase the permeability of the outer membrane. *Microbiological Reviews*, **56**(3):395–411, 1992.

VAN DEN ELZEN, P. J., KONINGS, R. N., VELTKAMP, E., AND NIJKAMP, H. J. Transcription of bacteriocinogenic plasmid CloDF13 in vivo and in vitro: structure of the cloacin immunity operon. *Journal of Bacteriology*, **144**(2):579–591, 1980.

VAN DER DONK, W. A. AND NAIR, S. K. Structure and Mechanism of Lanthipeptide Biosynthetic Enzymes. *Current opinion in structural biology*, **0**:58–66, 2014. doi: 10.1016/j.sbi.2014.09.006.

VAN DER LEY, P., DE GRAAFF, P., AND TOMMASSEN, J. Shielding of *Escherichia coli* outer membrane proteins as receptors for bacteriophages and colicins by O-antigenic chains of lipopolysaccharide. *Journal of Bacteriology*, **168**(1):449–451, 1986.

VAN OPIJNEN, T. AND CAMILLI, A. Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nature reviews. Microbiology*, **11**(7):10.1038/nrmicro3033, 2013. doi: 10.1038/nrmicro3033.

VAN RAAY, K. AND KERR, B. Toxins go viral: phage-encoded lysis releases group B colicins. *Environmental Microbiology*, **18**(5):1308–1311, 2016. doi: 10.1111/1462-2920.13246.

VANDERWAL, F. J., LUIRINK, J., OUDEGA, B., VAN DER WAL, F. J., LUIRINK, J., OUDEGA, B., VANDERWAL, F. J., LUIRINK, J., OUDEGA, B., VAN DER WAL, F. J., LUIRINK, J., AND OUDEGA, B. Bacteriocin release proteins: mode of action, structure, and biotechnological application. *Fems Microbiology Reviews*, **17**(4):381–399, 1995. doi: 0168644595000224[pii].

VANKEMMELBEKE, M., O SHEA, P., JAMES, R., AND PENFOLD, C. N. Interaction of nuclease colicins with membranes: insertion depth correlates with bilayer perturbation. *PloS one*, **7**(9):e46656, 2012. doi: 10.1371/journal.pone.0046656.

VANKEMMELBEKE, M., ZHANG, Y., MOORE, G. R., KLEANTHOUS, C., PENFOLD, C. N., AND JAMES, R. Energy-dependent Immunity Protein Release during tol-dependent Nuclease Colicin Translocation. *J Biol Chem*, **284**(28):18932–18941, 2009. doi: 10.1074/jbc.M806149200.

VARCAMONTI, M., NICASTRO, G., VENEMA, G., AND KOK, J. Proteins of the lactococcin A secretion system: lcnD encodes two in-frame proteins. *FEMS Microbiol Lett*, **204**(2):259–263, 2001.

VETTER, I. R., PARKER, M. W., TUCKER, A. D., LAKEY, J. H., PATTUS, F., AND TSERNOGLOU, D. Crystal structure of a colicin N fragment suggests a model for toxicity. *Structure*, **6**(7):863–874, 1998.

VIEJO, M. B., GARGALLO, D., FERRER, S., ENFEDAQUE, J., AND REGUE, M. Cloning and DNA sequence analysis of a bacteriocin gene of *Serratia marcescens*. *Journal of general microbiology*, **138 Pt 8**:1737–1743, 1992. doi: 10.1099/00221287-138-8-1737.

VOGLER, A. J., KEIM, P., AND WAGNER, D. M. A review of methods for subtyping Yersinia pestis: From phenotypes to whole genome sequencing. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, **37**:21–36, 2016. doi: 10.1016/j.meegid.2015.10.024.

VOLLMER, W., PILSL, H., HANTKE, K., HÖLTJE, J. V., BRAUN, V., HOLTJE, J. V., AND BRAUN, V. Pesticin displays muramidase activity. *Journal of Bacteriology*, **179**(5):1580–1583, 1997.

VOULHOUX, R., BOS, M. P., GEURTSEN, J., MOLS, M., AND TOMMASSEN, J. Role of a highly conserved bacterial protein in outer membrane protein assembly. *Science*, **299**(5604):262–265, 2003. doi: 10.1126/science.1078973.

WALKER, D., LANCASTER, L., JAMES, R., AND KLEANTHOUS, C. Identification of the catalytic motif of the microbial ribosome inactivating cytotoxin colicin E3. *Protein science : a publication of the Protein Society*, **13**(6):1603–1611, 2004. doi: 10.1110/ps.04658504.

WALKER, D., MOSBAHI, K., VANKEMMELBEKE, M., JAMES, R., AND KLEANTHOUS, C. The role of electrostatics in colicin nuclease domain translocation into bacterial cells. *J Biol Chem*, **282**(43):31389–31397, 2007.

WANG, Y., HUANG, S. H., WASS, C. A., STINS, M. F., AND KIM, K. S. The gene locus yijP contributes to *Escherichia coli* K1 invasion of brain microvascular endothelial cells. *Infect Immun*, **67**(9):4751–4756, 1999.

WATERS, V. L. AND CROSA, J. H. Colicin V virulence plasmids. *Microbiological Reviews*, **55**(3):437–450, 1991.

WATTAM, A. R., ABRAHAM, D., DALAY, O., DISZ, T. L., DRISCOLL, T., GABBARD, J. L., GILLESPIE, J. J., GOUGH, R., HIX, D., KENYON, R., MACHI, D., MAO, C., NORDBERG, E. K., OLSON, R., OVERBEEK, R., PUSCH, G. D., SHUKLA, M., SCHULMAN, J., STEVENS, R. L., SULLIVAN, D. E., VONSTEIN, V., WARREN, A., WILL, R., WILSON, M. J. C., YOO, H. S., ZHANG, C., ZHANG, Y., AND SOBRAL, B. W. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Research*, **42**(Database issue):D581—-91, 2014. doi: 10.1093/nar/gkt1099.

WAYNE, R., FRICK, K., AND NEILANDS, J. B. Siderophore protection against colicins M, B, V, and Ia in *Escherichia coli*. *Journal of Bacteriology*, **126**(1):7–12, 1976.

WEISS, J., HUTZLER, M., AND KAO, L. Environmental modulation of lipopolysaccharide chain length alters the sensitivity of *Escherichia coli* to the neutrophil bactericidal/permeability-increasing protein. *Infection and Immunity*, **51**(2):594–599, 1986.

WEST, S. A. AND GARDNER, A. Altruism, spite, and greenbeards. *Science*, **327**(5971):1341–1344, 2010. doi: 10.1126/science.1178332.

WHITNEY, J. C., BECK, C. M., GOO, Y. A., RUSSELL, A. B., HARDING, B. N., DE LEON, J. A., CUNNINGHAM, D. A., TRAN, B. Q., LOW, D. A., GOODLETT, D. R., HAYES, C. S., AND MOUGOUS, J. D. Genetically distinct pathways guide effector export through the type VI secretion system. *Molecular Microbiology*, **92**(3):529–542, 2014. doi: 10.1111/mmi.12571.

WIEDEMANN, I., BREUKINK, E., VAN KRAAIJ, C., KUIPERS, O. P., BIERBAUM, G., DE KRUIJFF, B., AND SAHL, H. G. Specific binding of nisin to the peptidoglycan precursor lipid II combines pore formation and inhibition of cell wall biosynthesis for potent antibiotic activity. *J Biol Chem*, **276**(3):1772–1779, 2001. doi: 10.1074/jbc.M006770200.

WIENER, M., FREYMANN, D., GHOSH, P., AND STROUD, R. M. Crystal structure of colicin Ia. *Nature*, **385**(6615):461–464, 1997.

WILLETT, J. L., GUCINSKI, G. C., FATHERREE, J. P., LOW, D. A., AND HAYES, C. S. Contact-dependent growth inhibition toxins exploit multiple independent cell-entry pathways. *Proc Natl Acad Sci USA*, **112**(36):11341–11346, 2015. doi: 10.1073/pnas.1512124112.

WILLEY, J. M. AND VAN DER DONK, W. A. Lantibiotics: peptides of diverse structure and function. *Annu Rev Microbiol*, **61**:477–501, 2007. doi: 10.1146/annurev.micro.61.080706.093501.

WILLIAMS, N., ACKERMAN, S. H., AND COLEMAN, P. S. Benzophenone-ATP: a photoaffinity label for the active site of ATPases. *Methods in enzymology*, **126**:667–682, 1986.

WILLIAMS, S. R., GEBHART, D., MARTIN, D. W., AND SCHOLL, D. Retargeting R-Type Pyocins To Generate Novel Bactericidal Protein Complexes . *Applied and Environmental Microbiology*, **74**(12):3868–3876, 2008. doi: 10.1128/AEM.00141-08.

WILSON, E. O. Sociobiology: The New Synthesis. Harvard Press, Cambridge, MA, 1975.

YAJIMA, S., NAKANISHI, K., TAKAHASHI, K., OGAWA, T., HIDAKA, M., KEZUKA, Y., NONAKA, T., OHSAWA, K., AND MASAKI, H. Relation between tRNase activity and the structure of colicin D according to X-ray crystallography. *Biochemical and biophysical research communications*, **322**(3):966–73, 2004. doi: 10.1016/j.bbrc.2004.07.206.

YANG, S.-C., LIN, C.-H., SUNG, C. T., AND FANG, J.-Y. Antibacterial activities of bacteriocins: application in foods and pharmaceuticals. *Front Microbiol*, **5**(May):241, 2014. doi: 10.3389/fmicb.2014.00241.

YAO, G. W., DUARTE, I., LE, T. T., CARMODY, L., LIPUMA, J. J., YOUNG, R., AND GONZALEZ, C. F. A Broad-Host-Range Tailocin from *Burkholderia cenocepacia*. *Applied and Environmental Microbiology*, **83**(10), 2017. doi: 10.1128/AEM.03414-16. URL `http://aem.asm.org/content/83/10/e03414-16.abstract`.

YORGEY, P., LEE, J., KÖRDEL, J., VIVAS, E., WARNER, P., JEBARATNAM, D., AND KOLTER, R. Posttranslational modifications in microcin B17 define an additional class of DNA gyrase inhibitor. *Proc Natl Acad Sci USA*, **91**(10):4519–4523, 1994.

ZERBINO, D. R. AND BIRNEY, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*, **18**(5):821–829, 2008. doi: 10.1101/gr.074492.107.

ZETH, K., ROMER, C., PATZER, S. I., AND BRAUN, V. Crystal structure of colicin M, a novel phosphatase specifically imported by *Escherichia coli*. *J Biol Chem*, **283**(37):25324–25331, 2008.

ZHAI, Y. F., HEIJNE, W., AND SAIER JR., M. H. Molecular modeling of the bacterial outer membrane receptor energizer, ExbBD/TonB, based on homology with the flagellar motor, MotAB. *Biochimica et biophysica acta*, **1614**(2):201–210, 2003. doi: S0005273603001767[pii].

ZHANG, D., DE SOUZA, R. F., ANANTHARAMAN, V., IYER, L. M., AND ARAVIND, L. Polymorphic toxin systems: Comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. *Biology Direct*, **7**(1):18, 2012. doi: 10.1186/1745-6150-7-18.

ZHANG, D., IYER, L. M., AND ARAVIND, L. A novel immunity system for bacterial nucleic acid degrading toxins and its recruitment in various eukaryotic and DNA viral systems. *Nucleic acids research*, **39**(11):4532–52, 2011. doi: 10.1093/nar/gkr036.

CHAPTER 7

Appendix

PLOS | COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Diversity and distribution of nuclease bacteriocins in bacterial genomes revealed using Hidden Markov Models

Connor Sharp[1], James Bray[2], Nicholas G. Housden[1], Martin C. J. Maiden[2]*, Colin Kleanthous[1]*

1 Department of Biochemistry, University of Oxford, Oxford, United Kingdom, **2** Department of Zoology, University of Oxford, Oxford, United Kingdom

* colin.kleanthous@bioch.ox.ac.uk (CK); martin.maiden@zoo.ox.ac.uk (MCJM)

Check for updates

OPEN ACCESS

Citation: Sharp C, Bray J, Housden NG, Maiden MCJ, Kleanthous C (2017) Diversity and distribution of nuclease bacteriocins in bacterial genomes revealed using Hidden Markov Models. PLoS Comput Biol 13(7): e1005652. https://doi.org/10.1371/journal.pcbi.1005652

Editor: Christine A. Orengo, University College London, UNITED KINGDOM

Received: January 25, 2017

Accepted: June 21, 2017

Published: July 17, 2017

Copyright: © 2017 Sharp et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant sequence files are available from public repositories. Specifically, the rMLST database (https://pubmlst.org/rmlst/), DOI:10.1099/mic.0.055459-0, and the PATRIC2 database (ftp://ftp.patricbrc.org/patric2/), DOI:10.1093/nar/gkt1099.

Funding: This work was supported in part by a Wellcome Trust collaborative award to CK and MCJM (201505/Z/16/Z). CS was supported by the Life Sciences Interface Doctoral Training Centre in Oxford, funded by the Engineering & Physical

Abstract

Bacteria exploit an arsenal of antimicrobial peptides and proteins to compete with each other. Three main competition systems have been described: type six secretion systems (T6SS); contact dependent inhibition (CDI); and bacteriocins. Unlike T6SS and CDI systems, bacteriocins do not require contact between bacteria but are diffusible toxins released into the environment. Identified almost a century ago, our understanding of bacteriocin distribution and prevalence in bacterial populations remains poor. In the case of protein bacteriocins, this is because of high levels of sequence diversity and difficulties in distinguishing their killing domains from those of other competition systems. Here, we develop a robust bioinformatics pipeline exploiting Hidden Markov Models for the identification of nuclease bacteriocins (NBs) in bacteria of which, to-date, only a handful are known. NBs are large (>60 kDa) toxins that target nucleic acids (DNA, tRNA or rRNA) in the cytoplasm of susceptible bacteria, usually closely related to the producing organism. We identified >3000 NB genes located on plasmids or on the chromosome from 53 bacterial species distributed across different ecological niches, including human, animals, plants, and the environment. A newly identified NB predicted to be specific for *Pseudomonas aeruginosa* (pyocin Sn) was produced and shown to kill *P. aeruginosa* thereby validating our pipeline. Intriguingly, while the genes encoding the machinery needed for NB translocation across the cell envelope are widespread in Gram-negative bacteria, NBs are found exclusively in γ-proteobacteria. Similarity network analysis demonstrated that NBs fall into eight groups each with a distinct arrangement of protein domains involved in import. The only structural feature conserved across all groups was a sequence motif critical for cell-killing that is generally not found in bacteriocins targeting the periplasm, implying a specific role in translocating the nuclease to the cytoplasm. Finally, we demonstrate a significant association between nuclease colicins, NBs specific for *Escherichia coli*, and virulence factors, suggesting NBs play a role in infection processes, most likely by enabling pathogens to outcompete commensal bacteria.

Sciences Research Council. The public MLST database, sited at the University of Oxford, was funded by the Wellcome Trust and European Union. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

Bacteria deploy a variety of antimicrobials to kill competing bacteria. Nuclease bacteriocins are a miscellaneous group of protein toxins that target closely related species, cleaving nucleic acids in the cytoplasm. It has proved difficult to establish how widespread bacteriocins are in bacterial populations due to the high diversity of bacteriocin-encoding genes. Here, we describe an *in silico* approach to identify nuclease bacteriocin genes in bacterial genomes and to distinguish them from other competition toxins. Bacteria that contain nuclease bacteriocin genes are found in many different types of environment but are prevalent in niches where interbacterial competition is likely to be high. Nuclease bacteriocins are found exclusively in γ-proteobacteria and are particularly abundant in the *Enterobacteriaceae* and *Pseudomonadaceae* families. Although the sequences we identify are indeed diverse (<20% sequence identity between protein families) we show that all nuclease bacteriocins contain an invariant motif, usually within a common structural scaffold, that is implicated in translocating the cytotoxic nuclease to the cytoplasm. Finally, we show that nuclease bacteriocins in pathogenic *E. coli* are strongly associated with virulence factors suggesting they play a role in pathogenicity mechanisms.

Introduction

Bacteriocins were the first of the interbacterial competition systems to be discovered [1]. They take the form of antimicrobial peptides or proteins released to the environment that have activity against closely related organisms. Experimental and theoretical work has suggested that bacteriocins are important agents of competition between microbial communities [2–6]. The potent antimicrobial activity of bacteriocins has generated interest in their application in agriculture [7], the food industry [8], and as potential clinical therapies for bacterial infections [9, 10]; however, without understanding the distribution of bacteriocins, the significance of their role in bacterial competition remains uncertain and their exploitation consequently limited. Here, we circumvent one of the problems in identifying protein bacteriocins in bacterial genomes, their sequence and domain diversity, by developing a bioinformatics pipeline that identifies enzymatic nuclease bacteriocins (NBs). NBs are one of the major classes of protein bacteriocins. The new tools we have developed allowed us to determine NB species distribution and evaluate current ideas as to how such molecules are deployed by Gram-negative bacteria.

The best studied of the Gram-negative protein bacteriocins are the colicins produced by *E. coli*. Colicin gene expression is typically coupled to the SOS stress response and nutrient status of the cell [11, 12]. Despite varying in size, colicins have a common tripartite structure, composed of an N-terminal translocation (T-) domain, a central receptor binding (R-) domain and a C-terminal cytotoxic domain. Colicins deliver one of five cytotoxic activities into cells; a pore-forming ionophore or lipid II hydrolase are delivered to the periplasm whereas DNases, tRNases or rRNases are delivered to the cytoplasm [13]. Colicin-producing bacteria protect themselves through the action of a small immunity protein that in the case of nuclease colicins is co-expressed and released from cells bound tightly to its colicin [14–16]. Colicins enter a target cell by binding an outer membrane receptor, typically an iron or vitamin transporter, and then parasitizing another outer membrane protein as a translocation portal for its T-domain to contact one of two trans-periplasmic proton motive force (pmf)-coupled systems: Tol for group A (e.g. ColE9) or Ton for group B (e.g. ColD) colicins [13]. The immunity proteins of nuclease colicins are dissociated at the cell surface in a pmf-dependent step during

translocation [17]. Colicin-like NBs have been identified in pathogens such as *Klebsiella pneumoniae* (klebicins) and *Pseudomonas aeruginosa* (pyocins), all having a distinct multi-domain architecture and associated immunity protein [18–20].

Colicins (and colicin-like proteins) are thought to play a role in the co-existence of mixed microbial communities, and possibly as virulence determinants during proliferation within such communities [2, 21]. ExPEC bacteria, for example, show a significant association between virulence factors and colicins in *E. coli* isolates that cause bacteraemia [22, 23]. In a mouse gut infection model, colicin production by *Salmonella enterica* serovar Typhimurium is exploited by the organism to outcompete commensal *E. coli* during inflammation [24]. A recent survey of *Shigella sonnei* isolates in Vietnam spanning three decades identified a colicin-producing plasmid as one of the factors that overcame a dispersal bottleneck, helping the organism become established in the population as the major cause of dysentery [25]. Hence identifying the ecological niches in which colicins and colicin-like proteins are found and ascertaining their prevalence amongst pathogens are missing pieces of the microbial ecology puzzle that has ramifications for clinical epidemiology, biotechnology and biomedicine.

Multiple factors make NB genes difficult targets for *in silico* identification. First, NBs are often (but not always) plasmid-encoded and only present in a fraction of isolates [26]. Second, extensive sequence diversification and domain rearrangements hamper identification of NBs. Finally, the nuclease domains of NBs are similar to the cytotoxic domains of other polymorphic toxin families, such as contact-independent inhibition (CDI) systems, further complicating their identification. In this study, we establish methods for the identification of NBs from large bacterial genome datasets. To overcome the low sequence identity between NBs, we use profile hidden markov models to identify conserved motifs in NB cytotoxic domains and their associated immunity proteins. We find that NBs are present at differing levels of abundance in bacterial species and that organisms exploit them in many varied ecological niches. Our analysis, validated for a novel NB shown to be active against *P. aeruginosa*, reveals the extent of domain rearrangements and sequence diversification that has occurred in NB genes, identifies a critical motif implicated in NB translocation and associates NBs in bacterial pathogenesis.

Results and discussion

Constructing Hidden Markov Models for the identification of NBs in bacterial genomes

Protein bacteriocins are difficult to identify in bacterial genomes using conventional pairwise alignment strategies. We therefore developed a bioinformatics pipeline to identify bacteriocins using profile Hidden Markov models (HMMer 3.1) [27] centred on the catalytic regions of the nuclease cytotoxic domain of the bacteriocin. Pore-forming bacteriocins were excluded from the study due to the absence of similarly conserved motifs. Five different NB families have been described in the literature, two DNases and three RNases (Fig 1 and S1 Fig) and the profile pairs were based on conserved motifs identified in their catalytic centres. For example, HNH-type endonuclease bacteriocins have a 30-residue motif (also referred to as the ββα-Me motif) as their active site ($HHX_{14}NX_8HX_3H$) (Fig 1). Conserved motifs are similarly identifiable in the cytotoxic domains of non-HNH DNase, tRNase and rRNase NBs [28, 29] (S1 Fig). These motifs are however frequently found in other bacterial contexts; for example, the HNH motif is also found in Type II restriction and homing endonucleases, mismatch repair enzymes, and is the catalytic core of the CRISPR-Cas9 complex [30–32]. Hence three additional criteria were used to distinguish NBs from other systems and to validate them as *bona fide* NBs. First, the presence of conserved motifs within the adjoining immunity protein. Immunity proteins are specific, tight-binding inhibitors of bacteriocins that protect the

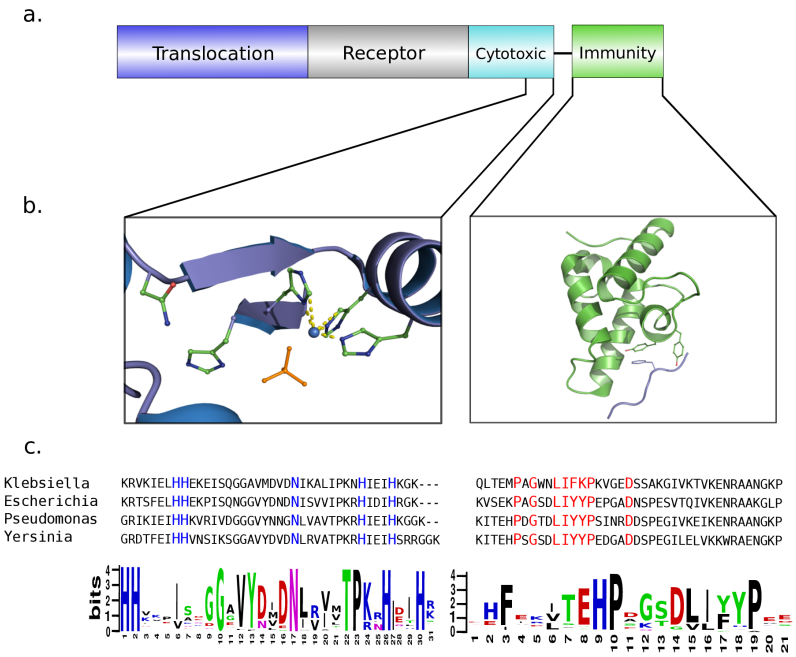


Fig 1. Identifying genetically linked conserved cytotoxic and immunity motifs is a powerful and accurate way to identify NB operons. ***a***, Gene/protein organisation of a typical nuclease bacteriocin from *E. coli*. ***b***, *Left-hand panel*, key interactions of conserved catalytic residues of the HNH motif of DNase bacteriocins. The two histidine residues of the HNH motif are involved in coordinating a divalent metal ion and the asparagine constrains the metal binding loop. The phosphate anion denotes the position of the scissile phosphate in substrate DNA (PDB code, 1V14 [34]). *Right-hand panel*, the helical immunity protein (*green*) showing the conserved aromatic residues of the α-helix III, which forms a critical part of the binding site for the DNase domain [15]. ***c***, Conserved residues used to form the HMM profile of each protein are highlighted in the sequence alignments.

https://doi.org/10.1371/journal.pcbi.1005652.g001

producing cell from the action of its own toxin [16]. In the case of NBs, immunity genes are co-expressed with the bacteriocin. The resulting heterodimeric complex is often released from the cell through the action of a bacteriocin-release protein or lysis protein, which is often encoded within bacteriocin operons [13]. In cases where a bacteriocin-release protein has not been identified latent prophage genes have been implicated in release [33]. NB-specific immunity proteins are only jettisoned during translocation of the toxin into a susceptible cell [17]. To increase the specificity of our searches we exploited the highly conserved genetic link between NB genes and their cognate immunity genes, creating 'profile pairs' where the immunity profile is within 60bps (or 200bps in the case of ColE5/pyocinS4 tRNase) downstream of the cytotoxic profile (S1 Table, S1 and S2 Figs). Second, the length of the NB sequence. NBs need multiple domains to translocate across the two membranes of the cell envelope and are typically >60 kDa in size. To be conservative, we filtered by sequence length, accepting sequences between 350–950 residues. Third, putative bacteriocins and the genes at the N-

terminus are analysed for the presence of domains readily identifiable through the PFAM database as linked to other polymorphic toxin systems are excluded; for example, TSS6 effector proteins, which are often NB-related nucleases, or hemagglutinin repeats that are associated with contact dependent inhibitors.

We subjected 100,154 PubMLST Multispecies isolate database, covering >3,000 species, to our pipeline and identified 3094 bacteriocin genes in 2479 isolates across 53 species (S3 Fig, S2 Table). The largest non-bacteriocin contaminants of the database were uropathogenic-specific protein (*usp*) from *E. coli* and various short DNase containing proteins predicted to be effectors of the T6SS. *E. coli usp* has 40–50% sequence identity to the HNH DNase cytotoxic domain of NBs and was identified in over 2000 *E. coli* and *Salmonella enterica* genomes [35]. They were not included in our analysis as Usp has been shown to have genotoxic effects towards mammalian not bacterial cells and also contain an N-terminal domain of unknown function (DUF769), predicted to be a member of the Hcp-1 proteins of T6SS [35, 36]. A group of HNH containing nucleases were also located in the Gram-positive *Bacillus* genera which are linked to the SUKH (Syd, US22, Knr4 homology) domain immunity protein superfamily that bind to a variety of bacterial toxins of diverse nuclease and nucleic acid deaminase families [37]. The HNH motif was dissimilar to the NB DNase motif, having an extended linker region between the final histidine residues, $HHX_{14}N_8HX_8H$. A non-HNH DNase T6SS effector was present in over half of *Klebsiella pneumoniae* strains. Other polymorphic toxin systems that share similar cytotoxic domains with NBs include the MAF proteins of *Neisseria* and a family of large T6SS effectors in the *Vibrio* genus that were identified by a LysM domain (PFAM 01476) and a recently described motif associated with type VI secretion, the MIX domain [38, 39]. Interestingly, we identified several bacterial species that utilize NBs as well as other competition systems such as T6SS, including *Pseudomonas aeruginosa*, *Escherichia coli* and *Yersinia pseudotuberculosis*. The dynamics of bacteria could utilise both diffusible and contact dependant competition systems is still unknown.

NBs are found in a variety of environments but are exclusive to γ-proteobacteria

We next looked for an association between NBs across different environmental niches. The PATRIC database contains 6256 genomes with associated environmental data [40]. The >1500 types of listed environment were grouped by an automated keyword search using custom python scripts. 120 unique environments were identified across 239 NB containing bacteria. Whilst it is not surprising that the majority of NB containing *E. coli* were identified in the GI tract and the majority of *Pseudomonas* species associated with either soil or plant based environments, we also observed these bacteria in the invasive environments of blood and urine including patients suffering septicemia (*P. aeruginosa* VRFPA01 and *P. aeruginosa* VRFPA02) or urinary tract infection (*E. coli* UMEA-3342-1) (S4 Fig).

NBs are found in significant proportions in the *Enterobacteriaceae* and *Pseudomonadaceae* families, however the proportion of a species that contained NB genes varied greatly; <5% *E. coli* strains encoded NBs compared to 85% for *Pseudomonas aeruginosa* and 31% for *Klebsiella pneumoniae*, which agrees with previous experimental work [41–43]. A prevalent feature of *Pseudomonas* and *Klebsiella* isolates is their possession of multiple NBs, which is likely to be a factor in NB evolution through increased rates of recombination, as has been proposed for pore-forming bacteriocins [41]. *Klebsiella pneumoniae* genomes commonly encode a DNase bacteriocin as well as cloacin, a rRNase bacteriocin. This trend has also been observed in *Pseudomonas* species [44]. A similar distribution is observed with *Yersinia mollaretii* genomes where 4 isolates contained both DNase and rRNase NBs.

An important aspect of bacteriocin biology is how bacterial genomes encode these toxic molecules since this will have an impact on their distribution in bacterial populations. NBs of *E. coli* (colicins) are known to be plasmid encoded whereas NBs of *P. aeruginosa* (pyocins) are located on the chromosome [13]. As part of our analysis we instigated a bioinformatics protocol for establishing whether NBs were chromosomal or plasmid-encoded as this has not been shown for many species containing NB-encoding genes. Distinguishing these gene locations is, however, challenging due to the differing levels of assembly of the genomes interrogated by our pipeline. Using the Plasmid finder database and the Carattoli typing scheme [45] we could identify plasmid replicons within NB containing contigs. Addition of a conserved region identified for pColE1 replicons [46] allowed the typing of additional plasmids including large tRNase containing plasmids in *E. coli* (S5a Fig). To demonstrate plasmid or chromosomal association of contigs containing NB genes for species not covered within this typing scheme, that could not be typed using the Plasmid Finder database, a python script was implemented to measure the percentage alignment of a contig to a database of known plasmids. As anticipated, NBs from *Escherichia* and *Pseudomonas* exhibited a strong association with plasmid and chromosomes, respectively (S5b Fig). NBs from *Yersinia* species, however, demonstrated little or no alignment to plasmid sequences, suggesting these are chromosomally-located. Interestingly, species from *Enterobacter* displayed a bimodal distribution suggesting their NBs could be plasmid encoded or chromosomal. For a small number of NBs, these predictions were confirmed using NCBI fully assembled genomes; both plasmid and chromosomal NB genes were identified for *Enterobacter* isolates using this approach. In total, chromosomal NBs were identified in four genera: *Pseudomonas*, *Serratia*, *Yersinia and Enterobacter*. In the fully assembled genome of *Klebsiella pneumoniae* KPNIH12, plasmid transfer genes and an NB gene (klebicin B) were incorporated into the genome and associated with a putative IS903 transposase. NBs with high sequence identity to that of *Enterobacter cloacae* (cloacin DF13; CloDF13) were identified in species across multiple genera. CloDF13 is a ~60 kDa NB that utilizes an rRNase cytotoxic domain. MAUVE 2.0 was used to identify a plasmid in seven species that included transfer and mobilisation genes with over 90% identity to pCloDF13, a broad range transmissible bacteriocinogenic plasmid [47]. pCloDF13 was identified in *Enterobacter cloacae*, *Enterobacter aerogenes*, *Escherichia coli*, *Klebsiella pneumoniae* and *Salmonella enterica*.

A surprising conclusion of our analysis is that NBs are found exclusively in γ-proteobacteria (Fig 2, S6 Fig). Indeed, only one enzymatic bacteriocin has been identified outside of the γ-proteobacteria, a colicin M like bacteriocin in the β-proteobacterium *Burkholderia* [48]. We note, however, that unlike NBs which translocate to the bacterial cytoplasm to cleave nucleic acids, colicin M kills bacteria by cleaving the lipid II precursor of peptidoglycan within the periplasm [49]. In trying to rationalise why NBs appear only to be part of the antimicrobial armoury of γ-proteobacteria we considered two possibilities, the environment in which these organisms typically reside and the machinery involved in their uptake. Environment can be dismissed as a cause of this exclusivity since α-, β- and ε-proteobacteria live in many of the same environments as γ-proteobacteria that contain NB genes; for example, NBs are abundant in soil dwelling γ-proteobacteria yet α-, β- and δ-proteobacteria are all more prevalent in soil [50]. The translocation machinery hijacked by NBs, the Ton and Tol systems, are found beyond the γ-proteobacteria (defined using a profile HMM strategy; S3 Table, S7 Fig), which suggests that the mechanism of NB import is also not a limiting factor in their species distribution. It remains unclear why NBs are exclusively found in γ-proteobacteria.

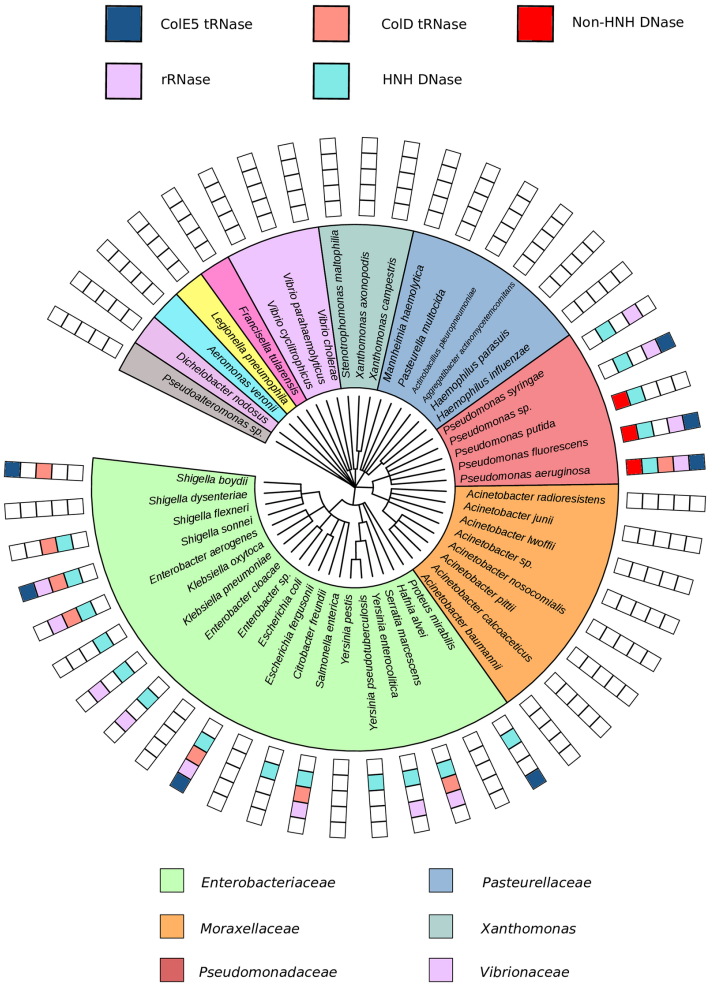


Fig 2. Distribution of NBs is restricted to the γ-proteobacteria. Taxonomic tree representing all γ-proteobacteria species in the pubMLST that have over 15 genomes, constructed using NCBI taxonomy commontree. The presence of different cytotoxic domains is indicated in the color key associated with each species. NBs are found throughout *Enterobacteriaceae* and *Pseudomonadaceae*.

https://doi.org/10.1371/journal.pcbi.1005652.g002

Sequence diversity and domain organisation of NBs

No systematic *in silico* analysis has been conducted to ascertain the diversity of bacteriocin protein sequences in bacteria. Domain organization in NBs has conventionally been based on that found in colicins and pyocins. Colicins are composed of an amino terminal T-domain, a central R-domain and a C-terminal cytotoxic domain (T-R-C), while pyocins are thought to have their T- and R- domains switched (R-T-C). Rather than assign functionality to domains, which is largely unknown in our dataset, we analyzed the NB protein sequences identified through our bioinformatics pipeline for structural elements/motifs commonly associated with NBs, including intrinsically disordered and coiled-coil regions and structured domains described in the PFAM database (Fig 3). For this analysis, the cytotoxic nucleases of the sequences, which are always at the C-termini of NBs, were excluded as they provide little information on domains involved in import. Our approach captured a large proportion of previously identified NBs reported in the literature thereby validating the pipeline. Our analysis revealed five principles that underpin the diversity of NB sequences involved in their import. First, with the exception of some *Pseudomonas* species, (which appear too diverse for meaningful analysis by CLANS therefore we present a phylogenetic analysis in S8 Fig) (see below), the majority of NB sequences (2866: ~93%) can be classified into eight groups (I-VIII) based on sequence identity and predicted domain organization. Second, six of the eight groups come from the *Enterobacteriaceae* (groups I, II, IV, V, VI and VII) and two from *P. aeruginosa* (groups III and VIII). Third, groups are generally not comprised of sequences from a single genus (e.g. group V includes *E. coli*, *Serratia marescens*, *Salmonella enterica* and *Klebsiella pneumoniae*) but can be populated entirely by a single genus (e.g. group VI contains only *Salmonella enterica* and group IV contains only *Yersinia* spp NBs). Conversely, some species appear in multiple groups (e.g. klebicins of *K. pneumoniae* are found in groups V and VII). Fourth, the degree of sequence identity within groups is high (40–90%) but is very low between groups (<20%). Fifth, only one sequence motif was common to all NBs across the eight groups, identified as the DPY motif in Fig 3 (see below). Four of the eight groups have duplicated DPY motifs (groups II, IV, V and VI) although the biological relevance of this is currently unknown. Below, we highlight some of the main structural features of NBs for a select few species that encode them.

E. coli NBs of the rRNase and DNase type were identified in the 1970s [54, 55] and are commonly referred to as E-type colicins [16]. The present analysis places these NBs in group I, which also includes NBs from *Klebsiella pneumoniae*, *Serratia marcescens*, *Shigella sonnei*, *Enterobacter cloacae*, *Xenorhabdus*, and *Citrobacter* (Fig 3). Structures for both the rRNase NB ColE3 and DNase NB ColE9 from this group have been reported [56], plus there have been numerous biophysical and structural studies of other NBs from this group [57]. The main structural features of group I NBs are a T-domain composed of a disordered N-terminal region adjoining a folded domain, the latter identified in the PFAM database as PFAM 03515, and a central coiled-coil region which constitutes the R-domain. In E-type colicins the R-domain binds the vitamin B_{12} transporter BtuB, while disordered regions (known as the Intrinsically Unstructured Translocation Domain or IUTD) are involved in binding OmpF in the outer membrane and TolB in the periplasm as part of their import mechanism [58, 59]. Group I NBs (apart from those within *Morganella*) have disordered regions at their N-terminus, which is also seen in five of the eight NB groups (Fig 3). It is likely these disordered regions are involved in the translocation mechanism of all NBs that contain them. Once in the periplasm the nuclease domains of NBs are translocated across the cytoplasmic membrane by the AAA+ ATPase/protease FtsH, which also proteolytically releases the nuclease to the cytosol [60, 61]. The

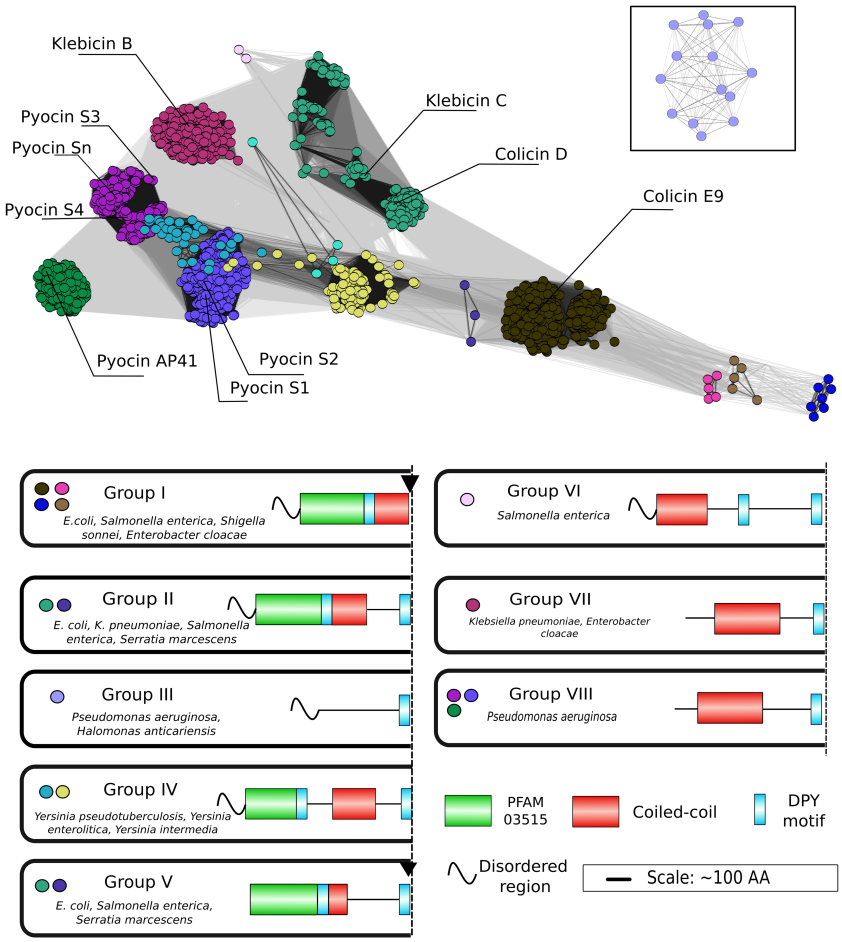


Fig 3. NBs cluster into 8 groups of differing domain arrangements. *Top panel*, sequences of NBs from *Enterobacteriaceae* and *Pseudomonas aeruginosa* had cytotoxic domains removed before clustering using CLANS [51]. *Bottom panel*, PFAM profiles, coiled-coil regions, disordered regions and the conserved DPY motif in the most common domain organisations observed for each species. Sequences within clades tend to share a similar predicted domain structure. Block arrows indicate proteolytic processing sites that have been identified for a select few NBs [52, 53]. Box inset contains pyocin S9 sequences which were too distantly removed to show at the correct distance. Connections show High Scoring Segment Pairs (HSP) of $<5x10^{-20}$. The figure also highlights previously identified NBs as reference points. In addition to identifying many more NBs within the groups containing such characterized NBs we also identify three NB groups (III, IV and VI) that have completely novel domain organisations not previously described in the literature.

https://doi.org/10.1371/journal.pcbi.1005652.g003

processing site identified by de Zamaroczy and co-workers [52] can be found in the NBs of groups I and V.

K. pneumoniae is an important human pathogen and a leading cause of drug resistant infections. Multiple NBs, including four well characterized klebicins (A-D), have been identified in *K. pneumoniae* isolates [62]. We identified klebicins A (equivalent to cloacin DF13, group I), klebicin B (group VII) and klebicin C (group II), as well as two novel DNases which share less than 50% similarity to klebicin B or C at the N-terminus (group VII). Sequences within group VII contained a large coiled-coil region at the N-terminus whereas group II klebicins have much shorter coiled-coil regions similar to group V NBs, which includes the well-studied colicin D [29, 63]. Group II and group V NBs only differ in the presence of N-terminal disorder and overlap in terms of their sequence similarity clustering. The klebicins of group II contain two copies of the DPY motif whereas group VII NBs only have one.

Serratia marcescens is a ubiquitous environmental organism that has become a major cause of healthcare associated infections, with increasing reports of antimicrobial resistance [64, 65]. Only one *Serratia* bacteriocin (the pore forming marcescin 28b) has been identified to-date [66]. Using our pipeline, 11 NB families with no significant similarity to 28b were identified in 44 *Serratia* isolates. Bacteriocins containing either DNase and tRNase cytotoxic domains appeared to be interchangeable suggesting recombination of cytotoxic domains, a trait that is common in polymorphic toxins [67]. 8 of the 11 clusters were predicted to be chromosomally encoded and are present in group V. Three of the NBs are more closely related to cloacin DF13, containing an identifiable N-terminal T-domain and 42% similarity to the T- and R-domains of cloacin DF13 (group I), but have a long linker region and a repeat of the DPY motif and therefore form a separate group (group II).

The genus *Yersinia* has previously been shown to contain a number of bactericidal toxins only two of which have been characterized as bacteriocins. Pesticins from *Yersinia pestis* degrade peptidoglycan using murimidase activity and colicin F_Y is a pore forming bacteriocin [68, 69]. We identified 18 NB sequence clusters at 90% sequence identity in 8 *Yersinia* species and identified both DNase and rRNase cytotoxic motifs. None of these sequences had significant similarity to pesticin or colicin F_Y. *Yersinia* NBs formed a separate group within the *Enterobacteriaceae*, the group IV NBs, although somewhat unexpectedly this did not include *Y. pestis*. These sequences have a ~90 residue unstructured region at the N-terminus similar to the IUTD of colicins and a long coiled-coil region, likely to be the receptor binding domain. 15 *Yersinia* NBs had a conserved region within the disordered N-terminal domain suggesting they may target the same translocation machinery.

Pseudomonas aeruginosa is an opportunistic pathogen and a leading cause of nosocomial infections [70, 71]. The S-type pyocins of *Pseudomonas aeruginosa* share a similar domain structure to colicins but with a rearrangement of their T- and R- domains and a small domain of unknown function between them [42]. An *in silico* investigation into the diversity of *Pseudomonas* toxins has previously identified several novel S-type pyocins [44]. Similar to this earlier study, we identified pyocins S1-9 (excluding the pore former pyocin S5) and S11-12, as well as two additional recombinations; pyocin S13 (or SD1) which contains the colicin D like cytotoxic domain of S11 and S12 associated with the N-terminus of S1 [72], and a pyocin we term S3C which contains the rRNase domain of colicin E3 associated with the pyocin S3 N-terminal domains and has not been reported previously in literature (Fig 3). Two novel HNH DNase motifs were also identified with 40–60% similarity to the S1, S2 and AP41 cytotoxic domains. Overall, 85% (874/1024 genomes) *P. aeruginosa* strains contained NB genes, which largely segregate to a single group in our analysis (group VIII). A second, much smaller, group (group III) is populated only by a single pyocin (S9), also identified by Ghequire et al [44]. These

proteins are predicted to have a disordered N-terminal domain, a single DPY motif and no identifiable coiled-coil region.

NBs were identified in 20 *Pseudomonas* species other than *P. aeruginosa*. NBs containing an rRNase domain were identified in *P. fluorescens*, *P. synxantha*, *P. poae*, *P. brassicacearum* and 10 *Pseudomonas* sp. isolates. In total, 175 *Pseudomonas* spp. (excluding *P. aeruginosa*) genomes contain a DNase or E3-type rRNase domain (51.5% and 21%, respectively). NBs with a tRNase cytotoxic domain were not identified outside of *P. aeruginosa* species which is in line with previous analysis [44] (S8 Fig). Many of the species observed were represented in the database by only a single genome suggesting the true extent of the sequence diversity within *Pseudomonas* spp. has yet to be established. Ghequire *et al* [44] also reported the presence of a group of NBs that contain a dual tandem repeat of the so-called pyocin_s domain (see below), followed by a DNase domain (either HNH or non-HNH) in 10 isolates. We observed a similar domain organisation but note that the additional DNase domain does not contain a functional HNH motif.

The DPY motif—A conserved epitope implicated in NB translocation to the cytoplasm

We analyzed the 3094 NB sequences captured by our bioinformatics pipeline using MEME [73] and identified a ~15-16-residue motif (D-X_4-FP-X_8-Y) located within the T-domain of nuclease colicins, which we refer to as the DPY motif (Fig 4a). T-domains of NBs such as

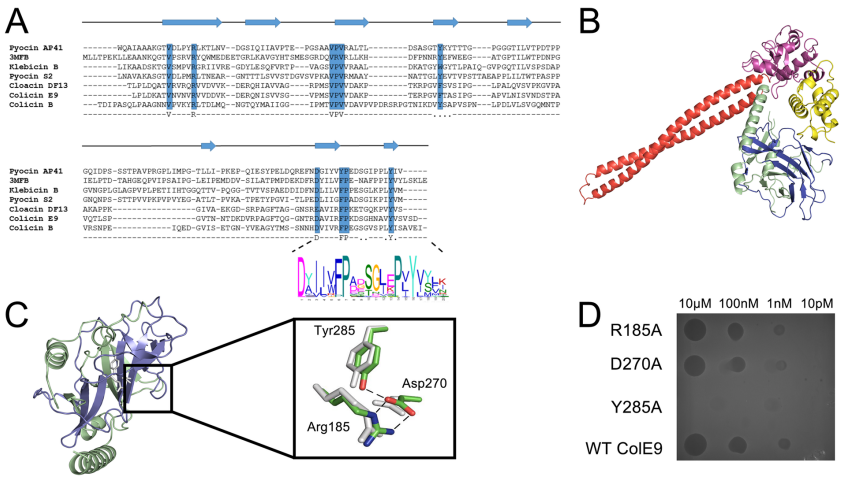


Fig 4. A conserved translocation motif was identified in all NBs. ***a***, Protein structure-based sequence alignment using PROMALS 3D indicates the conserved β-sheet secondary structure of a conserved domain identified in nuclease bacteriocins. Alignment features bacteriocins from *E. coli* (Colicin B, E9 and Cloacin DF13); *Klebsiella pneumoniae* (klebicin B), *Pseudomonas aeruginosa* (Pyocin AP41) and the pyocin_s domain from *Erwinia carotovora* The DPY motif was identified using MEME and is shown by a LOGO plot at the C-terminal end of the T-domain. ***b***, Crystal structure of colicin E9, with its constituent domains identified, in complex with its immunity protein Im9 (PDB code, 5EW5). ***c***, The conserved segment of the T-domain (*blue*) is annotated as the pyocin_S domain in the PFAM database (PFAM 06958), which is usually part of a larger T-domain, annotated as PFAM 03515 (*green*). *Inset*, Alignment of resides at the core of PFAM 06958 showing a conserved hydrogen bond network formed between the residues of the DPY motif; Asp270 and Tyr285 (colicin E9 numbering) and Arg185 at the beginning of PFAM 06958. ***d***, Cytotoxic plate killing assay of DPY motif mutations. 100-fold serial dilutions of colicin E9 and DPY motif alanine mutants were spotted onto a lawn of sensitive *E. coli* showing that only the Tyr285Ala mutant abolishes colicin activity.

https://doi.org/10.1371/journal.pcbi.1005652.g004

colicins are typically ~300 amino acids in length and share very low sequence identity (<15%), with many loop insertions/deletions between the β-sheet secondary structure elements. The DPY motif is found in all NB sequences we examined and as such is a clear identifier of T-domains, the first of its kind. Sano *et al*. [20] previously highlighted a similar region in a few nuclease pyocins and colicins the deletion of which abolished cytotoxic activity [74]. The structures of three NB T-domains are known (annotated in the PFAM database as PFAM 03515); colicin E3 (2B5U), colicin E9 (5EW5 [56]) and the S-type pyocin domain from *Erwinia carotovora* (3MFB) as well as the T-domain of the pore forming colicin B (1RH1) which shares high sequence identity to the T-domain of the NB colicin D. We performed structure-based sequence alignments and found that the DPY motif is integral to a much larger segment spanning the C-terminal half (~140 amino acids) of the T-domain (Fig 4b). This segment (coloured blue in Fig 4b) is identified in the PFAM database as the pyocin_s domain (PFAM 06958) [42]. Nine amino acids identified within PFAM 06958 are conserved across the T-domains of all NBs, which are coincident with five β-strands of the structure (Fig 4a). Finally, it is interesting to note that the location and indeed number of DPY motifs within NBs varies; it is generally placed at the C-terminal end of T-domains, which can be proximal or distal to the cytotoxic domain of NBs (Fig 3).

The DPY motif forms a buried hydrogen bond network within PFAM 06958 that in colicin E9 involves Tyr285 and Asp270, which is salt-bridged to Arg185 at the N-terminal end of PFAM 06958 (Fig 4c). We mutated these three residues to alanine separately and in combination and found that mutation of Tyr285 (by itself or in combination) abolished cell killing whereas Asp270 and Arg185 mutations had no observable effect (Fig 4d).

With the exception of colicin B, a pore-forming colicin that shares a common ancestry with the nuclease colicin D, the DPY motif is found exclusively in NBs. The motif is not present in bacteriocins that are active in the periplasm such as the pore-forming colicins E1, A, Ia, Ib and pyocin S5 and lipid II hydrolases such as colicin M. Hence, the DPY motif is specific to bacteriocins active in the cytoplasm. The involvement of the motif in translocation to the cytoplasm could be either at the outer membrane step, defining a path through the cell envelope, or the inner membrane, in a step prior to FtsH-mediated translocation of the nuclease to the cytoplasm.

Testing the activity of a novel nuclease bacteriocin

The ability of the pipeline to predict novel bacteriocins could improve our repertoire of potential species-specific antimicrobials for a number of clinically relevant species. As a test of the pipeline, we cloned and expressed a novel bacteriocin from *Pseudomonas aeruginosa*, which we call pyocin Sn. Pyocin Sn (832 residues) resides within group VIII, has no detectable disorder domain, is characterized by a large coiled-coil region and a C-terminal, non-HNH type DNase domain. Purified pyocin Sn had cytotoxic activity against three different strains of *P. aeruginosa* (S9 Fig). Hence our pipeline accurately identifies novel NBs in bacterial genomes.

Association of NBs with virulence factors in *Escherichia coli*

Recent work has highlighted the potential role of colicins as virulence factors in pathogenic bacteria, which can be used to displace commensal bacteria [24, 25]. As yet however there has been no assessment of whether specific bacteriocins such as colicins are associated with pathogenicity. To look for an association between colicinogenicity and virulence we calculated the pangenome of 357 *E. coli* strains (166 colicinogenic strains and 191 non-colicinogenic strains) using the roary pangenome. 41,723 gene groups were identified (1905 core genes

Table 1. Gene clusters identified by Roary pangenome software and shown as significantly associated with nuclease colicins.

Cluster	Description
Virulence factors	
yijP_3	Uncharacterized protein. Shown to be involved in invasion of brain endothelial cells [75]
stxA/B	Shigella toxin/antitoxin
terW/E	Tellurium resistance protein
ureC	Urease
hyl A/B/C/D	Hemolysin operon
efa1	*Escherichia coli* factor for adherence
O-antigen biosynthesis	
fnlB	UDP-l-FucNAc biosynthesis pathway [76]
fnl1	UDP-l-FucNAc biosynthesis pathway
fnl3	UDP-l-FucNAc biosynthesis pathway
wbuB	l-FucNAc transferase
Phage	
P	Phage terminase
X	Phage tail
S	Phage tail

https://doi.org/10.1371/journal.pcbi.1005652.t001

(100–99%), 739 soft core genes (99–95%), 3,500 shell genes (95–15%) and 35,579 cloud genes (15–0%)). Association was determined by a fisher's test followed by a Cochran-Mantel-Haenszel on BAPS clustered populations to correct for population stratification. 28 genes (excluding genes that code for a colicin, immunity protein or plasmid mobility genes) were found to be significantly associated with the trait of colicinogenicity (Table 1). These genes included virulence factors, toxin/anti-toxin modules, phages and genes involved in LPS and O-antigen biogenesis (Fig 5). Specific virulence factors are associated with pathotypes of colicinogenic *E. coli* such as the shigella toxin producing *E. coli* STEC and the enterohemorrhagic *E. coli* factor for adherence. Other genes associated with the carriage of colicin genes are general virulence factors such as hemolysins, which target the eukaryotic plasma membrane, and ureases that increase cytoplasmic pH to allow *E. coli* to survive in acidic environments. We also identify an association with phage genes suggesting either horizontal transfer of NB genes via bacteriophages or that associated phage lysis genes are being recruited for release of NBs, as has been proposed for other NBs [33]. The association of NBs with virulence factors suggests a role in the pathogenicity mechanisms of *E. coli*. This *in silico* evidence supports the hypothesis that bacteria use bacteriocins to displace native microflora in order to colonize and cause infection.

In summary, we have developed and validated a pipeline for the identification of NB genes in bacterial genomes. Using this pipeline, we are able to show that the five NB families are widespread in γ-proteobacteria and that bacteria containing these genes occupy diverse ecological niches. Most have yet to be characterized. We also show that while NBs have high sequence diversity and multiple domain arrangements, they retain a conserved motif that is likely required for translocation of the nuclease to the cytoplasm. Finally, using a pangenome analysis of *E. coli* isolates we show that NB genes are associated with virulence factors supporting the hypothesis that NBs are exploited by invasive bacteria to displace host microflora.

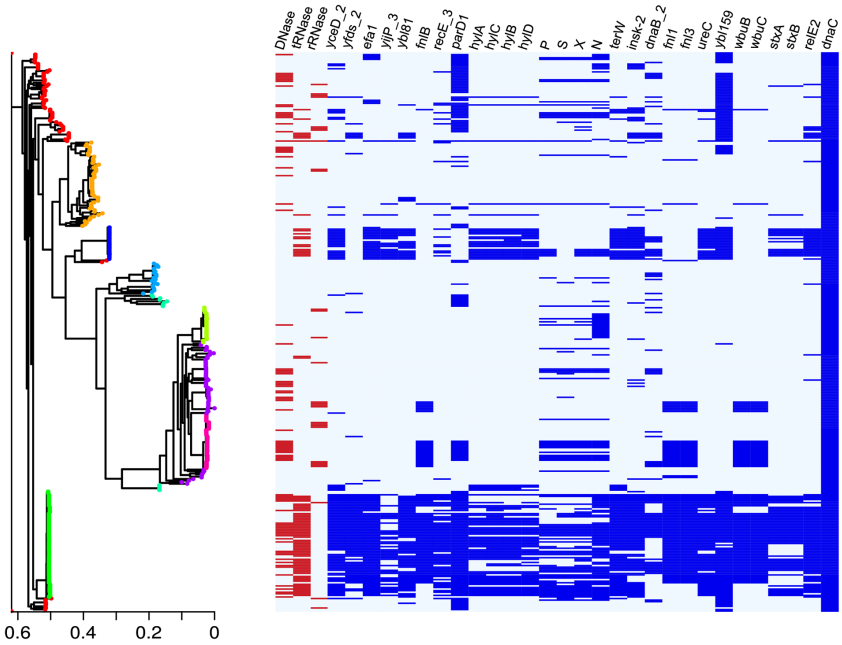


Fig 5. Pangenome analysis of colicinogenic bacteria shows evidence of an association between NBs and virulence factors. Association of pathogenicity and colicinogenicity genes based on a Cochran-Mantel-haenszel test. *Left-hand panel*, RAxML tree of a core genome alignment. Population structure was calculated using BAPS and tree nodes are coloured by cluster as predicted by BAPS. *Right-hand panel*, presence and absence of genes associated with colicinogenicity. Red genes show nuclease colicin of different cytotoxic domains. dnaC is a core gene and included as a control.

https://doi.org/10.1371/journal.pcbi.1005652.g005

Materials and methods

Genomic databases

Nucleotide fasta files for genomes from PATRIC [40] and NCBI assembled bacteria are freely available from ftp.patricbrc.org/patric2 (accessed 11/07/2014) and ftp.ncbi.nlm.nih.gov/genomes/ASSEMBLY_BACTERIA (accessed 14/05/2015), respectively. We also made use of the PubMLST Multispecies isolate database website (https://pubmlst.org/rmlst/ accessed 10/12/2014) [77]. All databases were subjected to the nuclease bacteriocin pipeline however due to the size and coverage of the pubMLST database (10-fold larger then PATRIC) results are only shown for pubMLST. PATRIC was used to calculate environmental association as it has greater metadata and NCBI used to identify Ton/Tol operons as genomes are assembled. PATRIC and NCBI databases did not produce any novel NBs that were not identified in pubMLST. The PubMLST Multispecies isolate database contains NCBI Nucleotide database (finished genomes) and community contributed NCBI assembled genomes (NCBI Assembly database) as well as genomes assembled in-house from the publically available raw read data at the European Nucleotide Archive (ENA-SRA). The PubMLST Multispecies isolate database

aims to collect a wide range of bacterial species from all bacterial Classes that have been NGS sequenced in order to reflect the current knowledge of bacterial genomes and does not explicitly bias itself towards any particular subset of bacteria; however, it is subject to any bias of sequences deposited in the ENA. PubMLST Multispecies isolate database genomes are assembled in-house from the ENA-SRA database. The short read sequences were assembled using the Velvet genome assembly program (v1.2.08) [78]. All odd-numbered kmer lengths between 21 and the read length were sampled using the VelvetOptimiser software (v2.2.4, bioinformatics.net.au/software.velvetoptimiser.shtml) to automatically calculate the optimal assembly parameters for Velvet. Default parameters were used throughout with the exception that no scaffolding was performed and only contigs with 200bp or more were included in the final assembly. Data from pubMLST used in this study including source (NCBI or ENA-SRA) and accession within source database is provided (S1 Data).

Developing a pipeline to detect nuclease NBs and associated immunity proteins

Profile pairs shown in S1 Table were used to search for homologues of a NB and immunity protein in each of the translations using HMMscan (HMMer3.1) with default search parameters (http://hmmer.org/). PFAM profiles were used for all five cytotoxic domains, HNH DNase, the non-HNH DNase pyocin S3 cytotoxic domain, tRNase (both Colicin D-like and Colicin E5-like) and rRNase [79]. To test for sensitivity, NBs identified in a first iteration of the pipeline were used to make secondary profile pairs (iterated HMM). Regions that matched the conserved motifs were extracted using the easel toolkit associated with HMMer3.1. Sequence alignments were performed using either MUSCLE [80] or CLUSTAL Omega [81].

Genomes from all databases were submitted through the pipeline as shown in S3 Fig. Briefly, genomes were translated into six frames. Each frame was aligned to the profile pairs in S1 Table using HMMer3.1 HMMscan with default parameters E-value cut off $5x10^{-2}$. Custom python scripts were used to find the intergenic distance, here defined as the bases between the end of the cytotoxic profile alignment and the start of the immunity profile alignment. We do not measure the actual intergenic region between two open reading frames, as ORFs for immunity genes are often not identified by annotation software. For all cytotoxic domains apart from the ColE5 tRNase profile pairs with intergenic distances of greater than 60 base pairs were discarded. For the ColE5 like tRNase cytotoxic domain the distance was extended to 200bp as the profiles did not cover the end of the cytotoxic domain and the beginning of the immunity leading to an extended region between the two (S2 Fig). Open reading frames containing the cytotoxic and immunity pairs were extracted using Prodigal v2.6.1. ORFs greater than 950 residues and less than 350 residues were discarded as non-NB ORFs [13]. Sequences were aligned to the PFAM-A version 27 profile database. ORFs with significant alignments (E-value $<5x10^{-5}$) to active secretion systems or profiles not biologically relevant to NBs were rejected. To correct for incorrectly predicted ORFs we cluster sequences using cd-hit to find the longest ORF prediction for each sequence. Two additional profiles were used to discriminate T6SS from *Vibrio* and *Klebsiella*. A profile built containing the recently described MIX domain of Vibrio T6SS. A second profile was generated to remove T6SS effectors from *Klebsiella pneumoniae* which were located downstream of a T6SS Rhs protein.

Determining chromosomal or plasmid association of NB genes

Conserved replicon sequences from the Carattoli typing scheme were downloaded from Plasmid Finder [45]. In addition, a conserved region from the ColE1 replicon were added [45]. An E-value cut off of $5x10^{-5}$ was used. Sequenced plasmids from NCBI and EBI were downloaded

to generate a custom blast database of 3603 plasmids. Contigs containing a suspected NB were aligned to the database using blastn with default settings. Regions of the contig that aligned to a plasmid (E value $< 5x10^{-7}$) were summed and overlaps removed to determine the percentage of plasmid association. A small number of contigs could be verified using the NCBI fully assembled genomes that contain fully assembled chromosomes and plasmids. To detect horizontal gene transfer by colicinogenic plasmids found across multiple genomes, sequences were clustered using cd-hit with sequence identity of 90% and genome comparisons were performed using MAUVE2 [82].

Identifying Ton and Tol operons in a diverse set of bacterial genomes

2785 fully assembled bacterial genomes with open reading frames predicted using Glimmer3.0 were downloaded from NCBI. S3 Table describes the PFAM profiles (Pfam 27) used to identify conserved regions within proteins of the Ton/Tol operons. HMMscan (HMMer3.1) was again used to identify profile matches in the open reading frames (E-value cutoff $5x10^{-5}$). Within each genome, regions of interest were identified by clustering co-ordinates of profile matches along the chromosome with a threshold of 2000 bases forming a new cluster. To differentiate between Ton, Tol and Mot operons, simple presence/absence rules were applied. For a region of interest to be classified as a Tol operon it needed to contain ExbD, MotA_ExbB and PD40. Ton operons were defined as containing ExbD, MotA_ExbB, TonB_C but not the PD40 repeat which is a structural motif within TolB that does not have a homologue in the Ton operon.

Clustering NB proteins by CLANS analysis

Cytotoxic domains were removed from bacteriocin sequences by removing the section of sequence that overlapped with the Pfam cytotoxic domains using HMMsearch and the easl toolkit. Sequences with cytotoxic domains removed were clustered using CLANS. Colored clusters were calculated using network approach where each sequence emits the clusters it is linked to weighted by the negative log of the p-value. This is iterated until clusters experience no further change. Only HSP hits $p<1x10^{-10}$ were used in the analysis [51]. Coiled-coil regions were predicted using COILS [83]. Disorder was predicted using IUPRED [84]. Disordered regions were defined as predictions of long disorder greater than 0.5 stretching for $\geq$ 30 residues.

Pangenome analysis

357 *E. coli* genomes were accessed from the ENA-SRA (S2 Data) and assembled using a Velvet based in—house assembly and improvement pipeline before annotation using Prokka [85]. Pangenome analysis was performed using the Roary pangenome software [86] and with similarity cut-off set at 95%. Phylogenetic trees were calculated from the core gene alignment produced by Roary using RAxML (WAG model with Gamma correction) [87]. The R stats package was used to perform the analysis. Gene clusters were tested for association with colicinogenic genomes by a fisher's exact test. Population stratification was evaluated using BAPS [88] clustering and a Cochran-Mantel-haenzsel test.

Prediction of PFAM domains

PFAM (Pfam 27) domains from were predicted using HMMscan (HMMer3.1).

Cytotoxic plate killing assay

Killing assays were performed by inoculating soft LB agar with either colicin sensitive *E. coli* JM83 for DPY mutations or one of three *P. aeruginosa* strains (PAO1, UCBPP-PA14 or PA14) for pyocin Sn sensitivity, and overlaid on LB agar. Serial dilutions were prepared using 20mM Tris-HCl pH 7 and 2 μl spotted onto the agar. Plates were incubated overnight at 37˚C and scored on presence of a zone of clearance in the soft LB agar.

Cloning and expression of the novel bacteriocin Pyocin Sn

Codon optimised synthetic genes encoding pyocin Sn (GenBank: AHA26272.1) and its immunity protein in series (Eurofins Genomics) were ligated into the NdeI / XhoI sites of pET24a (Novagen) to give pNGH252 such that the immunity protein contained a C-terminal hexa-histidine tag. A second copy of the immunity protein, again with a C-terminal hexa-histidine tag, was ligated into the NcoI / HindIII sites of pACYCDuet-1 (Novagen), to give pNGH260. Both plasmids were transformed into BL21 (DE3) cells to ensure the immunity protein expressed in excess of the bacteriocin. Cultures of pNGH252 pNGH260 BL21 (DE3) were grown at 28˚C in LB 50 μg/ml kanamycin, 34 μg/ml chloramphenicol to an OD600 nm of 0.7 upon which expression from both plasmids was induced through the addition of IPTG to a final concentration of 1 mM. Cells were grown for a further 16 hours, before being harvested by centrifugation, and lysed through sonication in 25 mM Tris-HCl, pH 7.5, 500 mM NaCl, and 1 mM PMSF. Cell debris was removed through centrifugation at 17,500 xg with the supernatant being passed through a 0.45 μm filter. The pyocin Sn-Im_{His6} complex was purified on a 5 ml HisTrap FF column (GE Healthcare) equilibrated in 25 mM Tris-HCl, pH 7.5, 150 mM NaCl eluting bound protein with a 0 to 250 mM imidazole gradient over 10 column volumes. The pyocin Sn-Im_{His6} complex was further purified on a HiLoad 26/60 Superdex 200 column (GE Healthcare) equilibrated in 25 mM Tris-HCl, pH 7.5, 150 mM NaCl. The purified complex was quantified through A280 nm using a sequence based extinction coefficient of 70,250 $M^{-1}.cm^{-1}$.

Supporting information

S1 Fig. Cytotoxic domains targeted by the pipeline. Multiple sequence alignments of cytotoxic domains analysed by the pipeline showing the conserved motifs of both the toxin nuclease domain and immunity proteins identified the HMM profiles. In total, five profile pairs, three for RNases and two for DNases, were created that captured all known NB types. Fig 1 (main text) shows the profile pair for HNH DNases and their immunity proteins.
(TIF)

S2 Fig. Distribution of intergenic regions between the cytotoxic HMM profiles and their associated HMM profiles of downstream immunity genes. Profile pairs with an intergenic region greater than 60bp (or 200bp for the case of ColE5 tRNases) were discounted from the analysis. a) Distribution for the HNH DNase and associated immunity gene. b) ColD tRNase, c) Non-HNH DNase, d) ColE5 tRNase, e) rRNase.
(TIF)

S3 Fig. Pipeline for identifying and validating bacteriocin/immunity profile pairs. Numbers shown indicate the number of pairs accepted or rejected at each step for the pubMLST database using all 5 PFAM bacteriocin/immunity profile pairs.
(TIF)

S4 Fig. The proportion of bacteria from each environment that were found to contain a NB. The PATRIC database contains metadata on the isolate environment of ~10% of bacteria. Environments are split to show contributions of the three most prevalent NB containing bacteria. (Sample sizes: Animal: 210, Blood: 527, Environmental: 641, GI/Stool: 1174, Human: 194, Lung: 385, Plant: 125, Soil: 116, Urine 778).
(TIF)

S5 Fig. Association of contigs to plasmid sequences. A) NB containing contigs were compared to the Plasmid finder database. Bars indicate the number of contigs which contained both an NB gene and a conserved plasmid sequence. B) Association of NB containing contigs to the EBI/NCBI plasmid database. Each dot represents one contig.
(TIF)

S6 Fig. Distribution of NBs is restricted to the γ-proteobacteria. Taxonomic tree representing all species in the pubMLST that have over 100 genomes, constructed using NCBI taxonomy commontree. *Red bar* indicates species that contained at least one NB.
(TIF)

S7 Fig. Distribution of the Ton and Tol operons throughtout the bacterial kingdom. Taxonomic tree of assembled bacterial genomes from NCBI. Tree was constructed using NCBI commontree, blue bars indicate the presence of a Ton or Tol operon.
(TIF)

S8 Fig. Phylogenetic analysis of NBs identified within *Pseudomonas* spp. Sequences from *Pseudomonas* spp. were aligned using MUSCLE and trees built using Mega 6.0 using the neighbour-joining algorithm with 1000 bootstrap replicates. Red boxes indicate the position of the pyocins of *P. aeruginosa*.
(TIF)

S9 Fig. Cytotoxic activity of a novel pyocin, pyocin Sn, targeting *P. aeruginosa* identified through the pipeline. Pyocin Sn (Group VIII, Fig 3) is a non-HNH type DNase bacteriocin (88.8 kDa) newly identified by our bioinformatics analysis. As a test of the validity of this identification, *pyocin Sn* in combination with its immunity protein was overexpressed and purified from *E. coli* extracts and its activity against three strains of *P. aeruginosa* demonstrated. See Materials & methods for further details.
(TIF)

S1 Table. HMM profiles used to identify the conserved motifs of NB cytotoxic domains and associated immunity proteins. PFAM profiles were first used to identify NBs. To test the robustness of our strategy, iterated profiles were created from the newly identified NBs and used in a second search.
(PDF)

S2 Table. Distribution of NBs amongst species. Species that were identified as encoding a NB within their genome and the frequency of occurrence.
(PDF)

S3 Table. HMM profiles used to identify the conserved regions of proteins in the Tol/Ton operons. Tol was differentiated from Ton by the presence of a PD40 repeat that is part of the TolB protein.
(PDF)

S1 Data. PubMLST Multispecies database genomes used in this study. Over 100,000 genomes from multiple data source were collected and assembled in-house to form the database.
(XLSX)

S2 Data. ENA accessions used in the pangenome analysis of *E. coli*. 357 *E. coli* accessions accessed from the ENA were used in the pangenome analysis.
(TXT)

Acknowledgments

We thank Renata Kaminska (Oxford) for help with purifying and assaying mutant colicins and Julian Parkhill (Wellcome Trust Sanger Institute, Cambridge) for helpful discussions.

Author Contributions

Conceptualization: Connor Sharp.

Data curation: Connor Sharp, Nicholas G. Housden.

Formal analysis: Connor Sharp, Colin Kleanthous.

Funding acquisition: Connor Sharp, Colin Kleanthous.

Investigation: Connor Sharp.

Methodology: Connor Sharp, Nicholas G. Housden.

Project administration: Martin C. J. Maiden.

Resources: James Bray, Martin C. J. Maiden.

Software: James Bray.

Supervision: Martin C. J. Maiden, Colin Kleanthous.

Validation: James Bray.

Visualization: Connor Sharp.

Writing – original draft: Connor Sharp, Colin Kleanthous.

Writing – review & editing: Connor Sharp, Colin Kleanthous.

References

1. Gratia A. Sur un remarquable exemple d'antagonisme entre deux souches de colibacille. CR SocBiol (Paris). 1925;(93):1040–1.
2. Kerr B, Riley MA, Feldman MW, Bohannan BJ. Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. Nature. 2002; 418(6894):171–4. https://doi.org/10.1038/nature00823 PMID: 12110887.
3. Chao L, Levin BR. Structured habitats and the evolution of anticompetitor toxins in bacteria. Proc Natl Acad Sci U S A. 1981; 78(10):6324–8. Epub 1981/10/01. PMID: 7031647.
4. Majeed H, Gillor O, Kerr B, Riley Ma. Competitive interactions in Escherichia coli populations: the role of bacteriocins. The ISME journal. 2011; 5:71–81. https://doi.org/10.1038/ismej.2010.90 PMID: 20664553.
5. Majeed H, Lampert A, Ghazaryan L, Gillor O. The weak shall inherit: bacteriocin-mediated interactions in bacterial populations. PloS one. 2013; 8:e63837. https://doi.org/10.1371/journal.pone.0063837 PMID: 23704942.
6. Riley Ma, Gordon DM. The ecological role of bacteriocins in bacterial competition. Trends in microbiology. 1999; 7:129–33. PMID: 10203843.

7. Grinter R, Milner J, Walker D. Bacteriocins active against plant pathogenic bacteria. Biochemical Society transactions. 2012; 40:1498–502. https://doi.org/10.1042/BST20120206 PMID: 23176505.

8. Schulz S, Stephan A, Hahn S, Bortesi L, Jarczowski F, Bettmann U, et al. Broad and efficient control of major foodborne pathogenic strains of Escherichia coli by mixtures of plant-produced colicins. Proceedings of the National Academy of Sciences of the United States of America. 2015; 112:E5454–60. https://doi.org/10.1073/pnas.1513311112 PMID: 26351689.

9. Cotter PD, Ross RP, Hill C. Bacteriocins—a viable alternative to antibiotics? Nature reviews Microbiology. 2013; 11:95–105. https://doi.org/10.1038/nrmicro2937 PMID: 23268227.

10. McCaughey LC, Ritchie ND, Douce GR, Evans TJ, Walker D. Efficacy of species-specific protein antibiotics in a murine model of acute Pseudomonas aeruginosa lung infection. Sci Rep. 2016; 6:30201. Epub 2016/07/23. https://doi.org/10.1038/srep30201 PMID: 27444885;

11. Gillor O, Vriezen JA, Riley MA. The role of SOS boxes in enteric bacteriocin regulation. Microbiology. 2008; 154(Pt 6):1783–92. Epub 2008/06/06. https://doi.org/10.1099/mic.0.2007/016139-0 PMID: 18524933;

12. Butala M, Sonjak S, Kamensek S, Hodoscek M, Browning DF, Zgur-Bertok D, et al. Double locking of an Escherichia coli promoter by two repressors prevents premature colicin expression and cell lysis. Molecular microbiology. 2012; 86(1):129–39. Epub 2012/07/21. https://doi.org/10.1111/j.1365-2958.2012.08179.x PMID: 22812562.

13. Cascales E, Buchanan SK, Duché D, Kleanthous C, Lloubès R, Postle K, et al. Colicin biology. Microbiology and molecular biology reviews: MMBR. 2007; 71:158–229.

14. Kleanthous C, Hemmings AM, Moore GR, James R. Immunity proteins and their specificity for endonuclease colicins: telling right from wrong in protein-protein recognition. Molecular microbiology. 1998; 28 (2):227–33. PMID: 9622349.

15. Kleanthous C, Kuhlmann UC, Pommer AJ, Ferguson N, Radford SE, Moore GR, et al. Structural and mechanistic basis of immunity toward endonuclease colicins. Nature structural biology. 1999; 6(3):243–52. https://doi.org/10.1038/6683 PMID: 10074943.

16. Papadakos G, Wojdyla Ja, Kleanthous C. Nuclease colicins and their immunity proteins. Quarterly reviews of biophysics. 2012; 45:57–103. https://doi.org/10.1017/S0033583511000114 PMID: 22085441.

17. Vankemmelbeke M, Zhang Y, Moore GR, Kleanthous C, Penfold CN, James R. Energy-dependent immunity protein release during tol-dependent nuclease colicin translocation. The Journal of biological chemistry. 2009; 284(28):18932–41. Epub 2009/05/22. PMID: 19458090.

18. Rakin A, Boolgakowa E, Heesemann J. Structural and functional organization of the Yersinia pestis bacteriocin pesticin gene cluster. Microbiology. 1996; 142:3415–24. https://doi.org/10.1099/13500872-142-12-3415 PMID: 9004504.

19. James R, Schneider J, Cooper PC. Characterization of three group A klebicin plasmids: localization of their E colicin immunity genes. Journal of general microbiology. 1987; 133:2253–62. https://doi.org/10.1099/00221287-133-8-2253 PMID: 3127541.

20. Sano Y, Kobayashi M, Kageyama M. Functional domains of S-type pyocins deduced from chimeric molecules. Journal of bacteriology. 1993; 175:6179–85. PMID: 8407790.

21. Hol FJH, Voges MJ, Dekker C, Keymer JE. Nutrient-responsive regulation determines biodiversity in a colicin-mediated bacterial community. BMC biology. 2014; 12:68. https://doi.org/10.1186/s12915-014-0068-2 PMID: 25159553.

22. Petkovsek Z, Zgur-Bertok D, Starcic Erjavec M. Colicin insensitivity correlates with a higher prevalence of extraintestinal virulence factors among Escherichia coli isolates from skin and soft-tissue infections. Journal of medical microbiology. 2012; 61:762–5. https://doi.org/10.1099/jmm.0.037234-0 PMID: 22403139.

23. Micenková L, Staudová B, Bosák J, Mikalová L, Littnerová S, Vrba M, et al. Bacteriocin-encoding genes and ExPEC virulence determinants are associated in human fecal Escherichia coli strains. BMC microbiology. 2014; 14:109. https://doi.org/10.1186/1471-2180-14-109 PMID: 24774171.

24. Nedialkova LP, Denzler R, Koeppel MB, Diehl M, Ring D, Wille T, et al. Inflammation fuels colicin Ib-dependent competition of Salmonella serovar Typhimurium and E. coli in enterobacterial blooms. PLoS pathogens. 2014; 10:e1003844. https://doi.org/10.1371/journal.ppat.1003844 PMID: 24391500.

25. Holt KE, Thieu Nga TV, Thanh DP, Vinh H, Kim DW, Vu Tra MP, et al. Tracking the establishment of local endemic populations of an emergent enteric pathogen. Proc Natl Acad Sci U S A. 2013; 110 (43):17522–7. https://doi.org/10.1073/pnas.1308632110 PMID: 24082120;

26. Riley MA, Gordon DM. A survey of Col plasmids in natural isolates of Escherichia coli and an investigation into the stability of Col-plasmid lineages. Journal of general microbiology. 1992; 138(7):1345–52. Epub 1992/07/01. https://doi.org/10.1099/00221287-138-7-1345 PMID: 1512564.

27. Eddy SR. Accelerated Profile HMM Searches. PLoS computational biology. 2011; 7:e1002195. https://doi.org/10.1371/journal.pcbi.1002195 PMID: 22039361.

28. Walker D, Lancaster L, James R, Kleanthous C. Identification of the catalytic motif of the microbial ribosome inactivating cytotoxin colicin E3. Protein science: a publication of the Protein Society. 2004; 13:1603–11. https://doi.org/10.1110/ps.04658504 PMID: 15133158.

29. Yajima S, Nakanishi K, Takahashi K, Ogawa T, Hidaka M, Kezuka Y, et al. Relation between tRNase activity and the structure of colicin D according to X-ray crystallography. Biochemical and biophysical research communications. 2004; 322:966–73. https://doi.org/10.1016/j.bbrc.2004.07.206 PMID: 15336558.

30. Keeble AH, Maté MJ, Kleanthous C. HNH endonucleases. In: Belfort M, Derbyshire V., Stoddard B. & Wood D., editor. Homing Endonucleases and Inteins. Nucleic acids and molecular biology. 16: Springer-Verlag; 2005.

31. Shen BW, Landthaler M, Shub DA, Stoddard BL. DNA binding and cleavage by the HNH homing endonuclease I-HmuI. J Mol Biol. 2004; 342(1):43–56. https://doi.org/10.1016/j.jmb.2004.07.032 PMID: 15313606.

32. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nature biotechnology. 2013; 31(3):233–9. https://doi.org/10.1038/nbt.2508 PMID: 23360965;

33. Nedialkova LP, Sidstedt M, Koeppel MB, Spriewald S, Ring D, Gerlach RG, et al. Temperate phages promote colicin-dependent fitness of Salmonella enterica serovar Typhimurium. Environmental microbiology. 2015. https://doi.org/10.1111/1462-2920.13077 PMID: 26439675.

34. Mate MJ, Kleanthous C. Structure-based analysis of the metal-dependent mechanism of H-N-H endonucleases. The Journal of biological chemistry. 2004; 279(33):34763–9. https://doi.org/10.1074/jbc.M403719200 PMID: 15190054.

35. Nipič D, Podlesek Z, Budič M, Črnigoj M, Žgur-Bertok D. Escherichia coli uropathogenic-specific protein, USP, is a bacteriocin-like genotoxin. Journal of Infectious Diseases. 2013; 208:1545–52. https://doi.org/10.1093/infdis/jit480 PMID: 23997234.

36. Armstrong GD. Uropathogenic Escherichia coli colicin-like usp and associated proteins: their evolution and role in pathogenesis. The Journal of infectious diseases. 2013; 208:1539–41. https://doi.org/10.1093/infdis/jit482 PMID: 23997232.

37. Zhang D, Iyer LM, Aravind L. A novel immunity system for bacterial nucleic acid degrading toxins and its recruitment in various eukaryotic and DNA viral systems. Nucleic acids research. 2011; 39:4532–52. https://doi.org/10.1093/nar/gkr036 PMID: 21306995.

38. Salomon D, Kinch LN, Trudgian DC, Guo X, Klimko Ja, Grishin NV, et al. Marker for type VI secretion system effectors. PNAS. 2014; 111:1406110111-. https://doi.org/10.1073/pnas.1406110111 PMID: 24927539.

39. Jamet A, Jousset AB, Euphrasie D, Mukorako P, Boucharlat A, Ducousso A, et al. A new family of secreted toxins in pathogenic Neisseria species. PLoS pathogens. 2015; 11(1):e1004592. Epub 2015/01/09. https://doi.org/10.1371/journal.ppat.1004592 PMID: 25569427;

40. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, et al. PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic acids research. 2014; 42:D581–91. https://doi.org/10.1093/nar/gkt1099 PMID: 24225323.

41. Riley Ma, Gordon DM. The ecology and evolution of bacteriocins. Journal of Industrial Microbiology & Biotechnology. 1996; 17:151–8. https://doi.org/10.1007/BF01574688

42. Michel-Briand Y, Baysse C. The pyocins of Pseudomonas aeruginosa. Biochimie. 2002; 84:499–510. PMID: 12423794.

43. Gordon DM, O'Brien CL. Bacteriocin diversity and the frequency of multiple bacteriocin production in Escherichia coli. Microbiology (Reading, England). 2006; 152:3239–44. https://doi.org/10.1099/mic.0.28690-0 PMID: 17074895.

44. Ghequire MGK, De Mot R. Ribosomally encoded antibacterial proteins and peptides from Pseudomonas. Fems Microbiol Rev. 2014; 38:523–68. https://doi.org/10.1111/1574-6976.12079 PMID: 24923764.

45. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrobial agents and chemotherapy. 2014; 58:3895–903. https://doi.org/10.1128/AAC.02412-14 PMID: 24777092.

46. Rijavec M, Budic M, Mrak P, Müller-Premru M, Podlesek Z, Zgur-Bertok D. Prevalence of ColE1-like plasmids and colicin K production among uropathogenic Escherichia coli strains and quantification of

inhibitory activity of colicin K. Applied and environmental microbiology. 2007; 73:1029–32. https://doi.org/10.1128/AEM.01780-06 PMID: 17122402.

47. Nijkamp HJ, de Lang R, Stuitje AR, van den Elzen PJ, Veltkamp E, van Putten AJ. The complete nucleotide sequence of the bacteriocinogenic plasmid CloDF13. Plasmid. 1986; 16(2):135–60. Epub 1986/09/01. PMID: 3749334.

48. Ghequire MGK, De Mot R. Distinct colicin M-like bacteriocin-immunity pairs in Burkholderia. Scientific reports. 2015; 5:17368. https://doi.org/10.1038/srep17368 PMID: 26610609.

49. Barreteau H, El Ghachi M, Barneoud-Arnoulet A, Sacco E, Touze T, Duche D, et al. Characterization of colicin M and its orthologs targeting bacterial cell wall peptidoglycan biosynthesis. Microbial drug resistance (Larchmont, NY). 2012; 18(3):222–9. Epub 2012/03/22. https://doi.org/10.1089/mdr.2011.0230 PMID: 22432709.

50. Spain AM, Krumholz LR, Elshahed MS. Abundance, composition, diversity and novelty of soil Proteobacteria. The ISME journal. 2009; 3(8):992–1000. Epub 2009/05/01. https://doi.org/10.1038/ismej.2009.43 PMID: 19404326.

51. Frickey T, Lupas A. CLANS: a Java application for visualizing protein families based on pairwise similarity. Bioinformatics. 2004; 20(18):3702–4. Epub 2004/07/31. https://doi.org/10.1093/bioinformatics/bth444 PMID: 15284097.

52. Chauleau M, Mora L, Serba J, de Zamaroczy M. FtsH-dependent processing of RNase colicins D and E3 means that only the cytotoxic domains are imported into the cytoplasm. The Journal of biological chemistry. 2011; 286(33):29397–407. https://doi.org/10.1074/jbc.M111.242354 PMID: 21700705;

53. Mora L, de Zamaroczy M. In vivo processing of DNase colicins E2 and E7 is required for their import into the cytoplasm of target cells. PLoS One. 2014; 9(5):e96549. Epub 2014/05/21. https://doi.org/10.1371/journal.pone.0096549 PMID: 24840776;

54. Schaller K, Nomura M. Colicin E2 is DNA endonuclease. Proc Natl Acad Sci U S A. 1976; 73(11):3989–93. PMID: 1069283.

55. Meyhack B, Meyhack I, Apirion D. Colicin E3: a unique endoribonuclease. Proc Natl Acad Sci U S A. 1973; 70(1):156–60. Epub 1973/01/01. PMID: 4567332;

56. Klein A, Wojdyla J, Joshi A, Josts I, McCaughey LC, Housden N, et al. Structural and biophysical analysis of nuclease protein antibiotics. Biochem J. 2016. Epub 2016/07/13. https://doi.org/10.1042/bcj20160544 PMID: 27402794.

57. Papadakos G, Housden NG, Lilly KJ, Kaminska R, Kleanthous C. Kinetic basis for the competitive recruitment of TolB by the intrinsically disordered translocation domain of colicin E9. J Mol Biol. 2012; 418(5):269–80. Epub 2012/02/09. https://doi.org/10.1016/j.jmb.2012.01.039 PMID: 22310049.

58. Housden NG, Hopper JTS, Lukoyanova N, Rodriquez-Larrea D, Wojdyla JA, Klein A, et al. Intrinsically disordered protein threads through the bacterial outer membrane porin OmpF. Science. 2013; 340:1570–4. https://doi.org/10.1126/science.1237864 PMID: 23812713

59. Housden NG, Wojdyla JA, Korczynska J, Grishkovskaya I, Kirkpatrick N, Brzozowski AM, et al. Directed epitope delivery across the Escherichia coli outer membrane through the porin OmpF. Proc Natl Acad Sci U S A. 2010; 107(50):21412–7. Epub 2010/11/26. https://doi.org/10.1073/pnas.1010780107 PMID: 21098297;

60. de Zamaroczy M, Buckingham RH. Importation of nuclease colicins into E coli cells: endoproteolytic cleavage and its prevention by the immunity protein. Biochimie. 2002; 84(5–6):423–32. PMID: 12423785.

61. Walker D, Mosbahi K, Vankemmelbeke M, James R, Kleanthous C. The role of electrostatics in colicin nuclease domain translocation into bacterial cells. The Journal of biological chemistry. 2007; 282(43):31389–97. Epub 2007/08/28. https://doi.org/10.1074/jbc.M705883200 PMID: 17720814.

62. Chavan M, Rafi H, Wertz J, Goldstone C, Riley Ma. Phage associated bacteriocins reveal a novel mechanism for bacteriocin diversification in Klebsiella. Journal of molecular evolution. 2005; 60:546–56. https://doi.org/10.1007/s00239-004-0263-9 PMID: 15883889.

63. Graille M, Mora L, Buckingham RH, van Tilbeurgh H, de Zamaroczy M. Structural inhibition of the colicin D tRNase by the tRNA-mimicking immunity protein. The EMBO journal. 2004; 23(7):1474–82. https://doi.org/10.1038/sj.emboj.7600162 PMID: 15014439.

64. Dessi A, Puddu M, Testa M, Marcialis MA, Pintus MC, Fanos V. Serratia marcescens infections and outbreaks in neonatal intensive care units. Journal of chemotherapy (Florence, Italy). 2009; 21:493–9. https://doi.org/10.1179/joc.2009.21.5.493 PMID: 19933039.

65. Gruber TM, Gottig S, Mark L, Christ S, Kempf VAJ, Wichelhaus TA, et al. Pathogenicity of pan-drug-resistant Serratia marcescens harbouring blaNDM-1. The Journal of antimicrobial chemotherapy. 2015; 70:1026–30. https://doi.org/10.1093/jac/dku482 PMID: 25468904.

66. Viejo MB, Gargallo D, Ferrer S, Enfedaque J, Regue M. Cloning and DNA sequence analysis of a bacteriocin gene of Serratia marcescens. Journal of general microbiology. 1992; 138 Pt 8:1737–43. https://doi.org/10.1099/00221287-138-8-1737 PMID: 1527512.

67. Jamet A, Nassif X. New Players in the Toxin Field: Polymorphic Toxin Systems in. 2015; 6:1–8. doi: 10.1128/mBio.00285-15. Copyright. PMID: 25944858

68. Bosák J, Laiblová P, Šmarda J, Dědičová D, Šmajs D. Novel colicin FY of Yersinia frederiksenii inhibits pathogenic Yersinia strains via YiuR-mediated reception, TonB import, and cell membrane pore formation. Journal of bacteriology. 2012; 194:1950–9. https://doi.org/10.1128/JB.05885-11 PMID: 22343298.

69. Vollmer W, Pilsl H, Hantke K, Holtje JV, Braun V. Pesticin displays muramidase activity. Journal of bacteriology. 1997; 179(5):1580–3. Epub 1997/03/01. PMID: 9045816.

70. Oliver A, Mulet X, Lopez-Causape C, Juan C. The increasing threat of Pseudomonas aeruginosa high-risk clones. Drug resistance updates: reviews and commentaries in antimicrobial and anticancer chemotherapy. 2015; 21–22:41–59. Epub 2015/08/26. https://doi.org/10.1016/j.drup.2015.08.002 PMID: 26304792.

71. Gellatly SL, Hancock RE. Pseudomonas aeruginosa: new insights into pathogenesis and host defenses. Pathogens and disease. 2013; 67(3):159–73. Epub 2013/04/27. https://doi.org/10.1111/2049-632X.12033 PMID: 23620179.

72. McCaughey L C, Josts I, Grinter R, White P, Byron O, Tucker N P, et al. Discovery, characterization and in vivo activity of pyocin SD2, a protein antibiotic from Pseudomonas aeruginosa. Biochem J. 2016; 473(15):2345–58. https://doi.org/10.1042/BCJ20160470 PMID: 27252387;

73. Bailey TL, Boden M, Buske Fa, Frith M, Grant CE, Clementi L, et al. MEME Suite: Tools for motif discovery and searching. Nucleic Acids Research. 2009; 37:202–8. https://doi.org/10.1093/nar/gkp335 PMID: 19458158.

74. Kageyama M, Kobayashi M, Sano Y, Masaki H. Construction and Characterization of Pyocin-Colicin Chimeric Proteins. 1996; 178:103–10.

75. Wang Y, Huang SH, Wass CA, Stins MF, Kim KS. The gene locus yijP contributes to Escherichia coli K1 invasion of brain microvascular endothelial cells. Infection and immunity. 1999; 67:4751–6. PMID: 10456927.

76. Beutin L, Tao J, Feng L, Krause G, Zimmermann S, Gleier K, et al. Sequence Analysis of the Escherichia coli O15 Antigen Gene Cluster and Development of a PCR Assay for Rapid Detection of Intestinal and Extraintestinal Pathogenic E. coli O15 Strains. Society. 2005; 43:703–10.

77. Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C, Colles FM, et al. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. Microbiology. 2012; 158(Pt 4):1005–15. https://doi.org/10.1099/mic.0.055459-0 PMID: 22282518;

78. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome research. 2008; 18(5):821–9. https://doi.org/10.1101/gr.074492.107 PMID: 18349386;

79. Gough J, Karplus K, Hughey R, Chothia C. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. Journal of molecular biology. 2001; 313:903–19. https://doi.org/10.1006/jmbi.2001.5080 PMID: 11697912.

80. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. 2004; 32:1792–7. https://doi.org/10.1093/nar/gkh340 PMID: 15034147.

81. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology. 2011; 7. https://doi.org/10.1038/msb.2011.75 PMID: 21988835.

82. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. Genome research. 2004; 14(7):1394–403. https://doi.org/10.1101/gr.2289704 PMID: 15231754;

83. Lupas A, Van Dyke M, Stock J. Predicting coiled coils from protein sequences. Science (New York, NY). 1991; 252:1162–4. https://doi.org/10.1126/science.252.5009.1162 PMID: 2031185.

84. Dosztanyi Z, Csizmok V, Tompa P, Simon I. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. J Mol Biol. 2005; 347(4):827–39. Epub 2005/03/17. https://doi.org/10.1016/j.jmb.2005.01.071 PMID: 15769473.

85. Moradigaravand D, Boinett CJ, Martin V, Peacock SJ, Parkhill J. Recent independent emergence of multiple multidrug-resistant Serratia marcescens clones within the United Kingdom and Ireland. Genome research. 2016; 26(8):1101–9. Epub 2016/07/20. https://doi.org/10.1101/gr.205245.116 PMID: 27432456;

86. Page AJ, Cummins Ca, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: Rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015; 31:btv421. https://doi.org/10.1093/bioinformatics/btv421 PMID: 26198102

87. Stamatakis a, Ludwig T, Meier H. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics (Oxford, England). 2005; 21:456–63. https://doi.org/10.1093/bioinformatics/bti191 PMID: 15608047.

88. Corander J, Waldmann P, Marttinen P, Sillanpää MJ. BAPS 2: enhanced possibilities for the analysis of genetic population structure. Bioinformatics (Oxford, England). 2004; 20:2363–9. https://doi.org/10.1093/bioinformatics/bth250 PMID: 15073024.

7.1 Genome assembly in the pubMLST Multispecies isolate database

The PubMLST Multispecies isolate database is comprised of NCBI finished and assembled genomes, as well as genomes in-house assembled from the European Nucleotide (ENA-SRA). Short reads are assembled using the Velvet assembly program (v1.2.08) (Zerbino and Birney, 2008). Odd-numbered kmer lengths between 21 and the read length were sampled using VelvetOptimiser(v2.2.4) to calculate the optimal parameters for Velvet. No scaffolding was performed and only contigs longer than 200bp were included in the final assembly.

The PubMLST Multispecies datbase was established to reflect the current knowledge of bacterial genomes. While the database does not bias itself to a specific sub-class of the bacterial kingdom, it is however, subject to any bias of sequences deposited in the ENA.

7.2 Conditions tested by the BIOLOG Phenotypic microarray

BIOLOG

Phenotype MicroArrays™

PM1 MicroPlate™ Carbon Sources

A1 Negative Control	A2 L-Arabinose	A3 N-Acetyl-D-Glucosamine	A4 D-Saccharic Acid	A5 Succinic Acid	A6 D-Galactose	A7 L-Aspartic Acid	A8 L-Proline	A9 D-Alanine	A10 D-Trehalose	A11 D-Mannose	A12 Dulcitol
B1 D-Serine	B2 D-Sorbitol	B3 Glycerol	B4 L-Fucose	B5 D-Glucuronic Acid	B6 D-Gluconic Acid	B7 D,L-α-Glycerol-Phosphate	B8 D-Xylose	B9 L-Lactic Acid	B10 Formic Acid	B11 D-Mannitol	B12 L-Glutamic Acid
C1 D-Glucose-6-Phosphate	C2 D-Galactonic Acid-γ-Lactone	C3 D,L-Malic Acid	C4 D-Ribose	C5 Tween 20	C6 L-Rhamnose	C7 D-Fructose	C8 Acetic Acid	C9 α-D-Glucose	C10 Maltose	C11 D-Melibiose	C12 Thymidine
D-1 L-Asparagine	D2 D-Aspartic Acid	D3 D-Glucosaminic Acid	D4 1,2-Propanediol	D5 Tween 40	D6 α-Keto-Glutaric Acid	D7 α-Keto-Butyric Acid	D8 α-Methyl-D-Galactoside	D9 α-D-Lactose	D10 Lactulose	D11 Sucrose	D12 Uridine
E1 L-Glutamine	E2 m-Tartaric Acid	E3 D-Glucose-1-Phosphate	E4 D-Fructose-6-Phosphate	E5 Tween 80	E6 α-Hydroxy Glutaric Acid-γ-Lactone	E7 α-Hydroxy Butyric Acid	E8 β-Methyl-D-Glucoside	E9 Adonitol	E10 Maltotriose	E11 2-Deoxy Adenosine	E12 Adenosine
F1 Glycyl-L-Aspartic Acid	F2 Citric Acid	F3 m-Inositol	F4 D-Threonine	F5 Fumaric Acid	F6 Bromo Succinic Acid	F7 Propionic Acid	F8 Mucic Acid	F9 Glycolic Acid	F10 Glyoxylic Acid	F11 D-Cellobiose	F12 Inosine
G1 Glycyl-L-Glutamic Acid	G2 Tricarballylic Acid	G3 L-Serine	G4 L-Threonine	G5 L-Alanine	G6 L-Alanyl-Glycine	G7 Acetoacetic Acid	G8 N-Acetyl-β-D-Mannosamine	G9 Mono Methyl Succinate	G10 Methyl Pyruvate	G11 D-Malic Acid	G12 L-Malic Acid
H1 Glycyl-L-Proline	H2 p-Hydroxy Phenyl Acetic Acid	H3 m-Hydroxy Phenyl Acetic Acid	H4 Tyramine	H5 D-Psicose	H6 L-Lyxose	H7 Glucuronamide	H8 Pyruvic Acid	H9 L-Galactonic Acid-γ-Lactone	H10 D-Galacturonic Acid	H11 Phenylethyl-amine	H12 2-Aminoethanol

PM2A MicroPlate™ Carbon Sources

A1 Negative Control	A2 Chondroitin Sulfate C	A3 α-Cyclodextrin	A4 β-Cyclodextrin	A5 γ-Cyclodextrin	A6 Dextrin	A7 Gelatin	A8 Glycogen	A9 Inulin	A10 Laminarin	A11 Mannan	A12 Pectin
B1 N-Acetyl-D-Galactosamine	B2 N-Acetyl-Neuraminic Acid	B3 β-D-Allose	B4 Amygdalin	B5 D-Arabinose	B6 D-Arabitol	B7 L-Arabitol	B8 Arbutin	B9 2-Deoxy-D-Ribose	B10 i-Erythritol	B11 D-Fucose	B12 3-O-β-D-Galacto-pyranosyl-D-Arabinose
C1 Gentiobiose	C2 L-Glucose	C3 Lactitol	C4 D-Melezitose	C5 Maltitol	C6 a-Methyl-D-Glucoside	C7 β-Methyl-D-Galactoside	C8 3-Methyl Glucose	C9 β-Methyl-D-Glucuronic Acid	C10 α-Methyl-D-Mannoside	C11 β-Methyl-D-Xyloside	C12 Palatinose
D1 D-Raffinose	D2 Salicin	D3 Sedoheptulosan	D4 L-Sorbose	D5 Stachyose	D6 D-Tagatose	D7 Turanose	D8 Xylitol	D9 N-Acetyl-D-Glucosaminitol	D10 γ-Amino Butyric Acid	D11 δ-Amino Valeric Acid	D12 Butyric Acid
E1 Capric Acid	E2 Caproic Acid	E3 Citraconic Acid	E4 Citramalic Acid	E5 D-Glucosamine	E6 2-Hydroxy Benzoic Acid	E7 4-Hydroxy Benzoic Acid	E8 β-Hydroxy Butyric Acid	E9 γ-Hydroxy Butyric Acid	E10 a-Keto-Valeric Acid	E11 Itaconic Acid	E12 5-Keto-D-Gluconic Acid
F1 D-Lactic Acid Methyl Ester	F2 Malonic Acid	F3 Melibionic Acid	F4 Oxalic Acid	F5 Oxalomalic Acid	F6 Quinic Acid	F7 D-Ribono-1,4-Lactone	F8 Sebacic Acid	F9 Sorbic Acid	F10 Succinamic Acid	F11 D-Tartaric Acid	F12 L-Tartaric Acid
G1 Acetamide	G2 L-Alaninamide	G3 N-Acetyl-L-Glutamic Acid	G4 L-Arginine	G5 Glycine	G6 L-Histidine	G7 L-Homoserine	G8 Hydroxy-L-Proline	G9 L-Isoleucine	G10 L-Leucine	G11 L-Lysine	G12 L-Methionine
H1 L-Ornithine	H2 L-Phenylalanine	H3 L-Pyroglutamic Acid	H4 L-Valine	H5 D,L-Carnitine	H6 Sec-Butylamine	H7 D,L-Octopamine	H8 Putrescine	H9 Dihydroxy Acetone	H10 2,3-Butanediol	H11 2,3-Butanedione	H12 3-Hydroxy 2-Butanone

BIOLOG

Phenotype MicroArrays™

PM9 MicroPlate™ Osmolytes

A1 NaCl 1%	A2 NaCl 2%	A3 NaCl 3%	A4 NaCl 4%	A5 NaCl 5%	A6 NaCl 5.5%	A7 NaCl 6%	A8 NaCl 6.5%	A9 NaCl 7%	A10 NaCl 8%	A11 NaCl 9%	A12 NaCl 10%
B1 NaCl 6%	B2 NaCl 6% + Betaine	B3 NaCl 6% + N-N Dimethyl Glycine	B4 NaCl 6% + Sarcosine	B5 NaCl 6% + Dimethyl sulphonyl propionate	B6 NaCl 6% + MOPS	B7 NaCl 6% + Ectoine	B8 NaCl 6% + Choline	B9 NaCl 6% + Phosphoryl Choline	B10 NaCl 6% + Creatine	B11 NaCl 6% + Creatinine	B12 NaCl 6% + L- Carnitine
C1 NaCl 6% + KCl	C2 NaCl 6% + L-Proline	C3 NaCl 6% + N-Acetyl L-Glutamine	C4 NaC1 6% + β-Glutamic Acid	C5 NaC1 6% + γ–Amino -N- Butyric Acid	C6 NaC1 6% + Glutathione	C7 NaCl 6% + Glycerol	C8 NaC1 6% + Trehalose	C9 NaC1 6% + Trimethylamine- N-oxide	C10 NaC1 6% + Trimethylamine	C11 NaCl 6% + Octopine	C12 NaC1 6% + Trigonelline
D1 Potassium chloride 3%	D2 Potassium chloride 4%	D3 Potassium chloride 5%	D4 Potassium chloride 6%	D5 Sodium sulfate 2%	D6 Sodium sulfate 3%	D7 Sodium sulfate 4%	D8 Sodium sulfate 5%	D9 Ethylene glycol 5%	D10 Ethylene glycol 10%	D11 Ethylene glycol 15%	D12 Ethylene glycol 20%
E1 Sodium formate 1%	E2 Sodium formate 2%	E3 Sodium formate 3%	E4 Sodium formate 4%	E5 Sodium formate 5%	E6 Sodium formate 6%	E7 Urea 2%	E8 Urea 3%	E9 Urea 4%	E10 Urea 5%	E11 Urea 6%	E12 Urea 7%
F1 Sodium Lactate 1%	F2 Sodium Lactate 2%	F3 Sodium Lactate 3%	F4 Sodium Lactate 4%	F5 Sodium Lactate 5%	F6 Sodium Lactate 6%	F7 Sodium Lactate 7%	F8 Sodium Lactate 8%	F9 Sodium Lactate 9%	F10 Sodium Lactate 10%	F11 Sodium Lactate 11%	F12 Sodium Lactate 12%
G1 Sodium Phosphate pH 7 20mM	G2 Sodium Phosphate pH 7 50mM	G3 Sodium Phosphate pH 7 100mM	G4 Sodium Phosphate pH 7 200mM	G5 Sodium Benzoate pH 5.2 20mM	G6 Sodium Benzoate pH 5.2 50mM	G7 Sodium Benzoate pH5.2 100mM	G8 Sodium Benzoate pH 5.2 200mM	G9 Ammonium sulfate pH8 10mM	G10 Ammonium sulfate pH 8 20mM	G11 Ammonium sulfate pH 8 50mM	G12 Ammonium sulfate pH8 100mM
H1 Sodium Nitrate 10mM	H2 Sodium Nitrate 20mM	H3 Sodium Nitrate 40mM	H4 Sodium Nitrate 60mM	H5 Sodium Nitrate 80mM	H6 Sodium Nitrate 100mM	H7 Sodium Nitrite 10mM	H8 Sodium Nitrite 20mM	H9 Sodium Nitrite 40mM	H10 Sodium Nitrite 60mM	H11 Sodium Nitrite 80mM	H12 Sodium Nitrite 100mM

PM10 MicroPlate™ pH

A1 pH 3.5	A2 pH 4	A3 pH 4.5	A4 pH 5	A5 pH 5.5	A6 pH 6	A7 pH 7	A8 pH 8	A9 pH 8.5	A10 pH 9	A11 pH 9.5	A12 pH 10
B1 pH 4.5	B2 pH 4.5 + L-Alanine	B3 pH 4.5 + L-Arginine	B4 pH 4.5 + L-Asparagine	B5 pH 4.5 + L-Aspartic Acid	B6 pH 4.5 + L-Glutamic Acid	B7 pH 4.5 + L-Glutamine	B8 pH 4.5 + Glycine	B9 pH 4.5 + L-Histidine	B10 pH 4.5 + L-Isoleucine	B11 pH 4.5 + L-Leucine	B12 pH 4.5 + L-Lysine
C1 pH 4.5 + L-Methionine	C2 pH 4.5 + L-Phenylalanine	C3 pH 4.5 + L-Proline	C4 pH 4.5 + L-Serine	C5 pH 4.5 + L-Threonine	C6 pH 4.5 + L-Tryptophan	C7 pH 4.5 + L-Citrulline	C8 pH 4.5 + L-Valine	C9 pH 4.5 + Hydroxy- L-Proline	C10 pH 4.5 + L-Ornithine	C11 pH 4.5 + L-Homoarginine	C12 pH 4.5 + L-Homoserine
D-1 pH 4.5 + Anthranilic Acid	D2 pH 4.5 + L-Norleucine	D3 pH 4.5 + L-Norvaline	D4 pH 4.5 + α- Amino-N- Butyric Acid	D5 pH 4.5 + p-Amino- Benzoic Acid	D6 pH 4.5 + L-Cysteic Acid	D7 pH 4.5 + D-Lysine	D8 pH 4.5 + 5-Hydroxy Lysine	D9 pH 4.5 + 5-Hydroxy Tryptophan	D10 pH 4.5 + D,L-Diamino pimelic Acid	D11 pH 4.5 + Trimethylamine- N-oxide	D12 pH 4.5 + Urea
E1 pH 9.5	E2 pH 9.5 + L-Alanine	E3 pH 9.5 + L-Arginine	E4 pH 9.5 + L-Asparagine	E5 pH 9.5 + L-Aspartic Acid	E6 pH 9.5 + L-Glutamic Acid	E7 pH 9.5 + L-Glutamine	E8 pH 9.5 + Glycine	E9 pH 9.5 + L-Histidine	E10 pH 9.5 + L-Isoleucine	E11 pH 9.5 + L-Leucine	E12 pH 9.5 + L-Lysine
F1 pH 9.5 + L-Methionine	F2 pH 9.5 + L-Phenylalanine	F3 pH 9.5 + L-Proline	F4 pH 9.5 + L-Serine	F5 pH 9.5 + L-Threonine	F6 pH 9.5 + L-Tryptophan	F7 pH 9.5 + L-Tyrosine	F8 pH 9.5 + L-Valine	F9 pH 9.5 + Hydroxy- L-Proline	F10 pH 9.5 + L-Ornithine	F11 pH 9.5 + L-Homoarginine	F12 pH 9.5 + L-Homoserine
G1 pH 9.5 + Anthranilic acid	G2 pH 9.5 + L-Norleucine	G3 pH 9.5 + L-Norvaline	G4 pH 9.5 + Agmatine	G5 pH 9.5 + Cadaverine	G6 pH 9.5 + Putrescine	G7 pH 9.5 + Histamine	G8 pH 9.5 + Phenylethylamine	G9 pH 9.5 + Tyramine	G10 pH 9.5 + Creatine	G11 pH 9.5 + Trimethylamine- N-oxide	G12 pH 9.5 + Urea
H1 X-Caprylate	H2 X-α-D- Glucoside	H3 X-β-D- Glucoside	H4 X-α-D- Galactoside	H5 X-β-D- Galactoside	H6 X-α- D- Glucuronide	H7 X-β- D- Glucuronide	H8 X-β-D- Glucosaminide	H9 X-β-D- Galactosaminide	H10 X-α-D- Mannoside	H11 X-PO4	H12 X-SO4