Natural antibiotic discovery

Thesis submitted by
Khadijah Mohammed Dashti
For the degree of
DOCTOR OF PHILOSOPHY
in the
Faculty of Medical Sciences
University College London

Department of Microbial Diseases
UCL Eastman Dental Institute
Royal Free Campus
London NW3 2PF
UK
2022

Declaration

I, Khadijah Mohammed Dashti confirm that the work presented in my thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis

Abstract

Currently, there is a decline in the discovery of new antibiotics. Several important habitats that have yet to be fully exploited for antibiotic-producing bacteria include water and the anterior nares. Therefore, antibiotic-producing bacteria isolated from these two environments were evaluated for antibiotic production, and using ribosome engineering, their activity was further enhanced. Five NW *Pseudomonas* sp. strains isolated from water were tested for antimicrobial activity against multidrug-resistant bacteria. The five strains were grown under various conditions, by which the growth medium significantly influenced the production of antimicrobials (*p*-value <0.001). All five strains were inhibitory to several indicator bacterium however, *Pseudomonas* sp. strain NW27 was inhibitory to *S. aureus*, *E. faecium* and *A. baumannii*. Bioinformatics analysis revealed that *Pseudomonas* sp. strain NW27 outperforms other strains in its ability to produce a variety of novel antimicrobials, including those related to viscosin, bananamide and syringomycin. Rifampicin-resistant *Pseudomonas* sp. strain NW27-A1 was isolated using 10 ug/ml rifampicin and showed significantly enhanced antimicrobial activity compared to the wild-type according to agar well diffusion assays (*p*-value <0.001). The MIC assay showed no significant change in the MIC of *Pseudomonas* sp. strain NW27-A1 compared to the wild-type (0.12 mg/ml versus 1.13 mg/ml respectively, *p*-value= 0.200). The active compounds were extracted from

the supernatant of *Pseudomonas* sp. strain NW27-A1 using acetonitrile, and 0.23% of the units were recovered with a 1.28 fold purification. Several polar, active fractions were identified by HPLC, however the least polar fraction G significantly reduced the cellular metabolism of *E. faecium* E1162 by 99.65%. Moreover, bioinformatic analysis of metagenomic data suggested the ability of the commensal bacterium *Dolosigranulum pigrum* to produce several novel bacteriocins and a novel polyketide, that can inhibit *S. aureus* growth in the anterior nares. The supernatant of *D. pigrum* ATCC 51524 did not demonstrate considerable antimicrobial activity however, production of hydrogen peroxide was one mechanism associated with the exclusion of *S. aureus* in the anterior nares. However, rifampicin-resistant *D. pigrum* DP8rif10 successfully inhibited *S. aureus* JE2 when collected at 30 hours post-inoculation. *Pseudomonas* sp. strain NW27-A1 and *D. pigrum* DP8rif10 therefore offer a potential source of novel therapeutic approaches in the future.

Impact statement

Infections caused by antibiotic resistant bacteria are on the rise and treatment options remain limited. Due to this global issue, there is an ever-increased demand for novel antimicrobials to treat infections caused by multi-drug resistant bacteria.

Pseudomonas sp. strain NW27-A1 isolated from water was likely to produce novel metabolites with antimicrobial properties against clinically significant multidrug-resistant bacteria. Such metabolites include siderophores, that can be useful in potentiating the activity of existing antibiotics, restoring their activity, and slowing the onset of resistance. It can therefore decrease deaths attributed to antibiotic-resistance and can reduce the length of hospital stay which will also cut down on health-care costs. It can also lead to stakeholders appreciating the chemical strategy to couple siderophores to antibiotics and can also attract potential investors and pharmaceutical companies. These companies are also interested in new products which prevent bacterial adhesion (antibiofilm). *Pseudomonas* sp. strain NW27-A1 is likely to produce novel surfactants exhibiting antibiofilm properties. The supernatant of this bacterium is active against carbapenem-resistant *A. baumannii*, commonly found in the matrix of biofilms of healthcare settings. These surfactants can therefore be used to treat this bacterium present on hospital surfaces, which will aid

in infection control in hospital settings, reducing the cost of treatment of problematic biofilm-related infections. An increased availability of these innovative approaches and products will be of great benefit to the antimicrobial research community worldwide and is likely to offer hope in a currently narrow pipeline of new antibiotics.

This thesis contributes to the general knowledge of antimicrobial drug discovery from bacteria in the human microbiome, a previously untapped source of antibiotics. This thesis has demonstrated the ability of nasal commensal bacterium *Dolosigranulum pigrum* to inhibit the growth of *S. aureus* in the anterior nares by producing novel antimicrobials, including several bacteriocins and a polyketide. *D. pigrum* DP8rif10, which showed appreciable antimicrobial production, can be used as an anterior nares decolonization probiotic prior to surgery, which will therefore improve patient's outcome. This pre-operative *S. aureus* nasal decolonization strategy can lead to lower mortality rates attributed to surgical site infections. This probiotic can cut down on healthcare costs required to control infections caused by *S. aureus*.

This thesis focuses on early antibiotic discovery and with bringing together experts from the academic and industrial sectors, can provide knowledge

of antibiotic production by *Pseudomonas* sp. strain NW27 and *D. pigrum* for diverse projects in the field of antibiotic development.

Covid-19 pandemic impact statement

I have been severely impacted by the Covid-19 pandemic. During the lockdown from 2020-2021, I had to homeschool my two children, which significantly reduced the time I could spend writing and analyzing my data. In addition to homeschooling, I also cared for a newborn as I also gave birth in March 2020, as soon as the pandemic started. Due to strict social distancing rules at the time, childcare was not available and my family and friends could not assist me, so I homeschooled while caring also for a newborn. Secondly, I contracted Covid-19 while 7 months pregnant during 2022, which resulted in severe illness, complications and ultimately induced abortion. This loss has significantly impacted my physical and mental health. Since then, I have also been experiencing long-term Covid symptoms, including headaches, insomnia, muscle soreness, concentration issues and depression.

Acknowledgment

First and foremost, I thank God for his abundant grace in getting me through all the challenging moments of completing this dissertation.

I would like to deeply thank my supervisors Dr Sean P. Nair and Prof Peter Mullany for all their guidance and valuable feedback throughout this research. Their encouragement and great supervision were impeccable and without their guidance, this dissertation would not have been possible.

This dissertation would not have been possible without the support from my beloved family and friends, especially my parents, my husband Saad and my children Mahdi, Fatemah and Ali. I thank you for your continuous support and for patiently dealing with my dissertation journey. I could not have completed this dissertation without your encouragement.

I would also like to thank laboratory members from Department of Microbial Diseases, including Dr Anna Tymon, Dr Haitham Hussain, Tracey Moss, Ingrid Green and to all my colleagues especially Arely, Shatha, Sophia and Hadeel. Your solidarity and support during this journey made this experience easier and enjoyable.

Finally, it is with much honor I would like to thank Department of Allied Health Sciences, Kuwait University, for providing me the scholarship and the financial support.

Table of contents

List of figures

List of tables

List of abbreviations

β	Beta
°C	Degrees Celsius
2,4-DAPG	2, 4 diacetylphloroglucinol
ABC	ATP-binding cassette
antiSMASH	antibiotics & Secondary Metabolite Analysis Shell
APE	Arylpolyene
APE Vf	Arylpolyene biosynthetic gene cluster from *Aliivibrio fischeri* ES114
ATA	Aurintricarboxylic acid
ATCC	American Type Culture Collection
BAGEL	Bacteriocin genome mining tool
BEI	Biodefense and Emerging Infections Research Resources Repository
BGC	Biosynthetic gene cluster
BLAST	Basic Local Alignment Search Tool
C8	Carbon 8
C18	Carbon 18
CDPS	tRNA-dependent cyclodipeptide synthases
CFU	Colony forming units
CLSI	Clinical and laboratory standards Institute
CRAB	Carbapenem-resistant *A. baumannii*
DAD	Diode array detector
DNA	Deoxyribonucleic acid

EARS-Net	European Antimicrobial Resistance Surveillance network
ESBL	Extended spectrum β-lactamase
FASTA	Fast-all
FDA	Food and Drug Administration
g	Grams
h	Hour
HGT	Horizontal gene transfer
HMMs	Hidden Markov models
HPLC	High-performance liquid chromatography
ICU	Intensive care unit
IMG/ABC	Integrated Microbial Genome Atlas of Biosynthetic Gene Clusters
IMG/M	Integrated Microbial Genomes and Microbiome
KD	Kilodalton
L	Litre
LB	Luria Bernati
LPS	Lipopolysaccharides
LTTR	LysR-Type Transcriptional Regulator
µm	Micrometer
M9	M9 minimal medium
µl	Microliter
MDR	Multidrug- resistant
mg/ml	Milligram per milliliter
MIC	Minimum inhibitory concentration
MiBIG	Minimum Information on Biosynthetic Gene Cluster

ml/min	Milliliter per minute
mL	Milliliter
mRNA	messenger RNA
MRSA	methicillin-resistant *Staphylococcus aureus*
NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
NAG	N-acetylglucosamine
NAGGN	N-acetylglutaminylglutamine amide
NAM	N-acetylmuramic acid
NARSA	Network on Antimicrobial Resistance in *Staphylococcus aureus*
NB	Nutrient broth
NCTC	National collection of Type Cultures
NHS	National Health Service
NIAID	National Institute of Allergy and Infectious Diseases
NIH	National Institute of Health
nm	Nanometre
NRP	Non-ribosomal peptide
NRPS	Non-ribosomal peptide synthetase
NW	Nair Water
OD	Optical density
OSMAC	One strain many compounds
ORF	Open read frame
ORs	Adjusted odds ratio
PABA	P-aminobenzoic acid

PBP	Penicillin binding proteins
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PEP	Phosphoethanolamine
pHMMs	Profile hidden Markov model
PK	Polyketide
PK/NRP	Polyketide/non-ribosomal peptide
PMBNP	Polymyxin B nonapeptide
ppGpp	Guanosine-5’-diphosphate-3’-diphosphate
PPY	Photopyrone
PpyS-KS	Photopyrone-like pyrone
PRISM	PRediction Informatics for Secondary Metabolomes
REDOX	Reduction-Oxidation
RFU	Relative fluorescent units
RiPP	Unspecified ribosomally synthesized and post-translationally modified peptide product
RNA	Ribonucleic acid
rRNA	Ribosomal ribonucleic acid
RNAP	RNA polymerase
rpm	Rotations per minute
RRE	RiPP precursor peptide recognition element
RT-qPCR	Reverse-transcriptase quantitative polymerase chain reaction
SARS-Cov-2	Severe Acute Respiratory Syndrome Coronavirus 2
SEM	Scanning electron microscopy
SNPs	Single nucleotide polymorphisms

SPE	Solid phase extraction
T1PKS	Type 1 polyketide synthase
T3PKS	Type 3 polyketide synthase
TFA	Trifluoroacetic acid
TH	Todd Hewitt
TS	Tryptic soy
U	Total activity
U/mg	Specific activity
UTI	Urinary tract infection
VISA	Vancomycin-intermediate *Staphylococcus aureus*
VRE	Vancomycin-resistant enterococci
VRSA	Vancomycin-resistant *Staphylococcus aureus*
v/v	volume/volume
WGS	Whole genome shotgun sequence
WHO	World Health Organization
WTA	Wall-teichoic acids
XDR	Extensively-drug resistant
xg	Gravitational force

Chapter 1.0

General Introduction

1.0 General Introduction

Throughout the ages, infectious diseases have been challenging to control (Glatter and Finkelman, 2021). The Bubonic plague, which peaked from 1347-1350, contributed to the death of one third of Europe's population (Slack, 1989). Up to the early 1900s, infectious diseases were the leading cause of death (CDC, 2000). The commercialization of antibiotics by the mid-1900s has contributed to a decrease in mortality caused by infectious diseases (Kugeler *et al*., 2015).

Antibiotics have become the foundation of treatment of infectious diseases, saving millions of lives (Hutchings *et al*., 2019). Selman Waksman introduced the term 'antibiotic' as any small molecule made by a microbe that antagonizes the growth of other microbes (Clardy *et al*., 2009). The term 'antibiotic' is derived from the Greek term 'biotikos', which is translated as 'against life' (Conly and Johnston, 2004).

1.1 History of antibiotic discovery

Paul Ehrlich discovered Salvarsan in 1909, the first synthetic arsenic-derived antibiotic (Bosch and Rosich, 2008). This antibiotic was effective in inhibiting *Treponema pallidum*, the causative pathogen of Syphilis (Bosch and Rosich, 2008). A less dangerous and more effective antibiotic to treat syphilis, Neosalvarsan, was introduced during the year 1913

(Rubin, 2007). However due to the presence of arsenic, these anti-syphilis drugs were associated with multiple side effects, including rashes and liver damage (McDonagh, 1912).

During 1930, Salvarsan and Neosalvarsan were replaced with a broad-spectrum antibacterial sulfonamide drug, Prontosil, which was discovered by German bacteriologist Gerhard Domagk (Sulek, 1968). This drug was one of the landmarks in antibiotic history. However, with time, bacteria became resistant to Prontosil, which was later replaced with penicillin (Skold, 2000).

The unintentional discovery by Scottish bacteriologist Alexander Fleming in 1928 has led to a major breakthrough in antibiotic discovery (Figure 1.1) (Tan and Tatsumura, 2015). He was able to isolate penicillin, the first true antibiotic (Tan and Tatsumura, 2015). The compound was purified, and production was initiated in 1939 due to the elucidation of the structure of penicillin G by Florey and Ernst Boris Chain (Tan and Tatsumura, 2015).

The reports of microorganism's ability to produce antimicrobial compounds, including penicillin, led Selman Waksman to study microbes as antimicrobial producers in the late 1930s (Comroe, 1978). By showcasing bacterial species with antagonistic relationships, he

discovered numerous antibiotics including actinomycin, neomycin and streptomycin (Waksman *et al*., 2010).

Most antibiotics commonly used today were discovered during the golden era of antibiotic discovery, which was between the 1940s and 1960s (Jackson *et al*., 2018). Since the discovery of daptomycin in the 1980s, there was a decline in the discovery of antibiotics (Hutchings *et al*., 2019). One of the antibiotics discovered after daptomycin include lipiarmycins (eg. fidaxomicin). Fidaxomicin is a new class of macrocyclic antibiotics for the treatment of colitis caused by *Clostridium difficile* (Venugopal *et al*., 2012). It is produced by the actinomycete *Dactylosporangium aurantiacum* subspecies *hamdenesis and* is able to kill bacteria by inhibiting transcription by bacterial RNA polymerase (RNAP) (Boyaci *et al*., 2018). Teixobactin is a novel cyclic-peptide based antibiotic discovered during 2015 through iChip technology (Piddock, 2015). It is effective against multidrug resistant *Staphylococcus aureus* and *Mycobacterium tuberculosis*, binding to cell wall precursors lipid II and lipid III (Piddock, 2015). Pseudouridimycin is another novel class of antibiotics discovered in 2017 (Chellat *et al*., 2017). Pseudouridimycin, produced by *Streptomyces* sp. ID38640, is a pseudouridine-containing peptidyl-nucleoside antibiotic that inhibits bacterial RNAP of several Gram positive and Gram negative bacteria through a unique mechanism different from other RNAP inhibitors such as rifamycin (Maffioli *et al*., 2019).

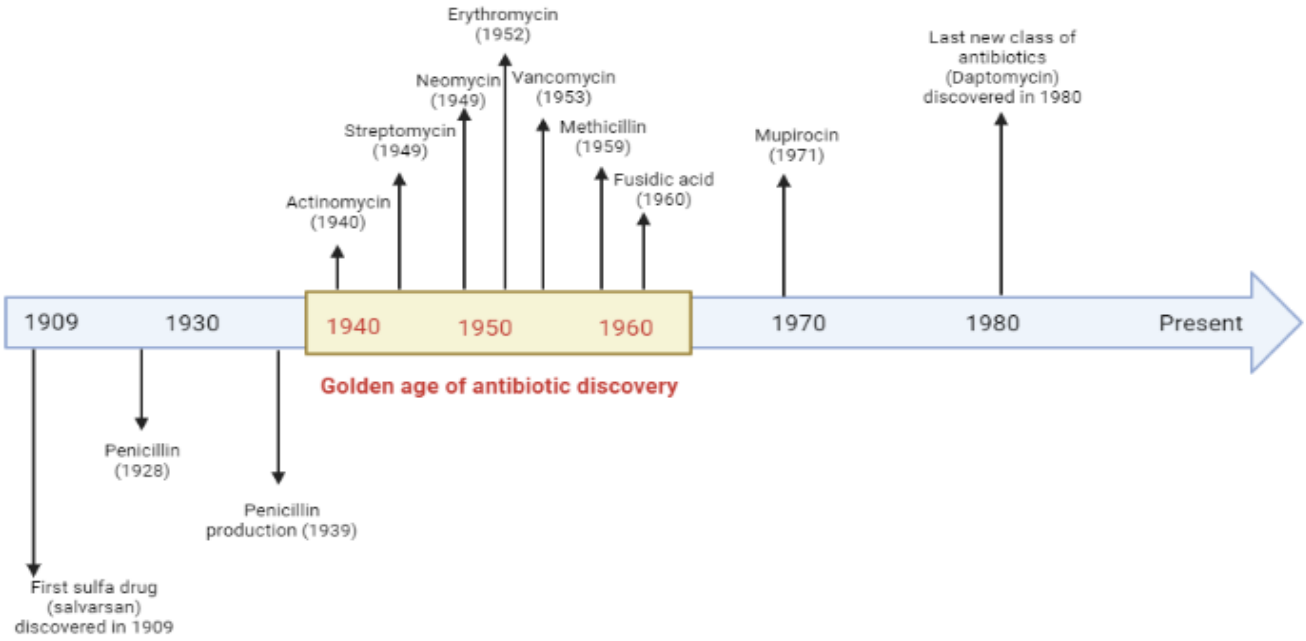


Figure 1.1 The timeline of antibiotic discovery.

The year represents the year of discovery of antibiotic. This figure was created using Biorender.

1.1.2 New approaches to discover antibiotics

In the golden age of antibiotic discover, new classes of antibiotics were being discovered by isolating antibiotic-producing bacteria from soil samples using culture-based techniques (Lewis, 2013). New approaches and sophisticated technology, together with the identification of new antibiotic-producing bacteria in previously under-explored environments, have contributed in the discover of new antibiotics in recent years (Miethke *et al*., 2021).

Innovative culturing techniques such as iChip, has led to the discovery of a new cyclic antibiotic teixobactin, able to inhibit cell wall synthesis of Gram positive bacteria (Piddock, 2015). This novel antibiotic was produced by a previously unculturable bacterium, *Eleftheria terrae* (Piddock, 2015). Recently, artificial intelligence has yielded halicin, a new antibiotic active against *Acinetobacter baumannii* (Stokes *et al*., 2020). Computational approaches were also used to discover the novel antibiotic, macolacin, which was able to inhibit multidrug-resistant *A. baumannii* (Wang *et al*., 2022a). The most recent class of antibiotics, named GSK3036656, is able to inhibit *Mycobacterium tuberculosis* by targeting the enzyme leucyl-tRNA synthetase (Carvalho, 2023). It was discovered by exploring boron-based compounds able to inhibit tRNA synthetase (Li *et al*., 2017).

Most antibiotics are produced by soil bacteria such as members of *Actinomyces* and *Lysobacter* (Butler *et al*., 2020). Even though soil is still receiving more attention as a source of antibiotic-producing bacteria, other sources, such as water, are also crucial in the discovery of novel antibiotics (Hamamoto *et al*., 2021, Piddock *et al*., 2015). It is well known that certain taxonomically diverse bacteria groups only found in marine environments are capable of producing novel secondary metabolites not observed in terrestrial microorganisms and possess distinctive physiological and structural traits that allow them to survive in harsh

environmental conditions (Radajewski *et al*., 2002). The aquatic environment is currently recognized as an unexplored source of unique natural products that deserves more attention (Tawiah et al., 2012).

Novel antibiotics have also been discovered in untapped human microbial environments, such as the anterior nares and the vaginal cavity (Donia *et al*., 2014, Zipperer *et al*., 2016). Lugdunin is a novel thiazolidine-containing cyclic peptide antibiotic isolated from *Staphylococcus lugdunensis* in the anterior nares (Zipperer *et al*., 2016). It is the first non-ribosomal peptide isolated from the human microbiome. Antibiotic-producing bacteria in underexplored habitats such as the anterior nares, combined with new tools for genome mining, can reinvigorate the natural product discovery field (Zipperer *et al*., 2016).

1.1.3 Development of a promising antibiotic

A compound with promising antimicrobial activity found in pre-clinical development would require extensive research and will ultimately move into clinical trials (Miethke *et al*., 2021). The main challenge is to optimize most of the drug's properties to be highly effective with minimal toxicity (Miethke *et al*., 2021). Such properties include selectivity, potency, cytotoxicity, and physiochemical parameters in addition to pharmacokinetics and pharmacodynamics (Miethke *et al*., 2021). To

prevent harmful therapeutic side effects, in vitro selectivity and cytotoxicity tests are crucial throughout the early discovery process (Dalhoff, 2021). In addition, damage to the microbiome is another consideration during drug development and can be modulated by selective drug design (Bhalodi *et al*., 2019). However, it is difficult to develop an antibiotic that fulfils all these requirements. Macolacin, a newly discovered naturally occurring colistin congener, proved extremely effective in animal tests against opportunistic bacteria such as colistin-resistant *Acinetobacter baumannii* (Wang *et al*., 2022a). However, it was too toxic for clinical use (Wang et al., 2022a). Moreover, the World Health Organization's (WHO) yearly analysis showed that compared to 31 antibiotics in 2017, there were only 27 novel ones in clinical development against priority pathogens in 2021. Only 6 of the 27 antibiotics in the clinical pipeline met at least one of the WHO's criteria for innovation (WHO, 2021).

1.2 Gram positive and Gram negative bacteria

Bacteria are usually protected from hostile environments by their cell envelope (Silhavy *et al*., 2010). The cell envelope is a complex multilayer structure by which most bacteria fall into two groups; Gram positive and Gram negative (Silhavy *et al*., 2010). The main difference between these two groups of bacteria is the thickness of the peptidoglycan layer (Beveridge, 2001). Apart from this structural difference, these two groups of bacteria are also molecularly different (Wang *et al*., 2022b). Gram

positive bacteria are genotypically a distinct group of bacteria (Kunder *et al*, 2015, Li *et al*., 2012b). Gram negative bacteria on the other hand, are genetically diverse and heterogenous (Oliveira and Reygaert, 2022).

Gram positive bacteria have a cell wall, mainly composed of thick layers of peptidoglycan and no outer lipid membrane (Figure 1.2) (Brown *et al*., 2013). Gram positive bacteria also possess a large volume of periplasm (Seltmann and Holst, 2002). The peptidoglycan layer is interspersed with wall-teichoic acids (WTA) (Brown *et al*., 2013). In addition to WTA, lipoteichoic acids are also present in the peptidoglycan layer of Gram positive bacteria, anchored to the cytoplasmic membrane via glycolipids (Ward, 1981). The cytoplasmic lipid membrane, or plasma membrane, is composed of a phospholipid bilayer (Barák and Muchová, 2013). This phospholipid bilayer is composed of two layers of phospholipid molecules, with a polar (hydrophilic) head and a nonpolar (hydrophobic) tail (Barák and Muchová, 2013). Protein molecules are embedded through this lipid bilayer (Pajerski *et al*., 2019).

Gram negative bacteria display three principal layers in the cell envelope; the outer membrane, a thin peptidoglycan layer and a cytoplasmic membrane (Glauert and Thornley, 1969). This outer membrane provides Gram negative bacteria resistance to some antibiotics. The outer

membrane is composed of surface proteins, lipoproteins, and lipopolysaccharide (LPS), consisting of lipid A and O antigen (Raetz and Whitfield, 2002). Gram negative bacteria have porin proteins in their outer membrane, providing access to certain substances including nutrients and water (Nikaido, 2003). Unlike Gram positive bacteria, Gram negative bacteria do not have teichoic acid present in their cell wall (Silhavy *et al*., 2010). The inner cytoplasmic membrane of Gram negative bacteria is mainly composed of phospholipids that acts as an electrochemical barrier (Silhavy *et al*., 2010).

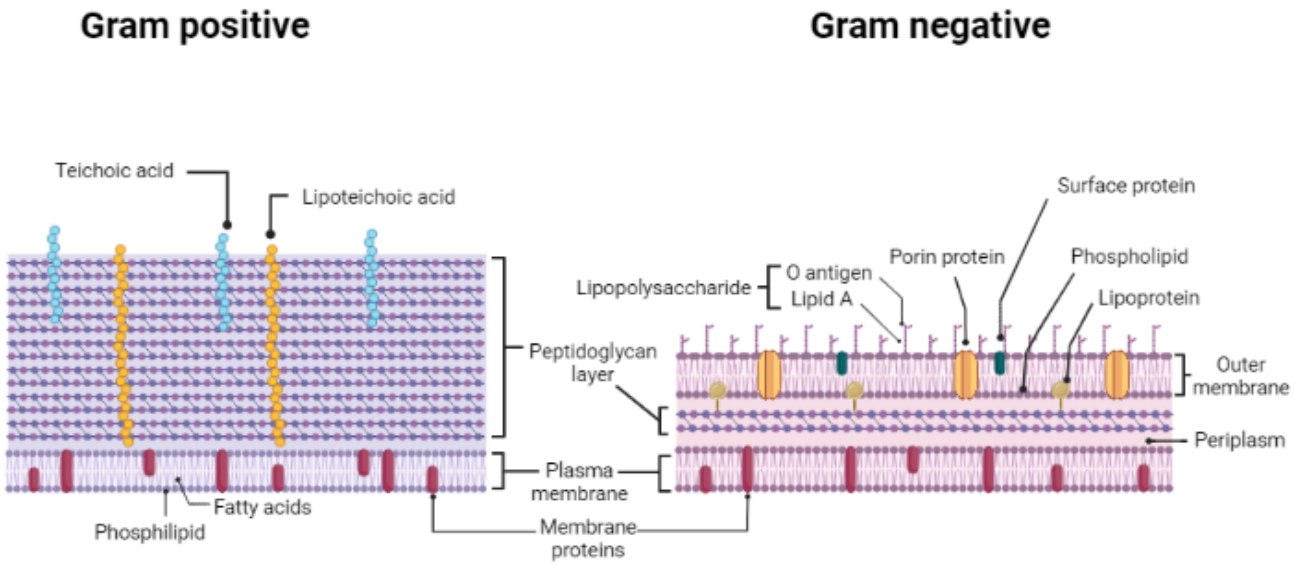


Figure 1.2 Differences between the cell wall of Gram positive and Gram negative bacteria.

A schematic figure comparing Gram positive (left) and Gram negative (right) bacterial cell membranes. This figure is based on Epand *et al*., 2016 and Pajerski *et al*., 2019. This figure was created using Biorender.

1.3 Modes of action of antibiotics

Most antibiotics fall into four main categories based on their site of activity (Thrum, 1977).

Table 1.1 Major antibiotic families and their modes of action

Mechanism of action	Antibiotic families
Inhibitors of cell wall synthesis	Penicillins (methicillin, oxacillin); carbapenems (impenem, meropenem); cephalosporins (cefalexin, cefoxitin); monobactams (aztreonam); glycopeptides (vancomycin, teicoplanin)
Inhibitors of protein synthesis	Aminoglycosides (streptomycin, gentamicin); tetracyclines (doxycycline, minocycline); macrolides (azithromycin)
Inhibitors of DNA synthesis	Fluoroquinolones (levofloxacin, ciprofloxacin), metronidazole
Inhibitors of folate pathway	Sulfonamides (sulfadoxine); trimethoprim
RNA synthesis inhibitors	Rifampicin, streptolydigin

List of major antibiotic families with some examples and their mechanisms of action

The categories include disruption of cell membrane integrity, inhibition of protein synthesis, inhibition of nucleic acid metabolism and inhibition of cell wall synthesis (Figure 1.3) (Hutching *et al*., 2019).

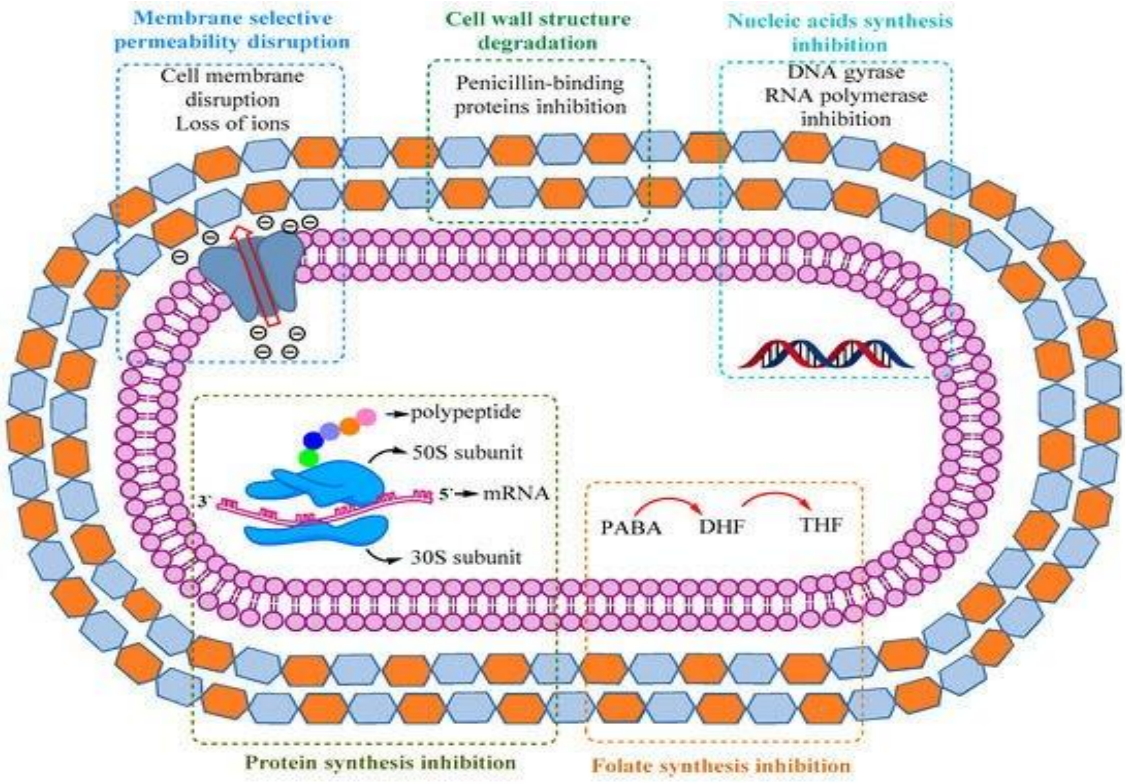


Figure 1.3 Mechanism of action of antibiotics.

Five basic mechanisms of antibiotic resistance include inhibition of cell wall synthesis, disruption of cell membrane integrity, inhibition of nucleic acid synthesis, inhibition of protein synthesis and inhibition of folate synthesis. This figure was adapted from Chiș *et al*., 2022.

1.3.1 Disruptors of cell membrane integrity

Some antibiotics disrupt the charge of phospholipids in the bacterial cell membrane (Velkov *et al*., 2013). The best known examples include polymyxin and daptomycin (Epand *et al*., 2016). Polymyxin, otherwise known as colisin, is a lipopeptide antibiotic that binds to lipid A in LPS in Gram negative bacteria (Martis *et al.*, 2014, Gogry *et al*., 2021). It displaces calcium and magnesium from the negatively charged phosphate groups of lipid A, leading to cell leakage and bacterial cell death (Gogry *et al*., 2021). However, colistin is considered as a last resort treatment due to its side effects such as renal toxicity (Spapen *et al*., 2011).

Daptomycin is another lipopeptide antibiotic that disrupts the cell membrane (Jung *et al*., 2004). This calcium-dependent antibiotic interacts with the negatively charged phospholipid (Jung *et al*., 2004). Gram negative bacteria have a lower content of negatively charged phospholipid than that of Gram positive bacteria (Epand, 2007). Due to this feature, daptomycin is ineffective against Gram negative bacteria (Tally *et al*., 1999). Daptomycin inserts into the cell membrane, where it then aggregates (Howe and Sofou, 2021). This leads to membrane leakage and leakage of ions, causing rapid depolarization and therefore loss of membrane integrity (Pogliano *et al*., 2012). Discovered in the early 1980s, daptomycin is the last novel structural class of cyclic lipopeptide antibiotics to be discovered (Tally and DeBruin, 2000).

1.3.2 Inhibitors of cell wall biosynthesis

Some antibiotics can inhibit the synthesis of bacterial cell wall. This type of inhibition is through three main mechanisms: inhibition of peptidoglycan synthesis, disruption of peptidoglycan cross-linkage and disruption of precursor movement. Peptidoglycan consists of glycan strands of repeating N-acetylglucosamine (NAG) and N-acetylmuramic acid (NAM) residues cross-linked via peptide side chains (Mihelič *et al*., 2017). The cell wall provides bacteria with mechanical strength and is correlated with structural integrity of the cell (Yadav *et al*., 2018).

β-lactam antibiotics (including penicillins, carbapenems and cephalosporins) are amongst the most common class of antibiotics that interfere with the biosynthesis of peptidoglycan (Aoki and Okuhara, 1980). β-lactam antibiotics have a β-lactam ring that mimics the peptidoglycan subunit component, D-alanyl D-alanine, the terminal amino acid residues on the precursor NAM/NAG peptide subunit of the peptidoglycan (Pandey and Cascella, 2022). β -lactams inhibit the peptide bond formation reaction catalyzed by transpeptidases, also known as penicillin binding proteins (PBP). This will therefore disrupt synthesis of new peptidoglycan, leading to lysis of the bacterium (Zeng and Lin, 2013).

Glycopeptide antibiotics are another major group of agents that inhibit cell wall synthesis. Glycopeptide antibiotics include vancomycin and teicoplanin (van Bambeke *et al*., 2017). Glycopeptides work by binding to the D-alanyl D-alanine terminus of the peptide side chain of the precursor peptidoglycan subunit, inhibiting peptidoglycan cross-linking and polymerization leading to inhibition of cell wall synthesis (van Bambeke *et al*., 2017). Gram negative bacteria are inherently resistant to these types of antibiotics due to the impermeability of the LPS membrane to the large glycopeptide molecules (Yarlagadda *et al*., 2016).

Some antibiotics block the movement of precursors required for the synthesis of peptidoglycan. Bacitracin is an antibiotic that prevents transport of peptidoglycan subunits from the cytoplasm to the cell wall through inactivation of the phospholipid carrier, bactoprenol (Scherer *et al*., 2013). Bactoprenol transports the peptidoglycan monomers through the cell membrane as they are inserted into the existing peptidoglycan (Münch *et al*., 2014). Bacitracin can sequester this lipid carrier, interrupting the flow of peptidoglycan precursors to the site of cell wall synthesis (Storm, 1974).

1.3.3 Inhibitors of folate pathway

Some classes of antibiotics inhibit the folate pathway responsible for the synthesis of folic acid (Bourne, 2014). Disruption of folic acid synthesis can be through inhibition of the enzyme dihydrofolate reductase (Wróbel *et al*., 2020). Such antibiotic classes include trimethoprim, which is commonly used as a first-line antibiotic for the treatment of adult urinary tract infections (UTI) (Crellin *et al*., 2018).

Folic acid inhibition can also be through substrate competition with P-aminobenzoic acid (PABA), preventing folic acid synthesis (Carter *et al*., 2007). Such group of antibiotics includes sulfonamides, which are competitive inhibitors and are structural analogues of PABA (Bernstein, 1982). They are an important class of antibiotics with a broad spectrum of activity against Gram positive and some Gram negative bacteria including *Klebsiella* and *Enterobacter* species (Ovung and Bhattacharyya, 2021).

1.3.4 Inhibitors of nucleic acid synthesis

1.3.4.1 Inhibitors of DNA replication

Bacterial deoxyribonucleic acid (DNA) replication involves multiple steps, starting with the opening of the double helix and separation of the DNA

strands. Briefly, the two strands are separated by helicase, which is a protein that breaks apart the hydrogen bonds between bases on the DNA strand (Lohman, 1992). This will separate the strands into a Y shaped replication fork, which is composed of a leading strand and a lagging strand (Lohman, 1992). DNA primase generates an RNA primer which provide DNA polymerase the required platform to start copying a DNA strand (Mott and Berger, 2007). DNA polymerase then extends these primers into new DNA strands (Prozorov, 2005). DNA polymerase attaches to the leading DNA strand and starts to attach new complementary nucleotides to the 3' end (Prozorov, 2005). DNA synthesis on this strand occurs continuously, unlike the lagging strand (Prozorov, 2005). On the lagging strand, DNA is synthesized discontinuously, generating a series of small fragments of new DNA (okazaki fragments) in the 5' to 3' direction (Oakley, 2019). These okazaki fragment are later joined to form a continuous chain of nucleotides (Oakley, 2019). DNA polymerase later removes the RNA primers and replaces them with new nucleotides (Oakley, 2019). Then, DNA ligase joins the disconnected Okazaki fragments of the lagging strand, leading a continuous DNA molecule (Prozorov, 2005).

Inhibitors of DNA replication include antibiotics such as metronidazole, which chemically modify DNA (Moreno *et al*., 1983). It inhibits nucleic acid synthesis by forming free radicals (Moreno *et al*., 1983). Metronidazole is

partially reduced by ferredoxin, creating a concentration gradient in the cell which promotes the formation of cytotoxic free radicals, which interact with bacterial DNA, causing strand breakage which will destabilize the DNA helix (Edwards *et al*., 1986).

Other antibiotics interfere with the enzymes associated with DNA replication. Quinolone class of antibiotics work by inhibiting DNA gyrase (topoisomerase II) activity (Fàbrega *et al*., 2009). Quinolones bind to the A subunit of the enzyme at the single strand ends of the DNA chain, causing inability of topoisomerase II to reseal the DNA (Aldred *et al*., 2014). This results in highly fragmented chromosome, which will therefore inhibit DNA synthesis (Malik *et al*., 2006).

1.3.4.2 Inhibitors of RNA polymerase

Antibiotics that inhibit enzyme function in nucleic acid synthesis can act on RNA polymerase, supressing the initiation of RNA transcription (Severinov, 2000). The best known is rifamycin, which acts by selectively binding to the β-subunit of bacterial RNA polymerase, inactivating mRNA synthesis (Mosaei *et al*., 2020). Rifamycins are the first-line antituberculosis agents and are also used in treatment of infections caused by Gram positive and Gram negative bacteria (Ho *et al*., 2009).

Other classes of antibiotics that inhibit RNA polymerase include sorangicin and streptolydigin (Ho *et al*., 2009).

1.3.5. Inhibitors of protein synthesis

Some classes of antibiotics exert their antimicrobial action by blocking the steps of protein synthesis that occur either on the 30S or 50S ribosomal subunit of the 70S ribosome (Figure 1.4). The three major families of antibiotics that affect protein synthesis are macrolides, tetracyclines and aminoglycosides (Leviton, 1999).

1.3.5.1 Inhibitors of 50S subunit

Antibiotics affecting early stages of protein synthesis include macrolides including azithromycin, clindamycin and erythromycin (Svetlov *et al*., 2021). They inhibit translocation by binding to the 50S subunit of bacterial ribosome at the peptidyl transferase center formed by the 23S rRNA, therefore inhibiting the elongation of peptide chains (Figure1.4) (Saini and Kumar, 2021).

1.3.5.2 Inhibitors of 30S subunit

Inhibitors of the 30S ribosome work by blocking the access of aminoacyl-tRNAs to the ribosome (Brodersen *et al*., 2000). Aminoglycoside

antibiotics for instance work by interacting with the 16S rRNA of the 30s subunit, giving rise to inaccurate mRNA translation and therefore biosynthesis of proteins with altered amino acid compositions (Cavallo and Martinetto, 1981). This subclass of antibiotics causes misreading and premature termination of translation mRNA (Cavallo and Martinetto, 1981).

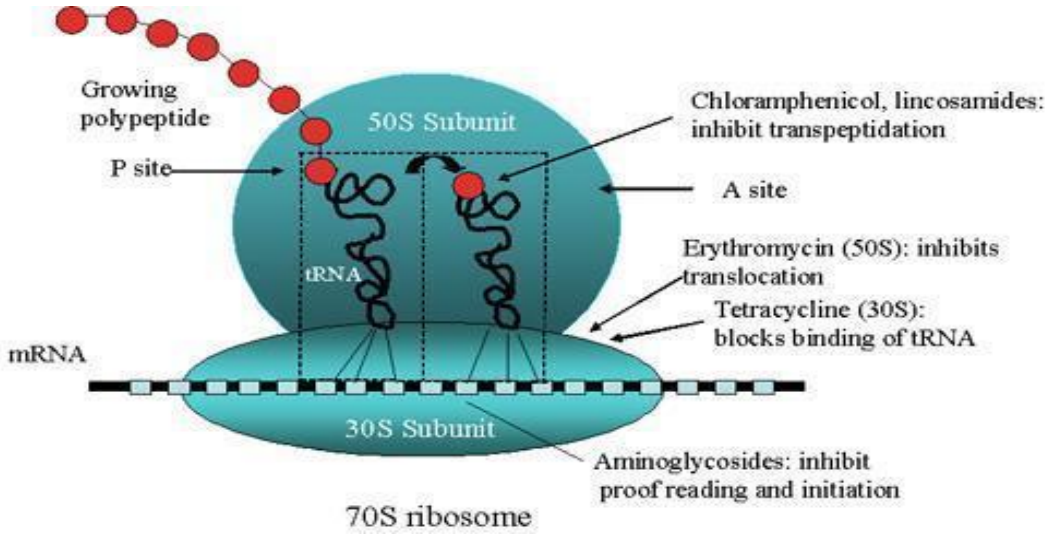


Figure 1.4 Activity of antibiotics that affect protein synthesis in bacteria.

The three steps in proteins synthesis include aminoacyl-tRNAs synthesis, attachment of tRNA molecules to the corresponding amino acids followed by transport to the ribosome, and initiation of ribosomal protein synthesis. Aminoglycosides bind to the 30S ribosome, affecting translation, while tetracyclines inhibit binding of tRNA to ribosomes. Erythromycins inhibit translocation by binding to 50S and block the exit of growing peptide chain, Chloramphenicol inhibits transpeptidation and translation by binding to the 50S subunit, inhibiting the action of peptidyl transferase, preventing peptide bond formation.

This figure was adapted from McDermott *et al*., 2003.

1.4 Antibiotic resistance

Infections caused by antibiotic resistant bacteria have become a global concern, with the WHO launching a Global Action Plan on antimicrobial resistance in 2015 (WHO, 2015). The WHO published a priority list of pathogens with multidrug resistance that cause deadly infections in healthcare settings (WHO, 2017). These pathogens are of concern because they are increasingly resistant to antibiotics and the development of new effective antibiotics is required to treat infections caused by these pathogens (WHO, 2017). Recently, researchers estimated 4.95 million deaths associated with antimicrobial resistance (AMR) including 1.27 million deaths attributable to bacterial AMR (Murray *et al*., 2022). This research has also suggested six pathogens, including *Escherichia coli*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Streptococcus pneumoniae*, *Acinetobacter baumannii* and *Pseudomonas aeruginosa*, to be responsible for more than 250,000 deaths associated with AMR in 2019 (Murray *et al*., 2022).

Antibiotic resistance is considered as a direct response to the exploitation of antibiotics in healthcare settings, animal industry and agriculture (Milani *et al*., 2019). It falls into two broad categories; intrinsic resistance and acquired resistance (Reygaert, 2018).

1.4.1 Intrinsic antibiotic resistance

Intrinsic antibiotic resistance refers to the innate ability of bacteria to resist the action of antibiotics due to the bacterial cell structure or functional characteristics (Cox and Wright, 2013). It is a type of resistance that is typical of all members of a bacterial species or genus (Woodford and Ellington, 2007).

One of the bacterial mechanisms involved in intrinsic resistance is the reduced outer membrane permeability (Uddin *et al*., 2021). Compared to Gram positive bacteria, Gram negative bacteria are intrinsically less permeable to certain antibiotics due to the permeability barrier formed by their outer membrane (Blair *et al*., 2015). Glycopeptide antibiotics such as vancomycin, are not effective against Gram negative bacteria due to lack of penetration through the outer membrane (Dhanda *et al*., 2019). Daptomycin is ineffective against Gram negative bacteria due to the lower composition of anionic phospholipids in the cytoplasmic membrane of Gram negative bacteria (Mishra *et al*., 2012). This feature therefore reduces the Ca^{2+}-mediated insertion of daptomycin into the cytoplasmic membrane of Gram negative bacteria (Kotsogianni *et al*., 2021).

Antibiotics are also inactivated by hydrolyzing enzymes, including β-lactamases, which hydrolyse the β-lactam ring present in the chemical

structure of β-lactam antibiotics (Davin-Regli and Pagès, 2015). Some enterococci are intrinsically resistant to β-lactams due to low affinity of their PBPs to β-lactam antibiotics and production of β-lactamases (Fontana *et al*., 1990).

Antibiotics can be actively transported out of the cell by bacterial efflux pumps (Van Bambeke *et al*., 2000). An example of this type of natural resistance mechanism is the AcrAB/TolC efflux pump in *Escherichia coli*, which cause this bacterium to be innately resistant to antibiotics such as tetracyclines, fluoroquinolones and macrolides (Piddock, 2006).

1.4.2 Acquired antibiotic resistance

Bacteria can become resistant to antibiotics by incorporating new genetic material (van Hoek *et al*., 2011). Acquired resistance mechanisms are obtained by transduction, conjugation and transformation, which are all termed as horizontal gene transfer (HGT) (Lerminiaux and Cameron, 2019). Mobile genetic elements including plasmids and integrative conjugative elements, are capable of intercellular transfer. Transposons and integrons depend on other mobile genetic element for inter-cellular transfer (Khedkar *et al*., 2022). Another form of acquired resistance is through mutations in the chromosome (Woodford and Ellington, 2007).

Conjugation is a cell to cell contact mechanism by which DNA transfer from donor to recipient cell (Thomas and Nielsen, 2005). It is the most common form of HGT (Huddleston, 2014). Transduction involves the transfer of DNA from donor to recipient cell via bacteriophages (Jiang and Paul, 1998). Transformation refers to the ability of the recipient cell to take up extracellular DNA (Thomas and Nielsen, 2005). Extracellular DNA can originate from the secretion of live cells or from the lysis of dead cells (Vlassov *et al,.* 2007) (Figure1.5).

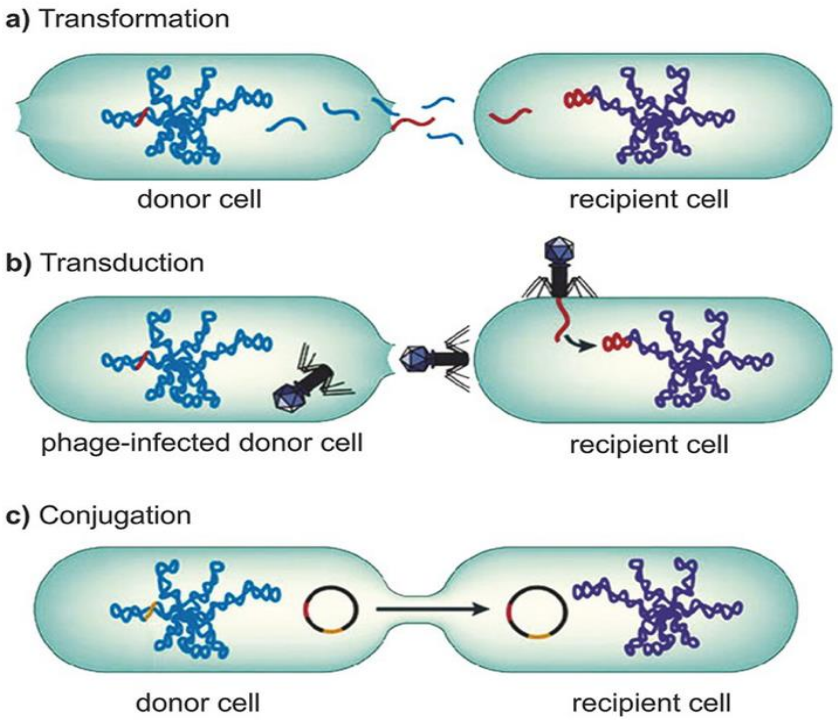


Figure 1.5 Mechanisms of horizontal gene transfer

a) Transformation occurs when the recipient bacterium uptakes free DNA from the environment and incorporates in into its chromosome b) Transduction occurs when bacterial DNA is being transferred by bacteriophages from the donor bacterium to the recipient bacterium c) Conjugation occurs when DNA transfer happens between bacteria through direct contact.

This figure was adapted from Fusco *et al*., 2019.

HGT contributes to the rising problem of antimicrobial resistance (Pilla *et al*., 2017). Microbes can share and sample a large gene pool through HGT, which can encode traits that can help them adapt in their environment (Pilla *et al*., 2017). HGT also contributes to spread of infections and outbreaks caused by antibiotic resistant bacteria (Hiller et al., 2010).

Plasmid-mediated resistance to β-lactam antibiotics is a classic example of how HGT contributes to the challenges of antimicrobial resistance (Lorme *et al.,* 2018). Plasmid-carried-β-lactamase genes, such as extended spectrum β-lactamases (ESBLs), confer resistance by hydrolysing β-lactam antibiotics, including penicillin, carbapenems and cephalosporins (Paterson and Bonomo, 2005). These genes disseminate by conjugation in bacteria such as *Acinetobacter*, *Pseudomonas* and *Enterobacteriaceae* (Kizny Gordon *et al*., 2017). Another prominent example for HGT through acquisition of antimicrobial resistance genes is vancomycin-resistant *Staphylococcus aureus* (VRSA). VRSA evolved from MRSA through the acquisition of a plasmid from enterococci (Gardete and Tomasz, 2014). The plasmid carries a *VanA* operon, which causes

structural change of peptidoglycan precursors, leading to lower binding capability of vancomycin (Cox and Wright, 2013).

Natural transformation contributes to the spread of antimicrobial resistance in many pathogens including clinically relevant antibiotic resistant bacteria (Johnston *et al*., 2014). These pathogens include *Pseudomonas*, *Acinetobacter*, *Staphylococcus* and *Streptococcus* (Johnston *et al*., 2014). Transduction by bacteriophages is another contributor to the spread of antimicrobial resistance genes by HGT. An example of this mechanism involves tetracycline and penicillin resistance genes, which are transduced between hospital isolates of *S. aureus* (Mašlaňová *et al*., 2016). Transduction in Gram negative bacteria has also been observed, where antimicrobial resistance genes, including ESBLs, transduce between *Pseudomonas* isolates (Blahová *et al*., 2000).

In addition to the mechanisms of HGT, spontaneous gene mutation in the chromosome also contributes to acquired antibiotic resistance (Woodford and Ellington, 2007). During bacterial cell replication, some bacteria develop mutations that make them resistant to antibiotics. Mutation based antibiotic resistant phenotypes are numerous, with some bacteria becoming resistant to sulfonamides and trimethoprim due to single nucleotide polymorphisms (SNPs) sequence (Huovinen *et al*., 1995). SNP

is a genetic variation at a single position in a DNA sequence (Huovinen *et al*., 1995). They may cause mutational changes that alter the proteins encoded by the genes, which can modify the proteins activity (Woodford and Ellington, 2007). Mutations can occur spontaneously or through exposure to mutagens such as chemicals and radiation (Bose, 2016). Some mutational changes for instance cause upregulations of multidrug efflux pumps, while other mutations change the outer membrane permeability barrier, both of which in turn increase resistance to antibiotics (Lister *et al*., 2009).

1.5 Nosocomial infections

'Nosocomial' or 'healthcare associated infections' are infections that appear in a patient undergoing medical care in the hospital or healthcare facility that were not present prior to a patient entering the hospital (Haque *et al*., 2018). These types of infection typically occur within 48 hours of hospital admission or 3 days after discharge of the patient or 30 days post operation (Horan *et al*., 2004, Haque *et al*., 2018).

Hospital acquired infections are caused by bacteria, fungi, viruses and parasites, with bacteria being the most common (Khan *et al*., 2017). Even though some bacteria belong to the normal human microbiome, they can

cause infection when the patient is immunocompromised, residing in the Intensive care unit (ICU) or is undergoing surgery (Vincent *et al*., 2009).

1.6 The ESKAPE pathogens

The most prevalent nosocomial pathogens are termed as the ‘‘ESKAPE’’ pathogens (Peters *et al*., 2008). This is an acronym for a group of bacteria representing both clinically significant Gram positive and Gram negative bacteria, made up of *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa* and *Enterobacter* species (Oliveira *et al*., 2020). Infections caused by the ESKAPE pathogens are a major global cause of morbidity and mortality (Denissen *et al*., 2022). The ESKAPE pathogens also represent 42.2% of species isolated from blood stream infections compared with pathogens not related to the ESKAPE ones (Marturano and Lowery, 2019). Moreover, infections caused by the ESKAPE pathogens are a leading cause of mortality worldwide (Murray *et al*., 2022). This group are increasingly resistant to many antibiotics and are increasingly prevalent in hospitals (Arbune *et al*., 2021).

1.6.1 *Enterococcus faecium*

Enterococci are Gram positive facultative anaerobic cocci, with *E. faecium* and *E. faecalis* being the most clinically significant (Arias and Murray,

2012). *E. faecium* is normally found in the gut as a part of the normal microbiome, however opportunistic infections due to this bacterium are common (Yang *et al*., 2009, Gao *et al*., 2018).

E. faecium has recently emerged as a significant threat in hospitals worldwide (van Hal *et al*., 2021). Many reports have revealed a significant increase in vancomycin-resistant enterococci (VRE), especially *E. faecium*, in health care settings (Mac *et al*., 2019, Büchler *et al*., 2022).The operons related to vancomycin resistance for enterococci include *VanA*, *VanB*, *VanC*, *VanD* and *VanE* (Ahmed and Baptiste, 2017). The major genotype for acquired vancomycin resistance in *E. faecium* is *VanA* (Werner *et al*., 2008). *VanA* substrates are d-alanine and d-lactate, which lead to the formation of depsipeptide d-Ala-d-Lac (Ahmed and Baptiste, 2017). Vancomycin resistance is caused due to altered peptidoglycan terminus (from the usual d-Ala-d-Ala to d-Ala-d-Lac) leading to reduced vancomycin binding and therefore, failure to prevent cell wall synthesis (Murray, 2000, Eliopoulos and Gold, 2001) (Figure 1.6). One alarming feature of VRE is its ability to transfer vancomycin resistant genes to other bacteria such as *S. aureus* (Leclercq *et al*., 1988, Cong *et al*., 2020). VRSA isolates involve the horizontal gene transfer of Tn1546 transposon from VRE to *S. aureus* by conjugation (de Niederhäusern *et al*., 2011).

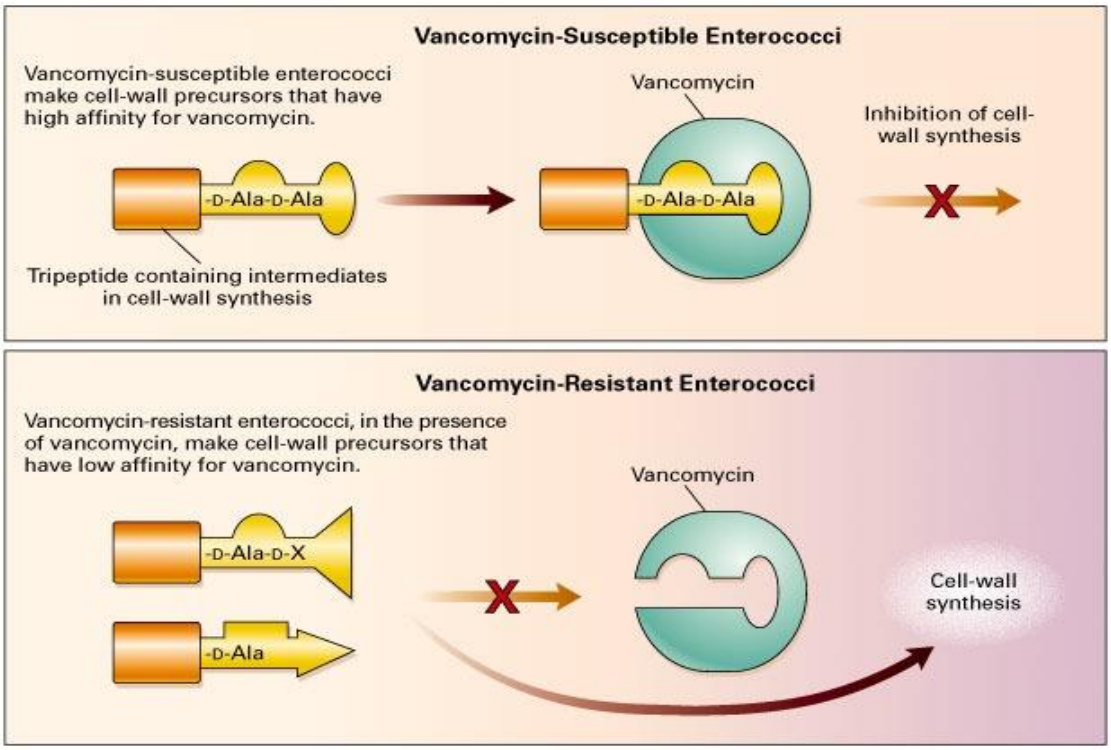


Figure 1.6 Schematic diagram of vancomycin resistance in vancomycin-resistant enterococci (VRE)

After translocation from the cytoplasm to the cell surface, cell-wall precursors ending in d-Ala-d-Ala in vancomycin-susceptible enterococci bind to vancomycin with high affinity. Once bound, cell wall synthesis is inhibited. Vancomycin-resistant enterococci (VRE) generate precursors with an altered peptidoglycan terminus (d-Ala-d-Lac, d-Ala or d-Ala-d-Ser), which has low affinity for vancomycin. Ala; alanyl or alanine, x; lactate for *VanA*, *VanB* and *VanD* types of resistance and Serine for *VanC* and *VanE* types of resistance. This figure was adapted from Murray, 2000.

1.6.2 *Staphylococcus aureus*

Staphylococcus aureus is a Gram positive, catalase positive, facultatively anaerobic bacterium (Rasigade and Vandenesch, 2014). It is both a commensal and pathogenic bacterium to humans, causing both hospital-acquired and community-acquired infections (Krismer *et al*., 2017, Wang *et al*., 2019, Rehman *et al*., 2020).

S. aureus is considered as a part of the normal human microbiome, colonizing the skin and upper respiratory tract with the anterior nares being the primary reservoir (von Eiff *et al*., 2001, Krismer *et al*., 2017). However, emergence and spread of MRSA is causing a serious global health threat and is considered as one of the top priorities for public health systems (Zhen *et al*., 2020, Jian *et al*., 2021). It is associated with high mortality rates, highlighting the desperate need to control bacterial antibiotic resistance (Sakamoto *et al*., 2021).

Some strains of *S. aureus* are resistant to methicillin due to the expression of penicillin-binding-proteins (PBP2a), which has low affinity for all β-lactam antibiotics (Shalaby *et al*., 2020). MRSA strains are characterized by the presence of *mecA* gene, which encodes this alternative PBP (Fishovitz *et al*., 2014). In addition, another mechanism of resistance of *S.*

aureus to β -lactam antibiotics is the production of penicillinases or β -lactamases which inactivate the antibiotic (Franciolli *et al*., 1991).

Cell wall synthesis in MRSA can be inhibited by glycopeptide antibiotic vancomycin, targeting the D-alanyl-D-alanine residues of peptidoglycan precursors (Barna and Williams, 1984). However, vancomycin intermediate *Staphylococcus aureus* (VISA) and VRSA infections are increasing (Shariati *et al*., 2020).

1.6.3 *Klebsiella pneumoniae*

Klebsiella pneumoniae is a non-fastidious, Gram negative rod-shaped bacterium that is the member of the family *Enterobacteriaceae* (Santajit and Indrawattana, 2016). This bacterium is a serious global pathogen commonly causing UTIs and pneumonia (Chapelle *et al*., 2021, Mohd Asri *et al*., 2021). An alarming feature of *K. pneumoniae* is its ability to rapidly disseminate in hospital environments (Malinga *et al*., 2022).

Many *K. pneumoniae* isolates have acquired a variety of genes for β-lactamase enzymes (Girlich *et al*., 2012). These enzymes include ESBLs, carbapenem-hydrolysing enzymes (Gupta *et al*., 2003). *K. pneumoniae* isolates can also become resistant to antibiotics due to the loss of porin

proteins (Hamzaoui *et al*., 2018). Loss of OmpK53, a major porin in *K. pneumoniae*, contributes to the development of multiple-drug resistant strains of *K. pneumoniae* (Sugawara *et al*., 2016). Moreover, *K. pneumoniae* can expel the drugs from the cell at a high rate by efflux pumps such as resistance-nodulation-division (RND)-type efflux pumps (Ni *et al*., 2020).

Infections caused by Gram negative bacteria, including *K. pneumoniae*, are usually treated with colistin, a last line treatment for multi-drug resistant Gram negative bacteria (El-Sayed Ahmed *et al*., 2020). However, colistin-resistant *K. pneumoniae* isolates are emerging (Di Tella *et al*., 2019). Colistin works by targeting the phosphate portion of the bacterial LPS, leading to disruption of negative charge of the outer membrane and cell death (Trimble *et al*., 2016). Resistance of *K. pneumoniae* to colistin is due to covalent modification of the bacterial LPS, leading to decreased affinity to colistin (Aghapour *et al*., 2019). These modifications are initiated by genetic mutations on chromosomal genes in addition to the acquisition of plasmid genes conferring colistin resistance (Liu *et al*., 2016, Poirel *et al*., 2017). One prominent example of these types of gene include the plasmid-mediated *mcr-1* gene (Figure 1.7) (Venter *et al*., 2017).

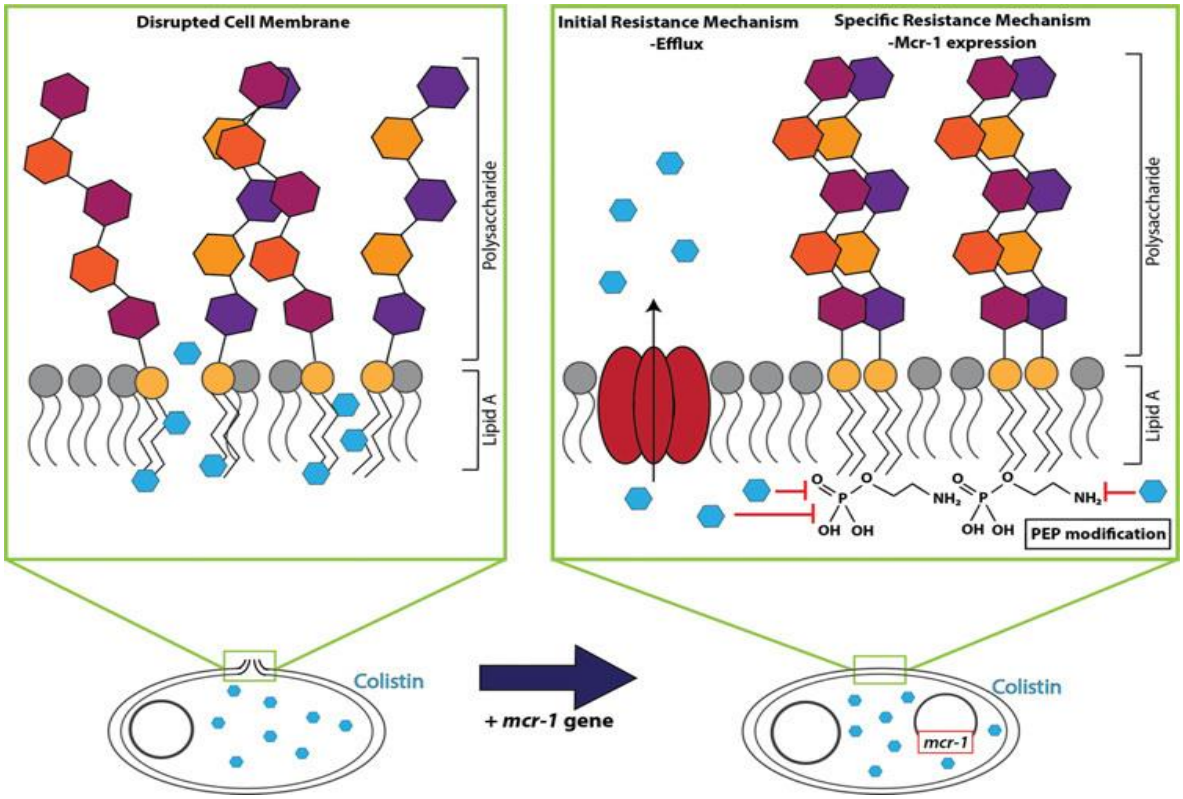


Figure 1.7 Mechanism of colistin resistance.

Colistin disrupts bacterial cell membrane by interacting with the lipid A portion of bacterial lipopolysaccharide. Colistin resistance is mediated by overexpression of drug efflux pumps. In addition, colistin resistance is facilitated through the acquisition of plasmid-mediated *mcr*-1 gene. This gene encodes a phosphoethanolamine transferase that modifies lipid A with a phosphoethanolamine (PEP) group. This will therefore prevent the interaction of colistin and lipid A.

This figure was adapted from Venter *et al*., 2017.

1.6.4 *Acinetobacter baumannii*

Acinetobacter baumannii is a Gram negative, rod-shaped bacterium commonly found in soil and is often isolated from hospital environments (Howard *et al*., 2012, Hrenovic *et al*., 2014). *A. baumannii* is a challenging pathogen in healthcare settings (Gottesman *et al*., 2021). This bacterium usually causes UTIs and respiratory tract infections (Jiménez-Guerra *et al*., 2018, Jain *et al*., 2019).

A. baumannii readily contaminates hospital environment, including hospital surface, equipment and indwelling device, suggesting that hospital surfaces serve as a reservoir for its transmission (Wieland *et al*., 2018). Transmission also frequently occurs via person-person contact, through hand contact of healthcare workers, visitors, and staff, contributing to the spread of this bacterium (Denton *et al*., 2004, Wieland *et al*., 2018).

This opportunistic pathogen is intrinsically resistant to some antibiotics due to the protection provided by the Gram negative outer membrane (Breijyeh *et al*., 2020). Carbapenem, a last resort antibiotic used for critically ill patients, is commonly prescribed for treatment of nosocomial infections caused by *A. baumannii* (Lima *et al*., 2019). Unfortunately, the global emergence of carbapenem-resistant *A. baumannii* (CRAB) is now

threatening human health and endangering healthcare systems (Gottesman *et al*., 2021).

Many mechanisms of CRAB have been described, including the horizontal acquisition of genes that code for the production of carbapenem-hydrolyzing enzymes (Hamidian and Nigro, 2019). These enzymes include Ambler class D (oxacillinases) or class B (metallo- β -lactamases). Some acquired oxacillinase enzymes include OXA-23, OXA-24, OXA-58 (Poirel *et al*., 2010). The main causes of carbapenem-resistance in *A. baumannii* is due to the genes that encode these enzymes (Mentasti *et al*., 2020). Genes encoding for intrinsic oxacillinase enzymes, such as intrinsic OXA (*oxaAb*), occur naturally in *A. baumannii* (Héritier *et al*., 2005). Other carbapenem resistance genes such as bla_{VIM}, bla_{IMP} and bla_{NDM}, encode metallo-β-lactamases in *A.baumannii* (Goudarzi *et al*., 2019). In addition to enzymatic-mediated resistance, alteration or loss of membrane-associated proteins such as OmpA, and modifications of efflux pumps, is associated to the resistance mechanism of CRAB (Coyne *et al*., 2011, Uppalapati *et al*., 2020).The combination of intrinsic resistance and multiple resistance mechanisms leads to difficulty in treating antibiotic resistant *A. baumannii*, further emphasizing the need for novel antimicrobials (Bartal *et al*., 2022).

1.6.5 *Pseudomonas aeruginosa*

Pseudomonas aeruginosa is a Gram negative rod-shaped bacterium that is found in soil, water and the environment (Wu and Li, 2015). *P. aeruginosa* commonly causes severe UTI (Jiménez-Guerra *et al*., 2018). This opportunistic pathogen can contaminate water networks in hospitals, making it one of the most problematic hospital-acquired pathogens (Hutchins *et al*., 2017). In the ICU, *P. aeruginosa* commonly causes ventilator-associated pneumonia and blood infections (Tumbarello *et al*., 2013, Dame *et al*., 2020).

P. aeruginosa shows very low permeability to antibiotics due to its outer membrane (Moradali *et al*., 2017). Another mechanism of intrinsic resistance in *P. aeruginosa* is the expression of efflux pumps and the production of antibiotic-inactivation enzymes, including ESBLs (Ferrer-Espada *et al*., 2019, Poole, 2011). *P. aeruginosa* develops resistance to antibiotics such as imipenem due to the reduction of the porin protein OprD (Li *et al*., 2012a). Some *P. aeruginosa* strains are becoming resistant to the last resort antibiotic, colistin (Dößelmann *et al*., 2017).

Another issue of antimicrobial resistance in healthcare associated *P. aeruginosa* is biofilm formation (Redfern *et al*., 2021). A biofilm is an aggregate of microbial communities that adhere to each other on a static

surface (Donlan, 2002). Biofilm are implanted in a matrix of extracellular polymeric substances, including proteins, metabolites, and extracellular DNA (Donlan, 2002). The matrix of biofilms provides a mechanical barrier, making it difficult to eliminate bacteria using antibiotics (Murray *et al*., 2017). Other pathogens found in biofilms of healthcare settings include *S. aureus*, *A. baumannii* and *K. pneumoniae* (Bales *et al*., 2013, Costa *et al*., 2019).

1.6.6 *Enterobacter* spp.

Enterobacter is a group of motile, facultatively anaerobic, Gram negative bacilli belonging to the family *Enterobacteriaceae* (Hormaeche and Edwards, 1960). *Enterobacter* spp. are commonly recovered from soil and exist as natural commensals in the human gut microbiome (Glushakova *et al*., 2022, Olsson *et al*., 2022). *E. cloacae* and *E.* aerogenes (now reclassified as *Klebsiella aerogenes*) are the most clinically important species from this genus (Tindall *et al*., 2017, Álvarez-Marín *et al*., 2021, Jeon *et al*., 2021) *Enterobacter* species commonly cause UTIs and respiratory infections (Álvarez-Marín *et al*., 2021, Ramirez and Giron, 2022). Bacteraemia is another clinical outcome seen in *E. aerogenes* and *E. cloacae*-associated infections, with a higher mortality caused by the latter (Jeon *et al*., 2021).

Infections due to *E. cloacae* are becoming difficult to treat due to the increased resistance to multiple antibiotics (Annavajhala *et al*., 2019). *E. cloacae* is intrinsically resistant to penicillins and second-generation cephalosporins (Davin-Regli and Pagès, 2015). Furthermore, colistin resistance in *E. cloacae* can be caused due to the overexpression of efflux pump *acrAB-tolC* (Telke *et al*., 2017). Overexpression is due to mutation in *AcrR* gene, the repressor of the AcrAB operon system (Pourahmad Jaktaji and Jazayeri, 2013). This bacterium is also resistant to carbapenem due to production of AmpC β-lactamase (Davin-Regli and Pagès, 2015). In addition, the decrease or loss of outer membrane proteins, including OmpA and OmpC, contribute to carbapenem resistance in *E. cloacae* (Mishra *et al*., 2020).

1.7 Aims of this study

The aims of this PhD project were to isolate antibiotic-producing bacteria from water sample in addition to the identification of an antibiotic-producing bacterium in the anterior nares using a bioinformatics-based approach. The objectives were:

1) To determine, by the use of various cultivation settings, the range of activity of five NW *Pseudomonas* sp. strains isolated from water against nosocomial pathogens.

2) To employ antiSMASH as a genome mining tool, to predict potential secondary metabolites with antimicrobial activities in the genome of NW *Pseudomonas* sp. strains.

3) To improve *Pseudomonas* sp. strain NW27's ability to produce antimicrobials utilizing ribosomal engineering techniques and to extract the active compound(s) using solvent extraction techniques

4) To interrogate the Integrated Microbial Genomes and Microbiome (IMG/M) database in an effort to identify antibiotic-producing bacteria in the anterior nares by searching for metagenomes with genes that are likely to code for proteins involved in antimicrobial production.

5) To evaluate the inhibitory activity of *Dolosigranulum pigrum* against *S. aureus* JE2 by cultivating in various media and to improve its antimicrobial potential by using ribosome engineering approaches.

Chapter 2.0

General Materials and Methods

2.0 General Materials and Methods

2.1 Bacterial strains and culture conditions

Indicator bacteria used in this study are mentioned in Table 2.1. All the indicators were grown using the conditions mentioned, unless stated otherwise.

Table 2.1 Indicator bacteria used in this study and their growth conditions.

Indicator Bacterium	Growth medium	Temp (°C)	Relevant characteristics	Phenotype of resistance[a]	Institute/Company
***Staphylococcus aureus* JE2**	LB	37	ST8 (CC8),laboratory strain CA-MRSA strain, Plasmid-cured derivative of USA300 LAC	MET, CIP, GM, VA	Network on Antimicrobial Resistance in *Staphylococcus aureus* (NARSA) by BEI Resources, NIAID, NIH
***Klebsiella pneumoniae* C3091**	LB	37	ST86, K2, 01, laboratory strain originally isolated from patient with UTI	AMP, TE, GM	(Oelschlaeger and Tall, 1997)
***Enterococcus faecium* E1162**	TS	37	ST17 (CC17), clinical strain, originally isolated from a patient with hospital acquired blood stream infection	AMP, TE, VA	American Type Culture Collection (ATCC)
***Acinetobacter baumannii* NCTC 12156**	NB	37	Laboratory strain, isolated from a patient with UTI	MER, IMP	National collection of Type Cultures (NCTC)
***Enterobacter cloacae* NCTC 10005**	NB	37	Laboratory strain, originally isolated from human cerebrospinal fluid	OX, AMP, TEC, CM	National collection of Type Cultures (NCTC)
***Pseudomonas aeruginosa* FRO1 (low mucoid)**	LB	37	Clinical strain, isolated from urine of a patient with urinary tract catheter	AMP, TMP	Sean Nair, University College London, Eastman Dental Institute
***Escherichia coli* NCTC 9081**	LB	37	Serotype O81, laboratory strain, isolated from a patients with UTI	VA, CM, TEC, L, AMP, OX	National collection of Type Cultures (NCTC)

Temp; temperature, LB; Luria Bernati broth, TS; Trypic soy broth, NB; Nutrient broth, NCTC; National Collection of Type Cultures, BEI; Biodefense and Emerging Infections Research Resources Repository, NIAID; National Institute of Allergy and Infectious Diseases, NIH; National Institute of Health. UTI; urinary tract infection, CC; clonal complex, ST-sequence type based on MLST- Multilocus sequence typing, K; capsular typing based on wzi gene, O; antigen typing based on wzm and wzt gene, CA-MRSA; Community-acquired methicillin resistant *Staphylococcus aureus.*

[a] Abbreviations are AMP; ampicillin, TE; tetracycline, CIP; ciprofloxacin, MET; methicillin, TMP; trimethoprim, VA; vancomycin, GM; gentamicin, TEC; teicoplanin, MER; meropenem, IMP; imipenem, OX; oxacillin, TEC; teicoplanin, CM; clindamycin, L; lincomycin

Each pathogen was distinguished as a control via biochemical tests, differential susceptibility to antibiotics and growth characteristics on

selective agar. Antibiotics in the antimicrobial panel were used at the standard concentrations on antibiotic disks for susceptibility testing (Table 2.2).

Table 2.2 Biochemical tests and selective medium used for the identification of bacterial isolates.

Test organism	Biochemical tests	Selective media	Susceptibility to antibiotics
***S. aureus* JE2**	Gram positive cocci (single pairs and clusters), catalase positive, coagulase positive	Chromogenic agar medium (MRSASelectII)	Oxacillin (1µg)
***E. faecium* E1162**	Gram positive, cocci (pairs and short chains), facultatively anaerobic catalase-negative	Chromogenic agar medium *(Enterococcus faecium* ChromoSelect agar)	Chloramphenicol (30 µg)
***A. baumannii* NCTC 12156**	Gram negative rod (single and pairs), catalase positive, citrate positive, coagulase negative, non-motile, urease negative	Chromogenic agar medium (CHROMagar *Acinetobacter*®)	Ciprofloxacin (15 µg)
***K. pneumoniae* C3091**	Gram negative rod, catalase positive, urease positive, citrate positive, negative for indole production	Cled agar	Ciprofloxacin (15 µg)
***P. aeruginosa* FRO1 (low mucoid)**	Gram negative rod, motile,catalase positive, coagulase negative, oxidase positive, negative for indole production	Cled agar	Ciprofloxacin (15 µg)
***E. cloaceae* NCTC 10005**	Gram negative rod (single and pairs), catalase positive, citrate positive, oxidase negative, urease negative	Violet red bile agar	Imipenem (10µg)
***E. coli* NCTC 9081**	Gram negative rod, catalase positive, urease negative, positive for indole production	EMB agar	Ciprofloxacin (15 µg)

Cled; Cystine Lactose Electrolyte deficient, EMB; Eosin methylene blue

2.1.1 Preparation of standard inoculum of indicator bacterium

One colony of each indicator bacterium mentioned in part 2.1 was inoculated in a 50 mL tube with 10 mL growth medium as mentioned in Table 2.1. These inoculums were incubated at 37°C in a 200 rpm shaking incubator for 16 hours. The 16 hour cultures were adjusted to a starting OD_{600nm} of 0.01 using spectrophotometry (OD_{600nm}) unless otherwise stated.

2.1.2 Determination of colony forming units (CFU)

A starter culture of indicator bacteria was prepared following the method mentioned in part 2.1.1. The optical density at 600nm (OD_{600nm}) of the 16-hour starter cultures were measured using a spectrophotometer, then diluted in the appropriate medium (table 2.1) to a calculated OD_{600nm} of 0.01 (equivalent to 10^6 CFU/ml^{-1}). Colony forming units was determined by spreading serial dilutions of the culture (100 µl of culture) onto appropriate agar plates and spread using an L-shaped spreader (VWR™). Plates were incubated at 37°C for 24 hours before the colonies were counted.

2.2 Assessing antimicrobial activity of filtered supernatant

2.2.1 Agar-well diffusion assays

Agar-well diffusion assays were performed to test the antimicrobial effect of supernatants prepared in this study. Agar-well diffusion assays were

performed as described by Quaiyum *et al.,* 2018. The indicator bacterium was prepared by inoculating each bacterium in their appropriate medium and growth conditions as mentioned in Table 2.1. The indicator bacterium was grown for 17 hours in a 37°C shaking incubator (200 rpm). The 17-hour cultures were adjusted to a density of 10^6 CFU/ml^{-1} equivalent to the turbidity of a 0.5 McFarland. Using a sterile cotton swab, a lawn of the culture of indicator bacterium was streaked onto agar (the type of agar will be mentioned in detail in Chapter 3 and Chapter 6) to create a uniform lawn of bacteria. The lawns were left to dry for 30 minutes at room temperature. After drying the plates, wells were prepared by punching wells in each plate using sterile 1mL pipette tips.

The filtered supernatant (100 µl) was pipetted into individual wells and was allowed to absorb into the wells for 15 minutes before incubating the plate at 30°C (when testing antibacterials produced by NW *Pseudomonas* sp. strains) and at 37°C (when testing antibacterials produced by *D. pigrum* ATCC 51524). Margins of inhibition were measured after 16 hours incubation. To measure the margins of inhibition, a ruler was used to measure (in millimeters) the well diameter (8mm). Then, the area with no bacterial growth is measured by measuring edge to edge across the margin of inhibition over the agar well. This measurement of antimicrobial activity minus the well diameter (8mm) is calculated as the margin of

inhibition. No margin of inhibition was reported as no inhibitory activity (0 mm).

2.2.2 Minimum inhibitory concentration of streptomycin and rifampicin

Minimum inhibitory concentrations (MICs) of rifampicin and streptomycin were determined in this research using broth microdilution assays according to CLSI guidelines (Garcia, 2010). A 10 mg/ml stock solution of both antibiotics was freshly prepared for this assay. Serial 2-fold broth dilutions starting at 256 µg/ml until 1.0 µg/ml rifampicin and streptomycin were made in LB broth using a flat bottom 96-well microtiter plate (VWR®). Both *Pseudomonas* sp. strain NW27 and *Dolosigranulum pigrum* ATCC 51524 form stationary-phase cultures were diluted to an $OD600_{nm}$ of 0.5 respectively in LB broth and Todd Hewitt broth growth medium and inoculated into each well to give 10^5 (50 µl). Microtiter plates were then incubated for 16 hours at 25°C (for *Pseudomonas* sp. strain NW27) and at 37°C (for *D. pigrum* ATCC 51524) with shaking at 200 rpm. The plates were observed for any cloudiness and turbidity after incubation. The lowest concentration of the tested compound that prevented growth was recorded as the MIC. All MIC determinations were performed in triplicate and median values were calculated for analysis.

2.3 In silico prediction of secondary metabolites biosynthetic gene clusters

To predict secondary metabolite biosynthesis pathways coding for molecules with putative antibiotic activity, a combination of genome mining platforms was used to identify conserved biosynthetic genes.

The whole genome sequences of the NW *Pseudomonas* sp. strains and the whole genome shotgun sequence of *D. pigrum* ATCC 51524 were submitted to 'antibiotics and secondary metabolites analysis shell' (antiSMASH). In addition, the WGS or large contigs containing genes with predicted protein products involved in antimicrobial synthesis (mentioned in detail in Chapter 6) were also submitted to antiSMASH, PRediction Informatics for Secondary Metabolomes (PRISM) and BActeriocin GEnome mining tool (BAGEL3).

2.3.1 antiSMASH

The antiSMASH is a biosynthetic gene cluster (BGC) mining platform that can predict genes coding for proteins involved in secondary metabolite synthesis, such as non-ribosomal peptides, polyketides, siderophores and bacteriocins (Blin *et al*., 2017, Blin *et al*., 2019). It can also detect other BGCs encoding biosynthetic pathways for molecules which do not have antimicrobial activity, for example LPS and capsular polysaccharides (Blin *et al*., 2017).

To run its profile-based BGC detection, the antiSMASH uses profile hidden Markov model (pHMMs) composed of protein sequences of experimentally characterized biosynthetic enzymes with profiles specific to conserved core enzymes of secondary metabolite biosynthesis pathways (Blin *et al*., 2017). The antiSMASH compares co-located core genes with a set of manually curated BGCs once the core enzymes have been identified (Blin *et al*., 2017).

The antiSMASH was accessed through the website: https://antismash.secondarymetabolites.org/. The detection strictness was set as “strict”, to detect well-defined clusters. All the extra parameters of antiSMASH were turned “on”, including KnownClusterBlast, ActiveSiteFinder, ClusterBlast, Cluster Pfam analysis, SubClusterBlast and Pfam-based GO term annotation.

2.3.2 PRISM

PRISM is a Java application and web server that can predict the chemical structure of genetically encoded nonribosomal peptides (NRPs) and type I and II polyketides (PKs) (Skinnider *et al*., 2015). PRISM searches the nucleotide sequence of interest against a library of hidden Markov models (HMMs) associated with secondary metabolism (Skinnider *et al*., 2017). It then clusters the identified biosynthetic genes and uses identified biosynthetic information for structure prediction (Skinnider *et al*., 2017).

PRISM was accessed through the webpage http://magarveylab.ca/prism/. The WGS potentially coding for proteins involved in antibacterial biosynthesis was submitted to PRISM by loading sample input as FASTA format. Since all the options of 'Advanced settings' were automatically selected by default, no further alterations were made.

2.3.3 BAGEL3

BAGEL3 is a web-based bacteriocin genome mining tool (van Heel *et al*., 2013, van Heel *et al*., 2018). BAGEL3 uses open read frames (ORF) prediction tools that detect genes possibly involved in bacteriocin biosynthesis and regulation, increasing the chance of identifying these loci in genomes (van Heel *et al*., 2018). In addition, this genome mining tool also identifies putative bacteriocins ORFs against the core peptide databases (van Heel *et al*., 2018). It can also identify biosynthetic gene clusters involved in the biosynthesis of Ribosomally synthesized and Post translationally modified Peptides (RiPPs) (van Heel *et al*., 2018).

BAGEL3 was accessed at: http://bagel.molgenrug.nl/. The WGS potentially coding for proteins involved in bacteriocin biosynthesis was submitted to BAGEL3 by loading the WGS file as FASTA format.

2.4 Statistical analysis

All experiments were replicated at least twice. Statistical analysis in Chapter 3 was performed by multiple regression analysis. For statistical evaluation in Chapter 4 and 5, all MIC data were log-transformed (base 2; expressed as log2 MIC). Nonparametric Mann-Whitney test was used to analyze differences between two different groups (Chapter 4 and 5). A *p*-value of <0.05 was considered statistically significant. All statistical analyses were conducted using Statistical Packages for Social Sciences (SPSS) software version 24.

Chapter 3

Primary screening of antibiotic production by *Pseudomonas* spp. isolated from water

3.0 Primary screening of antibiotic production by *Pseudomonas* spp. isolated from water

3.1 Introduction

3.1.1 Metabolic diversity of *Pseudomonas* spp.

Pseudomonas is a genus of aerobic, Gram negative, Gammaproteobacteria, belonging to the family Pseudomonadaceae (Özen and Ussery, 2012). The members of this genus are rod shaped, aerobic bacteria, commonly found in soil and water ecosystems (Iglewski, 1996).

Pseudomonads can grow with minimal nutritional requirements and can use a variety of environmental sources for nutrition (Wu *et al.,* 2020). For example, *Pseudomonas aeruginosa* only requires acetate and ammonia as a source of carbon and nitrogen, respectively (Szita *et al*., 1998). This species of *Pseudomonas* is also capable of responding to various environmental stresses, including pH (Eheth *et al.,* 2019).

Pseudomonas spp. can produce many types of secondary metabolites, including those with antimicrobial properties, (Pratiwi *et al*., 2017) and siderophores (Rieusset *et al*., 2020). Out of all the ESKAPE pathogens, the largest number of secondary metabolite biosynthetic gene clusters

identified was in *P. aeruginosa* (Tiwari *et al*., 2018). This diversity allows it to adapt to different environmental conditions (Riquelme *et al*., 2020, Thöming *et al*., 2020).

3.1.2. Secondary metabolites of some *Pseudomonas* spp.

Bacteria belonging to the *Pseudomonas* genus are known to generate different bioactive secondary metabolites that can exhibit antimicrobial activity (Matthijs *et al.*, 2014) (Table 3.1).

Table 3.1 Secondary metabolites produced by some *Pseudomonas* spp.

Secondary metabolite	**Producer strains**	**References**
Phenazines		
Phenazine-1-carboxylic acid	*P. aeruginosa*	(Simionato *et al*., 2017)
	P. fluorescens, *P. aureofaciens*	(Thomashow *et al*., 1990)
Phenazine-1,6-dicarboxylic acid	*P. chlororaphis*	(Guo *et al*., 2020)
Pyocyanin	*P. aeruginosa*	(Baron and Rowe, 1981)
Polyketides		
Mupirocin	*P. fluorescens*	(Sutherland *et al*., 1985)
2.4-diacetylphloroglucinol	*P. protegens*	(Julian *et al*., 2020)
Siderophores		
Pyochelin	*P. aeruginosa*	(Frangipani *et al*., 2014)
Pyoverdine	*P. aeruginosa*	(Schiessl *et al*., 2017)

Some *Pseudomonas* species can produce siderophores such as pyoverdines and pyochelins (Brandel *et al*., 2012). Siderophores are low molecular weight (< 10 KD) iron chelators produced by microorganisms to

eliminate the presence of iron (Brandel *et al*., 2012). Iron is one of the essential elements required for growth of bacteria (Hider and Kong, 2010). Bacteria acquire iron by releasing siderophores, which will chelate iron and the resulting complex will be internalized (Hider and Kong, 2010). Since bacterial resistance to antibiotics is still a major problem, development of new antibiotics and delivery systems is necessary, especially to treat bacterial infections caused by Gram negative bacteria (Ghosh *et al*., 2017). This has been previously done by coupling fimsbactin, a selective siderophore, to daptomycin, an antibiotic used against Gram positive bacteria (Ghosh *et al*., 2017). Researchers showed that this conjugate was active against antibiotic resistant strains of *A. baumannii* both *in vivo* and *in vitro*. The extended activity of daptomycin demonstrates the possibility of synthesizing novel antibiotic conjugates that can overcome the antibiotics lack of activity against Gram negative bacteria.

Pyoverdines are siderophores produced by certain pseudomonads such as *P. aeruginosa* (Schiessl *et al*., 2017). Pyoverdines also give the green, fluorescent colour to *Pseudomonas* culture (Barbhaiya and Rao, 1985). This siderophore binds to iron with a very high affinity (Brandel *et al*., 2012)

Pyochelin is another siderophore produced by *Pseudomonas* spp. (Schiessl *et al*., 2017). Compared to pyoverdine, pyochelin binds to iron with a low affinity (Brandel *et al*., 2012). Interestingly, Dumas *et al*., (2013)

have demonstrated the ability of *P. aeruginosa* to switch from initially producing pyochelin to pyoverdine only when the concentration of iron reduces substantially.

Pseudomonas sp. can also produce surfactants such as rhamnolipids (Thakur *et al*., 2021). Surfactants lower the surface tension of an aqueous solution and are commonly used as detergents and disinfectants (Falk, 2019). Biofilm-specific activity of some disinfectants have been shown to disrupt biofilm , therefore killing *S. aureus* strains, in the biofilm (Gregory *et al*., 2021). In addition, surfactant-based dressings have been tested for their anti-biofilm properties against Gram negative bacteria such as *Pseudomonas* (Das Ghatak *et al*., 2018). It is therefore interesting to identify a culture condition able to elicit the production of biosurfactants in this study.

3.2 Aims and objectives of the work in this chapter

The aim of this chapter was to compare the antimicrobial activity of five NW *Pseudomonas* sp. strains isolated from water by observing their inhibitory activity and their predicted natural products. The objectives were;

- Using agar well diffusion assays, to assess the spectrum of antibacterial activity of five NW *Pseudomonas* sp. strains cultivated under various growth conditions against nosocomial pathogens.
- To confirm the presence of biosurfactants in the supernatants of NW *Pseudomonas* sp. strains using drop-collapse assays.
- To predict candidate bacterial secondary metabolite biosynthetic gene clusters, using antiSMASH as a prediction tool, that may be responsible for antibiotic production in the five *Pseudomonas* sp. strains.

3.3. Materials and methods

3.3.1. Preparation of filtered supernatants to be tested for antimicrobial activity

Initially, supernatants of NW *Pseudomonas* sp. strains that might contain antimicrobials were prepared in small amounts. This step was done to identify the optimum conditions required for antimicrobial production. The *Pseudomonas* isolates used in this study are named as follows; NW stands for, N=Nair and W=water isolate, the number refers to the order in which the bacterium was originally isolated. Isolation of NW *Pseudomonas* sp. strains was performed by Dr Sean Nair, Eastman Dental Institute, UCL. Briefly, the water sample was diluted and plated onto LB agar with an overlay of 0.5% LB agar seeded with MRSA. Out of approximately 5000 colonies, 30 showed MRSA inhibitory activity and distinctive colony morphologies. On secondary screening, only five of the 30 showed inhibitory activity and different colony morphologies. The isolates were identified as *Pseudomonas* species by 16S rRNA sequencing. The five NW *Pseudomonas* sp. strains were also identified as novel strains within the genus *Pseudomonas* by Dr Sean Nair. Glycerol stocks of these bacterial strains were stored at -80°C for further analysis. The NW *Pseudomonas* sp. strains were streaked out from frozen stocks on LB plates and were incubated at 30°C for 24 hours.

Growth of the NW *Pseudomonas* sp. strains was studied at two temperatures (17°C and 25°C), at different time points (24, 48 and 72 hours) and three media were used to culture them (Nutrient broth, Luria Bernati and M9 minimal medium). A temperature of 17°C was chosen to represent the temperature of the water when the bacteria were isolated, while a temperature of 25°C was chosen as it was found to be the optimum growth temperature in nutrient broth by Dr Nair.

Starter cultures were prepared by inoculating a single colony of each NW *Pseudomonas* sp. strain from the resulting plates in 10 mL nutrient broth, Luria bernati broth or M9 minimal medium in a 50 mL tube and was incubated at either 17°C or 25°C with shaking at 150 rpm for 17 hours. This was used to prepare the starter culture.

After completion of 17 hours, the optical density at 600nm (OD_{600nm}) was measured using a spectrophotometer. The starter cultures were then diluted in appropriate medium to a calculated OD of 0.05 (OD_{600nm}). These tubes were then incubated for 24 hours, 48 hours or 72 hours at either 17°C or 25°C.

After completion of required incubation time, supernatants were collected by centrifuging at 4193 xg for 15 minutes. The supernatant was then

filtered using 0.22 µm pore size syringe filters (Fisher Scientific) and were stored at -20°C before further analysis. This assay was done in triplicate.

3.3.2. Agar well diffusion assay

The inhibitory activity of the supernatants against indicator bacteria were tested using agar well diffusion assays performed following Chapter 2 section 2.2.1. Indicator bacteria used were *Staphylococcus aureus* JE2, *Enterococcus faecium* E1162, *Klebsiella pneumoniae* C3091, *Acinetobacter baumannii* NCTC 12156*, Pseudomonas aeruginosa* FRO1 (low mucoid) and *Enterobacter cloacae* NCTC 10005.

3.3.3 Drop-collapse assay

The drop collapse test was performed by following the procedure described by Tugrul and Cansunar., 2005 with minor alterations. Sterile distilled water (25 µl) was pipetted onto the hydrophobic side of a piece of parafilm. Then, 10 µl of methylene blue was pipetted on top of this water droplet. The filtered NW *Pseudomonas* sp. strains supernatant (25 µl) was gently dropped on top of the drop of water and methylene blue. The diameter of the drop size of each droplet (in mm) was measured using a gridline and was compared to the negative control. The negative control was 25 µl of sterile LB broth, nutrient broth or M9 minimal medium instead of supernatant. The increase in drop size was due to the collapse of drop, which was in turn due to the presence of a surfactant. The diameter should

be at least 1 mm larger than that of the negative control to be scored as a 'collapsed droplet' (Thavasi *et al*., 2013).

3.3.4 In silico prediction of secondary metabolite biosynthetic gene clusters in the whole genome sequence of the five NW *Pseudomonas* sp. strains

To predict secondary metabolite biosynthesis pathways coding for molecules with putative antibiotic activity in the whole genome sequence of *Pseudomonas* sp. strain NW7, NW9, NW10, NW16 and NW27, antiSMASH was used as described in Chapter 2, part 2.3.1.

3.3.5 Statistical analysis

Multiple regression analysis was used to identify if there was a relationship between the margins of inhibition (dependent variable) and the three parameters: time, cultivation broth and cultivation temperature (independent variables). A *p*-value of <0.05 was considered statistically significant.

3.4 Results

3.4.1 Determining the spectrum of activity of antimicrobial compounds produced by 5 NW *Pseudomonas* sp. strains

To validate this experiment, each ESKAPE pathogen was grown selectively on its appropriate agar as described in Chapter 2, part 2.1 and biochemical tests were carried out. In addition, all the ESKAPE pathogens behaved as expected in the antibiogram. A total of 270 supernatants from five NW *Pseudomonas* sp. strains were prepared and tested (Figure 3.1).

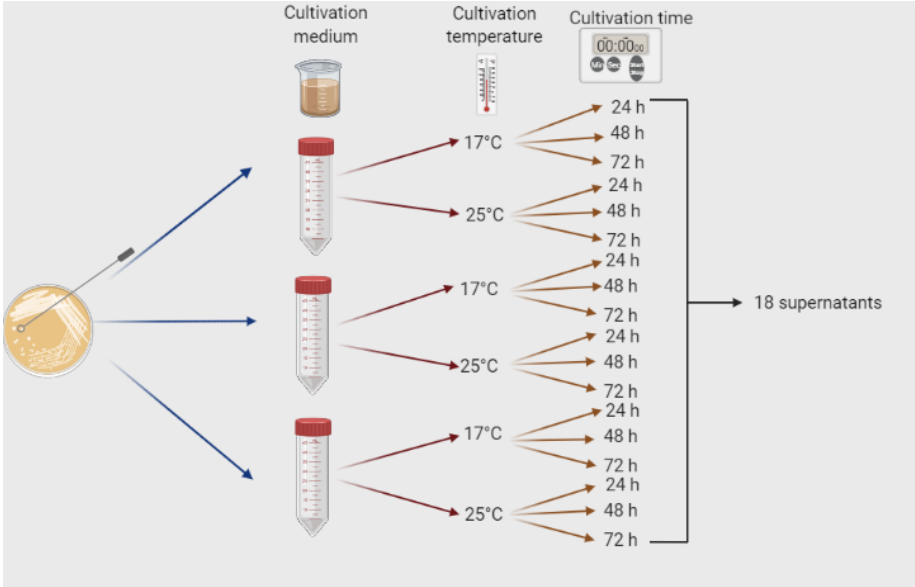


Figure 3.1 Preparation of 18 supernatants using one NW *Pseudomonas* sp. strain.

This figure shows how 18 supernatants were prepared using one NW *Pseudomonas* sp. strain. To test the supernatant of all five NW *Pseudomonas* sp. strains in triplicates, a total of 270 supernatants were prepared. Each *Pseudomonas* sp. NW strain was grown in three different media (nutrient broth, Luria Bernati broth and M9 minimal medium), using 2 different temperatures (17°C and 25°C) at different time points (24, 48 and 72 hours). This figure was created using Biorender.

Out of these supernatants, only 44 were able to elicit antimicrobial activity. Out of these 44 supernatants, 14 of them which were produced by *Pseudomonas* sp. strain NW7, NW9, NW16 and NW27 incubated at 17°C, were able to elicit antimicrobial activity against indicator bacteria (Table 3.2). *Pseudomonas* sp. strain NW10 incubated at 17°C did not inhibit the growth of any indicator bacterium.

A significant relationship between the margins of inhibition formed by *Pseudomonas* sp. strain NW7 incubated at 17°C and the cultivation time and growth medium were found for each (p-value <0.001). Similarly, a significant difference in the margins of inhibition were found when *Pseudomonas* sp. strain NW9 was cultivated at 17°C in different growth medium and different cultivation time (p-value <0.001).

Table 3.2 Spectrum of activity of antimicrobial(s) produced by NW *Pseudomonas* sp. strains incubated at 17°C under different conditions against indicator bacteria.

NW *Pseudomonas* sp. strains	Growth conditions		Indicator bacterium		
	Medium	Time (h)	*S. aureus* JE2	*A. baumannii* NCTC 12156	*K. pneumoniae* C3091
NW7	LB	24	0(0)	0(0)	2 (2)
	M9	24	11 (11-12)	6 (6)	2 (2)
		48	14 (11-17)	9 (9-11)	0(0)
		72	15 (14-16)	9 (8-10)	0(0)
NW9	LB	24	0(0)	0(0)	2 (2)
	NB	24	0(0)	0(0)	3 (1-3)
	M9	24	12 (11-13)	9 (9-10)	0(0)
		48	13(13)	9 (9-11)	0(0)
		72	15 (14-15)	9 (9-10)	0(0)
NW16	NB	24	0(0)	0(0)	5 (4-6)
NW27	LB	48	0(0)	5 (4-8)	3 (3)
	NB	24	0(0)	5 (5-6)	0(0)
		48	0(0)	2 (2-3)	0(0)
	M9	24	0(0)	0(0)	2 (2-3)
Oxacillin (1 µg)			9 (7-8)	-	-
Vancomycin (30 µg)			0(0)	0(0)	-
Ciprofloxacin (15 µg)			0(0)	14 (14-16)	13 (13)
Ampicillin (10 µg)			-	-	0(0)

The mean margins of inhibition (mm) of the filtered supernatants of 4 NW *Pseudomonas* sp. strains grown under different conditions against *S. aureus* JE2, *A. baumannii* NCTC 12156 and *K. pneumoniae* C3091. Results of *E. faecium* E1162, *P. aeruginosa* FRO1 (low mucoid) and *E. cloacae* NCTC 10005 are not mentioned because they were resistant to all the supernatants. Supernatant produced by *Pseudomonas* sp. strain NW 10 was not able to inhibit the growth of any of the ESKAPE pathogens. h; hours, LB; Luria Bernati broth, NB; Nutrient broth, M9; M9 minimal medium, -; not tested. Values represent a set of triplicate results and are expressed as median (range).

The remaining 30 supernatants produced by all NW *Pseudomonas* sp. strains incubated at 25°C, were able to elicit antibacterial activity against the indicator bacteria (Table 3.3).

According to statistical analysis, type of growth medium played a significant role in the margins of inhibition formed by *Pseudomonas* sp. strain NW7 grown at 25°C (*p*-value <0.001). However, no significant difference between the margins of inhibition produced by *Pseudomonas* sp. strain NW7 grown at 25°C and the cultivation time was observed (*p*-value >0.05). Moreover, type of growth medium showed a significant effect on the margins of inhibition formed by *Pseudomonas* sp. strain NW9 (*p*-value <0.001). In contrast, there was no significant difference in the margins of inhibition formed by *Pseudomonas* sp. strain NW27 incubated at 25°C and the type of growth medium (*p*-value = 0.200). Additionally, cultivation time showed a significant difference in the margins of inhibition formed by *Pseudomonas* sp. strain NW27 incubated at 25°C (*p*-value <0.007).

Table 3.3 Spectrum of activity of antimicrobial(s) produced by NW *Pseudomonas* sp. strains incubated at 25°C under different conditions against indicator bacteria.

Pseudomonas sp. strains NW	Growth conditions		Indicator bacterium			
	Medium	Time (h)	*S. aureus* JE2	*E. faecium* E1162	*A. baumannii* NCTC 12156	*K. pneumoniae* C3091
NW7	LB	24	9 (7-12)	10 (10)	0(0)	0(0)
		48	8 (7-9)	8 (8-9)	0(0)	0(0)
		72	11 (8-12)	10 (10)	0(0)	0(0)
	NB	24	9 (8-10)	11(11-12)	1 (1)	0(0)
		48	10 (7-12)	13 (13-16)	1 (1)	0(0)
		72	10 (7-12)	13 (11-14)	2 (1-2)	0(0)
	M9	24	10 (10-11)	14 (13-15)	3 (1-4)	1 (1-2)
		48	12 (12-13)	0(0)	0(0)	0(0)
		72	13 (13)	0(0)	0(0)	0(0)
NW9	LB	24	4 (4)	10 (7-11)	0(0)	1 (1)
		48	9 (8-10)	8 (8-9)	0(0)	0(0)
		72	7 (6-8)	8 (8-9)	0(0)	0(0)
	NB	24	11 (11-12)	12 (11-12)	1 (1-2)	0(0)
		48	11 (11-12)	15 (15)	0(0)	2 (1-2)
		72	12 (11-12)	13 (13-14)	2 (1-3)	0(0)
	M9	24	11 (11-12)	14 (14-15)	0(0)	0(0)
		48	13 (12-13)	0(0)	0(0)	0(0)
		72	12 (12-13)	0(0)	0(0)	0(0)
NW10	LB	24	8 (3-12)	0(0)	3 (1-6)	0(0)
		48	2 (1-4)	0(0)	0(0)	0(0)
NW16	LB	24	0(0)	0(0)	0(0)	2 (2-3)
NW27	LB	24	10 (10-11)	12 (12-13)	4 (3-6)	2(2-3)
		48	11 (11-12)	11 (7-13)	5 (5-6)	0(0)
		72	8 (8-9)	9 (7-10)	3 (2-3)	0(0)
	NB	24	9 (8-10)	6 (4-7)	0(0)	0(0)
		48	12 (11-12)	15 (13-17)	0(0)	0(0)
		72	10 (9-12)	12 (12-13)	0(0)	0(0)
	M9	24	6 (5-7)	0(0)	0(0)	0(0)
		48	12 (11-12)	0(0)	0(0)	0(0)
		72	11 (10-12)	0(0)	0(0)	0(0)
Oxacillin (1 µg)			9 (7-8)	-	-	-
Chloramphenicol (30 µg)			-	16(15-17)	-	-
Vancomycin (30 µg)			0(0)	0(0)	0(0)	-
Ciprofloxacin (15 µg)			0(0)	-	14 (14-16)	13 (13)
Ampicillin (10 µg)			-	0(0)	-	0(0)

The mean margins of inhibition (mm) of the filtered supernatants of 5 NW *Pseudomonas* sp. strains grown under different conditions against *S. aureus* JE2, *E. faecium* E1162, *A. baumannii* NCTC 12156 and *K. pneumoniae* C3091. Results of *P. aeruginosa* FRO1 (low mucoid) and *E. cloacae* NCTC 10005 are not mentioned because they were resistant to all the supernatants. h; hours, LB; Luria Bernati broth, NB; Nutrient broth, M9; M9 minimal medium, -; not tested. Values represent a set of triplicate results and are expressed as median (range).

The growth of *S. aureus*, *E. faecium*, *A. baumannii* and *K. pneumoniae* were inhibited by the supernatant of *Pseudomonas* sp. strain NW27 grown in LB broth for 24 hours at 25°C. According to statistical analysis, incubation temperature of *Pseudomonas* sp. strain NW27 grown in LB broth significantly affected the margins of inhibition (*p*-value <0.001). Moreover, cultivation temperature displayed a significant effect on margins of inhibition formed by *Pseudomonas* sp. strain NW27 (*p*-value <0.005). In addition, cultivation time showed a significant effect in the margins of inhibition formed by *Pseudomonas* sp. strain NW27 grown at 25°C (*p*-value <0.001). Unlike this strain, no significant relationship between the margins of inhibition formed by *Pseudomonas* sp. strain NW9 and cultivation time was found (*p*-value= 0.848).

Another supernatant that produced broad antimicrobial activity was that of *Pseudomonas* sp. strain NW7 grown in M9 minimal media for 24 hours at 25°C. This supernatant was able to inhibit *S. aureus, E. faecium*, *A. baumannii* and *K. pneumoniae*. A significant relationship between the margins of inhibition formed by *Pseudomonas* sp. strain NW7 grown in M9 and the cultivation time were found (*p*-value <0.005). The broad antimicrobial activity could signify the presence of either multiple antimicrobial compounds in these supernatants or one antibacterial

compound with broad antimicrobial activity. *P. aeruginosa* and *E. cloaceae* were not inhibited by any of the supernatants.

S. aureus and *E. faecium* were inhibited by multiple supernatants. These include those of *Pseudomonas* sp. strain NW7 grown in LB for 24, 48 and 72 hours at 25°C, *Pseudomonas* sp. strain NW9 grown in LB for 48 and 72 hours at 25°C, *Pseudomonas* sp. strain NW27 grown in NB for 24, 48 and 72 hours 25°C. This can indicate the presence of an antimicrobial compound in these supernatants specifically targeting Gram positive bacteria.

3.4.2 Production of surfactant by NW *Pseudomonas* sp. strains

The drop collapse assay relies on destabilizing a droplet of water by surfactant. The water droplet and the supernatant are mixed and placed on an oil coated or hydrophobic surface (Bodour and Miller-Maier, 1998). If the drop does not contain surfactants, the droplet will remain stable because the polar water molecules are repelled from the hydrophobic surface. If the liquid mixture contains a biosurfactant, the drop will collapse because the force or interfacial tension between the droplet and the hydrophobic surface is reduced (Bodour and Miller-Maier, 1998). Of the different supernatants tested, some were found to collapse the water droplet, indicating the presence of a surfactant.

All the supernatants of NW *Pseudomonas* sp. strains grown at 17°C were unable to collapse the droplets. However, five supernatants harvested at 25°C were able to collapse the water droplet. The largest droplet collapse occurred with the cell free supernatants of *Pseudomonas* sp. strain NW10 grown in LB for 24 hours at 25°C (Table 3.4). Supernatant of *Pseudomonas* sp. strain NW16 grown in LB for 24 hours at 25°C was also able to collapse droplets, producing a median drop size of 9 mm. There was a collapse that occurred with the supernatant of *Pseudomonas* sp. strain NW27 grown in LB for 48 and 72 hours at 25°C. These results

indicate that *Pseudomonas* sp. strain NW10, NW16 and NW27 have the potential to produce surfactants when grown under the conditions mentioned in Table 3.4.

Table 3.4. Drop-collapse assay using supernatants of NW *Pseudomonas* sp. strains grown at 25°C under different incubation conditions.

Growth conditions			Spectrum of antimicrobial activity according to agar well diffusion assays	Drop collapse assays (mm)
NW *Pseudomonas* sp. strains	**Medium**	**Time (h)**		
NW10	LB	24	*S. aureus* JE2, *A. baumannii* NCTC 1256	9 (9)
		48	*S. aureus* JE2	9 (8.5-9)
NW16	LB	24	*K. pneumoniae* C3091	9 (8.5-9)
NW27	LB	48	*S. aureus* JE2, *E. faecium* E1162, *A. baumannii* NCTC 1256	7 (7)
	LB	72	*S. aureus* JE2, *E. faecium* E1162, *A. baumannii* NCTC 1256	7 (7)

This table represents drop collapse assay results and the spectrum of antimicrobial activity of each supernatant. Results of drop collapse assays are represented by median drop size of the cell free supernatants (in mm) compared to the drop size of the negative control (6 mm). Values represent a set of three replicates and are expressed as median (range). Other supernatants were not mentioned because they did not collapse the drop. LB; Luria Bernati broth, mm; millimeter.

3.4.3 antiSMASH predictions of BGCs in the whole genome sequence of five NW *Pseudomonas* sp. strains

3.4.3.1 *Pseudomonas* sp. strain NW7

The antiSMASH predicted seventeen BGCs in the whole genome sequence of *Pseudomonas* sp. strain NW7.

Some of these BGCs had genes that encode for proteins involved in the production of known products, including 2,4-diacetylphloroglucinol and sessilin A (Table 3.5). Other BGCs had genes that have the potential to code for the production of metabolic products related to lankacidin C, pyoverdines, fengycin, mitomycin, thiazostatin/watasemycin A/ watasemycin B/ 2-hydroxyphenylthiazoline enantiopyochelin/isopyochelin, methanobactin and arylpolyene. Seven other BGCs were also predicted by antiSMASH but the metabolic product of these BGCs were too dissimilar to known BGCs for antiSMASH to suggest similar products. This suggests the metabolic products of these 7 BGCs to be novel.

Table 3.5 Summary of biosynthetic gene clusters in the genome of *Pseudomonas* sp. strain NW7 that have the potential to code for the production of secondary metabolites.

Label of region in the genome	Type of BGC	Similarity of putative BGC to other BGC (%)	Label of similar BGC	Appendices in this chapter	
				Diagrammatical representation of putative BGC	Comparison of known BGCs
1.1	T3PKS	2,4-diacetylphloroglucinol (100%)	BGC0000281	1a	1b
1.2	Redox-cofactor	Lankacidin C (13%)	BGC00001100	2a	2b
1.3	NAGGN	-	-	3	N/A
1.4	NRPS	Pyoverdine (21%)	BGC0000413	4a	4b
1.5	ranthipeptide, NRPS	Pyoverdine (33%)	BGC0000413	5a	5b
1.6	Betalactone	Fengycin (13%)	BGC0001095	6a	6b
1.7	Acyl amino acids	-	-	7	N/A
1.8	NRPS	Mitomycin (5)	BGC0000915	8a	8b
1.9	NRPS	thiazostatin/watasemycin A/ watasemycin B/ 2-hydroxyphenylthiazoline enantiopyochelin/isopyochelin (27%)	BGC0001801	9a	9b

Label of region in the genome	Type of BGC	Similarity of putative BGC to other BGC (%)	Label of similar BGC	Appendices in this chapter	
				Diagrammatical representation of putative BGC	Comparison of known BGCs
1.10	NRPS	Sessilin A (100%)	BGC0000425	10a	10b
1.11	RiPP-like	Methanobactin (50%)	BGC0002004	11a	11b
1.12	CDPS	-	-	12	N/A
1.13	RiPP-like	-	-	13	N/A
1.14	RiPP-like	-	-	14	N/A
1.15	RRE-containing	-	-	15	N/A
1.16	arylpolyene	Arylpolyene (40%)	BGC0000837	16a	16b
1.17	RiPP-like	-	-	17	N/A

Region represents the region of the BGC as predicted by antiSMASH. The antiSMASH predicted a percentage similarity of the genes in the putative BGC to be present in other known BGCs. The percentage similarity is based on the Minimum Information on Biosynthetic Gene Cluster (MiBIG) repository which is interconnected with antiSMASH. All BGCs can be found in the appendices.T3PKS; type III polyketide synthase, NAGGN; N-acetylglutaminylglutamine amide, NRPS; Non-ribosomal peptide synthetase cluster, RiPP-like; Other unspecified ribosomally synthesized and post-translationally modified peptide product (RiPP) cluster, CDPS; tRNA-dependent cyclodipeptide synthases, RRE-containing; RRE-element containing cluster, -; not mentioned because it was too dissimilar to predict by antiSMASH, N/A; not applicable.

3.4.3.2 *Pseudomonas* sp. strain NW9

The antiSMASH predicted seventeen BGCs in the whole genome shotgun sequence of *Pseudomonas* sp. strain NW9.

The antiSMASH predicted some of these BGCs to have genes that encode for proteins involved in the biosynthesis of known metabolic products, including sessilin A and 2,4-diacetylphloroglucinol (Table 3.6). Other BGCs had genes that have the potential to code for the production of metabolic products related to lankacidin C, pyoverdines, fengycin, mitomycin, thiazostatin/watasemycin A/ watasemycin B/ 2-hydroxyphenylthiazolin enantiopyochelin/ isopyochelin, methanobactin, and arylpolyene. Seven other BGCs were also predicted by antiSMASH but the metabolic product of these BGCs were too dissimilar to known BGCs for antiSMASH to suggest similar products. This suggests the metabolic products of these 7 BGCs to be novel.

Table 3.6 Summary of biosynthetic gene clusters in the genome of *Pseudomonas* sp. strain NW9 that have the potential to code for the production of secondary metabolites.

Label of region in the genome	Type of BGC	Similarity of putative BGC to other BGC (%)	Label of similar BGC	Label of appendices in this chapter	
				Diagrammatical representation of putative BGC	Comparison of known BGCs
2.1	Redox-cofactor	Lankacidin C (13%)	BGC0001100	18a	18b
2.2	NAGGN	-	-	19	N/A
2.3	NRPS	Pyoverdine (21%)	BGC0000413	20a	20b
2.4	Ranthipeptide, NRPS	Pyoverdine (33%)	BGC0000413	21a	21b
2.5	Betalactone	Fengycin (13%)	BGC0001095	22a	22b
2.6	Acyl amino acids	-	-	23	N/A
2.7	NRPS	Mitomycin (5%)	BGC0000915	24a	24b
2.8	NRPS	thiazostatin/watasemycin A/ watasemycin B/ 2-hydroxyphenylthiazolin enantiopyochelin/ isopyochelin (27%)	BGC0001801	25a	25b

Label of region in the genome	Type of BGC	Similarity of putative BGC to other BGC (%)	Label of similar BGC	Label of appendices in this chapter	
				Diagrammatical representation of putative BGC	Comparison of known BGCs
2.9	NRPS	Sessilin A (100%)	BGC0000425	26a	26b
2.10	RiPP-like	Methanobactin (50%)	BGC0002004	27a	27b
2.11	CDPS	-	-	28	N/A
2.12	RiPP-like	-	-	29	N/A
2.13	RiPP-like	-	-	30	N/A
2.14	RRE-containing	-	-	31	N/A
2.15	Arylpolyene	Arylpolyene (40%)	BGC0000837	32a	32b
2.16	RiPP-like	-	-	33	N/A
2.17	T3PKS	2,4-diacetylphloroglucinol (100%)	BGC0000281	34a	34b

Region represents the region of the BGC as predicted by antiSMASH. The antiSMASH predicted a percentage similarity of the genes in the putative BGC to be present in other known BGCs. The percentage similarity is based on the Minimum Information on Biosynthetic Gene Cluster (MiBIG) repository which is interconnected with antiSMASH. All BGCs can be found in the appendices. NAGGN; N-acetylglutaminylglutamine amide, NRPS; Non-ribosomal peptide synthetase cluster, RiPP-like; Other unspecified ribosomally synthesised and post-translationally modified peptide product (RiPP) cluster, T3PKS; type III polyketide synthase, RRE-containing; RRE-element containing cluster, CDPS; tRNA-dependent cyclodipeptide synthases, -; not mentioned because it was too dissimilar to predict by antiSMASH, N/A; not applicable.

3.4.3.3 *Pseudomonas* sp. strain NW10

The antiSMASH predicted eleven BGCs in the whole genome shotgun sequence of *Pseudomonas* sp. strain NW10.

Some BGCs had genes that have the potential to code for proteins involved in the biosynthesis of metabolic products related to fragin, arylpolyene, lankacidin C, pyoverdines and lokisin (Table 3.7). The antiSMASH predicted another five BGCs, but their metabolic products were too dissimilar to known BGCs for antiSMASH to suggest similar products. The metabolic products of these 5 BGCs are suggestive to be novel.

Table 3.7 Summary of biosynthetic gene clusters in the genome of *Pseudomonas* sp. strain NW10 that have the potential to code for the production of secondary metabolites.

Label of region in the genome	Type of BGC	Similarity of putative BGC to other BGC (%)	Label of putative BGC	Appendices in this chapter	
				Diagrammatical representation of putative BGC	Comparison of known BGCs
1.1	NRPS-like	Fragin (37%)	BGC0001599	35a	35b
1.2	Arylpolyene	Arylpolyene (40%)	BGC0000837	36a	36b
2.1	PpyS-KS	-	-	37	N/A
4.1	Redox-cofactor	Lankacidin C (13%)	BGC0001100	38a	38b
13.1	NRPS	Pyoverdin (11%)	BGC0000413	39a	39b
15.1	NRPS	Pyoverdin (17%)	BGC0000413	40a	40b
17.1	NRPS	Lokisin (85%)	BGC0001980	41a	41b
22.1	NAGGN	-	-	42	N/A

Label of region in the genome	Type of BGC	Similarity of putative BGC to other BGC (%)	Label of putative BGC	Appendices in this chapter	
				Diagrammatical representation of putative BGC	Comparison of known BGCs
25.1	hsterlactone, butyrolactone	-	-	43	N/A
28.1	RRE-containing	-	-	44	N/A
58.1	RiPP-like	-	-	45	N/A

Region represents the region of the BGC as predicted by antiSMASH. The antiSMASH predicted a percentage similarity of the genes in the putative BGC to be present in other known BGCs. The percentage similarity is based on the Minimum Information on Biosynthetic Gene Cluster (MiBIG) repository which is interconnected with antiSMASH. All BGCs can be found in the appendices.

NRPS; Non-ribosomal peptide synthetase cluster, PpyS-KS; Photopyrone-like pyrone, NAGGN; N-acetylglutaminylglutamine amide RiPP-like; Other unspecified ribosomally synthesised and post-translationally modified peptide product (RiPP) cluster, RRE-containing; RRE-element containing cluster, -; not mentioned because it was too dissimilar to predict by antiSMASH, N/A; not applicable.

3.4.3.4 *Pseudomonas* sp. strain NW16

The antiSMASH predicted nine BGCs in this whole genome shotgun sequence of *Pseudomonas* sp. strain NW16.

The antiSMASH predicted some BGCs to have genes that have the potential to encode proteins involved in the biosynthesis of metabolic products related to lankacidin C, arylpolyene, pyoverdines, lokisin, fengycin and fragin (Table 3.8). Two BGCs were also predicted but their metabolic products were not similar to known BGCs for antiSMASH to suggest similar products. This suggests the metabolic product of these 2 BGCs to be novel.

Table 3.8 Summary of biosynthetic gene clusters in the genome of *Pseudomonas* sp. strain NW16 that have the potential to code for the production of secondary metabolites.

Label of region in the genome	Type of BGC	Similarity of putative BGC to other BGC (%)	Label of similar BGC	Appendices in this chapter	
				Diagrammatical representation of putative BGC	Comparison of known BGCs
1.1	Redox-cofactor	Lankacidin C (13%)	BGC0001100	46a	46b
2.1	Arylpolyene	Arylpolyene (40%)	BGC0000837	47a	47b
4.1	NAGGN	-	-	48	N/A
4.2	NRPS	Pyoverdin (17%)	BGC0000413	49a	49b
8.1	RiPP-like	-	-	50	N/A
10.1	Betalactone	Fengycin (13%)	BGC0001095	51a	51b
11.1	NRPS	Lokisin (92%)	BGC0001980	52a	52b
15.1	NRPS	Pyoverdin (11%)	BGC0000413	53a	53b
24.1	NRPS-like	Fragin (37%)	BGC0001599	54a	54b

Region represents the region of the BGC as predicted by antiSMASH. The antiSMASH predicted a percentage similarity of the genes in the putative BGC to be present in other known BGCs. The percentage similarity is based on the Minimum Information on Biosynthetic Gene Cluster (MiBIG) repository which is interconnected with antiSMASH. All BGCs can be found in the appendices.NAGGN; N-acetylglutaminylglutamine amide, NRPS;, Non-ribosomal peptide synthetase cluster, RiPP-like; Other unspecified ribosomally synthesised and post-translationally modified peptide product (RiPP) cluster, -; too dissimilar to predict by antiSMASH, N/A; not applicable.

3.4.3.5 *Pseudomonas* sp. strain NW27

The antiSMASH predicted seventeen BGCs in the whole genome sequence of *Pseudomonas* sp. strain NW27.

Some BGCs were predicted to have genes that have the potential to code for the production of metabolic products related to pyoverdine, bananamide, fengycin, viscosin, lokisin, lankacidin, arylpolyene, fragin, a lipopolysaccharide and syringomycin (Table 3.9). Six BGCs were also predicted by antiSMASH but the metabolic product of these BGCs were too dissimilar to known BGCs for antiSMASH to suggest similar products. This suggests the metabolic products of these 6 BGCs to be novel

Table 3.9 Summary of biosynthetic gene clusters in the genome of *Pseudomonas* sp. strain NW27 that have the potential to code for the production of secondary metabolites.

Label of region in the genome	Type of BGC	Label of putative BGC	Similarity of putative BGC to other BGC (%)	Appendices in this chapter	
				Diagrammatical representation of putative BGC	Comparison of known BGCs
1.1	Siderophore	-	-	55	N/A
1.2	NRPS	BGC0000413	Pyoverdine (9%)	56a	56b
1.3	NAGGN	-	-	57	N/A
1.4	Siderophore	-	-	58	N/A
1.5	NRPS	BGC0001346	bananamide 1/ bananamide 2/ bananamide 3 (62%)	59a	59b
1.6	Betalactone	BGC0001095	Fengycin (13%)	60a	61a
1.7	RiPP-like	-	-	61	N/A
1.8	NRPS	BGC0001312	Viscosin (43%)	62a	62b
1.9	Thioamidites, RiPP-like	BGC0001980	Lokisin (14%)	63a	63b
2.1	RiPP-like	-	-	64	N/A
2.2	Redox-cofactor	BGC0001100	Lankacidin C (13%)	65a	65b

Label of region in the genome	Type of BGC	Label of putative BGC	Similarity of putative BGC to other BGC (%)	Appendices in this chapter	
				Diagrammatical representation of putative BGC	Comparison of known BGCs
3.1	Arylpolyene	BGC0000837	Arylpolyene (40%)	66a	66b
3.2	NRPS-like	BGC0001599	Fragin (25%)	67a	67b
4.1	Thiopeptide	BGC0000774	Lipopolysaccharide (5%)	68a	68b
9.1	NRPS	BGC0000413	Pyoverdin (10%)	69a	69b
11.1	NRPS	BGC0000437	Syringomycin (58%)	70a	70b
13.1	RiPP-like	-	-	71	N/A

Region represents the region of the BGC as predicted by antiSMASH. The antiSMASH predicted a percentage similarity of the genes in the putative BGC to be present in other known BGCs. The percentage similarity is based on the Minimum Information on Biosynthetic Gene Cluster (MiBIG) repository which is interconnected with antiSMASH. All BGCs can be found in the appendices. NRPS; Non-ribosomal peptide synthetase cluster, RiPP-like; Other unspecified ribosomally synthesized and post-translationally modified peptide product (RiPP) cluster, NAGGN; N-acetylglutaminylglutamine amide, -; too dissimilar to predict by antiSMASH, N/A; not applicable.

3.5 Discussion

In this chapter, different conditions were used to prepare the NW *Pseudomonas* sp. strains supernatants. Since some of these were able to inhibit both Gram positive and Gram negative bacteria, it suggested the possibility that: 1) The antibacterial compound(s) in the supernatants could target a common site present in both Gram positive and negative indicator bacteria suggesting the presence of a broad spectrum antibiotic, 2) The antibacterial compound(s) in the supernatants targeted different sites in Gram positive and negative indicator bacteria. To determine which of these possibilities accounts for the antimicrobial activity would require the number of active compounds to be identified by extracting them from the supernatants using solvent extraction methods, followed by column chromatography.

The growth medium plays a vital role in the production of secondary metabolites with antimicrobial properties (Sajeed *et al*., 2011). This study demonstrated a significant relationship between the growth medium and the margins of inhibition formed by most strains, including *Pseudomonas* sp. strain NW27 and NW9 (p-value <0.001). This emphasized the role of growth medium in secondary metabolite production by bacteria (Romano *et al.*, 2018). Sasirekha *et al*., have demonstrated that increased siderophore production was obtained when *P. aeruginosa* FP6 was grown in succinate medium, which is an iron-deficient medium. This was followed

by King's B, succinate, glutamate and glucose medium (Sasirekha and Srividya, 2016). In agreement with this study, Sajeed *et al*., 2011 have demonstrated that maximum siderophore production by a clinical isolate of *P. aeruginosa* was achieved when it was grown in succinate medium compared to other media, including King's B medium. Moreover, regulation of secondary metabolite production is influences by nutrients in the growth media (Santamaria *et al*., 2022). For example, the production of rhamnolipid, a surfactant produced by *Pseudomonas aeruginosa*, is catalyzed by the RhlA enzyme (Zhu *et al*., 2008). The *rhlAB* operon is regulated by a network with at least three quorum sensing signals and is reliant on nutrients present in the media, requiring a high carbon-to-nitrogen or carbon-to-iron ratio (Santamaria *et al*., 2022). In this study, a nutrient rich medium (nutrient broth), a complex medium (Luria bernati medium) and a medium with minimal nutrients (M9 minimal media) were used to grow the 5 NW *Pseudomonas* sp. strains. This was carried out to understand the influence of media composition on secondary metabolite production by the bacteria.

The results of this chapter demonstrated the broad antimicrobial activity of the supernatant of *Pseudomonas* sp. strain NW7 grown in M9 minimal medium for 24 hours at 25°C. It was able to inhibit methicillin-resistant *S. aureus*, vancomycin-resistant *E. faecium*, *A. baumannii* and ampicillin-resistant *K. pneumoniae*. However, drop collapse assay results suggest

the inability of this bacterium to produce surfactants. Moreover, bioinformatics analysis of *Pseudomonas* sp. strain NW7 revealed its genome to code for BGCs involved in the production 2,4-DAPG and sessilin A and other metabolic products that are related to lankacidin C, pyoverdine, fengycin, mitomycin, methanobactin, thiazostatin and arylpolyene. It is therefore possible that 2,4-DAPG and sessilin A were the active antimicrobial compounds in these supernatants.

2,4-DAPG and sessilin A were metabolic products common to both *Pseudomonas* sp. strain NW7 and NW9. In this research, these two strains also exhibited antibacterial activity against methicillin-resistant *S. aureus* JE2 and vancomycin-resistant *E. faecium* E1162. 2,4-DAPG is an antibiotic compound produced by some *Pseudomonas fluorescens* isolates (Almario *et al*., 2017). Previously, 2,4-DAPG was shown to have moderate antibacterial activity against VRSA and some vancomycin-resistant *Enterococcus* spp. (VRE) strains (Isnansetyo *et al*., 2003). 2,4-DAPG was also seen to inhibit *E. coli* (Julian *et al*., 2020). 2,4-DAPG was recently isolated from *Pseudomonas fluorescens* when grown in King's B broth at 37°C at 120 rpm for 24 hours (Suresh *et al*., 2021). This condition does not match the growth conditions used in this research and in addition, King's B broth was not used as a cultivation medium. However nutrient media, which was one of the cultivation medium used in this research, has the same composition of peptones as Kings B broth. Therefore, the

peptone composition could be attributable to production of 2,4-DAPG, which is reflected by the inhibitory effect of the supernatant of *Pseudomonas* sp. strain NW7 (grown at 25°C in nutrient broth for 24,48 and 72 in a 150 rpm shaking incubator) against *S. aureus*, *E. faecium* and *A. baumannii*.

Sessilin is a cyclic lipopeptide produced by some *Pseudomonas* sp. and exhibits antifungal and anti-oomycetes properties (Ferrarini *et al*., 2022). In this research, sessilin BGC was predicted in the genome of *Pseudomonas* sp. strain NW7 and NW9. Unlike a previous report which demonstrate the production of sessilin A by *Pseudomonas* sp. CMR12a when cultured in King's B broth for 72 hours at 28°C, the *Pseudomonas* sp. strains NW7 and NW9 were grown in different culture media (including luria bernati broth, nutrient broth and M9 medium) and different cultivation temperatures (Geudens *et al*., 2018). The antibacterial activity of sessilin has not yet been investigated, however this research provided insight on the conditions under which *Pseudomonas* sp. strains NW7 and NW9 can possibly produce sessilin.

In addition, seven BGCs were predicted in the genome of *Pseudomonas* sp. strain NW7, but their metabolic products were not predicted by antiSMASH, possibly indicating novelty of these products. These possible

novel products might be responsible for the inhibitory activity displayed by the supernatants of this bacterium.

This research revealed the ability of the *Pseudomonas* sp. strain NW9 to inhibit *S. aureus, E. faecium* and *A. baumannii* when grown NB broth for 24 and 72 hours at 25°C. In addition, the supernatant of this bacterium grown in NB broth for 48 hours at 25°C, was able to inhibit *S. aureus*, *E. faecium* and *K. pneumoniae*. Bioinformatics analysis revealed 17 BGCs in the genome of *Pseudomonas* sp. strain NW9, with some carrying genes to encode for known compounds such as 2,4-DAPG and sessilin A. Bioinformatics analysis showed other BGCs that can encode proteins involved in the production of metabolic products related to pyoverdin, mitomycin, arylpolyene and lankacidin C.

Bioinformatics analysis demonstrated the presence of BGCs in the genome *Pseudomonas* sp. strain NW9 that carry genes possibly coding for the production of a fengycin-like compound, a lipopeptide biosurfactant with antifungal properties produced by *Bacillus* sp. (González-Jaramillo *et al*., 2017). However, *Pseudomonas* sp. strain NW9 was unable to produce surfactants as demonstrated by drop collapse assays. This indicates that growth conditions used to prepare the supernatants of *Pseudomonas* sp. strain NW9 were unable to activate the BGC with genes similar to those

coding for a fengycin-like compound. Bioinformatics analysis also revealed the genome of *Pseudomonas* sp. strain NW9 to code for 7 BGCs potentially involved in the production of metabolic products that are not similar to any compound in the MiBIG database. Since the supernatants of this bacterium displayed broad antimicrobial activity, the active compounds might be the metabolic products of these unique BGCs.

This study has revealed the inhibitory activity of *Pseudomonas* sp. strain NW10 supernatants to be limited to *S. aureus* and *A. baumannii*. This supernatant was prepared by growing *Pseudomonas* sp. strain NW10 in LB broth for 24 hours at 25°C. Interestingly, this supernatant also contained surfactants according to drop collapse assays. Bioinformatics analysis revealed the potential of this bacterium to code for BGCs involved in the biosynthesis of metabolic products related to lokisin, a cyclic lipopeptides with surfactant activity (Nielsen *et al*., 2002). It also displays an antifungal effect, controlling rice blast disease caused by the fungus *Magnaporthe oryzae* (Omoboye *et al*., 2019b). Based on these results, the surfactant in this supernatant could be a compound similar to lokisin, even though the genome of this strain had BGCs with genes that have the potential to code for the production of compounds related to fragin, arylpolyene, lankacidin and pyoverdin. Fragin has been previously isolated from *Pseudomona mediterannea* EDOX cultivated in King's B media in a 210 rpm 28°C shaking incubator for 24 (Zhou *et al*., 2021).

Moreover, previous researchers were unable to purify lankacidin from three *Pseudomonas* strains cultivated in King's B medium and luria bernati broth at 28°C, even though lankacidin BGC was predicted in their genomes (Rieusset *et al*., 2020). This can demonstrate how difficult it is to isolate lankacidin from different *Pseudomonas* strains.

In addition, bioinformatics analysis predicted 5 BGCs in the genome of *Pseudomonas* sp. strain NW10 that code for metabolic products that are too dissimilar for antiSMASH to predict, indicating possible novelty of these products.

The results in this chapter revealed the narrow antimicrobials activity of *Pseudomonas* sp. strain NW16 when grown in nutrient broth for 24 hours at 17°C and in luria bernati broth for 24 hours at 25°C, where they were able to inhibit *K. pneumoniae*. Conditions used to prepare the latter supernatant were also able to collapse the droplet, indicating the presence of surfactants. The genome of *Pseudomonas* sp. strain NW16 is likely to carry genes coding for proteins involved in the production of compounds related to fengycin and lokisin, which can indicate these compounds to potentially be responsible for the collapsed drop. Moreover, the latter supernatant produced one of largest drop collapse activity in this study, which could have been attributed to a high concentration of surface active

compounds, multiple surfactants or to different hydrophobicity of compound(s) produced. Quantification of surfactants can be further investigated using high performance liquid chromatography- mass spectrometry (HPLC-MS) (Wang and Kasperski, 2018). This can provide us a better understanding of the surfactant-producing capability of these bacteria.

The supernatants of *Pseudomonas* sp. strain NW27 grown in LB broth for 24 hours at 25°C had broad spectrum antimicrobial activity. This could either be due to the production of more than one antimicrobial compound or due to the antimicrobial targeting a common target in these indicator bacteria. However, this supernatant did not contain surfactants according to drop collapse assays. Moreover, *Pseudomonas* sp. strain NW27 was able to produce surfactants only when grown in luria bernati broth for 48 and 72 hours at 25°C. Bioinformatics analysis has also predicted its whole genome sequence to carry genes involved in the biosynthesis of metabolic products that might be related to lokisin, fengycin and viscosin. Viscosin is a secondary metabolite produced by some *Pseudomonas* spp., including *P. fluorescens* (Alsohim *et al*., 2014). It is a peptidolipid biosurfactant with antifungal properties (Martin *et al*., 2019).These three metabolic products could be novel as they were not identical to any compounds in the MiBIG database. It is possible that these were the active biosurfactants in this supernatant. This finding also suggests that culture conditions in this

study were successful in activating these BGCs which led to biosurfactant production and therefore, collapse of the drop.

Bioinformatics analysis also predicted BGCs in the genome of *Pseudomonas* sp. strain NW27 to code for proteins involved in the production of metabolic products related to pyoverdin and fragin. Pyoverdine is a siderophore necessary for *Pseudomonas aeruginosa* virulence and biofilm development (Peek *et al*., 2012). Pyoverdine inhibits the growth of several Gram negative bacteria and positive bacteria including *Bacillus megaterium* (Liu *et al*., 2021). No study has determined the antibacterial effect of pyoverdine-like compounds against one of the most problematic Gram negative bacteria *Acinetobacter baumannii*. In addition, two siderophores with no known similar clusters were predicted by antiSMASH in the genome of *Pseudomonas* sp. strain NW27. These are not similar to any compounds in the MiBIG database, indicating that they are novel.

Some BGCs, including those that have the potential to code for the production of compounds related to viscosin, bananamide and syringomycin, were exclusively predicted in the genome of *Pseudomonas* sp. strain NW27. This range of possible secondary metabolites could have contributed to the susceptibility of indicator bacteria to the supernatant of this strain in addition to its surfactant properties. Syringomycin is a

lipodepsipeptidic phytotoxin produced by *Pseudomonas* spp., including *Pseudomonas syringae* pv. *atrofacciens* (Vassilev *et al*., 1996). Syringomycin has antifungal activity (Sorensen *et al*., 1996). However, less efforts have been made to investigate its antibacterial properties against the ESKAPE pathogens.

Bananamide is a cyclic lipopeptide identified in *Pseudomonas*, which exhibits antifungal activity (Omoboye et al., 2019a). However, no studies demonstrate the capability of bananamide-like metabolites to inhibit the ESKAPE pathogens. An antifungal bananamide-type compound was previously isolated from *Pseudomonas* sp. COW3 isolate cultured in King's B broth under shaking conditions at 28°C for 24 h (Omoboye *et al*., 2019a). This growth condition was not comparable to the growth conditions in this research. In addition, bioinformatics analysis revealed the metabolic products of 6 BGCs in the genome of *Pseudomonas* sp. strain NW27 to be too dissimilar for antiSMASH to predict, indicating the potential of this strain to produce novel metabolic compounds.

It is noteworthy that all the NW *Pseudomonas* sp. strains used in this investigation produced secondary metabolites in response to growth conditions, including cultivation temperature. Maximum antibiotic production by *Pseudomonas* spp. is strain-dependent, however in our study, cultivation temperature was a determinant of antibiotic production

by the NW *Pseudomonas* sp. strains. A cultivation temperature of 25°C was significantly better in initiating antimicrobial production by the NW *Pseudomonas* sp. strains, even though the temperature of the natural habitat of these strains was 17°C. It could suggest the temperature of 25°C has led to increased transcription of the genes involved in the production of secondary metabolites. Previously, temperate soil *Pseudomonas* spp. were cultivated in luria bernati broth for 7 days at four different temperatures, 4, 17, 30 and 37°C (Mitsutomi *et al*., 2017). Optimal antibiotic production was at 4°C by which two compounds related to beta lactam antibiotics, sulfazecin and isosulfazecin, were isolated.

In this chapter *Pseudomonas* sp. strain NW27 was shown to produce surfactants, broad-spectrum antimicrobial compounds and many possible novel metabolic products. In the work I describe in the next chapter, antimicrobial production by this strain will be enhanced using ribosomal engineering strategies.

Chapter 4

Improvement of antimicrobial production by *Pseudomonas* sp. strain NW27

4.0 Improvement of antimicrobial production by *Pseudomonas* sp. strain NW27

4.1 Activation of cryptic biosynthetic gene clusters

Biosynthetic gene clusters (BGCs) are a clustered group of two or more genes in a genome that together can encode a biosynthetic pathway to produce a secondary metabolite (Cimermancic *et al*., 2014). Most biosynthetic gene clusters in the microbial genome are 'silent' or weakly expressed under conventional culture conditions, suggesting that many potential secondary metabolites are yet to be identified (Rutledge and Challis, 2015, Buijs *et al*., 2020). Several approaches can unlock these cryptic BGCs, including a cultivation based approach, the 'OSMAC' (one strain many compounds) approach, in addition to a molecular based approach such as ribosome engineering (Reen *et al*., 2015).

4.1.1 OSMAC approach

Production of secondary metabolites by microbes has been shown to vary with culture media and culture conditions (Neve *et al*., 2021). Stimulation of secondary metabolite production through changing culture conditions is a classic approach termed 'OSMAC' (one strain many compounds) (Bode *et al*., 2002). This term describes the ability of a single microbial strain to produce multiple compounds when grown in different culture conditions (Bode *et al*., 2002, Schwarz *et al*., 2021). In addition, changes in cultivation

parameters, such as flask shapes, aeration and temperature have been shown to activate cryptic biosynthetic pathways in several microbes (Geng and Belas, 2010, Sanchez *et al*., 2013, Veerabadhran *et al*., 2018). As an example, Sharma *et al*., (2017) used different cultivation conditions such as different media and cultivation temperatures to increase the amount of valinomycin produced by *Streptomyces lavendulae* (Sharma *et al*., 2017).

4.1.2 Ribosome engineering

Ribosome engineering is a molecular approach to discover microbes with spontaneous mutations in their ribosome or RNAP by screening for mutants that are resistant to antibiotics, including streptomycin, gentamycin and rifampicin, on petri-dishes (Zhu *et al*., 2019). Ribosome or RNAP mutations can result in activation of silent BGCs which in turn increases the amount of secondary metabolites or can produce novel natural products (Zhu *et al*., 2019).

Ribosome modulation can be performed using drugs that attach to the ribosome, one of which is streptomycin. Streptomycin is a protein synthesis inhibitor, binding to the small 16S rRNA of the 30S ribosomal subunit (Springer *et al*., 2001). It will interfere with the binding of formyl-methionyl-tRNA to the 30S subunit, causing inhibition of protein synthesis (Springer *et al*., 2001). Certain streptomycin resistance mutations can

result in antibiotic overproduction. For instance, mutations associated with streptomycin in *rpsL*, a gene encoding the ribosomal protein S12, lead to an error-restrictive phenotype (Carter *et al*., 2000). Aberrant protein synthesis activity has been demonstrated by *rpsL* mutant of *Streptomyces coelicolor*, which showed enhanced actinorhodin production due to increased expression of ribosome recycling factor (Hosaka *et al*., 2006). Moreover, a wild type *Streptococcus lividans* TK21 contains a complete actinorhodin *act* BGC, but does not normally produce actinorhodin (Shima *et al*., 1996). Streptomycin-resistant *Streptomyces lividans* strain TK24 was able to produce large quantities of actinorhodin, indicating the dependence of actinorhodin-activation on a mutated or 'engineered' ribosome (Shima *et al*., 1996). Another study demonstrated the use of ribosomal engineering to activate weakly expressed BGC pathways by *Streptomyces diastatochromogenes*. The streptomycin-resistant strain of *S. diastatochromogenes* had increased production of toyocamycin (Shentu *et al*., 2016).

In addition to ribosomal mutations, rifampicin- selected mutations in RNAP are shown to play a role in antibiotic overproduction (Tanaka *et al*., 2013). Rifampicin is an inhibitor of RNA polymerase (Zhu *et al*., 2019). Rifampicin resistance can arise due to mutations in *rpoB,* the gene encoding its β-subunit of RNAP (Ochi and Hosaka, 2013). Rifampicin

selects for bacteria with mutations in the β-subunit of the RNAP, increasing the transcription level of the silent BGC (Baral *et al*., 2018).

When bacteria encounter adverse environmental conditions such as amino acid scarcity, they exert a stringent response (Ochi, 2007). Repression of stable RNA synthesis is one of the most prominent elements of the stringent response (Ochi, 2007). A key mediator of this response is the bacterial alarmone guanosine-5'-diphosphate-3'-diphosphate (ppGpp). The ppGpp can bind to RNAP and affect the production of antibiotics (Artsimovitch *et al*., 2004). Thus, modifying the RNAP by introducing a rifampicin resistance mutation may mimic the ppGpp-bound form of RNAP with enhanced affinity to the promoter region of the secondary metabolite synthesis regulatory genes (Ochi *et al*., 2014). This will therefore activate dormant genes, which in turn leads to the expression of silent BGCs (Shima et al., 1996).

Using rifampicin in a ribosome engineering strategy, fredericamycin A, an anticancer agent, was produced by rifampicin resistant *Streptomyces somaliensis* ZH66-RIF_1, but not its wild-type strain (Zhang *et al*., 2015). This is suggestive of cryptic gene cluster activation in the mutant strain, with the mutation of R444H in the β-subunit of RNAP (Zhang *et al*., 2015). The effectiveness of selecting for drug-resistance mutations in favor of

strain improvement was also investigated by Tamehiro *et al*., (2003) who demonstrated the improvement of salinomycin production by *Streptomyces albus* isolates resistant to rifampicin (Tamehiro *et al*., 2003). Interestingly, other studies demonstrated the production of a family of novel antibiotics, piperidamycins, by two *Streptomyces mauvecolor rpoB* mutants (H437D or H437L) (Tanaka *et al*., 2013).

4.2 Aims and objectives of the work in this chapter

The aim of this chapter was to compare the antimicrobial activity of the supernatant of mutant *Pseudomonas* p. strain NW27-A1 to that of the wild-type and to investigate the uptake barrier. The objectives were:

- To enhance antibiotic production by *Pseudomonas* sp. strain NW27 utilizing ribosomal engineering techniques and to compare its antibacterial activity to that of the wild-type using agar well diffusion assays.
- To improve uptake of antimicrobial compounds, present in *Pseudomonas* sp. strain NW27-A1 supernatant by permeabilization of *Escherichia coli* NCTC 9081 cell envelope.

4.3 Materials and methods

4.3.1 Isolation and analysis of spontaneously rifampicin resistant mutants of *Pseudomonas* sp. strain NW27

4.3.1.1 Antibiotic susceptibility assay

In an effort to enhance antimicrobial production by *Pseudomonas* sp. strain NW27, I isolated mutants that were resistant to rifampicin. In order to achieve this, the MIC of rifampicin against *Pseudomonas* sp. strain NW27 was determined by the broth microdilution assay as described in Chapter 2 section 2.2.2. The MIC was determined as 1 µg/ml.

4.3.1.2 Isolation of *Pseudomonas* sp. strain NW27 resistant to rifampicin

One single colony of *Pseudomonas* sp. strain NW27 was grown in 10 mL LB broth in a 50 mL tube. The inoculum was incubated for 17 hours at 25°C with shaking at 150 rpm. Ten LB broth agar plates containing 10 µg/ml rifampicin and another ten LB broth agar plates containing 50 µg/ml rifampicin were used for this assay. 100 µl of the 17-hour *Pseudomonas* sp. strain NW27 culture was spread onto these agar plates using an L-shaped spreader. The plates were incubated at 30°C and were monitored for 24, 48 and 72 hours for the presence of colonies. Putative rifampicin-resistant *Pseudomonas* sp. strain NW27 colonies were grown in 10 mL LB broth for 17 hours and were used to prepare a bacterial stock. One mutant

which was named as *Pseudomonas* sp. strain NW27-A1, was used for further investigation.

To isolate single bacterial colonies of *Pseudomonas* sp. strain NW27-A1, streak-plating method was performed by streaking from bacterial stock onto LB agar containing 10 µg/ml rifampicin. The culture plate was left to incubate at 30°C for 24 hours until colonies were formed.

The mutant frequency is the ratio of mutants divided by the total number of bacteria in the population. The mutation frequency was estimated by dividing the number of rifampicin-resistant colonies on rifampicin-containing plates by the number of CFU on antibiotic-free plates (Sheng *et al*., 2020).

4.3.2 Large scale preparation of *Pseudomonas* sp. strain NW27 and *Pseudomonas* sp. strain NW27-A1 supernatants

For analysis of secondary metabolites with antimicrobial properties, large volumes of filtered supernatant were prepared and concentrated. This concentrated supernatant was also used for further analysis in the work described in Chapter 5.

Initially, *Pseudomonas* sp. strain NW27 was streaked from frozen stock onto LB broth agar plates and *Pseudomonas* sp. strain NW27-A1 was streaked onto an LB plate containing 10 µg/ml rifampicin. Both plates were incubated at 30°C for 24 hours. After 24 hours, one colony of each strain was cultured in 10 mL LB broth in a 50 mL tube. These starter cultures were incubated for 17 hours at 25°C on a rotary shaker set at 150 rpm.

To prepare large volumes of supernatant, 200mL of LB broth were prepared and autoclaved in a 1 liter glass bottle. The 17-hour starter cultures of *Pseudomonas* sp. strain NW27 and *Pseudomonas* sp. strain NW27-A1 were used to adjust the optical density of the 200mL LB broth to a final optical density of 0.05 (OD_{600nm}). This assay was done in triplicate.

The 200 ml cultures were then incubated for 48 hours at 25°C on a rotary shaker set at 150 rpm. After 48 hours incubation, supernatants were collected by centrifuging in the Sorvall RC-5B Plus centrifuge for 30 minutes at 19592 xg. These supernatants were filtered using Nalgene Rapid flow disposable filter units (0.45 µm pore size) using a vacuum pump. The inhibitory activity of these supernatants against indicator bacteria (*S. aureus* JE2, *E. faecium* E1162, *Klebsiella pneumoniae* C3091, *Acinetobacter baumannii* NCTC12156*,*
Pseudomonas aeruginosa FRO1 (low mucoid), *Enterobacter*

cloacae NCTC 10005) were tested using agar well diffusion assays as described in Chapter 2 section 2.2.1. The filtered supernatants were stored at -20°C until needed.

4.3.2.1 Concentrating large volumes of *Pseudomonas* sp. strain NW27 and *Pseudomonas* sp. strain NW27-A1 supernatants

Large volumes of *Pseudomonas* sp. strain NW27 and *Pseudomonas* sp. strain NW27-A1 supernatants were concentrated to 140 mg/ml using a freeze dryer. The 200 mL of supernatant was initially frozen at -80°C in a 1L freeze drying plastic flask. The frozen sample was placed onto the freeze dryer with the temperature adjusted to -120°C.The sample was removed from the freeze dryer when a powder was formed (after approximately 72 hours). The dry powder was scraped off the flask using a blunt scalpel. The weight of an empty 50mL tube was measured. The dry powder was added to this tube and the weight was measured. The mass of the empty tube and the tube after adding the powder were subtracted to give the weight of the powder. To achieve a high stock concentration of the active compounds, the dry powder was measured and reconstituted with sterile distilled water to a concentration of 140 mg/ml.

4.3.3 Evaluating the antimicrobial activity produced by *Pseudomonas* sp. strain NW27 and *Pseudomonas* sp. strain NW27-A1

4.3.3.1 Minimum inhibitory concentration of *Pseudomonas* sp. strain NW27 and *Pseudomonas* sp. strain NW27-A1 supernatants against indicator bacteria

The concentrated supernatant (140 mg/ml) prepared in large volumes was used in this assay. The MIC of *Pseudomonas* sp. strain NW27 supernatant and *Pseudomonas* sp. strain NW27-A1 supernatant against *S. aureus* JE2, *E. faecium* E1162 and *A. baumannii* NCTC 12156, was determined using the broth microdilution assay as described in Chapter 2 section 2.2.2. To achieve a high primary concentration of the concentrated supernatant, serial 2-fold broth dilutions starting at 17.5 mg/ml until 0.28 mg/ml of supernatant were made in LB broth in a 96-well microtiter plate. This assay was done in triplicates using three biological repeats and the median values were determined.

4.3.3.2 Agar-well diffusion assays

The inhibitory activity of the supernatants against *S. aureus* JE2, *E. faecium* E1162 and *A. baumannii* NCTC 12156 and *E. coli* NCTC 9081 were tested using agar-well diffusion assays as described in Chapter 2 section 2.2.1. The concentrated supernatants (140 mg/ml) prepared in

large volumes were diluted to 10.52 mg/ml and used in this assay. These assays were done in triplicates using three biological replicates.

4.3.3.3 Alamar blue cellular metabolism assay

Alamar blue is a redox indicator used to evaluate metabolic activity and cellular health of many cell types, including bacteria (Yajko *et al*., 1995). If growth and bacterial metabolism is sufficient, the oxidation-reduction indicator alamar blue changes color from blue to bright pink (Yajko *et al*., 1995) (Figure 4.1). As cells being tested grow, metabolic activity will lead to a chemical reduction of resazurin to resorufin (Page *et al*., 1993). Resazurin acts as an electron acceptor in the electron transport chain (Page *et al*., 1993). Alamar blue can be reduced by nicotinamide adenine dinucleotide (NADH) and nicotinamide adenine dinucleotide phosphate (NADPH) (Page *et al*., 1993). Once the REDOX indicator accepts electrons, it changes from oxidized (non-fluorescent, blue) state to reduced (fluorescent, pink) state (Page *et al*., 1993). Once it enters a living cell, resazurin will be reduced to resorufin, emitting a highly fluorescent pink color (Pettit *et al*., 2009). A decreased rate of reduction, which is demonstrated by less fluorescence, therefore suggests an impairment of cellular metabolism (Pettit et al., 2009). This assay was used in this study to determine if the supernatants affected cellular metabolism of indicator bacteria.

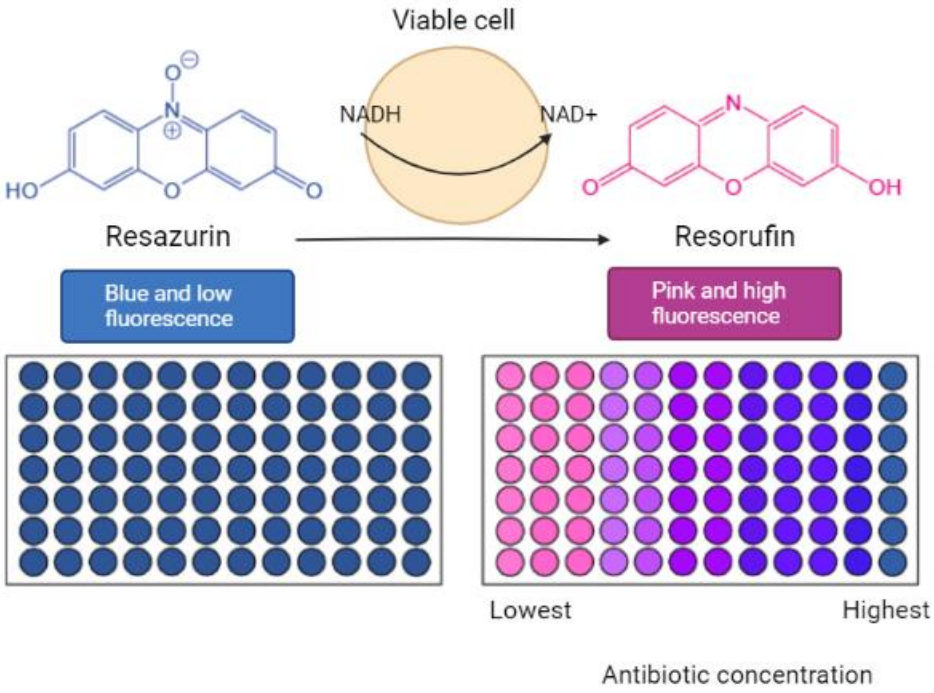


Figure 4.1 The assay principle of Alamar blue.
A schematic figure showing reduction of resazurin to resorufin. This figure was created using Biorender.com.

Indicator bacteria (*S. aureus* JE2, *E. faecium* E1162 and *A. baumannii* NCTC 1256) were grown as described in Chapter 2 section 2.1.1, except that the 16 hour cultures were adjusted to an OD_{600nm} of 0.05 (equivalent of 10^8 CFU/mL). The required volume of starter culture of indicator bacteria to be used in this assay was 5 mL. To the 5 mL of starter culture, 20 µl of Alamar blue (4 mg/ml) was added. Using a 96-well black plate (Thermo Scientific, UK), 100 µl of this mixture was pipetted into each well. For the test wells, the MICs of both *Pseudomonas* sp. strain NW27 and *Pseudomonas* sp. strain NW27-A1 supernatants were used in this assay.

100 µl of supernatant was pipetted into the wells containing indicator bacteria. Control wells were prepared by adding 100 µl sterile LB broth instead of supernatants. The fluorescence was monitored by incubating the 96 well plate at 30°C for 12.75 hours in a Fluoroskan Ascent ™ Microplate Fluorometer using excitation set at 530 nm and emission set at 590 nm. The metabolic activity was plotted using OriginPro2019.

4.3.4 Sensitization of *E. coli* NCTC 9081 to the supernatant of *Pseudomonas* sp. strain NW27-A1 using polymyxin B nonapeptide (PMBNP)

Some of the Gram negative bacteria mentioned in this chapter were resistant to inhibition by the supernatant of *Pseudomonas* sp. strain NW27 and *Pseudomonas* sp. strain NW27-A1.

Polymyxin B nonapeptide (PMBNP) is a permeabilizing agent, making Gram negative bacteria more permeable to other antibiotics by binding to bacterial lipopolysaccharides (Viljanen and Vaara, 1984, Tsubery *et al.*, 2000). The aim of this assay was to sensitize *E. coli* NCTC 9081 to treatment with *Pseudomonas* sp. strain NW27-A1 supernatant using PMBNP as the permeabilizing agent. The supernatant of *Pseudomonas* sp. strain NW27-A1 was included in this assay because it showed enhanced antimicrobial activity compared to the wild-type strain.

A stock solution of PMBNP (100 µg/ml) was prepared using sterile distilled water and was filtered using 0.22 µm syringe filters (Thermo Scientific, UK). The MIC of PMBNP was determined by broth microdilution as described in Chapter 2 section 2.2.2 using a range of concentrations (1.57-100 µg/ml).

The supernatant of *Pseudomonas* sp. strain NW27-A1 (140 mg/ml) was serially diluted two-folds in 100 µl volumes in a non-treated 96-well plate (VWR, UK), after which 100 µl of indicator bacterium (*E. coli* NCTC 9081) was added. The indicator bacteria were prepared as described in Chapter 2 section 2.1. Then, 50 µl of 100 µg/ml PMBNP was added to each well in the 96-well plate. Growth of the indicator bacteria in the presence of each supernatant was determined using a spectrophotometer which measured the optical density (OD_{600nm}) of each sample every 6 minutes for 13 hours. Growth of non-treated, non-sensitized *E. coli* was also determined (control 1) in addition to growth of non-treated, sensitized *E. coli* (control 2). This assay was performed in biological triplicates.

4.4 Results

4.4.1 Enhancement of antibiotic production by *Pseudomonas* sp. strain NW27

4.4.1.1 Isolation of *Pseudomonas* sp. strain NW27 resistant to 10 ug/ml rifampicin

The MIC of rifampicin against *Pseudomonas* sp. strain NW27 was 1 ug/ml. I therefore chose to select for resistant *Pseudomonas* sp. strain NW27 mutants using 10 ug/ml rifampicin and 50 ug/ml rifampicin (10X and 50X the MIC). These mutants were required in this study in order to increase antibiotic production by *Pseudomonas* sp. strain NW27.

Colonies resistant to 50 ug/ml rifampicin were not isolated. However, I managed to isolate ten *Pseudomonas* sp. strain NW27 colonies that were resistant to 10 ug/ml rifampicin. The frequency at which the spontaneous rifampicin mutants arose was 1×10^{-6}.

One interesting mutant, which was named *Pseudomonas* sp. strain NW27-A1, produced larger margins of inhibition, in the agar well diffusion assays, when compared to wild-type *Pseudomonas* sp. strain NW27. Upon statistical analysis using Mann-Whitney non-parametric test, a significant difference was found between the margins of inhibition formed by *Pseudomonas* sp. strain NW27 and NW27-A1 (p-value <0.001).

Moreover, a statistically significant association between the time of harvesting the supernatant and the zones of inhibition formed by *Pseudomonas* sp. strain NW27 and NW27-A1 was also demonstrated (*p*-value <0.005).

Inhibitory activity of the supernatants of *Pseudomonas* sp. strain NW27-A1 was observed against *S. aureus* JE2, *E. faecium* E1162 and *A. baumannii* NCTC 12156 (Table 4.1). However, *Pseudomonas aeruginosa* FRO1 (low mucoid), *Enterobacter cloacae* NCTC 10005, and *Klebsiella pneumoniae* C3091 were resistant to these supernatants. The mutant, *Pseudomonas* sp. strain NW27-A1, was used for further analysis in this research.

Table 4.1 Comparison of margins of inhibition formed by the supernatants of *Pseudomonas* sp. strain NW27 and *Pseudomonas* sp. strain NW27-A1 using agar well diffusion assays.

Indicator bacterium	***Pseudomonas* sp. strain NW27**			***Pseudomonas* sp. strain NW27-A1**		
	Time of harvesting supernatant (h)					
	24	48	72	24	48	72
***S. aureus* JE2**	9 (9-11)	11 (11)	8 (8-9)	13 (13)	14 (14)	10 (9-10)
***E. faecium* E1162**	11 (10-13)	11 (11-12)	9 (9-10)	15 (13-15)	15 (15)	12 (9-13)
***A. baumannii* NCTC 1256**	3 (3-5)	6 (6)	2 (2-3)	7 (7-8)	7 (6-10)	4 (4)
***K. pneumoniae* C3091**	3 (3)	0 (0)	0 (0)	(0)	0 (0)	0 (0)

The mean margins of inhibition (mm) of filtered supernatants of *Pseudomonas* sp. strain NW27 and *Pseudomonas* sp. strain NW27-A1 (grown in Luria Bernati broth at 25°C on a rotary shaker set at 150 rpm) against indicator bacteria *S. aureus* JE2, *E. faecium* E1162, *A. baumannii* NCTC 12156 and *K. pneumoniae* C3091. Supernatants were collected at different time points (24, 48 and 72 hours) and were tested using agar well diffusion assays. Results of *P. aeruginosa* FRO1 (low mucoid), *E. cloacae* NCTC 10005 and *E. coli* NCTC 9081 are not shown because they were resistant to all the supernatants. Margins of inhibition were measured as millimeters (mm). h; hours. Values represent a set of triplicate results and are expressed as mean (with range).

4.4.1.2 Minimum inhibitory concentration of *Pseudomonas* sp. strain NW27 and *Pseudomonas* sp. strain NW27-A1 supernatants

The MIC of the supernatants of *Pseudomonas* sp. strain NW27 and *Pseudomonas* sp. strain NW27-A1 against *S. aureus* JE2, *E. faecium* E1162 and *A. baumannii* NCTC 12156 are shown in Table 4.2. The MIC values were log-transformed (base 2; expressed as log2 MIC. Upon statistical analysis using Mann-Whitney test, no significant difference was observed between the MICs of *Pseudomonas* sp. strain NW27 and NW27-A1 (*p*-value=0.200).

Table 4.2 Minimum inhibitory concentration of *Pseudomonas* sp. strain NW27 and *Pseudomonas* sp. strain NW27-A1 supernatants against indicator bacteria.

Indicator bacterium	Minimum inhibitory concentration (MIC) mg/ml	
	***Pseudomonas* sp. strain NW27**	***Pseudomonas* sp. strain NW27-A1**
S. aureus JE2	1.13 (1.13)	0.12 (0.12)
E. faecium E1162	1.13 (1.13)	0.12 (0.12)
A. baumannii NCTC 12156	2.13 (2.13)	1.13 (1.13)

Comparison of MIC values of *Pseudomonas* sp. strain NW27 and *Pseudomonas* sp. strain NW27-A1 supernatants against the representative indicator bacteria. The values represent a set of triplicate results as a median value (with range).

4.4.2 Metabolic function assay using Alamar blue dye

The supernatants were tested at different concentrations against *S. aureus* JE2, *E. faecium* E1162 and *A. baumannii* NCTC 12156.

Cellular metabolism of *S. aureus* was reduced after treatment with the MICs of the supernatants of *Pseudomonas* sp. strain NW27 (1.13 mg/ml) and NW27-A1 (0.12 mg/ml), respectively demonstrated as a semi-flattened ascending curve and a flat curve (Figure 4.2). Treatment using the latter supernatant was more effective in reducing *S. aureus* cellular metabolism as demonstrated by the lower relative fluorescent units (RFUs). Treatment using *Pseudomonas* sp. strain NW27 supernatant (0.12 mg/ml) also reduced cellular metabolism of *S. aureus* JE2 as demonstrated by the flattened curve. Metabolic activity of *S. aureus* was also affected after treatment with 0.08 mg/ml of *Pseudomonas* sp. strain NW27 and NW27-A1 supernatants, demonstrated as ascending curves. The latter treatment displayed lower RFUs compared to the former.

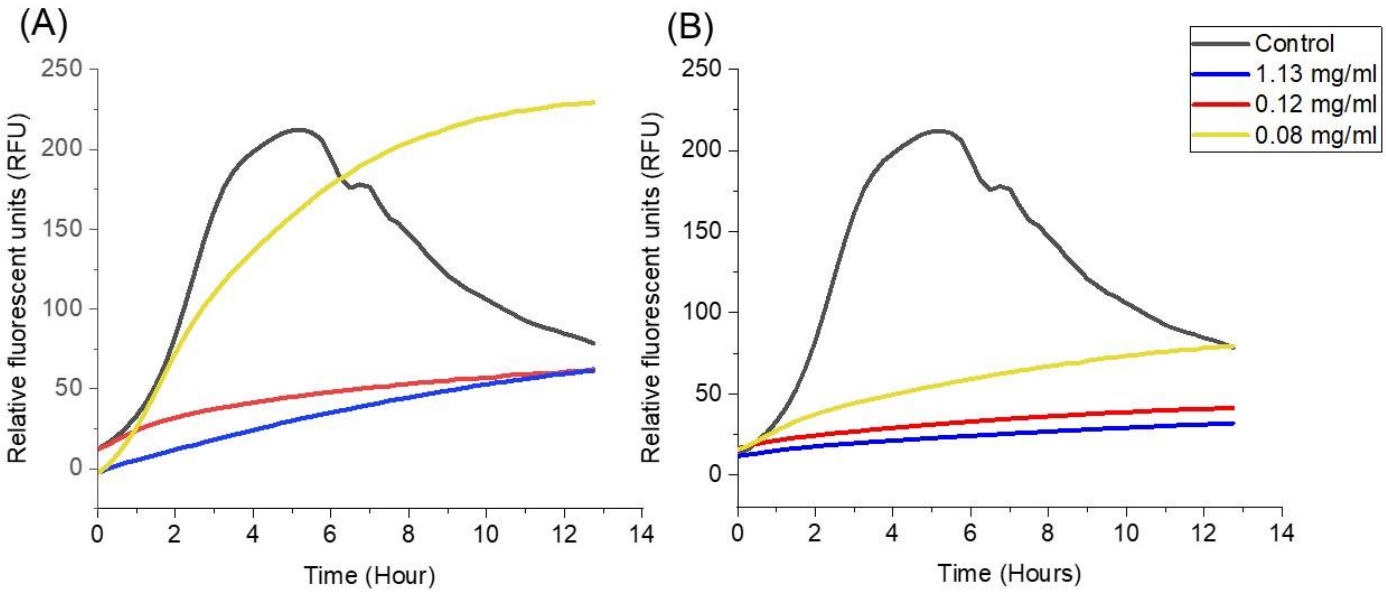


Figure 4.2 Metabolic activity of *S. aureus* JE2 after treatment with different concentrations of *Pseudomonas* sp. strain NW27 supernatants and *Pseudomonas* sp. strain NW27-A1 supernatants.

(A) Treatment of *S. aureus* JE2 using three concentrations of *Pseudomonas* sp. strain NW27 (B) Treatment of *S. aureus* JE2 using three concentrations of *Pseudomonas* sp. strain NW27-A1. Three different concentrations (1.13, 0.12 and 0.08 mg/ml) of both supernatants were used. Control represents metabolic activity of *S. aureus* without any treatment. Three replicates of each control and treatment were performed, and their mean values were plotted as the relative fluorescent units.

The metabolic activity of *E. faecium* was reduced after treatment with the supernatants of *Pseudomonas* sp. strain NW27 and NW27-A1 at the MICs of 1.13 mg/ml and 0.12 mg/ml respectively (Figure 4.3). Treatment using the latter was better in limiting *E. faecium* cellular metabolism, as demonstrated by the flatter curve. Treatment using lower concentrations of both supernatants NW27 (0.12 mg/ml) and NW27-A1 (0.08 mg/ml) also affected cellular metabolism of *E. faecium*. The latter treatment produced low RFUs which rapidly increased after 10 hours of treatment as demonstrated by the ascending curve. Treatment of *E. faecium* using the former supernatant displayed low RFUs and an ascending curve. Metabolic activity of *E. faecium* was minimally affected after treatment using the supernatant of *Pseudomonas* sp. strain NW27 (0.08 mg/ml) as demonstrated by the curve which was similar to the control curve. It indicates the inability of this low concentration of supernatant to reduce cellular metabolism.

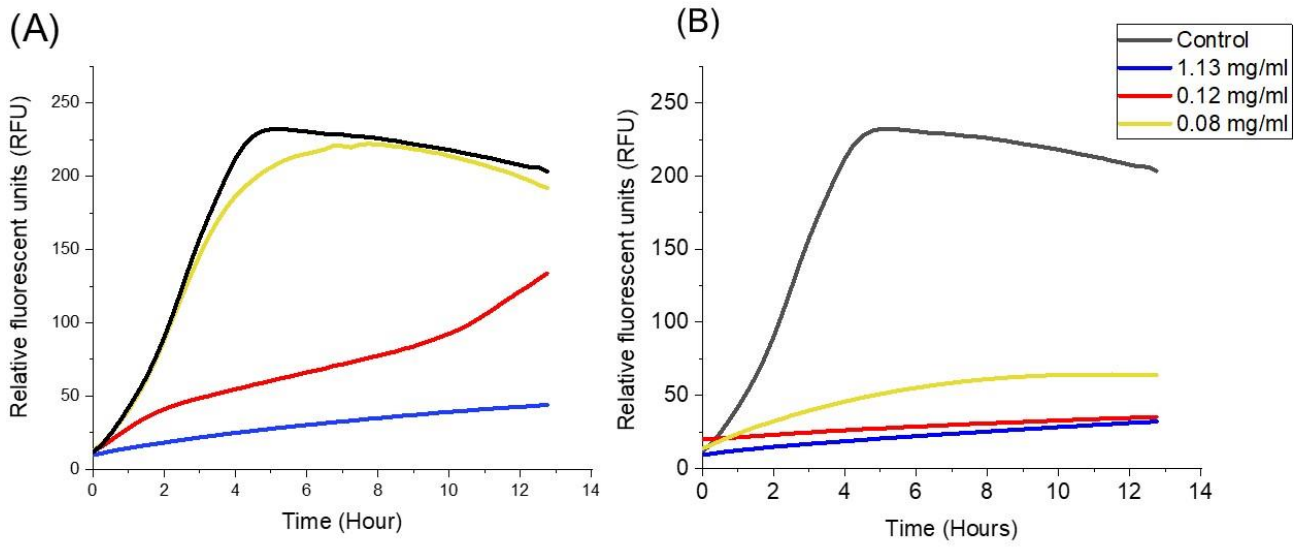


Figure 4.3 Metabolic activity of *E. faecium* E1162 after treatment with different concentrations of *Pseudomonas* sp. strain NW27 supernatants and *Pseudomonas* sp. strain NW27-A1 supernatants.

(A) Treatment of *E. faecium* E1162 using three concentrations of *Pseudomonas* sp. strain NW27 (B) Treatment of *E. faecium* E1162 using three concentrations of *Pseudomonas* sp. strain NW27-A1. Three different concentrations (1.13, 0.12 and 0.08 mg/ml) of both supernatants were used. Control represents metabolic activity of *E. faecium* E1162 without any treatment. Three replicates of each control and treatment were performed, and their mean values were plotted as the relative fluorescent units.

Cellular metabolism of *A. baumannii* was reduced after treatment with the MICs of the supernatants of *Pseudomonas* sp. strain NW27 (2.13 mg/ml) and NW27-A1 (1.13 mg/ml). The former treatment produced a flattened curve with an ascending gradient (Figure 4.4). The latter treatment was better in limiting cellular metabolism as it formed a flattened curve with low RFUs. Treatment using lower concentrations of both supernatants *Pseudomonas* sp. strain NW27 (1.13 mg/ml) and NW27-A1 (0.12 mg/ml) also affected cellular metabolism of *A. baumannii*. Resurgence of cellular metabolism was respectively noticed 3 hours and 6 hours after treatment with these supernatants. Cellular metabolism of *A. baumannii* was minimally affected after treatment with the supernatant of *Pseudomonas* sp. strain NW27 (0.12 mg/ml). It indicates the inadequate concentration of supernatant to reduce cellular metabolism.

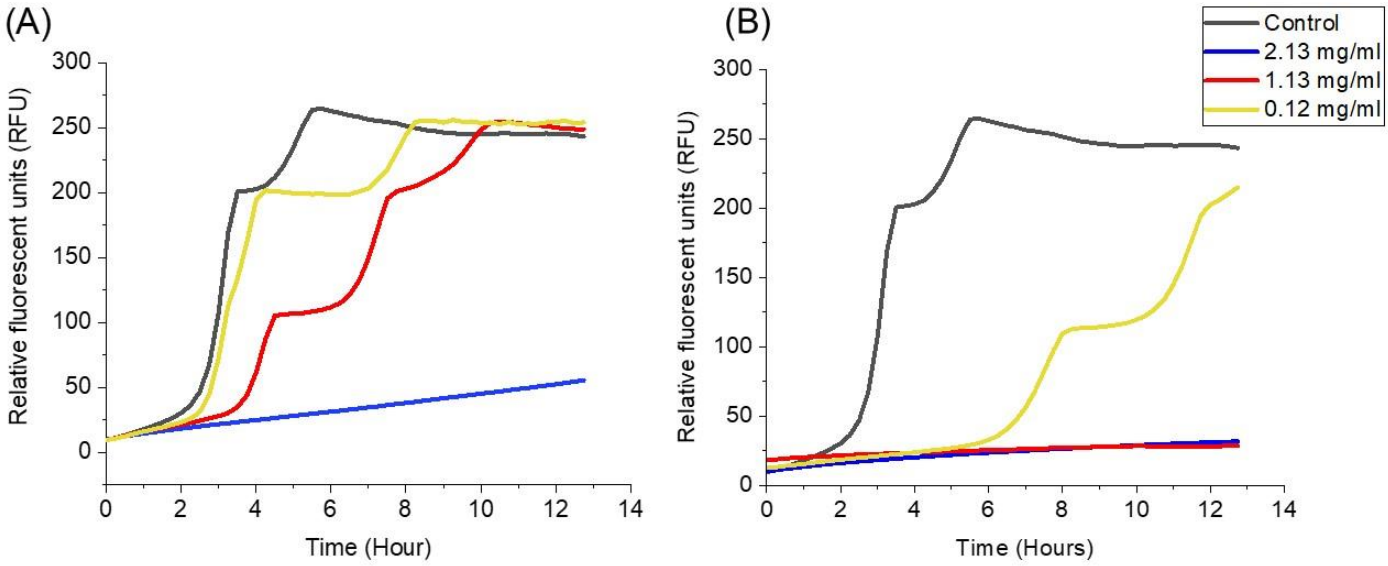


Figure 4.4 Metabolic activity of *A. baumannii* NCTC 12156 after treatment with different concentrations of *Pseudomonas* sp. strain NW27 supernatants and *Pseudomonas* sp. strain NW27-A1 supernatants.

(A) Treatment of *A. baumannii* NCTC 12156 using three concentrations of *Pseudomonas* sp. strain NW27 (B) Treatment of *A. baumannii* NCTC 12156 using three concentrations of *Pseudomonas* sp. strain NW27-A1. Three different concentrations (2.13, 1.13 and 0.12 mg/ml) of both supernatants were used. Control represents metabolic activity of *A. baumannii* NCTC 12156 without any treatment. Three replicates of each control and treatment were performed, and their mean values were plotted as the relative fluorescent units.

4.4.3 Growth of sensitized *E. coli* NCTC 9081 in the presence of *Pseudomonas* NW27-A1 supernatants.

PMBNP (100 µg/ml) was used to increase the outer membrane permeability of *E. coli* NCTC 9081 and facilitate the entrance of components of *Pseudomonas* sp. strain NW27-A1 supernatant into *E. coli* NCTC 9081.

The absorbance values of non-treated sensitized *E. coli* (+PMBNP-NW27A1) were lower than the non-treated non-sensitized *E. coli* (-PMBNP-NW27A1) (Figure 4.5). PMBNP therefore affected the growth of *E. coli* in this assay, delaying its growth as demonstrated by the curve in Figure 4.5. Non-treated sensitized *E. coli* (+PMBNP-NW27A1) reached log phase 1 hour post-incubation, and non-treated non-sensitized *E. coli* (-PMBNP-NW27A1) reached log phase of growth at approximately 2 hours post-incubation.

Treatment with 5.12 mg/ml of *Pseudomonas* sp. strain NW27-A1 was able to inhibit the growth of sensitized *E. coli* NCTC9081 (+PMBNP+NW27A1) (Figure 4.5A). This concentration was also affecting the growth of non-sensitized *E. coli* as demonstrated by the lower OD_{600nm} compared to the non-sensitized *E. coli* control (-PMBNP+NW27A1).

As demonstrated by the ascending curve, sensitized *E. coli* grew after 11 hours of incubation when treated with a lower concentration of *Pseudomonas* sp. strain NW27-A1 supernatant (4.12 mg/ml) (+PMBNP+NW27A1) (Figure 4.5B). In addition, non-sensitized *E. coli* started to regain its growth after treatment with this supernatant (-PMBNP+NW27A1).

Sensitized *E. coli* grew 8.5 hours after treatment using 3.13 mg/ml *Pseudomonas* sp. strain NW27-A1 supernatant, which was demonstrated by the ascending curve (+PMBNP+NW27A1) (Figure 4.5C). Non-sensitized *E. coli* was able to grow in the presence of this supernatant (-PMBNP+NW27A1).

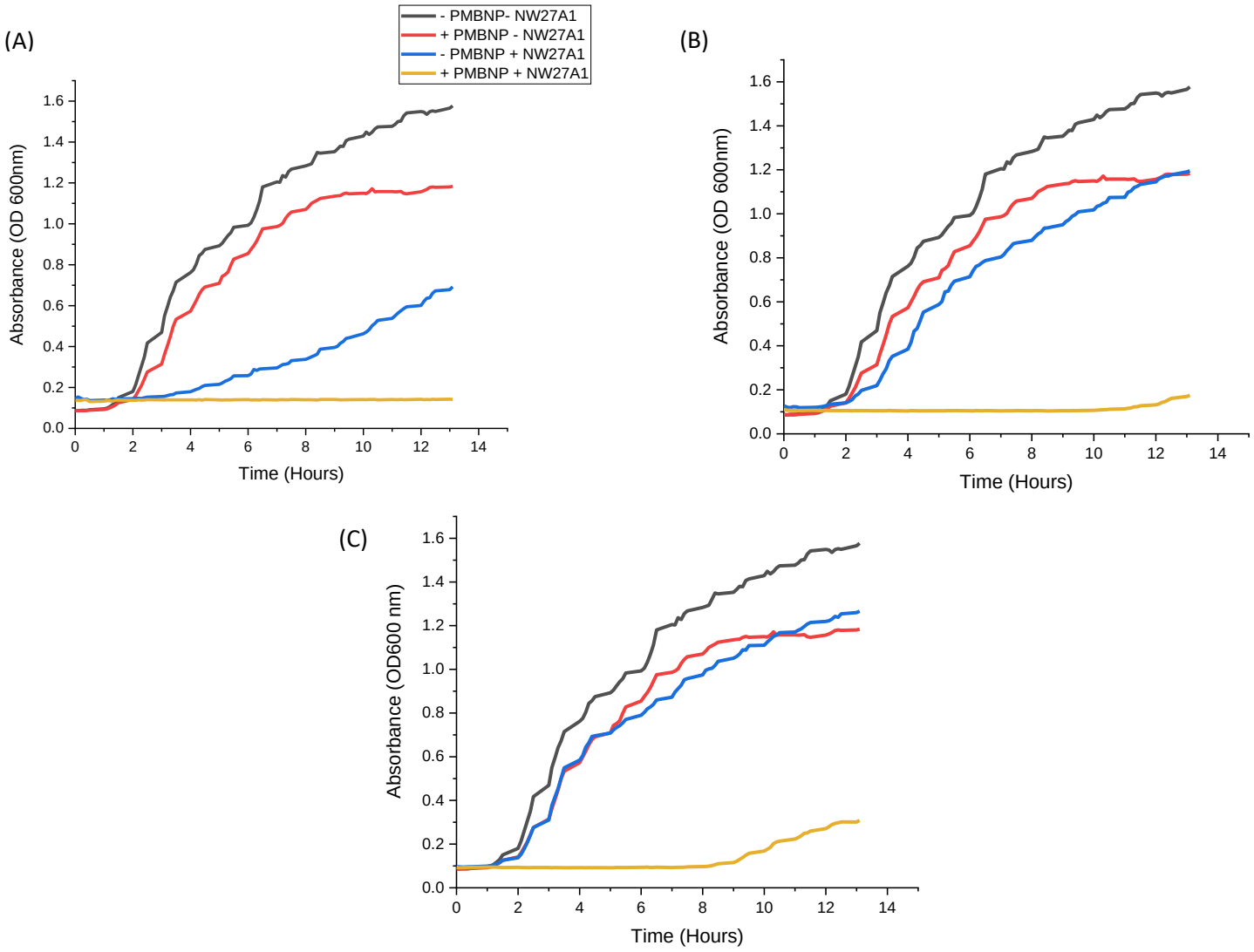


Figure 4.5 Sensitization of *E. coli* NCTC 9081 to *Pseudomonas* sp. strain NW27-A1 supernatant using Polymyxin B nonapeptide (100 µg/ml).

(A) Growth curve of *E. coli* NCTC 9081 when incubated with 5.12 mg/ml *Pseudomonas* sp. strain NW27-A1 in the presence and absence of PMBNP (100 µg/ml).

(B) Growth curve of *E. coli* NCTC 9081 when incubated with 4.12 mg/ml *Pseudomonas* sp. strain NW27-A1 in the presence and absence of PMBNP (100 µg/ml).

(C) Growth curve of *E. coli* NCTC 9081 when incubated with 3.13 mg/ml *Pseudomonas* sp. strain NW27-A1 in the presence and absence of PMBNP (100 µg/ml).

Growth was measured using a spectrophotometer. The experiment was performed three times using three biological replicates.

-PMBNP-NW27A1; non-sensitized *E. coli* without treatment using supernatant

+PMBNP-NW27A1; sensitized *E. coli* without treatment using supernatant

-PMBNP+NW27A1; non-sensitized *E. coli* after treatment with supernatant

+PMBNP+NW27A1 sensitized *E. coli* after treatment with supernatant

4.5 Discussion

Natural antibiotic discovery and yield improvement of antibiotic-producing bacterial strains is useful for obtaining more or different antimicrobial compounds (Zhu *et al*., 2019). Such approaches include ribosomal engineering using rifampicin to screen for mutants with increased secondary metabolite production (Zhu *et al*., 2019). This approach is effective in improving the production of antibiotics from wild-type strains (Shentu *et al*., 2016). Ribosome engineering is a rapid, cost-effective way to enhance secondary metabolite production, which is why I chose to employ this technique (Zhu *et al*., 2019).

Even though *Pseudomonas* sp. strain NW27 showed inhibitory activity against *S. aureus* JE2, *E. faecium* E1162 and *A. baumannii* NCTC1256 as described in Chapter 3, the production of antimicrobial compounds was further enhanced by selecting for rifampicin-resistant mutations. One of the previous studies demonstrated the effect of rifampicin resistance (*rpoB*) mutations in actinomycetes including *Streptomyces* and *Amycolatopsis*, producing multiple metabolites that were undetectable in the wild-type strains (Tanaka *et al*., 2013). These *rpoB* mutations were effective in activating silent secondary metabolite BGCs not normally active under general laboratory conditions (Tanaka *et al*., 2013). These mutations also enhanced the production of erythromycin, indicating the activation of weakly expressed BGCs (Tanaka *et al*., 2013).

This study was successful in isolating mutants of *Pseudomonas* sp. strain NW27 resistant to 10 µg/ml rifampicin. The margins of inhibition against *S. aureus* JE2, *E. faecium* E1162 and *A. baumannii* NCTC1256 were significantly increased, indicating the possible activation of weakly expressed BGCs. Interestingly, *K. pneumoniae* was resistant to the supernatants of *Pseudomonas* sp. strain NW27-A1, while sensitive to the parental strain NW27. This can suggest the supernatant of *Pseudomonas* sp. strain NW27-A1 had different antimicrobial compounds to the wild-type *Pseudomonas* sp. strain NW27 as opposed to more of the same compound originally produced by the latter. The mutation might have led to downregulation of certain regulatory genes responsible for the regulation of secondary metabolite synthesis (Covington *et al.,* 2021). Moreover, transcriptional regulators can upregulate or downregulate required gene sets depending on environmental ques (Covington *et al*., 2021). The growth conditions in this study might have triggered regulators to downregulate gene sets that were upregulated in the wild-type.

It is hypothesized that the observed enhanced effect could be a consequence of activation of cryptic or weakly expressed BGCs, in the genome of *Pseudomonas* sp. strain NW27-A1. It would be of interest to determine this by gene expression analysis using reverse-transcriptase quantitative PCR (RT-qPCR) (Shostak *et al*., 2020). In addition, the

secretory system that controls the secretion of antibiotics may be responsible for the difference in antimicrobial production between mutant *Pseudomonas* sp. strain NW27-A1 and its parental strain. The growth conditions used in this study may have allowed *Pseudomonas* sp. strain NW27-A1 to express the efflux mechanism that is encoded by genes present in the BGC. Molecular genetic studies at the transcriptional or post-transcriptional level could provide a better understanding to the secretory system of both strains (Lazzaro *et al*., 2017). Moreover, further analysis is required to determine the molecular basis of rifampicin resistance that caused antibiotic overproduction by *Pseudomonas* sp. strain NW27-A1.

In this study, there was a reduction in cellular metabolism of *S. aureus*, *E. faecium* and *A. baumannii*, after treatment with the MICs of *Pseudomonas* sp. strain NW27 and NW27-A1 supernatants, with the latter more effectively reducing the cellular metabolism of the indicator bacterium. Treatment of indicators using the MIC of the wild type *Pseudomonas* sp. strain NW27 supernatant demonstrated ascending metabolic activity, which can indicate that the concentration of antimicrobial compounds was insufficient to quench all metabolic activity of indicator bacteria. This effect could be caused by increased futile cycling, which accelerates cellular metabolism while reducing bacterial growth (Stokes *et al*., 2019). The energetically wasteful futile cycle occurs when two metabolic pathways run

simultaneously in opposite directions leading to neither consumption nor production of the metabolites (Chao *et al*., 1994). Some antibiotics such as β-lactams, inhibit PBP2, but initialize a futile cycle of cell wall biosynthesis and degradation, which increases the metabolic rate of bacteria (Lobritz *et al*., 2022). To have a further understanding of the metabolic consequences of treatment of the indicator bacteria (*S. aureus*, *E. faecium* and *A. baumannii*) with the supernatants, gene-knockout can be performed by which the genes that express proteins involved in futile cycle can be deleted (Liu *et al*., 2016).

Growth of *A. baumannii* after treatment with the supernatant of *Pseudomonas* sp. strain NW27-A1 (0.12 mg/ml) showed a pattern of growth that represents a stringent response. This response is a survival mechanism used by bacteria to cope with stress such as nutrient starvation and other stress conditions (Aedo, 2016). The stringent response is mediated by the accumulation of ppGpp, which controls broad metabolic alterations necessary for survival under adverse conditions (Gaca *et al*., 2013). This alarmone is synthesized by two distinct enzymes, RelA and SpoT (Ono *et al*., 2020). *A. baumannii* Δ*relA* deletion mutant and a Δ*relA*Δ*spoT* double mutant can be generated in the future to gain a further understanding of the metabolic effect of ppGpp post-supernatant treatment (Bai *et al*., 2021).

Moreover, treatment of indicator bacteria with the MIC of *Pseudomonas* sp. strain NW27-A1 supernatant did not demonstrate ascending resurgence in metabolic activity. This can be due to the presence of more and/or different antimicrobials in this supernatant. This could also indicate that the compound(s) in these supernatants are bacteriostatic rather than bactericidal. Time-kill kinetics assays could be used in future work to determine if the antimicrobial compound(s) in the supernatant were bactericidal or bacteriostatic.

Cellular metabolism of the indicator bacteria was also affected when treated using concentrations lower than the MICs, however the reduced metabolic activity was not maintained. The recovery of cellular metabolism of *E. faecium* and *S. aureus* following treatment with *Pseudomonas* sp. strain NW27-A1 (0.08 mg/ml) supernatant may indicate a reduced concentration of active compound(s) that caused the reduction of cellular metabolism.

The reduced metabolic state of the indicator bacteria demonstrated in metabolic function assays could be due to some secondary metabolites in the supernatant of *Pseudomonas* sp. strain NW27-A1, altering the metabolic state of bacteria. This altered cellular metabolism can indicate

bacterial cells entering a state of low metabolic activity or suppression of bacterial metabolic activity by the compound(s) present in the supernatant.

Data shows that the supernatant of *Pseudomonas* sp. strain NW27-A1 was most active against Gram positive bacteria and was less active against *A. baumannii*. However, most of the Gram negative bacteria used in this study were resistant to this supernatant. Resistance can be attributed to the large size of the antimicrobial compounds, unable to penetrate the outer membrane of Gram negative bacteria (Aqib *et al*., 2022). An analogous illustration is the fact that vancomycin is ineffective against Gram negative bacteria due to its large size and cannot enter bacterial cells due to the lipid bilayer (Silhavy *et al*., 2010).

In this study, *E. coli* NCTC 9081 was sensitized to this supernatant by using PMBNP, an outer membrane permeabilizer. Typically, polymyxin has a fatty acid tail that binds to lipid A of the lipopolysaccharide molecule, displacing magnesium and calcium (Trimble *et al*., 2016). After this displacement, the membrane weakens and permeability barrier is disrupted, enabling uptake of molecules that were previously impermeable to enter, and periplasmic proteins to leak out (Trimble *et al*., 2016). PMBNP lacks the fatty acid tail and therefore has no antibacterial activity but is still able to compromise the outer membrane (Daugelavicius *et al*.,

2000). PMBNP therefore works by permeabilizing the outer membrane of Gram negative bacteria through charge: charge interaction (Trimble *et al*., 2016).

Non-treated, sensitized *E. coli* displayed delayed growth, where it reached log phase 2 hours post-incubation. This could be due to the effect of PMBNP, which permeabilized *E. coli*, possibly allowing the leakage of periplasmic proteins essential for cellular growth (Dixon *et al*., 1986). PMBNP can damage the outer membrane sufficiently to enable molecules to diffuse into the periplasm (Ofek *et al*., 1994). PMBNP could have mediated the loss of low molecular weight substances (eg. potassium) from *E. coli* (Dixon *et al*., 1986).

This study demonstrated an increased susceptibility of *E. coli* to *Pseudomonas* sp. strain NW27-A1 supernatant in the presence of PMBNP, suggesting that the outer membrane of Gram negative bacteria plays a role in preventing the entry of antimicrobial compounds produced by *Pseudomonas* sp. strain NW27-A1 into the cell of Gram negative bacteria (Breijyeh *et al*., 2020). Data showed that sensitized *E. coli* was inhibited by treatment using the highest concentration of *Pseudomonas* sp. strain NW27-A1 supernatant (5.12 mg/ml). It is likely that PMBNP increased the permeability of the outer membrane of *E. coli*, allowing

antimicrobial compounds in *Pseudomonas* sp. strain NW27-A1 supernatant to diffuse into the periplasm and eventually, find their target in the bacterium (Detweiler, 2020).

One of the first changes made by bacteria in response to antibiotic stress occurs in the lag phase, by which it extends in addition to a slowed or reduced growth rate (Fridman *et al*., 2014). This “tolerance by lag” allows bacteria to survive under high antibiotic concentrations, allowing them to develop tolerance to antibiotics (Fridman *et al*., 2014). In addition, extended lag time can contribute to enhancement of bacteria survivability and regrowth when suitable conditions arise (Fridman *et al*., 2014, Li *et al*., 2016). In fact, tolerance to antibiotics can be achieved by slowed growth of the bacterium (Gilbert *et al*., 1990).

It was interesting to observe that when treated using a lower concentration of *Pseudomonas* sp. strain NW27-A1 (4.12 mg/ml), sensitized *E. coli* NCTC 9081 remained in lag phase until 11 hours post-inoculation, where it began to grow. This phenomenon was also observed after treatment using a lower concentration of supernatant (3.13 mg/ml), where sensitized *E. coli* started growing after 8.5 hours. Lag time extension could be employed by sensitized *E. coli* to overcome elevated

antibiotic stress by developing strategies to resist killing by antimicrobial compounds in the supernatant (Bernier *et al*., 2013).

It seemed that the extension of lag time was concentration-dependent; the higher the concentration of antibiotic, the longer the lag phase time extension. The duration of lag phase can be a meaningful indicator of dose-dependent antibiotic inhibition, which can mediate different bacterial responses (Bernier *et al*., 2013). At a high concentration (5.12 mg/ml), the lethal action of the antimicrobial compounds in the supernatant continued, leading to *E. coli* death or growth arrest. At a lower concentration (4.12 mg/ml), the antimicrobial compounds could have acted as stress inducers. Various cellular responses of *E. coli* could have been affected in addition to altered gene expression, leading to an adaptive response impacting antibiotic tolerance (Bernier *et al*., 2013). A previous study has shown extended lag phase of *E. faecium* after exposure to antibiotic, including vancomycin and fosfomycin (Theophel et al., 2014). The effect of these antibiotics on the growth dynamics of *E. faecium* included induced oxidative stress as shown by intracellular reactive oxygen species (ROS) accumulation, which can ultimately disrupt metabolic pathways and inhibit various physiological functions of the cell (Marambio-Jones *et al*., 2010).

Antibiotic stability can be influenced by multiple factors, including pH, temperature, CO2 levels and time. In the supernatant of *Pseudomonas* sp. strain NW27-A1, the antimicrobial compounds could have been degraded with time. For instance, the degradation of β-lactam antibiotics in aqueous solutions are known to be strongly pH-dependent with maximal stability around pH 4-5 for amoxicillin and ampicillin (Hou, 1971). Others investigated mecillinam degradation, by which its half-life was as short as 2 hours in (3-(N-Morpholino) propane sulfonic acid) MOPS-based medium and 4-5 hours in LB medium at 37°C, however adjustment of temperature increased its stability to a half-life of around 6 hours (Brouwers *et al*., 2020). Thermal stability of the antimicrobial compounds of *Pseudomonas* sp. strain NW27-A1 was not investigated in this research. To investigate this aspect, a delay-time bioassay can be used to estimate the degradation half-life of the antimicrobial compounds in the growth medium at different temperatures (Brouwers *et al*., 2020).

In the work described in the next chapter, an attempt was made to extract the antimicrobial compound(s) responsible for inhibiting the growth of indicator bacteria from the supernatant of *Pseudomonas* sp. strain NW27-A1.

Chapter 5.0

Extraction of antimicrobial compound(s) produced by *Pseudomonas* sp. strain NW27-A1

5.0 Extraction of antimicrobial compound(s) produced by *Pseudomonas* sp. strain NW27-A1

5.1 Introduction

Supernatants collected from bacteria are generally a complex solution, which contains a mixture of carbohydrates, organic acids, proteins, fatty acids in addition to bacterial outer membrane vesicles and bacteriophages (Champagne-Jorgensen *et al*., 2021, Yang *et al*., 2021, Lim *et al*., 2018). Techniques for separation of these compounds commonly include multiple stages, including use of organic solvents and solid phase extraction (Ndlovu *et al*., 2017, Schwarz *et al*., 2021). Extraction and purification methods are chosen based on physiochemical properties of the target compound (Ndlovu *et al*., 2017).

5.1.1 Some methods used for the extraction of secondary metabolites

5.1.1.1 Liquid-liquid extraction

Liquid-liquid extraction, also known as solvent extraction, is a selective separation procedure to separate compounds based on their solubility in two immiscible solvents, usually water (aqueous phase) and an organic solvent (organic phase) (Hammad *et al*., 2021). The organic solvent should have high affinity for some components in the aqueous phase, which contains the components to be separated (Hammad *et al*., 2021).

These two immiscible solvents are shaken together, allowing for the selective transfer of solutes from one phase to the other, followed by separation of the immiscible phases (Hammad *et al*., 2021).

The choice of solvent for this type of extraction technique is based on considerations such as 1) selectivity of the solvent, which is based on its intermolecular interactions to the target compounds (Tshepelevitsh *et al*., 2017, Lee *et al*., 2020) 2) solubility of the compounds in the extraction solvent (Lee *et al*., 2020) 3) polarity of the solvent (Kislik, 2012).

For this type of extraction technique, a standard solvent composition is difficult to establish since solvent selection varies with each target compound (Cequier-Sánchez *et al*., 2008). Some of the solvents used in solvent extraction techniques include ethyl acetate, dichloromethane, chloroform and acetonitrile (Cequier-Sánchez *et al*., 2008, Kumar *et al*., 2014).

5.1.1.2 Solid-liquid extraction

Solid-liquid extraction techniques involve extraction of compounds from a solid sample using a solvent in which the analyte is soluble (Qureshi *et al*., 2021). This type of extraction process, also known as leaching, involves

contacting a solid material of interest with a liquid solvent that dissolves the constituents (Acquaro Junior *et al*., 2019). Initially, the solid material is contacted with the extraction solvent. The resulting insoluble solvent is separated by physical means such as sonication (Qureshi *et al*., 2021). This extraction process will generate an extract comprised of a large amount of solvent in addition to a residual, insoluble solid (Chanioti *et al*., 2014).

5.1.1.3 Solid phase extraction (SPE)

Solid phase extraction (SPE) is a liquid chromatography separation technique which separates compounds suspended in a sample from other compounds depending on their chemical and physical properties (Fagerquist *et al*., 2005, Maranata *et al*., 2021). It is an extraction process involving a liquid (mobile phase) and solid phase (stationary phase). SPE uses selective adsorption and elution to separate and purify samples (Walker and Mills, 2002). The analytes suspended in the mobile phase adsorb to the solid stationary phase depending on the affinity, followed by eluting them with an elution solution (Hennion, 1999).

The first step of SPE is conditioning of the column by pre-treating the dry packing material with a solvent or buffer to prepare for binding of the analyte (Figure 5.1) (Ötles and Kartal, 2016). The next step involves loading of the analyte-containing mixture and running it through a column.

The analyte should selectively bind to the column packaging material, along with undesired impurities. As such, the third step involves running a solvent through the column, removing the impurities from the solution but leaving the bound analyte in the column material. Finally, the elution step involves the addition of an elution solution, which will remove the desired analyte from the stationary phase into a receiving tube (Ötles and Kartal, 2016).

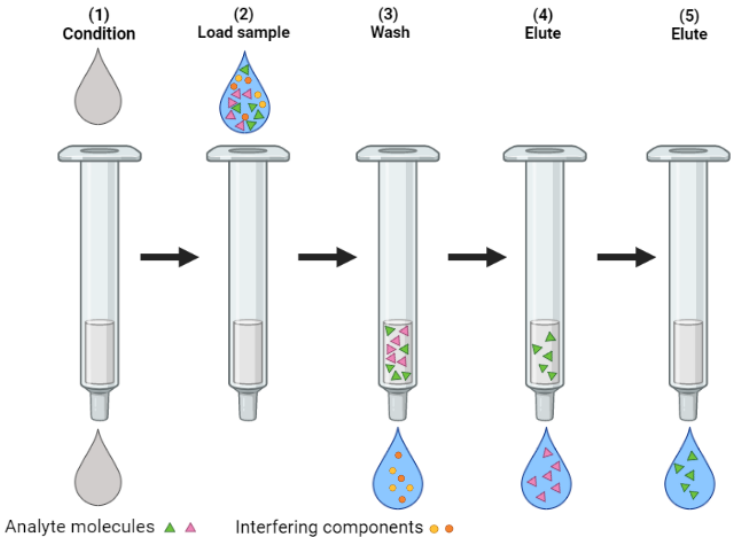


Figure 5.1 A schematic diagram demonstrating the extraction of two analytes using a solid phase extraction cartridge.

Step 1 involves conditioning the column using a solvent or buffer solution. Step 2 involves loading the sample containing the analyte molecules. Step 3 involves washing the undesired components while the analyte selectively binds to the column packaging material. Step 4 involves eluting the desired analyte using an elution solvent. The elution step can be done more than once as depicted in step 5. This figure was created using Biorender.

The types of SPE methodologies can fall into several categories, including normal phase, reversed phase and ion exchange SPE.

5.1.1.3.1 Normal phase SPE

Normal phase SPE uses a polar, hydrophilic stationary phase with a non-polar mobile phase (Mallet, 1998). The column is filled with polar, silica particles (Mallet, 1998). Normal phase SPE uses a non-polar, non-aqueous mobile phases such as chloroform (Walker and Mills, 2002). As the mobile phase passes through the column, the polar analytes will bind/adsorbs to the polar sorbent (Mallet, 1998). The non-polar analytes will pass faster through the stationary phase (Mallet, 1998). Non-polar analytes in the mobile phase will therefore elute quicker than polar analytes. Analytes can be eluted by increasing the polarity of the mobile phase (Walker and Mills, 2002).

5.1.1.3.2 Reverse phase SPE

The chromatography mode of reverse phase SPE is the opposite of normal phase SPE. In reverse phase SPE, the mobile phase is polar while the stationary phase is hydrophobic and non-polar (Mallet, 1998). Polar mobile phases in reverse phase SPE include water, methanol or acetonitrile (Mallet, 1998). The silica substrate in the reverse-phase SPE column is modified to make it non-polar. Such modifications include adding

8 or 18 carbons (C8 or C18 respectively) to the silica (Walker and Mills, 2002). The hydrocarbon chains retain compounds of low polarity due to the hydrophobic effect. The non-polar molecules will bind/adsorb strongly to the non-polar stationary phase, while the weakly retained polar molecules will pass more quickly through the column and elute earlier (Mallet, 1998). By washing the cartridge with a non-polar solvent, the interaction of the analyte and stationary phase is disrupted, leading to the elution of the analyte. Analytes can be eluted faster by decreasing the polarity of the mobile phase (Walker and Mills, 2002) .

5.1.1.3.3 Ion exchange SPE

The principle of separation in ion exchange SPE is based on electrostatic interactions between the analytes and the charged groups (positively or negatively charged) on the stationary phase . Analytes are extracted by a high-energy ionic interaction with the sorbent (Walker and Mills, 2002).

There are two types of ion exchange SPE techniques: anion exchange SPE and cation exchange SPE. Cation-exchange sorbents contain negatively charged surface groups that interact and retain positively charged analytes such as bases via electrostatic (ionic) interactions (Mallet, 1998). In contrast, sorbents used in anion-exchange SPE contain positively charged surface groups, retaining negatively charged analytes such as acids (Mallet, 1998) .

To elute the analyte from the sorbent used in ion-exchange SPE, the stationary phase is washed with a solvent that neutralizes the charge of the analyte and or/ stationary phase (Mallet, 1998). The electrostatic interaction between the analyte and the stationary phase will no longer exist due to this neutralization and hence, the analyte will elute from the cartridge (Mallet, 1998).

5.2 Aims and objectives of this chapter

The aims of this chapter were to extract antibacterial compound(s) from *Pseudomonas* sp. strain NW27-A1, evaluate the antimicrobial potential of the acetonitrile extract followed by determination of the active fractions. The objectives were;

- To extract bioactive secondary metabolites from the supernatant of *Pseudomonas* sp. strain NW27-A1 using solvent extraction techniques and to assess the antimicrobial activity of the extracts.
- To identify the active fractions by subjecting the active acetonitrile extract to high-performance-liquid-chromatography (HPLC).

5.3 Materials and methods

The supernatant of *Pseudomonas* sp. strain NW27-A1 (grown in LB broth for 48 hours at 25°C) was used for extraction procedures in this chapter.

5.3.1 Extraction of antimicrobial compounds(s) using acetonitrile

Extraction of antimicrobial(s) from *Pseudomonas* sp. strain NW27-A1 supernatant was performed using acetonitrile as the extraction solvent. Acetonitrile was used as the extraction solvent due to its moderate polarity and its ability to recover polar and non-polar compounds (Ian *et al*., 2012).

For purification, a large amount of the inhibitory compound was required. Therefore, three volumes of reconstituted, freeze-dried *Pseudomonas* sp. strain NW27-A1 (140 mg/ml) supernatant prepared in Chapter 4 were used. This extraction procedure was performed three times using three independent *Pseudomonas* sp. strain NW27-A1 supernatants.

The three supernatants' volumes, their masses after freeze drying and their corresponding volume of acetonitrile are mentioned in Table 5.1.

Table 5.1 Summary of three *Pseudomonas* sp. strain NW27-A1 supernatants used in large scale extraction using acetonitrile as the extraction solvent.

Replicate	Supernatant volume (mL)	Mass (g) of freeze-dried supernatant	Volume of acetonitrile (mL)	Total volume (mL)
1	58.48	8.24	526.52	585
2	50.65	7.121	455	505
3	40	6.958	360	400

This table represents three *Pseudomonas* sp. strain NW27-A1 replicates used for large-scale extraction using acetonitrile as the extraction solvent. Replicate number represents the supernatant replicates. The non-concentrated supernatants (200 mL per replicate) were freeze-dried, and the masses were recorded. The supernatant volumes in this table were prepared by reconstituting each mass of freeze-dried supernatant with sterile distilled water to make a concentration of 140 mg/ml.

Each volume of *Pseudomonas* sp. strain NW27-A1 supernatant was left to shake in a rolling shaker at 14°C for 24 hours in its appropriate volume of acetonitrile, which was 9X the volume of the supernatant as mentioned in Table 5.1. For example in replicate 1, the first volume of supernatant (58.48 mL) was left to shake with 526.52 mL of acetonitrile in a 1L borosilicate glass bottle.

After 24 hours, the phases were allowed to separate, and the acetonitrile phase (upper phase) was centrifuged at a low speed of 581 xg for 5 minutes. The acetonitrile phase was collected after centrifugation. 10 mL of each acetonitrile extract from each replicate was dried using a stream of nitrogen and the dried product was reconstituted using 1 mL HPLC

grade water. The remaining acetonitrile extract volumes were stored for further downstream experiments.

A negative control was also prepared, where a volume of LB broth (50mL) was mixed with a volume of acetonitrile (450 mL). This mixture was put through the same extraction, drying and reconstitution steps as the experimental sample. The purpose of this negative control was to determine whether the presence of acetonitrile residues possibly left over from the drying process could have an antibacterial effect.

5.3.2 Minimum inhibitory concentration of the acetonitrile extracts

To test the inhibitory activity of these extracts against *S. aureus* JE2 and *E. faecium* E1162 and *A. baumannii* NCTC 12156, the extracts were diluted 13.3 times using HPLC grade water and the MIC of the extracts was determined using broth microdilution assays as described in Chapter 2 section 2.2.2.

5.3.3 Growth of *E. faecium* E1162 in the presence of the acetonitrile extracts

The growth of *E. faecium* E1162 after treatment with acetonitrile extracts was determined using a spectrophotometer which measured the optical density ($OD600_{nm}$) over time for 13 hours. 100 µl of either supernatant of

Pseudomonas sp. strain NW27-A1 (before extraction) at the MIC or the acetonitrile extracts were added to a non-treated 96-wellplate (VWR, UK). Then, 100 µl of indicator bacteria (*E. faecium* E1162) was added to the wells. *E. faecium* E1162 was prepared as described in Chapter 2 section 2.1.The experiment was done using three replicates. The negative control prepared in part 5.3.1 of this chapter, which was found to have no effect on this bacterium's growth, was also applied *to E. faecium* in the same way.

5.3.4 Construction of a purification table

To monitor the progress of the extraction method, a purification table was constructed by calculating the units of activity, total activity (U), specific activity (U/mg), recovery of units (%) and fold purification of the acetonitrile extract (Burgess, 2009).

Units of activity are arbitrarily defined as the minimum inhibitory concentration per milligram. It is calculated by dividing the mass of the sample by its MIC. Total activity is a measure of the total units of activity in the volume of a given sample. It is calculated by multiplying the units of activity by the volume of a given sample. Specific activity is a measure of the amount of analyte recovered at a given step divided by the mass of total analytes in the sample. It is measured by dividing the total activity by the mass of the sample. Recovery of units refers to the ratio of the amount

of analyte of interest obtained at a given step divided by the original amount present in the original sample (ie. the supernatant before extraction), converted to percent by multiplying by 100. Fold purification is the ratio of the MIC of the purified analyte to that of the original sample (ie. supernatant pre-extraction).

5.3.5 High performance liquid chromatography (HPLC) analysis

The first replicate of the large-scale acetonitrile extract was used for HPLC analysis. The HPLC analysis of this acetonitrile extract was conducted using Agilent 1260 Infinity II LC System (Agilent Technologies, USA), equipped with a diode array detector (DAD). The separation was conducted using a Poroshell 120 EC-C18 reverse-phase chromatography column, 150 mm length, 4.6 mm width and particle size 4 µm. A gradient elution system was chosen in this study using a mixture of 0.1% Trifluoroacetic acid (TFA) (HPLC grade ≥ 99.9%; Sigma-Aldrich, UK) and acetonitrile/0.1% TFA (HPLC, gradient grade ≥ 99.9%; Sigma-Aldrich, UK). The total run time was 50 minutes, and the concentration gradient was varied as follows; a) 0.1% TFA for the first 0-5 minutes, b) a concentration gradient starting from 0-100% acetonitrile/0.1% TFA for 40 minutes and c) 100% acetonitrile/0.1% TFA for 5 minutes. The flow rate was 1 ml/min with 1 mL fractions collected.

5.3.5.1 Alamar blue cellular metabolism assay

To test the inhibitory activity by fractions collected, cellular metabolism of *E. faecium* E1162 in the presence of each fraction was monitored using a fluorometer. Alamar blue cellular metabolism assay was done as described in Chapter 4 section 4.3.3.1. The RFU of *E. faecium* E1162 treated with the fractions was compared to non-treated *E. faecium* E1162 at 10 hours incubation. The reduction in cellular metabolism (%) was calculated and plotted in the HPLC chromatogram.

5.4 Results

5.4.1 Extraction of antimicrobial compounds inhibitory to *S. aureus* JE2 and *E. faecium* E1162

The three large scale acetonitrile extracts were able to inhibit the growth of *S. aureus* and *E. faecium*, but not *A. baumannii*. The MIC of the extracts against *S. aureus* JE2 and *E. faecium* E1162 was 0.016 mg/ml (Table 5.2). Upon statistical analysis using Mann-Whitney test, a significant difference was observed between the MICs of *Pseudomonas* sp. strain NW27-A1 and the acetonitrile extract (p-value <0.001).

Table 5.2 Minimum inhibitory concentration of acetonitrile extracts

Indicator bacterium	Minimum inhibitory concentration (MIC) mg/ml	
	***Pseudomonas* sp. strain NW27-A1 supernatant**	**Acetonitrile extract**
S. aureus JE2	0.12 (0.12)	0.016 (0.015-0.018)
E. faecium E1162	0.12 (0.12)	0.016 (0.015-0.018)

Comparison of MIC values of *Pseudomonas* sp. strain NW27-A1 supernatant and acetonitrile extracts against the representative indicator bacteria. The values represent a set of triplicate results as a median value (with range).

Since activity was present, a purification table was constructed to evaluate the purity and yield of the acetonitrile extract derived from replicate 1 which was later subjected to HPLC analysis.

The units of activity in the supernatant (pre-extraction) was 68666.6 (ie. 68666.6 active arbitrary units are present in the mass (8240 mg)). The total arbitrary active units (Total activity) in the 58.48 mL supernatant (pre-extraction) was 4015626.7. The active arbitrary units (Specific activity) in the mass of this samples was 487.3.

The units of activity in the acetonitrile extract was 937.50 (ie. 937.50 active units are present in the mass (15 mg)). Since the volume of this extract was 10mL, total active units (Total activity) was 9375. Specific activity of this extract was 625. The recovery of units in the acetonitrile extract was 0.23% with a 1.28 fold purification (Table 5.3).

Table 5.3 Purification table for *Pseudomonas* sp. strain NW27-A1 supernatant using acetonitrile

Sample	Minimum inhibitory concentration (MIC) mg/ml	Volume of sample (mL)	Mass of sample (mg)	Unit of activity	Total activity (U)	Specific activity (U/mg)	Recovery of units (%)	Fold purification
Supernatant (pre-extraction)	0.12	58.48	8240	68666.6	4015622.7	487.3	N/A	N/A
Acetonitrile extract	0.016	10	15	937.50	9375	625	0.23	1.28

Supernatant (pre-extraction) refers to the supernatant of *Pseudomonas* sp. strain NW27-A1.

Units of activity is calculated by dividing mass of sample by MIC of sample

Total activity is calculated by multiplying units of activity by the volume of the sample

Specific activity is calculated by dividing the total activity by mass of sample

Recovery of units is calculated by dividing total activity of acetonitrile extract by total activity of the supernatant pre-extraction multiplied by 100

Fold purification= MIC of supernatant before extraction÷ MIC of acetonitrile extract

N/A; Not applicable, mg; milligram, mL; millilter, U; Units

5.4.2 Inhibition of *E. faecium* E1162 by the acetonitrile extract

E. faecium E1162 was grown selectively on its appropriate agar as described in Chapter 2, part 2.1 and behaved as expected when tested using biochemical tests. The positive control (*E. faecium* treated with the MIC of 4µg/ml rifampicin) displayed no growth as represented by the low absorbance values. The acetonitrile extract was able to inhibit the growth of *E. faecium* E1162. As shown in Figure 5.2, the absorbance of *E. faecium* after treatment with acetonitrile extract did not increase in comparison to the positive control.

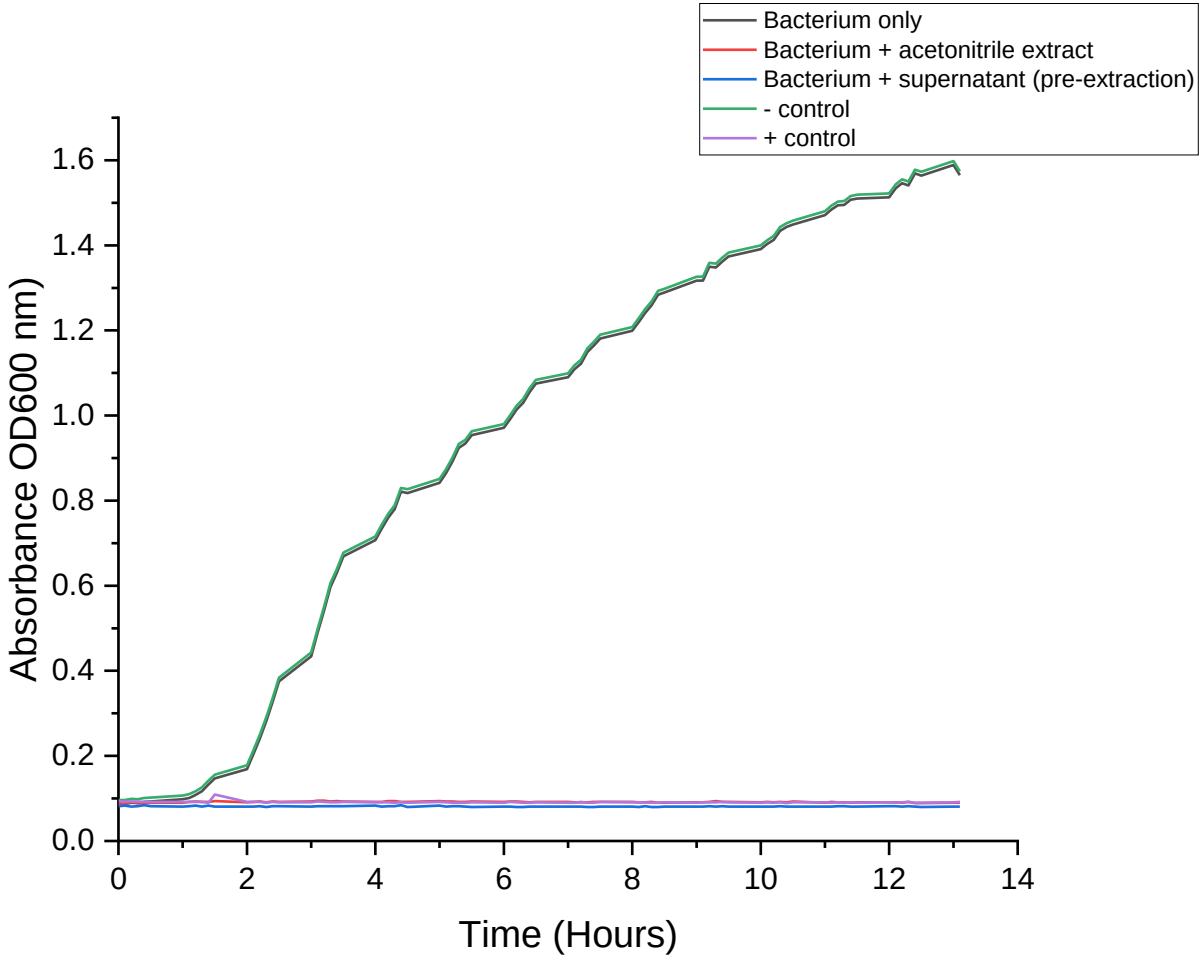


Figure 5.2 Growth curve of *E. faecium* E1162 after treatment with acetonitrile extract

Growth of *E. faecium E1162* was monitored in the presence and absence of the acetonitrile extracts.

Black line represents the mean absorbance of the indicator bacterium (*E. faecium*) without any treatment; 100 µl of indicator bacterium + 100 µl of sterile distilled water. The red line represents the mean absorbance of the indicator bacterium after treatment with the pre-extraction *Pseudomonas* sp. strain NW27-A1 supernatant at the MIC (1.09 mg/ml). The blue line represents the mean absorbance of the indicator bacterium after treatment of with acetonitrile extracts at the MIC (0.31 mg/ml). The green line (- control) represents the indicator bacterium treated with broth extracted with acetonitrile. The purple line (+ control) represents the indicator bacterium treated with 4µg/ml rifampicin.

5.4.3 Reverse-phase high performance liquid chromatography (HPLC) results of acetonitrile extract

Analysis by reverse-phase HPLC showed multiples peaks detected at 254 nm (Figure 5.3). The active fractions that were inhibiting *E. faecium* E1162 by alamar blue cellular metabolism assays were labelled alphabetically in ascending order. Seven active fractions were detected in this assay. The active fractions A, B, C and D, E, F and G were eluted at retention times of 3, 8, 11, 12, 13, 17 and 44 minutes respectively.

Active fractions A, B, C, D, E, F and G were able to reduce the cellular metabolism of *E. faecium* E1162 by 71.19%, 86.10%, 89.38%, 91.95%, 85.85%, 95.91% and 99.65% respectively. Active fraction G demonstrated highest ability to reduce the cellular metabolism of *E. faecium* E1162.

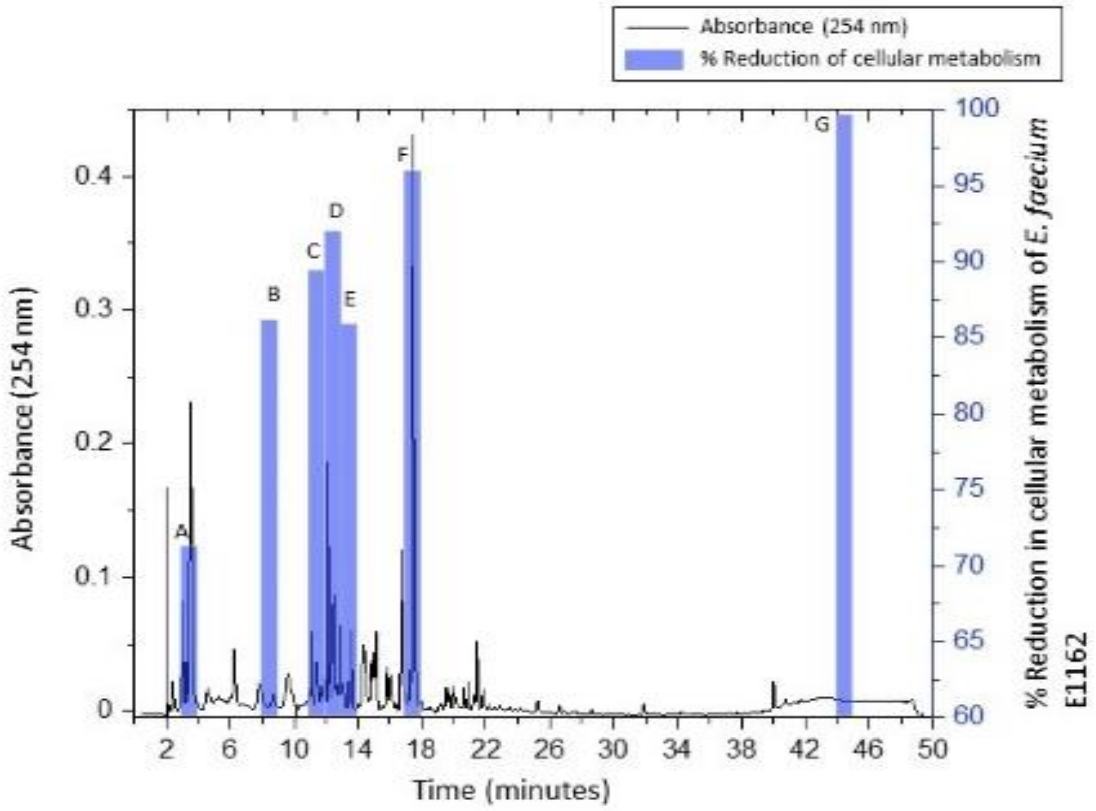


Figure 5.3 Reverse-phase HPLC chromatogram of large scale acetonitrile extract.
Active fractions A, B, C, D, E, F and G are displayed in this chromatogram. The blue bars represent the reduction of *E. faecium* E1162 cellular metabolism by the compound(s) in the active peaks.

5.5 Discussion

In this chapter, extraction using acetonitrile was successful in inhibiting *S. aureus* and *E. faecium*. Acetonitrile is a polar-aprotic solvent (Huffman *et al*., 2012). It is a hydrophilic solvent able to recover lipopeptides (Wang *et al*., 2008, Biniarz and Lukaszewicz, 2017). Being predominantly acidic, acetonitrile is considered as a hydrogen bond donor (Hunter, 2004). It is expected that acetonitrile has extracted polar compounds responsible for inhibitory activity.

The acetonitrile extracts were not able to inhibit the growth of *A. baumannii*. It can indicate that the concentration of the extracts was not sufficient to inhibit the growth of *A. baumannii* NCTC 12156. Moreover, *A. baumannii* may show resistance to the extract through various combined mechanisms including decreased permeability and efflux pump overexpression (Uppalapati *et al*., 2020).

Resistance could be due to the structure of Gram negative cell wall. The compound(s) in the extracts might have not been active against *A. baumannii* due to their inability to penetrate the outer layer of this bacterium (Glauert and Thornley, 1969). The main differences between Gram positive and Gram negative bacteria are related to their cell wall composition. The cell wall of Gram positive bacteria is mainly composed

of a thick peptidoglycan layer (Rohde *et al*., 2019). On the other hand, Gram negative bacteria have cell walls with a thin peptidoglycan layer and an outer membrane with a lipopolysaccharide component (Breijyeh *et al*., 2020). The outer membrane is external to the peptidoglycan cell wall and can provide Gram negative bacteria resistance to certain antibiotics by acting as a barrier (Miller, 2016). This outer membrane is not present in Gram positive bacteria, which makes Gram negative bacteria more resistant to antibiotics (Breijyeh *et al*., 2020).

Efflux pumps can contribute to the resistance of *A. baumannii* to the acetonitrile extract (Uppalapati *et al*., 2020). The active compounds in the extract can be actively transported out of the bacterial cell by efflux pumps, including adeABC efflux pump, encoded by *adeABC* efflux gene in *A. baumannii* NCTC 12156. This gene could be overexpressed in this strain, enhancing the expression of adeABC efflux pump and reduced the susceptibility of *A. baumannii* to the active compounds in the extract (Xu *et al*., 2019). Loss of membrane-associated proteins such as OmpA, one of the most abundant porins in *A. baumannii*, can also contribute to resistance (Uppalapati *et al*., 2020). OmpA could possibly couple with efflux pumps and force out the active compound in the extract from the periplasm (Tsai *et al*., 2020). An *A. baumannii* OmpA deletion mutant can be generated in the future and determine if it can confer this bacterium

susceptible to the active compound(s) in the acetonitrile extract (Uppalapati *et al*., 2020).

In addition, the acetonitrile extract could have been ineffective against *A. baumannii* due to the molecular size and hydrophobicity of the antimicrobial compound(s) present in the extracts. Some charged molecules cannot passively diffuse into the lipid bilayer which has polar and non-polar components (Livermore, 1990). For instance, vancomycin is hydrophilic and cannot penetrate Gram negative bacterial cells due to its net positive charge and large molecular mass (Dinu *et al*., 2020). Acetonitrile in this thesis is likely to have extracted polar analytes, which are hydrophilic. In addition, they were possibly unable to pass the outer membrane of *A. baumannii* due to their large size. To enable its delivery across the cell envelope, conjugation of the active compound to a cell-penetrating molecular transporter can be studied in the future (Antonoplis *et al*., 2019). Moreover, if the active compound(s) in the extracts were small hydrophilic molecules, they can enter *A. baumannii* through porin proteins, which produce transmembrane diffusion channels. The resistance of *A. baumannii* to antimicrobial analytes in the extracts could be explained by the presence of porin channels that were not functional or "open" under *A. baumannii* growth conditions used in this thesis (Nikaido and Rosenberg, 1981). A mutant *A. baumannii* strain that produces fewer porin molecules than that of the wild type could be generated in the future

to have a better understanding of *A. baumannii* resistance mechanisms demonstrated in this study (Nikaido and Rosenberg, 1981).

Recovery of large scale extraction using acetonitrile was low (0.23%). This could be due to incomplete transfer of analytes to the organic phase during agitation. Moreover, a rolling shaker set at low speed was used to agitate the two phases during large scale acetonitrile extraction procedures. The speed of agitation, could have had a negative effect on extraction efficiency as others have demonstrated higher agitation can lead to higher transfer of analytes from the aqueous phase to the organic phase, leading to the increase of extraction yields (Pinto *et al*., 2004). Even though recovery of units was low, fold purification was relatively high (1.28). This fold purification can be due to the presence of impurities in the extract prepared using a one-step extraction method (Peypoux *et al*., 1999). Therefore, to further purify and concentrate the antimicrobial compound(s), adjustments to the extraction technique can be performed in the future by which solid-phase extraction using graphitized carbon black (GCB) column can be used to further purify the active compound(s).

HPLC results revealed the presence of multiple peaks in the chromatogram. Some of the peaks collected as single fractions showed

reduced cellular metabolism of *E. faecium* E1162. A fraction detected at 2 minutes was presented with a peak but was unable to inhibit *E. faecium*. Other fractions such as those detected at approximately 6, 16 and 21 minutes were also inactive against *E. faecium*. This suggests that the acetonitrile extract contained multiple polar compound(s), but not all compounds have antimicrobial properties.

Active fraction A was able to reduce the metabolic activity of *E. faecium* E1162 by 71.19%. Fraction A was recovered from the column using 0.1% trifluoroacetic acid. The ion-pairing agent, trifluoroacetic acid (TFA), is a common solvent used in reverse-phase HPLC to release peptides from the mobile phase material (Chen *et al*., 2004). It can bind to the positive charge and polar groups on the polypeptide (Chen *et al*., 2004). It is therefore anticipated that the active fraction was highly polar.

The varying concentration of the TFA adjusts the selectivity of polypeptides on reverse phase chromatography (Chakraborty and Berger, 2005). A peak was recovered using <12.5% acetonitrile/0.1% TFA. Fraction B was able to reduce cellular metabolism of *E. faecium* by 86.10% with a peak detected. The peak was broad compared to other peaks observed by active fractions. Like active fraction A, fraction B was detected

at the start of the HPLC, indicating the polar nature of the active compound.

Other peaks of active fractions C, D and E were able to reduce the metabolism of *E. faecium* E1162 by 89.38%, 91.95% and 85.85% respectively. Active fraction F was able to reduce the metabolic activity of *E. faecium* E1162 by 95.91%, displaying a high peak compared to other peaks. The compound in this active fraction is therefore likely to be responsible for inhibition of *E. faecium* in this research.

Even though there was no visible sharp peak in active fraction G, the compound in this fraction was able to reduce the cellular metabolism of *E. faecium* E1162 by 99.65%. This fraction was recovered using ~100% acetonitrile/0.1% TFA. The compound in this active fraction is suspected to be less polar than the other active fractions.

Lack of peak in active fraction G may result from the DAD detector's inability to identify this peak or the use of innapropriate mobile (Caltabiano *et al*., 2018, Chen *et al*., 2022). Lack of peak can also indicate wrong polarity in the detector output signal (Baranowska, 2012). To resolve these possible issues in the future, fitting and tubings between the injector and detector can be checked prior to HPLC run followed by changing the

polarity of the mobile phase. High pressure can also be attributed to the absence of peak (Shalaan *et al*., 2010). High pressure can be caused by partial blockages in HPLC components including the filters, column inlets or detector (Wolfender, 2009). To resolve this issue in the future, the faulty component can either be repaired or replaced, followed by another HPLC run with a different replication of acetonitrile extract.

As future work, the *Pseudomonas* sp. strain NW27-A1- derived HPLC fractions could be further characterised and structurally analysed, which could lead to the application of these compounds as antibiotics to treat nosocomial infections caused by *S. aureus* and *E. faecium*. Other purification techniques can also be applied in the future to extract and purify *Pseudomonas* sp. strain NW27-A1-derived antimicrobials which may target *A. baumannii* NCTC 12156.

Chapter 6: An approach to identifying bacteria involved in competitive exclusion using *in silico* predicted biosynthetic gene clusters from metagenome datasets

6.0 An approach to identifying bacteria involved in competitive exclusion using *in silico* predicted biosynthetic gene clusters from metagenome datasets

6.1 Introduction

6.1.1 The human nasal microbiome and *Staphylococcus aureus* nasal carriage

The anterior nares are a complex ecosystem characterized by the presence of a diverse microbial community (Frank *et al*., 2010). After birth, different microbial species colonize the upper respiratory tract, which can be affected by factors including mode of delivery and feeding method (formula-fed versus breastfed) (Shilts *et al*., 2016, Kumpitsch *et al*., 2019).

Nasal microbiome composition changes with age (Oh *et al*., 2012). In 1.5-month-old infants, the anterior nares are dominated by the genera *Staphylococcus*, *Moraxella*, *Streptococcus*, *Corynebacterium* and/or *Dolosigranulum* (Biesbroek *et al*., 2014). The microbiome of breastfed infants is composed of *Dolosigranulum*/*Corynebacterium* profiles while formula-fed infants show *S. aureus* predominance (Shilts *et al*., 2016, Kumpitsch *et al*., 2019). It has been shown that breast-fed infants exhibit a microbial profile that provides them protection against respiratory tracts infections (Biesbroek *et al*., 2014). The microbiome of breastfed infants

has a more stable pattern, which is marked by early presence and high abundance of *Moraxella* and *Corynebacterium*/*Dolosigranulum* (Biesbroek *et al*., 2014).

During childhood, the nasal microbiome is dominated by members of the phyla Firmicutes, Proteobacteria, Actinobacteria and Bacteriodetes (Oh *et al*., 2012). The most common genus of Actinobacteria detected during childhood is *Corynebacterium* (Oh *et al*., 2012). The nasal microbiome of children is more dense but less diverse in comparison with adults nasal microbiomes (Camarinha-Silva *et al*., 2014).

During puberty, the nasal microbiome changes and persists throughout the healthy adult's life (Oh *et al*., 2012). It is dominated by Actinobacteria with the genera *Corynebacterium* and *Propionibacterium* being the significant colonizers (Oh *et al*., 2012). Firmicutes also colonize the anterior nares of healthy adults, with *Staphylococcus* being the most common genus detected (Schenck *et al*., 2016). Microbial communities of the anterior nares of adults are mainly dominated by *Cutibacterium*, *Corynebacterium* and *Staphylococcus* (Schenck *et al*., 2016).

S. aureus nasal carriage can be divided into three patterns in the healthy population: persistent carriage (~20%), intermittent carriage (60%) and non-carriage (20%) (Kluytmans *et al*., 1997). Nasal carriage of *S. aureus* contributes to an increased risk of developing an infection with the same bacterial strain (Perl *et al*., 2002). Moreover, persistent carriers have a higher chance of acquiring a *S. aureus* infection, including MRSA (Nouwen *et al.*, 2005).

6.1.2 Factors associated with *S. aureus* nasal colonization

S. aureus carriage is influenced by many factors, including host, environment, bacterial factors and in addition, by interaction with other commensal bacteria within the anterior nares (Zipperer *et al*., 2016). According to a previous study done on 617 twin pairs, host genetic factors have limited influence on *S. aureus* nasal carriage (Andersen *et al*., 2012). In addition, smoking is inversely associated with *S. aureus* nasal carriage, with males more likely to be carriers than females (Olsen *et al*., 2012). Other host factors that affect *S. aureus* nasal carriage include oral contraceptives. Women taking hormonal contraceptives are more likely to be persistent carriers of *S. aureus* in the anterior nares (Zanger *et al*., 2012). It has been proposed that high estrogen levels change the mucosal surfaces, making them more hospitable to *S. aureus* (Sugarman and Epps, 1982).

S. aureus is also acquired from sources in the environment, which is an important determinant for carriage (Olsen *et al*., 2012). Transmission of *S. aureus* through crowding and between household members of large families has been documented (Mollema *et al*., 2010). Healthcare workers, who switch between hospital and community, also have the tendency to carry *S. aureus* in the anterior nares (El Aila *et al*., 2017).

Bacterial factors play a major role in *S. aureus* nasal colonization with recent studies demonstrating the importance of the human microbiome as a source of novel antibiotics (Krismer *et al*., 2017, Brugger *et al*., 2020). For instance, the commensal bacterium *S. lugdunensis* prevents *S. aureus* from growing in the anterior nares by producing lugdunin (Zipperer *et al*., 2016). More recently, *Dolosigranulum pigrum*, a commensal bacterium in the anterior nares, was recently demonstrated to limit the growth of *S. aureus* in vitro, however the causative agent has not yet been identified and isolated (Brugger *et al*., 2020).

6.1.3 Nasal decolonization of *Staphylococcus aureus*

Pseudomonic acid, now marketed under the trade name Mupirocin, is commonly used as a nasal ointment to eliminate *Staphylococcus aureus* nasal carriage (Bode *et al*., 2010). Mupirocin, is a polyketide antibiotic produced by the soil bacterium *Pseudomonas fluorescens* (Matthijs *et al.*,

2014). It works by inhibiting bacterial isoleucyl-tRNA (Fuller *et al.,* 1971). To significantly reduce carriage rate, mupirocin must be applied twice daily for five days (Wertheim *et al*., 2005). Unfortunately, mupirocin resistance is on the rise with a prevalence of >13% for MRSA (Dadashi *et al*., 2020).

Previously, several candidate interventions for the decolonization of *S. aureus* in the anterior nares have been investigated, including antibiotics such as neomycin (Blanchard *et al*., 2016), antibacterial enzymes such as lysostaphin and bacteriocins such as nisin (Kokai-Kun *et al*., 2003). However, none of these interventions were able to eliminate *S. aureus* nasal colonization. The most recent candidate, aurintricarboxylic acid (ATA), also failed to decolonize *S. aureus* in the anterior nares in a novel persistent *S. aureus* nasal colonization model (Fernandes de Oliveira *et al*., 2021).

6.2 Aims and objectives of this chapter

The aim of this study was to discover antibiotic-producing bacteria in the anterior nares by searching for metagenomes containing genes coding for proteins probably involved in antimicrobial biosynthesis. The objective was:

- To query the Integrated Microbial Genomes and Microbiome (IMG/M) database in order to search for genes in the metagenome of those who do not carry *S. aureus* in the anterior nares for proteins involved in the production of antibiotics that may be preventing the growth of *S. aureus* in the anterior nares.

6.3 Materials and methods

6.3.1 Collecting metagenomic data of the anterior nares from the IMG/M database

This chapter was focused on identifying bacteria in the anterior nares capable of inhibiting *S. aureus* through the production of secondary metabolites with antimicrobial properties.

The Integrated Microbial Genomes (IMG) system is a data management and analysis platform for publicly available genomes from all three domains of life (Markowitz *et al*., 2014). In this study, nasal metagenomic data was obtained from the IMG/M website https://img.jgi.doe.gov/cgi-bin/m/main.cgi . Under the dataset options of IMG content, 'All Metagenome' dataset option was selected. The IMG/M metagenome database was filtered to search for 'anterior nares', using 'All Columns' options as the filter column. This retrieved 96 genome samples. Each of these samples had metagenomic IDs, which were used to query the integrated Microbial Genome Atlas of Biosynthetic Gene Clusters (IMG/ABC) database to check for the presence of BGCs in the individual metagenomes. The IMG/ABC is a publicly available database of predicted BGCs (Hadjithomas *et al*., 2017). The IMG/ABC website is accessed through https://img.jgi.doe.gov/cgi-bin/abc-public/main.cgi. The bioinformatics workflow is presented in Figure 6.1.

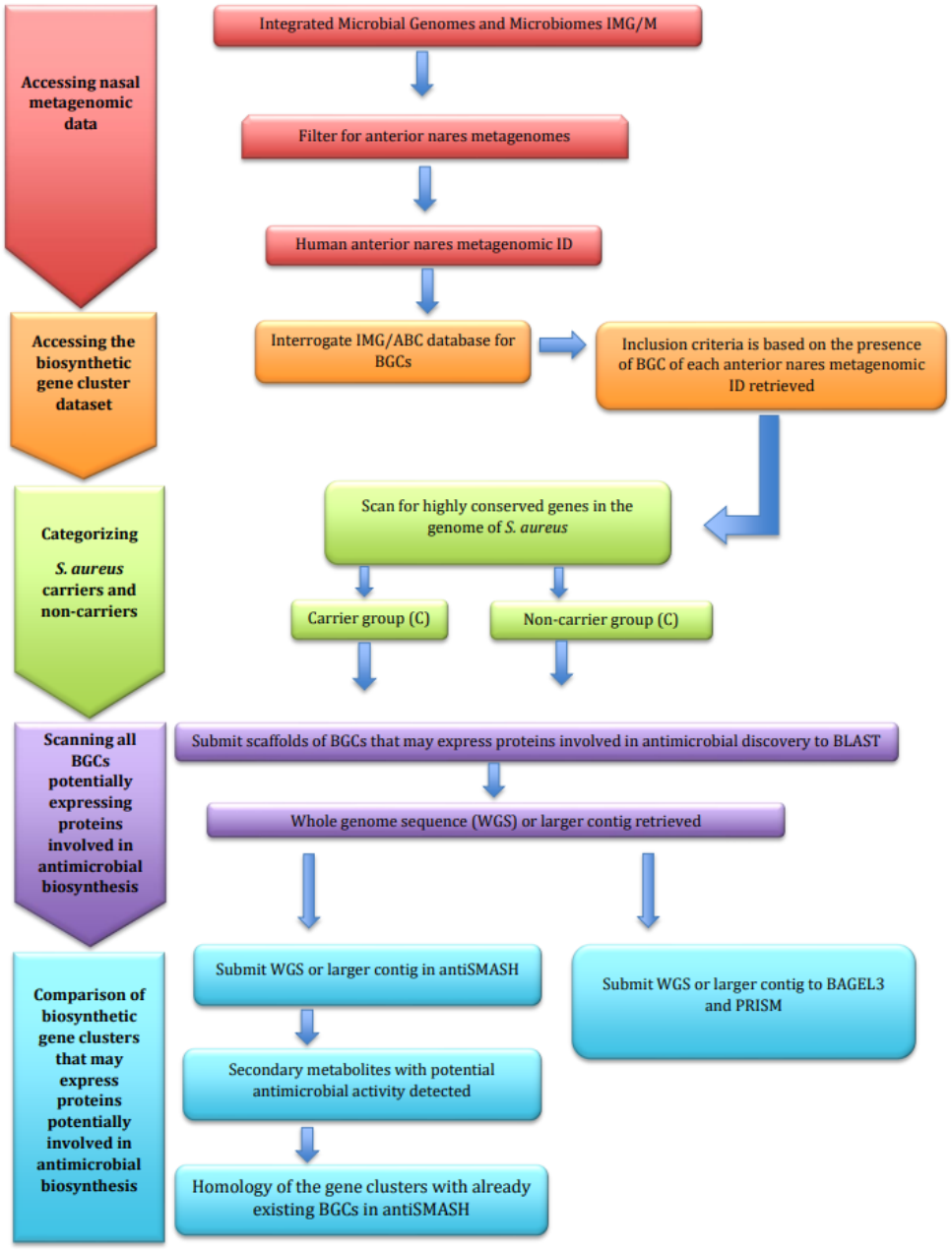


Figure 6.1 Summary of the bioinformatics workflow.

6.3.2 Tentative categorization of metagenomes as being from *S. aureus* carriers and non-carriers

The 36 data sets were categorized into *S. aureus* carriers (C) and non-carriers (NC) by searching the metagenome database of each individual for the presence or absence of the highly conserved sequences of *S. aureus* genes, including *tuf*, *sodA,* and *rpoB.* These genes can differentiate between different staphylococcal species (Ghebremedhin *et al*., 2008). Data sets were also screened for the presence of specific *S. aureus* genes including *fmhA*, *clfB, isdA, fnbA, sceD, oatA, femX, is*dA, and *arcC*. The nucleotide sequence of these genes, if found, was submitted in BLAST. If one of these genes in Table 6.1 was present in the metagenome of the individual, the individual was categorised as a tentative carrier of *S. aureus*.

Table 6.1 Genes used to tentatively categorize individuals as being carriers or non-carriers of *S. aureus*.

Gene	Gene product	Function
clfB	Clumping factor B	Cell adhesion to human fibrinogen and epithelial cells
fnbA	Fibronectin-binding protein A	Bacterial attachment to multiple substrates
sodA	Superoxide dismutase	Destroys superoxide anion radicals
arcC	Carbamate kinase	ATP binding
isdA	Iron-regulated surface determinant protein A	Metal ion binding and enhances bacterial cellular hydrophobicity
sceD	Transglycosylase	Cleave peptidoglycan and affects clumping of bacterial cells
oatA	O-acetyltransferase	Integral membrane protein involved in peptidoglycan synthesis
rpoB	DNA-directed RNA polymerase subunit beta	Catalyzes the transcription of DNA into RNA
tuf	Elongation factor	Binding of aminoacyl-tRNA to the ribosomes during protein biosynthesis
femX	Lipid II: glycine glycyltransferase	Peptidoglycan biosynthesis and regulation of cell shape
fmhA	Aminoacyltransferase FemA	Peptidoglycan biosynthesis and cell wall organization

6.3.3 Determination of biosynthetic gene clusters that express proteins potentially involved in antimicrobial biosynthesis.

Analysis of BGCs in each data set was achieved by accessing the Atlas of Biosynthetic Gene Clusters within the Integrated Microbial Genomes system (IMG/ABC). The IMG metagenomic ID was inserted in the IMG/ABC Biosynthetic gene clusters statistics option, choosing 'Total clusters in Metagenome' option (https://img.jgi.doe.gov/cgi-bin/abc-public/main.cgi). The IMG/ABC platform has been updated in 2019, so this option has now been replaced with "Metagenome Bin Biosynthetic gene clusters" option.

The metagenome of each individual was analyzed to see how many BGCs were present. The predicted protein products of all the genes in the BGCs of each subject were analyzed to determine if they potentially coded for proteins involved in antimicrobial biosynthesis. Analysis of the nucleotide sequences and amino acid sequences was conducted using BLASTn and BLASTp respectively. If the description using BLAST included terms pertaining to antibiotic production (for example, type A2 lantipeptide), then the BGC was considered as potentially coding for proteins involved in antimicrobial biosynthesis.

The scaffolds of these BGCs were submitted to BLAST to see if a whole genome sequence or a large contig containing these genes could be identified using BLASTn with Whole-genome shotgun contigs (WGS) database as the standard search database. The WGS or large contigs were then submitted to antiSMASH, PRISM and BAGEL3.

6.4 Results

6.4.1 Tentative categorization of metagenomes as being from *S. aureus* carriers and non-carriers

This analysis identified 36 metagenome datasets with predicted BGCs. However, further analysis of the 36 datasets showed that they came from 32 individuals included in this study (Table 6.2). This was due to some individuals having an additional data set from a second visit. Individuals 2 and 7 had datasets for the first and second visits, while individual 4 had a dataset for the first visit and two datasets for the second visit.

Out of the 32 individuals 5 were tentatively categorized as *S. aureus* carriers, while 27 were non-carriers. Individuals 5, 11, 13, 28 and 30 were categorized as tentative carriers of *S. aureus*. The metagenome of individual 5 had *fnbA S. aureus* gene, while the metagenome of individual 11 had *clfB*, *fnbA* and *sodA S. aureus* genes. Individual 13 metagenome carried *arcC S. aureus* gene. Individual 28 metagenome carried *sodA*, *clfB*, *fnbA*, *isdA*, *oatA* and *sceD S. aureus* genes. Individual 30's metagenome had *fnbA* and *arcC S. aureus* gene. The details of these genes are shown in Appendix 72.

Table 6.2 Informative summary of the metagenomic datasets of the 32 individuals included in this study and their *S. aureus* carriage status.

Subject	Gender	Metagenome ID	Visit	Number of BGC	Tentative *S. aureus* genes present in the metagenome
1	M	7000000567	1	34	-
2	M	7000000273	1	8	-
		7000000711	2	21	-
3	F	7000000569	1	7	-
4	M	7000000314	1	13	-
		7000000415 (replicate 1)	2	4	-
		7000000035 (replicate 2)	2	4	-
5	M	7000000280	1	4	*fnbA*
6	F	7000000066	1	5	-
7	M	7000000023	1	29	-
		7000000427	2	1	-
8	F	7000000413	1	32	-
9	F	7000000620	2	5	-
10	M	7000000662	1	6	-
11	M	7000000629	1	8	*fnbA, clfB, sodA*
12	M	7000000529	1	1	-
13	F	7000000487	1	11	*arcC*
14	F	7000000520	1	6	-
15	F	7000000578	1	11	-
16	M	7000000477	1	11	-

Subject	Gender	Metagenome ID	Visit	Number of BGC	Tentative *S. aureus* genes present in the metagenome
17	F	7000000705	1	4	-
18	F	7000000209	1	6	-
19	F	7000000021	2	21	-
20	F	7000000728	1	19	-
21	M	7000000424	1	9	-
22	M	7000000220	1	8	-
23	F	7000000227	1	5	-
24	F	7000000476	1	10	-
25	M	7000000493	1	5	-
26	F	7000000613	2	23	-
27	M	7000000384	1	4	-
28	F	7000000602	2	22	*fnbA, clfB, isdA, oatA, sceD, sodA*
29	M	7000000141	1	4	-
30	F	7000000387	1	24	*fnbA, arcC*
31	M	7000000566	2	15	-
32	M	7000000679	2	4	-

This table describes the metagenomic dataset of each individual with regards to the gender, metagenome ID in IMG/M, the number of BGCs identified using IMG/ABC, the presence of tentative *S. aureus* genes and the visit when sample was taken. Visit 1 represent the metagenomic data of the individual from the anterior nares sample of the first visit, while visit 2 represents the metagenomic data of the individual from the anterior nares sample of the second visit. Individuals carrying one or more *S. aureus* genes in their metagenome are tentatively labelled as carriers of *S. aureus*. BGC; Biosynthetic gene clusters, M; Male, F; Female IMG/M; Integrated microbial genomes and microbiomes, -; no genes tentatively belonging to *S. aureus* were found in the metagenome of the individual.

6.4.2 Determination of biosynthetic gene clusters that have the potential to code for proteins involved in the production of antimicrobials

After searching the metagenome of all 32 individuals in the IMG/ABC database, 404 BGCs were found in total. Of the 404 BGCs identified, 13 could potentially code for proteins involved in the production of antimicrobials; 11 in the non-carrier group and 2 in the carrier group (Table 6.3). The raw data was recorded along with the designated number of the BGCs as depicted by IMG/ABC. This analysis was performed between years 2017 and 2018. The IMG/ABC website was updated as of September 2019 and the data for these BGCs are no longer available (Palaniappan *et al*., 2020). It is unclear if the numbering system of the BGCs has been changed however the IMG/ABC platform mentioned that some BGCs are “hidden”. It is not possible to retrieve the raw data of the BGCs using the designated numbering system recorded in this thesis. The genes that led to the identification of these 13 BGCs are detailed in appendix 73. This appendix contains the raw data. To search for these genes (eg. C176097__gene_11238) in the IMG/M website, the IMG metagenome ID of interest (eg. 7000000567) was searched in the IMG/M webpage. After retrieving the Microbiome details webpage, the “Explore” options was selected followed by the “Chromosomal maps” option. The scaffold ID was searched amongst the list of scaffolds in the metagenome. However due to the new numbering system, retrieving the gene using the

complete gene ID cannot be performed. If the gene ID in IMG/ABC is C176097__gene_11238, the scaffold to search for should be C176097. After the scaffold ID is selected, a new page will open with a map of C176097. Clicking anywhere on this map will open a “Chromosomal viewer page”, by which the “scaffold:C176097” option can be selected. Selecting this option will open a Metagenome scaffold detail page, which contains details such as gene count. For this scaffold, 5 genes were present and by selecting this feature, it will open a page with the genes in the scaffold by which the gene C176097__gene_11238 can be found.

Table 6.3 Detection of biosynthetic gene clusters that have the potential to encode for proteins involved in the production of antimicrobials as predicted by antiSMASH, BAGEL and PRISM

Individual	BGC	NCBI accession of WGS	Bacterium	Nasal carriage	Visit	Family of secondary metabolite predicted in WGS
1	161402349	MUYF01000003.1	*Dolosigranulum pigrum*	NC	1	Bacteriocin DpA, DpB, DpC, DpD, DpE and polyketide DpA
1	161402358	MUYF01000003.1	*Dolosigranulum pigrum*	NC	1	Bacteriocin DpA, DpB, DpC, DpD, DpE and polyketide DpA
1	161402355	CP010827.1	*Corynebacterium singulare*	NC	1	NRP CsinA and CsinB, Bacteriocin CsinF, Terpene CsinA, Type 1 polyketide synthase CsinA, polyketide CsinB
1	161402357	MUYF01000003.1	*Dolosigranulum pigrum*	NC	1	Bacteriocin DpA, DpB, DpC, DpD, DpE and polyketide DpA
2	161380740	JVSN01000021.1	*Corynebacterium propinquum*	NC	1	Siderophore CpA
2	161380741	MUYF01000003.1	*Dolosigranulum pigrum*	NC	1	Bacteriocin DpA, DpB, DpC, DpD, DpE and polyketide DpA
8	161390751	CP014634.1	*Corynebacterium simulans*	NC	1	NRP CsimC, CsimD and CsimE and Type 1 polyketide synthase CsimB
12	161400179	MUYF01000003.1	*Dolosigranulum pigrum*	NC	1	Bacteriocin DpA, DpB, DpC, DpD, DpE and polyketide DpA

Individual	BGC	NCBI accession of WGS	Bacterium	Nasal carriage	Visit	Family of secondary metabolite predicted in WGS
21	161391586	MUYF01000003.1	*Dolosigranulum pigrum*	NC	1	Bacteriocin DpA, DpB, DpC, DpD, DpE and polyketide DpA
28	161404210	CP015646.1	*Staphylococcus aureus*	C	2	Siderophore SaB, Siderophore SaB, NRP SaD, Terpene SaB
30	161388644	CP003604.1	*Staphylococcus aureus*	C	1	Siderophore SaD, Siderophore SaE, NRP SaE, Terpene SaB
31	161402347	MUYF01000003.1	*Dolosigranulum pigrum*	NC	2	Bacteriocin DpA, DpB, DpC, DpD, DpE and polyketide DpA
31	161402345	CP009312.1	*Lawsonella clevelandensis*	NC	2	Nonribosomal Peptide/Type 1 polyketide LcA

Potential antimicrobials are labelled by a lettering format. The first two or three letters pertain to the bacterium ie, Dp is *D. pigrum*, Csin is *C. singulare*, Csim is *C.simulans*, Sa is *S. aureus* and Lc is *L. cleveandensis*. The last letters, (A-F) are assigned to demonstrate any difference between the potential antimicrobials predicted. BGC; Biosynthetic gene cluster, WGS; Whole genome shotgun sequence, NC; non-carrier of *S. aureus*, C; carrier of *S. aureus*, NRP; Non-ribosomal peptide.

Overall, BGCs that have the potential to encode proteins involved in antimicrobial production were detected in the whole genome sequence of Actinobacteria (mainly *Corynebacterium singulare*, *Corynebacterium simulans* and to a lesser extent, *Lawsonella clevelandensis*), and Firmicutes (mainly *Dolosigranulum pigrum* and *Staphylococcus aureus*).

Tools used in this analysis have predicted that some BGCs in the metagenomes of the non-carrier group are likely to encode proteins involved in the production of bacteriocins, non-ribosomal peptides (NRP), type 1 polyketide/nonribosomal peptides (PK/NRP), siderophores and polyketides. BGCs that are likely to encode for proteins involved in bacteriocin biosynthesis were predicted in *D. pigrum* and *C. singular* exclusively in the individuals who do not carry *S. aureus*.

In contrast, the individuals who tentatively carried *S. aureus* were also predicted to carry *S. aureus* BGCs that have the potential to encode for proteins involved in the production of siderophores, NRPs and terpenes. The focus of this research was on antimicrobials that were involved in the exclusion of *S. aureus* from the anterior nares, so these BGCs were not further explored.

6.4.2.1 Potential secondary metabolites produced by *Dolosigranulum pigrum*

The antiSMASH predicted the genome of *D. pigrum* (accession number MUYF01000003.1) to have genes that could be involved in the production of three different bacteriocins and one polyketide. The three bacteriocins are labelled as bacteriocin DpA, bacteriocin DpB and bacteriocin DpC in this study. The polyketide is labelled as polyketide DpA in this study.

Bacteriocin DpA BGC is likely to carry genes that code for a lantipeptide bacteriocin, as predicted by antiSMASH (Figure 6.2). Bacteriocin DpA BGC is likely to contain genes that code for the production of a bovicin-like or a thermophilin-like compound (appendix 75).

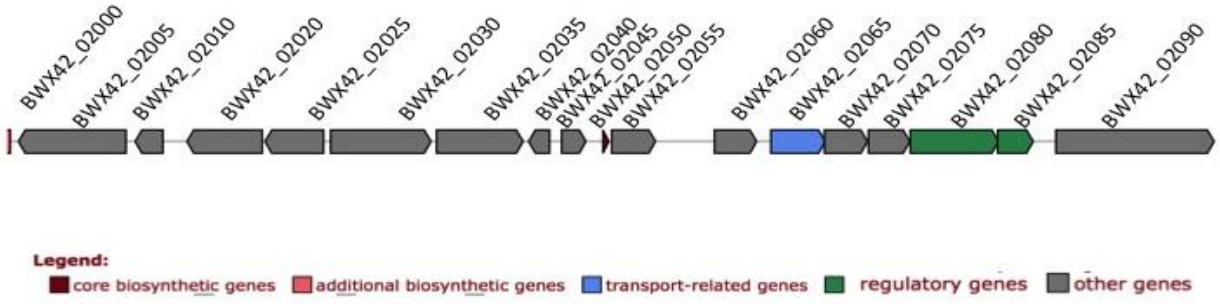


Figure 6.2 Diagrammatic representation of bacteriocin DpA BGC in the genome of *D. pigrum* (accession number MUYF01000003.1) as predicted by antiSMASH. Information about the genes that form this BGC are present in appendix 74.

As predicted by antiSMASH, bacteriocin DpB BGC has the potential to carry genes that code for a bacteriocin (Figure 6.3). This BGC contains a core biosynthetic gene (Locus tag BWX42_08270 in appendix 76) that can code for the production of an ATP-binding cassette (ABC) transporter. The antiSMASH has also predicted this putative BGC to have genes that could be present in other BGCs. The antiSMASH predicted 8% of the genes in these BGC to be likely present in the bacteriocin DpB BGC (Appendix 77).

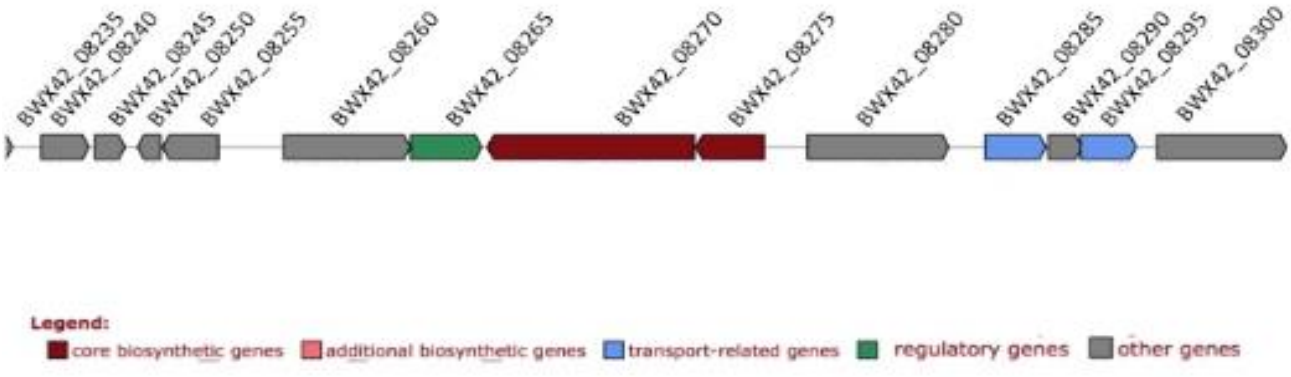


Figure 6.3 Diagrammatic representation of bacteriocin DpB BGC in the genome of *D. pigrum* (accession number MUYF01000003.1) as predicted by antiSMASH.

Information about the genes that form this BGC are present in appendix 76.

Bacteriocin DpC BGC has the potential to carry genes that code for the production of a bacteriocin (Figure 6.4). This BGC contains a core biosynthetic gene and a transport-related gene (locus tag BWX42_08385 and locus tag BWX42_08405 in appendix 78) that respectively code for the production of an ABC transporter and a multidrug ABC transporter. The antiSMASH has also predicted this putative BGC to have genes that could be present in other BGCs, but their metabolic products were not predicted by antiSMASH (Appendix 79).

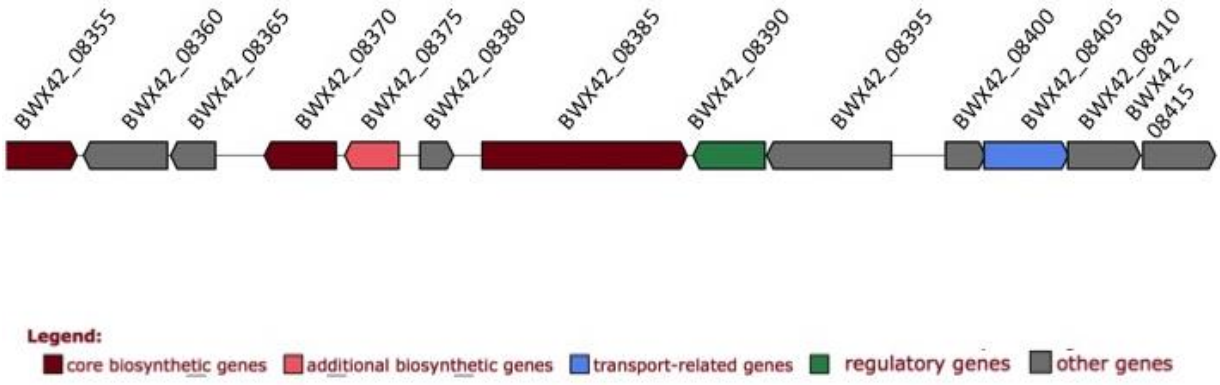


Figure 6.4 Diagrammatic representation of bacteriocin DpC BGC in the genome of *D. pigrum* (accession number MUYF01000003.1) as predicted by antiSMASH. Information about the genes that form this BGC are present in appendix 78.

As predicted by antiSMASH, polyketide DpA BGC (Figure 6.5) has the potential to code for the production of a polyketide. Two core biosynthetic genes in this BGC (locus tag BWX42_08665 and locus tag BWX42_08685 in appendix 80) respectively code for the production of acyl carrier protein and beta-ketoacyl-[acyl-carrier-protein] synthase II. Polyketide DpA BGC has the potential to code for the production of a compound related to xantholipin. The antiSMASH predicted 4% of the genes in the xantholipin BGC to possibly be present in the putative polyketide DpA BGC (appendix 81).

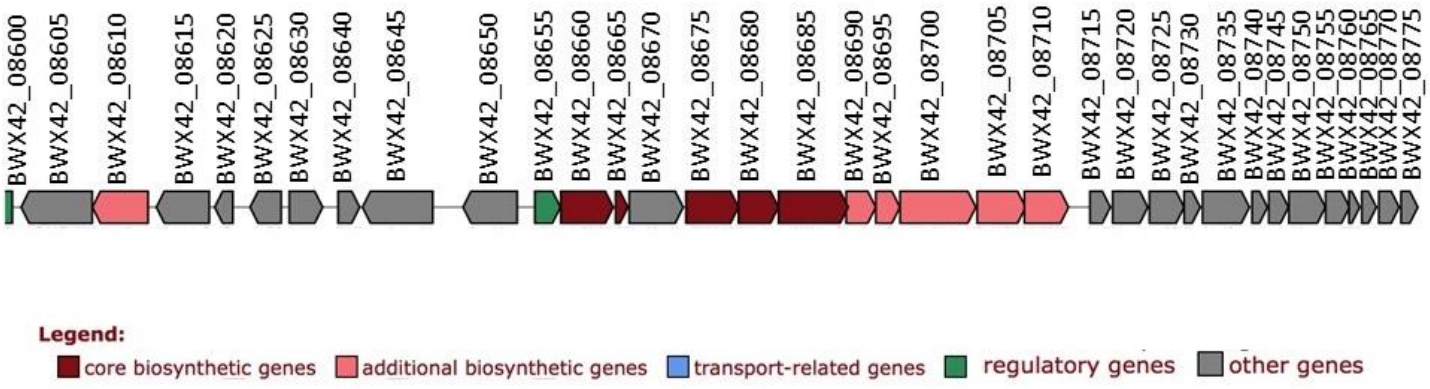


Figure 6.5 Diagrammatic representation of polyketide DpA BGC in the genome of *D. pigrum* (accession number MUYF01000003.1) as predicted by antiSMASH. Information about the genes that form this BGC are present in appendix 80.

PRISM predicted one BGC to have genes involved in polyketide production (Figure 6.6). It consisted of three biosynthetic domains: thiolation, acyltransferase and ketosynthase. PRISM predicted an amino acid sequence (labelled as orf1832 in appendix 82) positioned from 1822383-1822608 in the genome of *D. pigrum* (accession number MUYF01000003.1). It was 98.65% similar to the amino acid sequence annotated by the locus tag BWX42_08665 in antiSMASH (positioned from 1822384-1822608). This polyketide BGC therefore corresponded to polyketide DpA BGC predicted by antiSMASH.

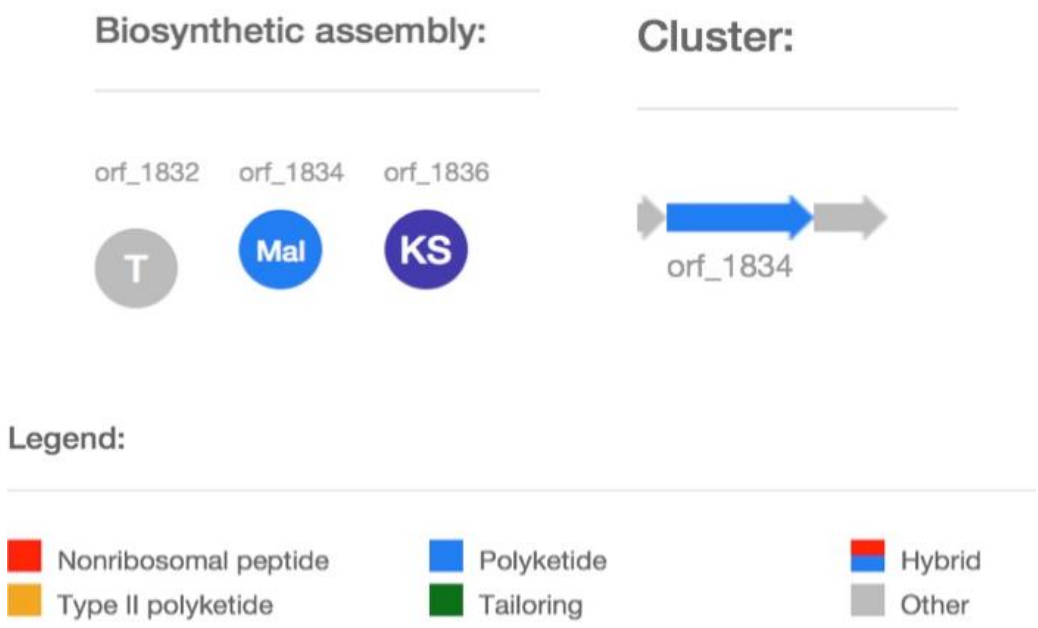


Figure 6.6 Predicted region of biosynthetic genes involved in polyketide DpA production in the genome of *D. pigrum* (MUYF0100003.1) predicted using PRISM. Domains in biosynthetic assembly are T; thiolation, Mal; acyltransferase, KS; ketosynthase. Details of genes are presented in appendix 82.

BAGEL predicted 2 bacteriocin BGCs, labelled as bacteriocin DpD BGC and bacteriocin DpE BGC, in the genome of *D. pigrum* (accession MUYF0100003.1). These bacteriocins were not detected by antiSMASH. Bacteriocin DpD BGC was predicted to possibly belong to the haloduracin subgroup of bacteriocins (Figure 6.7). This BGC was located from 1610103-1633628 in the genome of *D. pigrum* (accession MUYF0100003.1). One gene in the bacteriocin DpD BGC (locus tag ABC in appendix 83) can code for the production of putative lantibiotic ABC transporter. This BGC also carries a gene (locus tag LanM in appendix 83) that has the potential to code for the production of lantibiotic mersacidin modifying enzyme.

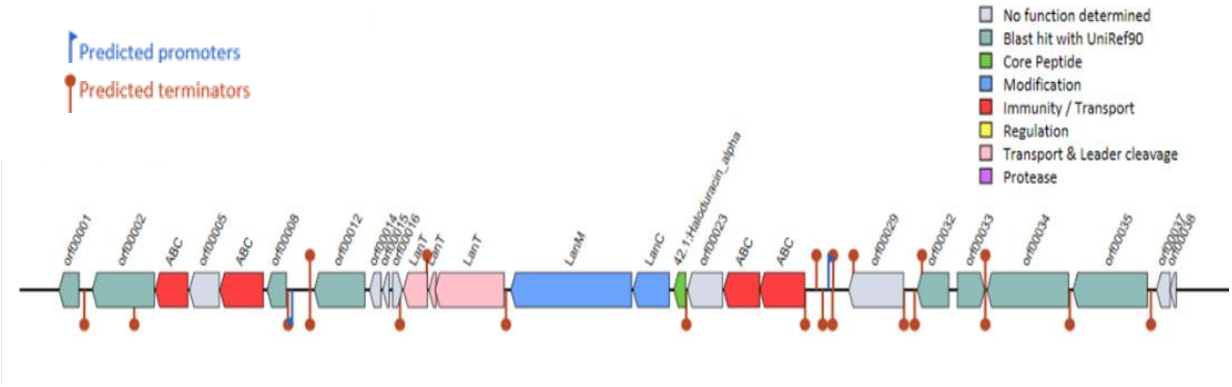


Figure 6.7 Diagrammatic representation of bacteriocin DpD BGC in the genome of *D. pigrum* (accession number MUYF01000003.1) as predicted by BAGEL. Information about the genes that form this BGC are present in appendix 83.

Bacteriocin DpE was predicted to possibly belong to the propionicin subgroup of bacteriocins (Figure 6.8). Bacteriocin DpE BGC was located from 1898483-1918843 in the genome of *D. pigrum* (accession MUYF0100003.1).

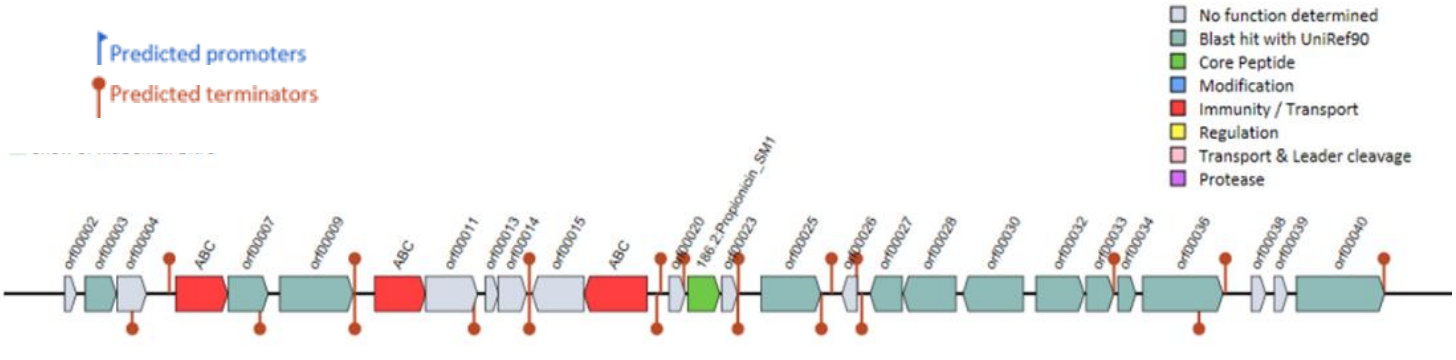


Figure 6.8 Diagrammatic representation of bacteriocin DpE BGC in the genome of *D. pigrum* (accession number MUYF01000003.1) as predicted by BAGEL. Information about the genes that form this BGC are present in appendix 84.

6.4.2.2 Potential secondary metabolites produced by *Corynebacterium* sp.

C. singulare, *C. propinquum* and *C. simulans* were the *Corynebacterium* strains identified to carry genes that have the potential to code for antimicrobial production in the individuals who tentatively did not carry *S. aureus*.

6.4.2.2.1 *Corynebacterium singulare*

The antiSMASH predicted the genome of *C. singulare* to carry BGCs with genes that have the potential to code for the production of one bacteriocin labelled as bacteriocin CsinF, one non-ribosomal peptide labelled as NRP CsinA, one terpene labelled as terpene CsinA and one type 1 polyketide synthase (T1PKS) labelled as T1PKS CsinA.

The antiSMASH predicted a BGC that has the potential to encode for proteins involved in the production of a NRP, labelled as NRP CsinA BGC (Figure 6.9). This BGC contains a gene (Locus tag CSING_01115 in appendix 85) that codes for the production of a protein involved in colicin V production. Another gene (locus tag CSING_01310 in appendix 85) codes for protein F, a protein involved in the type II secretion system. The antiSMASH predicted 34-42% of the genes in other BGCs that could also be present in this NRP CsinA BGC (Appendix 86).

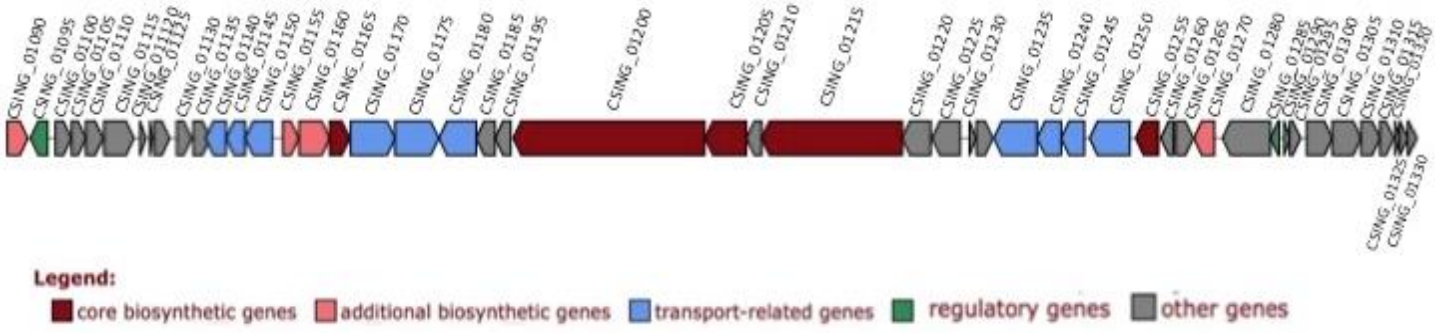


Figure 6.9 Diagrammatic representation of non-ribosomal peptide CsinA BGC in the genome of *C. singulare* (accession number CP010827.1) as predicted by antiSMASH. Information about the genes that form this BGC are present in appendix 85.

As predicted by antiSMASH, one BGC in the genome of *C. singulare* carries genes that have the potential to encode for the production of a bacteriocin, labelled as bacteriocin CsinF in this study (Figure 6.10). One of the core biosynthetic genes in this BGC (locus tag CSING_11435) can code for the production of a protein involved in the biosynthesis of a compound related to lactococcin 972 family bacteriocin. The antiSMASH predicted 16-19% of the genes of other BGCs to possibly be present in bacteriocin CsinF BGC (a diagram of these similar BGCs can be found in appendix 88).

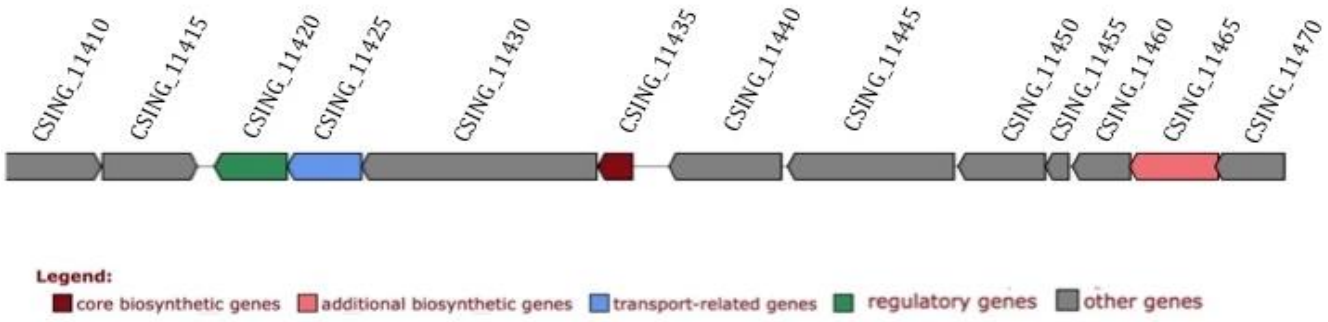


Figure 6.10 Diagrammatic representation of bacteriocin CsinF BGC in the genome of *C. singulare* (accession number CP010827.1) as predicted by antiSMASH. Information about the genes present in this BGC are present in appendix 87.

The antiSMASH predicted a BGC to carry genes that are likely to code for proteins involved in the production of T1PKS, with the BGC named as T1PKS CsinA BGC in this study (Figure 6.11). One biosynthetic gene in this BGC (CSING_12275 in appendix 89) codes for methyltransferase which can be a component of the modification domains present in the T1PKS CsinA module. One biosynthetic gene in this BGC (CSING_12295 in appendix 89) codes for a putative RND superfamily drug exporter, which can be involved in the export of T1PKS CsinA. A transport-related gene in this BGC (CSING_12380 in appendix 89) codes for an ABC-type cobalamin/Fe3+-siderophore transporter. The antiSMASH predicted 66-93% of the genes in other BGCs that could also be present in T1PKS CsinA BGC (Appendix 90).

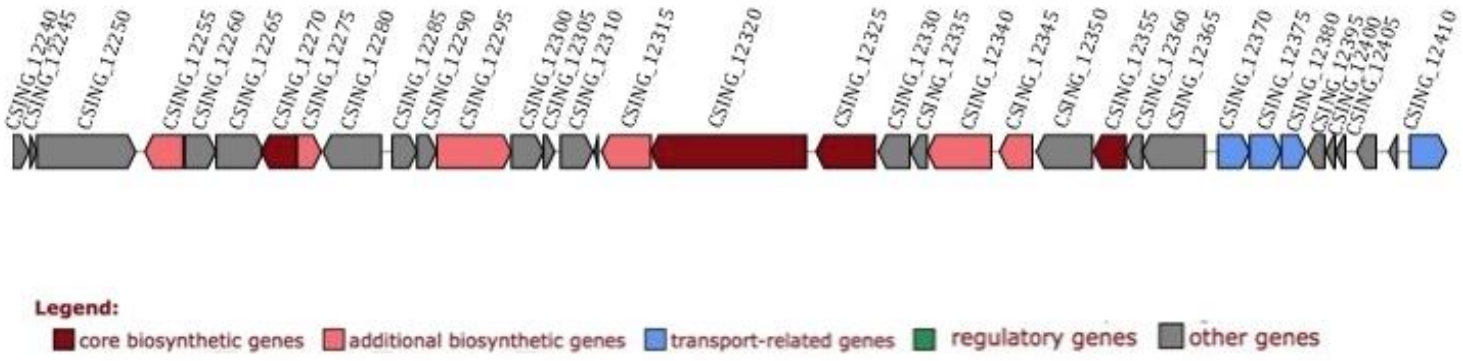


Figure 6.11 Diagrammatic representation of T1PKS CsinA BGC in the genome of *C. singulare* (accession number CP010827.1) as predicted by antiSMASH. Information about the genes present in this BGC are present in appendix 89.

The antiSMASH predicted a BGC that is likely to carry genes that code for the production of a terpene, labelled as terpene CsinA in this study (Figure 6.12). The antiSMASH predicted 25-47% of the genes in other BGCs to possibly be present in the putative terpene CsinA BGC (Appendix 92).

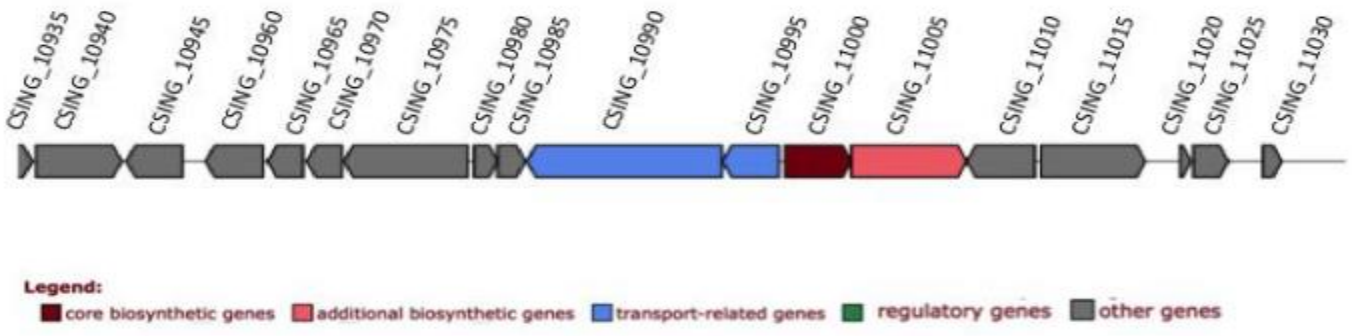


Figure 6.12 Diagrammatic representation of terpene CsinA BGC in the genome of *C. singulare* (accession number CP010827.1) as predicted by antiSMASH. Information about the genes present in this BGC are present in appendix 91.

BAGEL predicted one BGC to contain genes possibly involved in bacteriocin biosynthesis (Figure 6.13). BAGEL predicted an amino acid sequence (labelled as orf00016 in appendix 93) that was 100% identical to the amino acid sequence with locus tag CSING_11430 in antiSMASH. This bacteriocin BGC therefore corresponded to bacteriocin CsinF BGC predicted by antiSMASH.

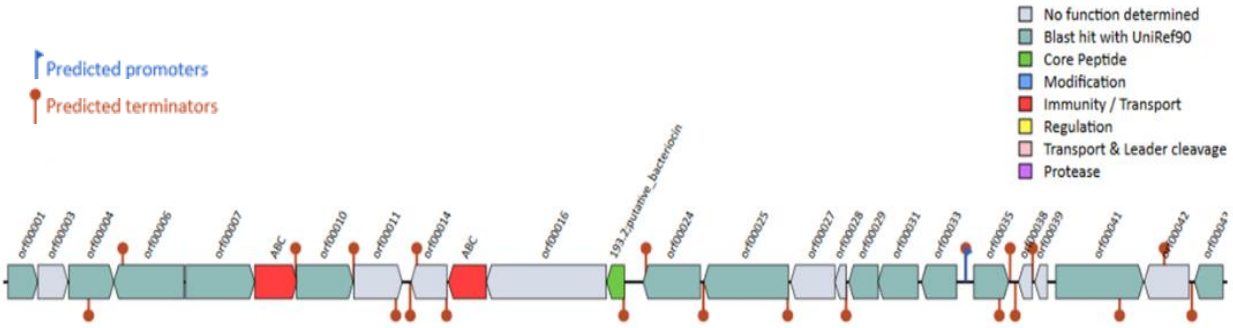


Figure 6.13 Diagrammatic representation of bacteriocin CsinF BGC in the genome of *C. singulare* (accession number CP010827.1) as predicted by BAGEL. Information about the genes present in this BGC are present in appendix 93.

PRISM has predicted the presence of two BGCs with genes possibly involved in the biosynthesis of NRPs. One of these NRP BGCs corresponded to non-ribosomal peptide CsinA BGC predicted by antiSMASH. The amino acid sequence with locus tag CSING_01210 in antiSMASH was 99.95% similar to amino acid sequence (labelled as orf_232 in appendix 94) predicted by PRISM.

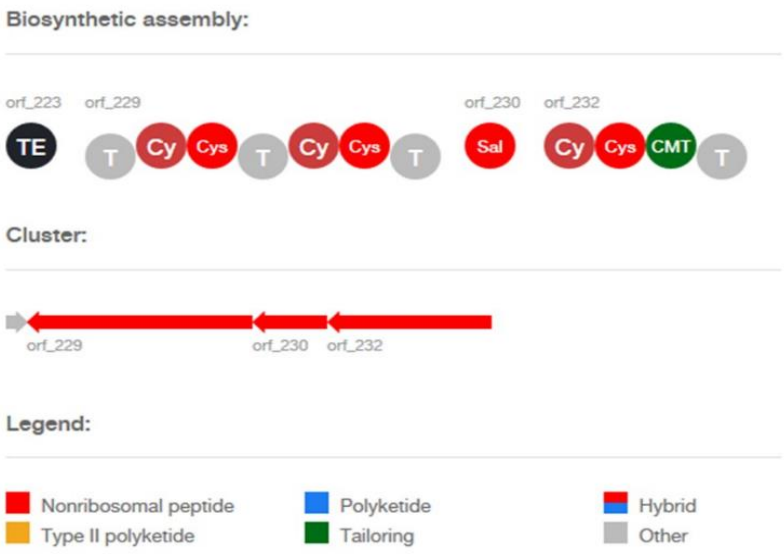


Figure 6.14 Predicted genes in the BGC of NRP CsinA in the genome of *C. singulare*.

Image acquired using PRISM. TE; thioesterase, T; Thiolation, Cy; condensation, Cys; adenylation, T; thiolation, Sal; Acyl adenylation enzyme, CMT; C-methyltransferase. Information about the genes present in this BGC are present in appendix 94.

The other NRP BGC predicted by PRISM was not predicted by antiSMASH. It was therefore labelled as NRP CsinB BGC (Figure 6.15).

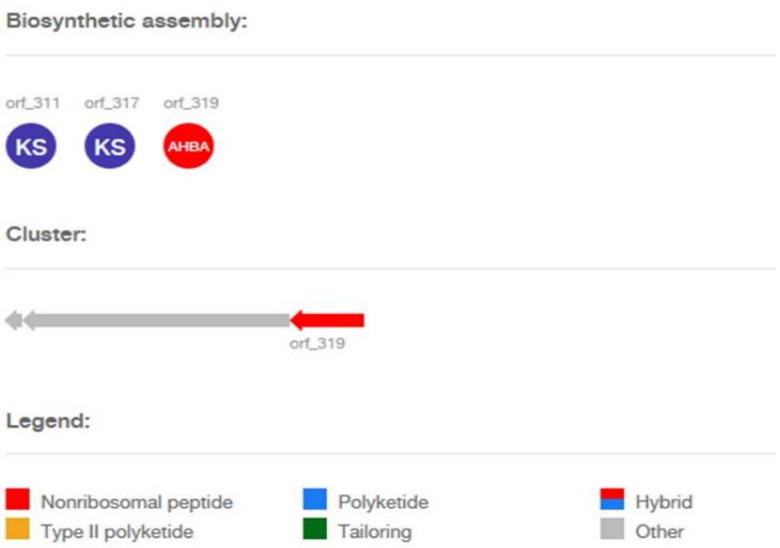


Figure 6.15 Predicted genes in the BGC of NRP CsinB in the genome of *C. singulare*.

Image acquired using PRISM. Domains in biosynthetic assembly are KS; Ketoynthase, AHBA; acyl adenylation enzyme. Information about the genes present in this BGC are present in appendix 95.

6.4.2.2.2 *Corynebacterium propinquum*

The antiSMASH predicted that the genome of *C. propinquum* carried a BGC with genes coding for the production of a siderophore (labelled as siderophore CpA (Figure 6.16). The antiSMASH also predicted some genes in other BGCs to possibly be present in the putative siderophore CpA BGC (Appendix 97). BAGEL did not predict bacteriocin BGCs in the genome of *C. propinquum*. In addition, no BGCs were predicted by PRISM.

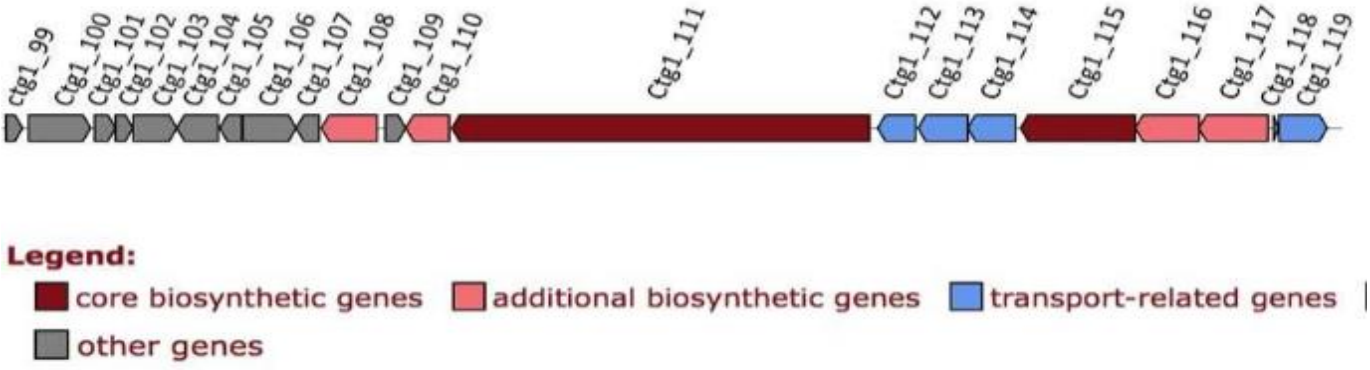


Figure 6.16 Diagrammatic representation of siderophore CpA BGC in the genome of *C. propinquum* (accession number JVSN01000021.1) as predicted by antiSMASH. Information about the genes present in this BGC are present in appendix 96 .

6.4.2.2.3 *Corynebacterium simulans*

The antiSMASH predicted BGCs that carry genes potentially coding for two different NRPs and one type of T1PKS in the genome of *C. simulans* (accession number CP014634.1).

The first NRP BGC was arbitrarily named as NRP CsimC BGC in this study (Figure 6.17). This BGC has the potential to contain genes that code for the production of a coelichelin-like or a scabichelin-like compound. The antiSMASH predicted 46% of the genes in the coelichelin BGC and 27% of the genes in the scabichelin BGC to possibly be present in the putative NRP CsimC BGC (Appendix 98).

Figure 6.17 Diagrammatic representation non-ribosomal peptide CsimC BGC in the genome of *C. simulans* as predicted by antiSMASH. Information about the genes present in this BGC are present in appendix 99.

A second NRP BGC, which was called NRP CsimD BGC in this study, was also detected in the genome of *C. simulans* (Figure 6.18). One of the transport-related genes in this BGC (locus tag WM42_1803) code for an ABC-transporter protein. Moreover, this BGC carries a gene (locus tag WM42_1828 in appendix 100) involved in the regulation of NRP CsimD production. These types of genes are typically found in the vicinity of cyclic lipopeptide BGCs and can code for LuxR-type proteins, involved in cyclic lipopeptide production. Thus it can be assumed that they play a role in the regulation of NRP CsimD (D'aes *et al*., 2014).

This BGC has the potential to code for the production of a metabolic compound related to glycopeptidolipid. The antiSMASH predicted 12% of the genes of the glycopeptidolipid BGC to possibly be present in the putative NRP CsimD BGC (appendix 101).

Figure 6.18 Diagrammatic representation non-ribosomal peptide CsimD BGC in the genome of *C. simulans* as predicted by antiSMASH. Information about the genes present in this BGC are present in appendix 100.

The antiSMASH also predicted a T1PKS BGC, named T1PKS CsimB BGC in this study, in the genome of *C. simulans* (Figure 6.19). The antiSMASH predicted this BGC to have genes that could be present in other BGCs. It was predicted that 61-80% of the genes in other BGCs could also be present in the putative T1PKS CsimB BGC (Appendix 103).

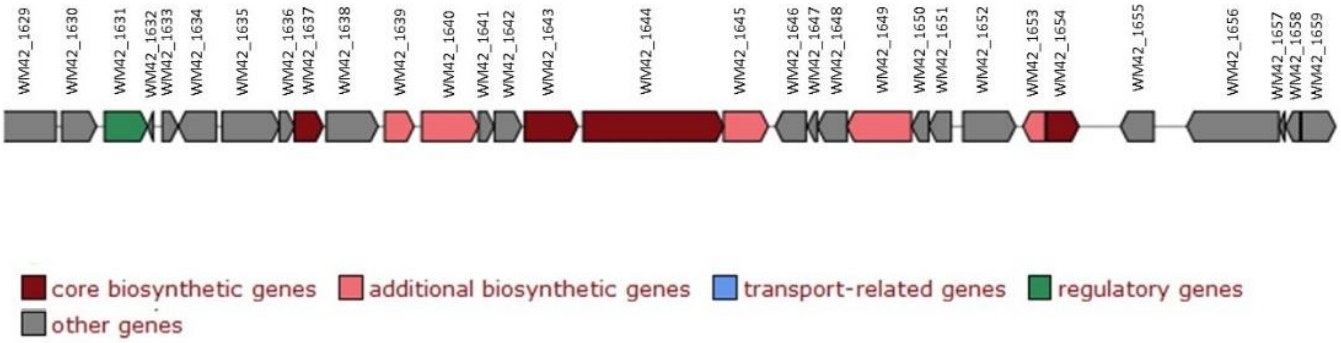


Figure 6.19 Diagrammatic representation T1PKS CsimB BGC in the genome of *C. simulans* as predicted by antiSMASH. Information about the genes present in this BGC are present in appendix 102.

BAGEL did not predict any BGCs that carry genes involved in bacteriocin production. Prism predicted three different types of NRPs BGCs, two of which were already predicted by antiSMASH, and one PK/NRP BGC, which was also predicted as a T1PKS BGC in antiSMASH.

PRISM predicted a BGC that corresponds to the NRP CsimC BGC predicted using antiSMASH (Figure 6.20). PRISM predicted an amino acid sequence (labelled as orf_1332 in appendix 104) which was 99.97% similar to the amino acid sequence with locus tag WM42_1378 in antiSMASH.

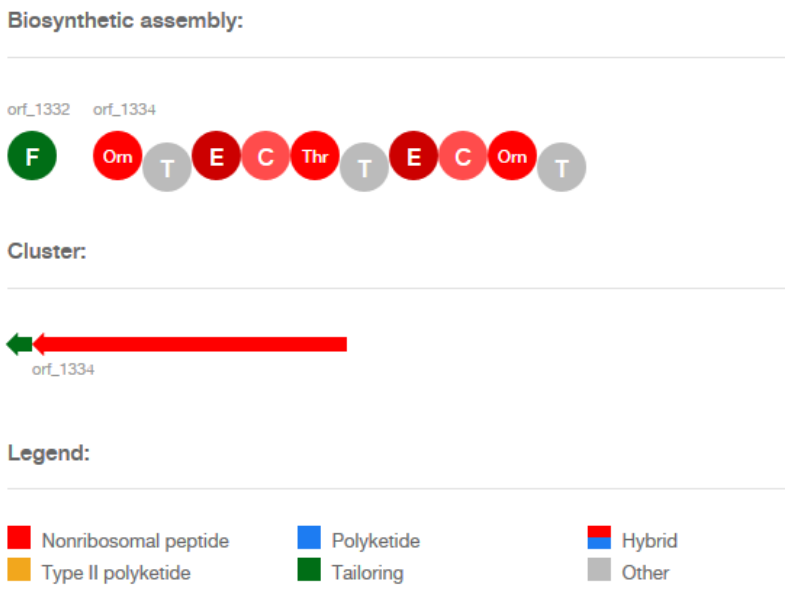


Figure 6.20 Predicted genes in the NRP CsimC BGC in the genome of *C. simulans*.

Image was acquired using PRISM. Domains in biosynthetic assembly are F; formyltransferase, Orn; Adenylation, T; Thiolation, E; Epimerization, C; condensation, Thr; adenylation, Orn; Adenylation, KS; Ketoynthase, AHBA; acyl adenylation enzyme. Information about the genes present in this BGC are present in Appendix 104.

PRISM predicted a BGC corresponding to NRP CsimD BGC predicted using antiSMASH (Figure 6.21). The amino acid sequences with locus tags WM42_1807, WM42_1813, WM42_1814, WM42_1816 predicted in antiSMASH were 99.10%, 99.96%, 100% and 99.95% similar to amino acid sequences respectively labelled as orf_1759, orf_1765, orf_1766 orf_1768 in PRISM (Appendix 105).

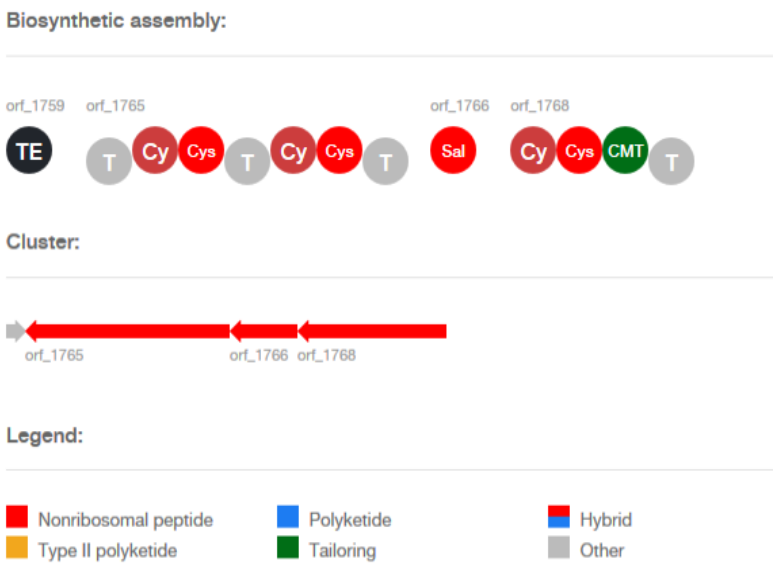


Figure 6.21 Predicted genes in the NRP CsimD BGC in the genome of *C. simulans*. Image acquired using PRISM. Domains in biosynthetic assembly are TE; thioesterase, T; thiolation, Cy; condensation, Cys; adenylation, Sal; Acyl adenylation enzyme, CMT; C-methyltransferase. Information about the genes present in this BGC are present in Appendix 105.

PRISM also predicted another BGC that carries genes with the potential to code for a NRP. The BGC was named as NRP CsimE BGC in this thesis (Figure 6.22).

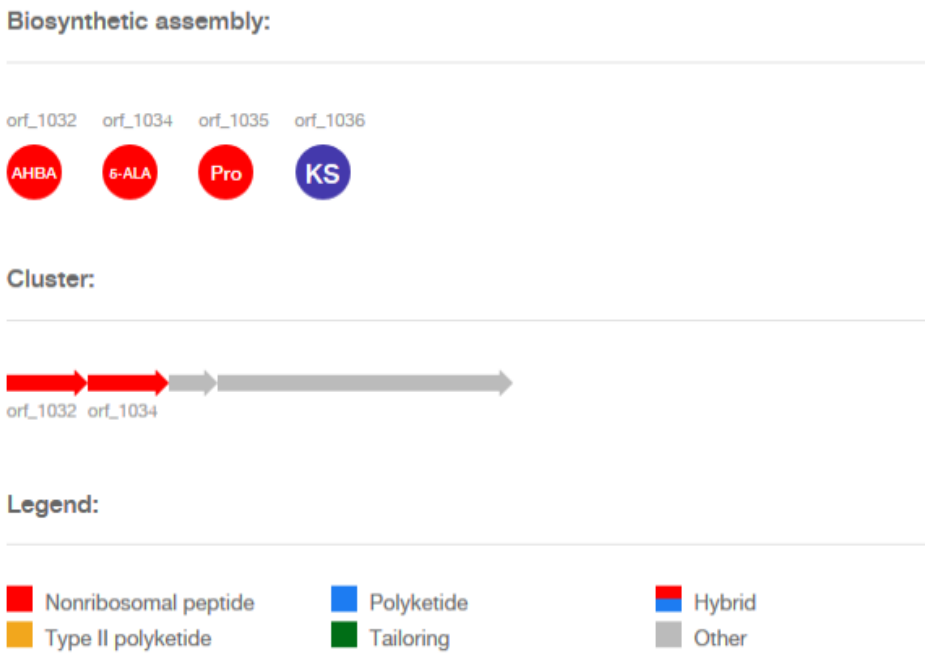


Figure 6.22 Predicted genes in the NRP CsimE BGC in the genome of *C. simulans*. Image acquired using PRISM. Domains in biosynthetic assembly are AHBA; acyl adenylation enzyme, 5-ALA; Acyl adenylation enzyme, Pro; adenylation, KS; ketosynthase. Information about the genes present in this BGC are present in Appendix 106.

PRISM predicted a BGC to be similar to T1PKS CsimB BGC that was predicted using antiSMASH (Figure 6.23). The amino acid sequences with locus tags WM42_1643 and WM42_1644 predicted in antiSMASH were 99.84% and 99.94% similar to the amino acid sequences respectively labelled as orf_1592 and orf_1593 in PRISM.

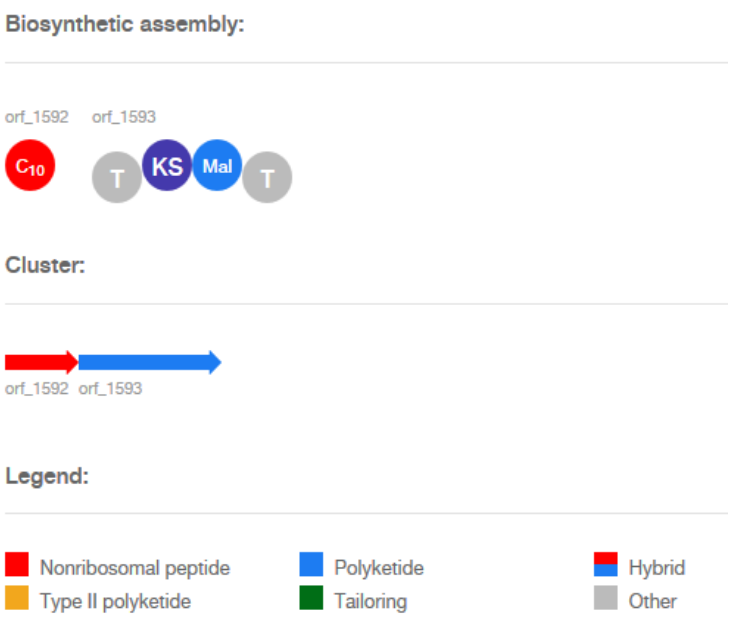


Figure 6.23 Predicted genes in the T1PKS CsimB BGC in the genome of *C. simulans*. Image acquired using PRISM. Domains in biosynthetic assembly are C10; acyl adenylation enzyme, T; thiolation, KS; ketosynthase, Mal; acyltransferase, T; thiolation. Information about the genes present in this BGC are present in Appendix 107.

6.4.2.3 Potential secondary metabolites produced by *Lawsonella clevelandensis*

The antiSMASH predicted one NRPS/ Type 1 PKS (NRPS/T1PKS) BGC, arbitrarily named NRP/T1PKS LcA BGC in this study (Figure 6.24). The antiSMASH predicted 37-100% of genes in other BGCs to possibly be present in the putative NRP/T1PKS LcA BGC (a diagram of these similar BGCs can be found in appendix 109). One of these BGCs was 100% identical to the NRP/T1PKS LcA BGC predicted in this study.

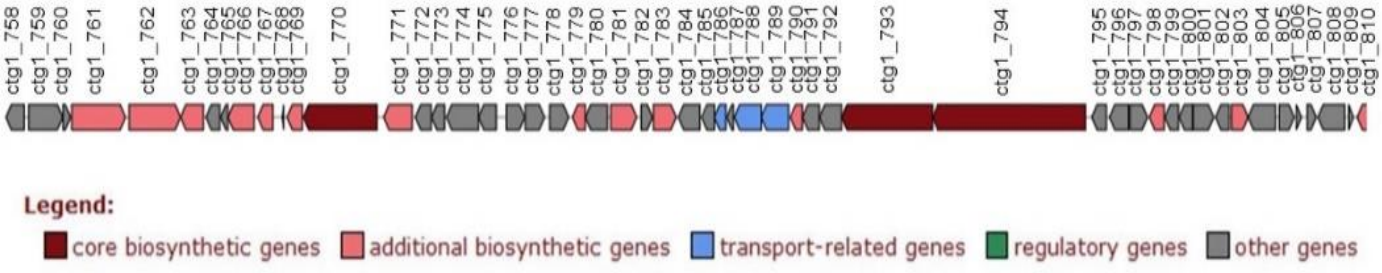


Figure 6.24 Diagrammatic representation of NRP/T1PKS LcA BGC in the genome of *L. clevelandensis* as predicted by antiSMASH. Information about the genes present in this BGC are present in appendix 108.

BAGEL did not predict any BGCs that have the potential to carry genes involved in bacteriocin production in the genome of *Lawsonella clevelandensis*. Prism predicted one NRP BGC and one non-ribosomal peptide/polyketide synthase (NRP/PKS) BGC, both of which had been predicted by antiSMASH.

The NRPS/PKS BGC predicted by PRISM corresponded to the NRP/T1PKS LcA BGC predicted by antiSMASH (Figure 6.25). The amino acid sequences with locus tags ctg1_770 and ctg1_771 predicted in antiSMASH were 77.87% and 100% similar to amino acid sequences respectively labelled as orf_770 and orf_771 in PRISM (appendix 110). PRISM and antiSMASH therefore identified the same NRP/T1PKS LcA BGC.

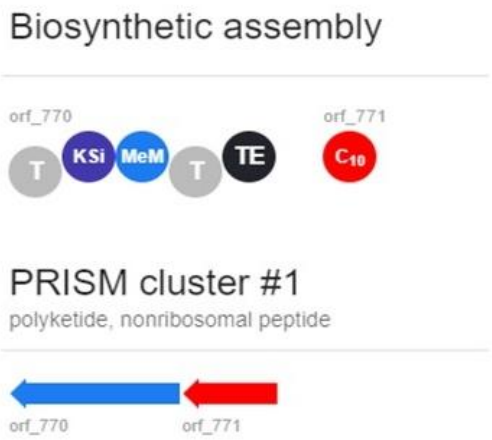


Figure 6.25 Predicted genes in the NRPS/PKS LcA BGC in the genome of *Lawsonella clevelandensis* as predicted by PRISM. Image acquired using PRISM. Domains in biosynthetic assembly are T; thiolation, KSi; ketosynthase, MEM; acyltransferase, T; thiolation, TE; thioesterase. Information about the genes present in this BGC are present in Appendix 110.

The NRP BGC predicted by PRISM corresponds to NRP/T1PKS LcA BGC predicted by antiSMASH (6.26). The amino acid sequences with locus tags ctg1_790 and ctg1_793 predicted in antiSMASH were 99.63%and 99.94% similar to amino acid sequences respectively labelled as orf_790 and orf_794 in PRISM (appendix 111). The BGCs predicted by PRISM

were therefore identical to NRP/T1PKS LcA BGC predicted by antiSMASH.

Figure 6.26 Predicted genes in NRP LcA BGC in the genome of *Lawsonella clevelandensis* as predicted by PRISM. Image acquired using PRISM. Domains in biosynthetic assembly are TE; thioesterase, Cy; condensation, Cys; Adenylation, CMT; C-methyltransferase, T; thiolation, TE; thioesterase, C10; acyl adenylation enzyme. Information about the genes present in this BGC are present in Appendix 111.

6.4.3 Presence of potential BGCs in the whole genome shotgun sequence of *D. pigrum* ATCC 51524

The whole genome shotgun sequence of *D. pigrum* (ATCC 51524) was publicly available at the time of this research. Future work using this strain was anticipated, so it was necessary to check if this strain carries the BGCs predicted in this research. The whole genome shotgun sequence of

D. pigrum ATCC 51524 (NCBI accession JH601103.1) was therefore submitted to antiSMASH, PRISM and BAGEL3. This was performed to identify if the whole genome shotgun sequence of *D. pigrum* ATCC 51524 carries genes involved in the production of bacteriocin DpA, bacteriocin DpB, bacteriocin DpC, bacteriocin DpD, bacteriocin DpE and polyketide DpA, which were predicted in the genome of *D. pigrum* (NCBI accession MUYF0100003.1) in section 6.4.2.1 of this chapter.

All were present. An additional bacteriocin, initially not predicted in the genome of *D. pigrum* (NCBI accession MUYF0100003.1) by the platforms, was predicted in the genome of *D. pigrum* NCTC 51524. After searching for the BGC in the genome of *D. pigrum* (NCBI accession MUYF0100003.1), it showed that it carried genes that have the potential to be involved in the biosynthesis of an additional bacteriocin . This bacteriocin was arbitrarily named bacteriocin DpG in this study.

The antiSMASH predicted the whole genome shotgun sequence of *D. pigrum* ATCC 51524 to have genes involved in the production of 2 different bacteriocins and one polyketide. The bacteriocin BGCs predicted were bacteriocin DpD BGC and bacteriocin DpG BGC. Bacteriocin DpD BGC was predicted in the genome of *D. pigrum* (NCBI accession MUYF01000003.1). In addition, the polyketide BGC predicted in the whole

genome shotgun sequence of *D. pigrum* ATCC 51524 was similar to polyketide DpA BGC predicted in the genome of *D. pigrum* (NCBI accession MUYF01000003.1). This polyketide BGC was also predicted in the whole genome shotgun sequence of *D. pigrum* ATCC 51524 using PRISM, which was the only BGC predicted by this platform. BAGEL was able to predict bacteriocin DpD BGC, bacteriocin DpE BGC and bacteriocin DpG BGC in the whole genome shotgun sequence of *D. pigrum* ATCC 51524.

Since bacteriocin DpA BGC, bacteriocin DpB BGC and bacteriocin DpC BGC were not predicted by any of the platforms, the gene sequences that form these BGCs were searched in the whole genome shotgun sequence of *D. pigrum* ATCC 51524. This was done by searching for the genes that form bacteriocin DpC BGC, bacteriocin DpB BGC and bacteriocin DpC BGC in the whole genome shotgun sequence of *D. pigrum* ATCC 51524.

6.4.3.1 Bacteriocin DpA

Bacteriocin DpA BGC was not predicted in the genome of *D. pigrum* ATCC 51524 by antiSMASH, PRISM and BAGEL. However, a gene involved in the production of bacteriocin DpA was identified in the whole genome shotgun sequence of *D. pigrum* ATCC 51524. This was done by searching for core biosynthetic gene sequences predicted in bacteriocin DpA BGC

in the genome of *D. pigrum* (accession number MUYF01000003.1). One core biosynthetic gene in bacteriocin DpA BGC (locus tag BWX42_02000) in the genome of *D. pigrum* (accession number MUYF01000003.1) was 99.69% similar to a gene in the whole genome shotgun sequence of *D. pigrum* ATCC 51524 (Appendix 112).

6.4.3.2 Bacteriocin DpB

Bacteriocin DpB BGC was not predicted in the whole genome shotgun sequence of *D. pigrum* ATCC 51524 by antiSMASH, PRISM and BAGEL. However, a gene involved in the production of bacteriocin DpB was identified in the whole genome shotgun sequence of *D. pigrum* ATCC 51524. This was done by searching for gene sequences predicted in bacteriocin DpB BGC in the genome of *D. pigrum* (accession number MUYF01000003.1). One gene in bacteriocin DpB BGC (locus tag BWX42_08265) in the genome of *D. pigrum* (accession number MUYF01000003.1) was 97.14% similar to a gene in the whole genome shotgun sequence of *D. pigrum* ATCC 51524 (Appendix 113).

6.4.3.3 Bacteriocin DpC

Bacteriocin DpC BGC was not predicted in the genome of *D. pigrum* ATCC 51524 by antiSMASH, PRISM and BAGEL. However, a gene involved in the production of bacteriocin DpC was identified in the whole genome

shotgun sequence of *D. pigrum* ATCC 51524. This was done by searching for all the biosynthetic gene sequences predicted in bacteriocin DpC BGC in the genome of *D. pigrum* (accession number MUYF01000003.1). One core biosynthetic gene in bacteriocin DpC BGC (locus tag BWX42_08375) in the genome of *D. pigrum* (accession number MUYF01000003.1) was 96.81% similar to a gene in the whole genome shotgun sequence of *D. pigrum* ATCC 51524 (appendix 114).

6.4.3.4 Bacteriocin DpD

The antiSMASH and BAGEL were able to predict genes involved in the biosynthesis of bacteriocin DpD in the whole genome shotgun sequence of *D. pigrum* ATCC 51524.

Bacteriocin DpD BGC was predicted using antiSMASH (Figure 6.27) and BAGEL (Figure 6.28). The amino acid sequence (labelled as ABC in BAGEL) was 244 amino acids in length. This sequence was 99.59% similar to the amino acid sequence labelled as Ctg1_543 in antiSMASH (Appendix 117). One gene in this BGC (locus tag orf00008 in appendix 116) codes for an iron-dependent repressor IdeR. Other genes in this BGC (locus tag ABC and LanM in appendix 116) respectively code for a putative lantibiotic ABC transporter and a lantibiotic mersacidin modifying enzyme.

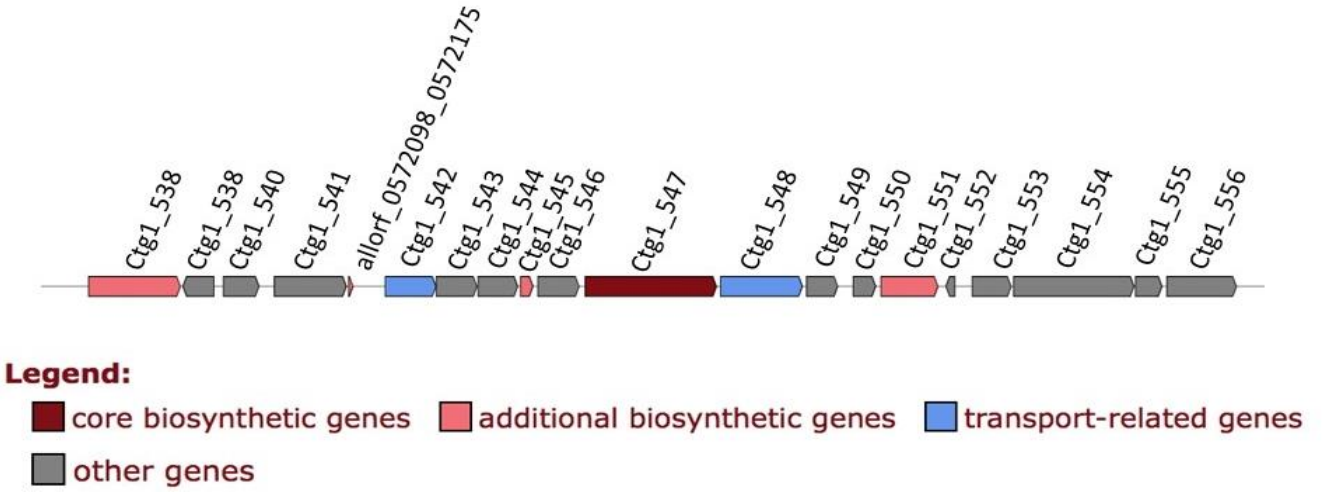


Figure 6.27 Diagrammatic representation of bacteriocin DpD BGC in the whole genome shotgun sequence of *D. pigrum* ATCC 51524 as predicted by antiSMASH. Information about the genes present in this BGC are present in appendix 115.

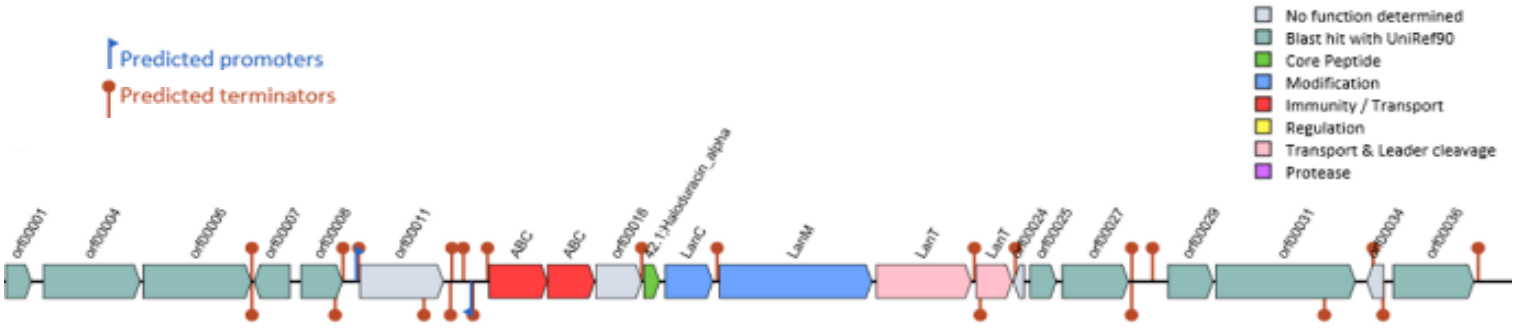


Figure 6.28 Predicted bacteriocin DpD BGC in the whole genome shotgun sequence of *D. pigrum* ATCC 51524 as predicted by BAGEL. Information about the genes present in this BGC are present in appendix 116.

6.4.3.5 Bacteriocin DpE

A gene involved in the biosynthesis of bacteriocin DpE was predicted in the whole genome shotgun sequence of *D. pigrum* ATCC 51524 by BAGEL. The amino acid sequence (labelled as ABC in BAGEL) is 245 amino acids in length (Figure 6.29). It was 100% identical to a gene in the whole genome shotgun sequence of *D. pigrum* ATCC 51524 (Appendix 118). One gene in this BGC (locus tag orf00007 in appendix 119) codes for a transcriptional repressor YvoA, which can control transcription of bacteriocin DpE BGC. A transport-related gene in this BGC (locus tag ABC in appendix 119) codes for a putative lantibiotic ABC transporter.

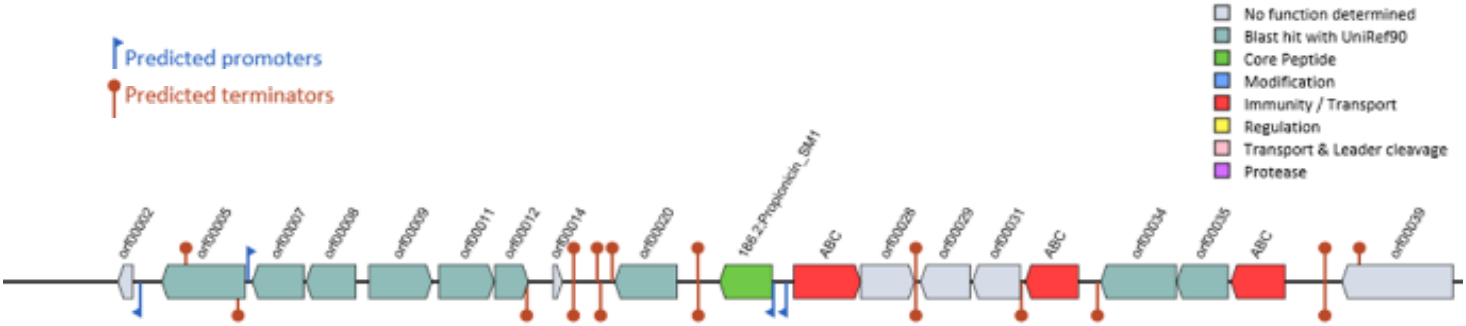


Figure 6.29 Predicted bacteriocin DpE BGC in the whole genome shotgun sequence of *D. pigrum* ATCC 51524 as predicted by BAGEL. Information about the genes present in this BGC are present in appendix 119.

6.4.3.6 Bacteriocin DpG

An additional bacteriocin BGC, initially not predicted in the of whole genome of *D. pigrum* (NCBI accession MUYF0100003.1) by all the platforms, was predicted in the whole genome shotgun sequence of *D. pigrum* NCTC 51524 by antiSMASH. This bacteriocin BGC was arbitrarily named bacteriocin DpG BGC in this study (Figure 6. 30).

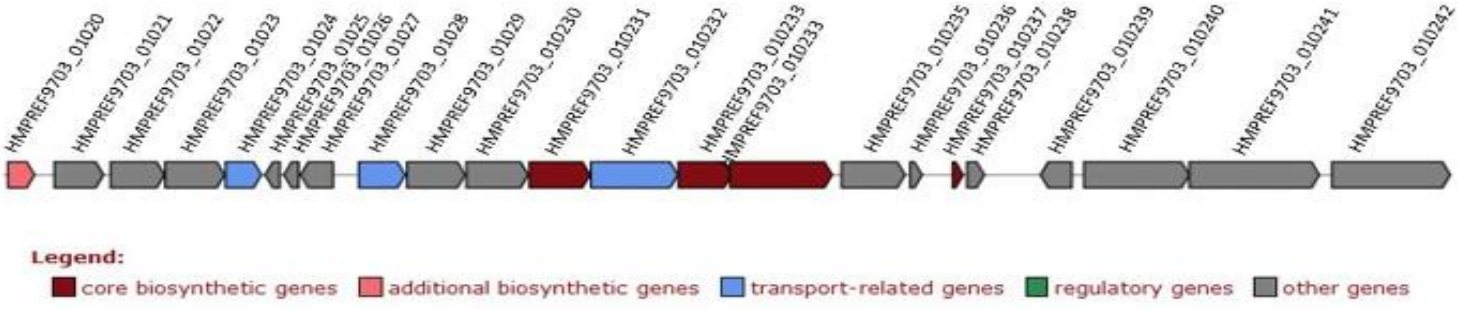


Figure 6.30 Predicted bacteriocin DpG BGC in the whole genome shotgun sequence of *D. pigrum* ATCC 51524 as predicted by antiSMASH. Information about the genes present in this BGC are present in appendix 120.

Bacteriocin DpG BGC was also predicted in the whole genome shotgun sequence of *D. pigrum* ATCC 51524 by BAGEL (Figure 6.31). The amino acid sequence (labelled as orf00001 in BAGEL) was 100% identical to an amino acid sequence labelled as HMPREF9703_01020 in antiSMASH. One of the predicted genes in this BGC (locus tag ABC in appendix 121) codes for a putative bacteriocin ABC transporter protein. Another gene

(locus tag LanB in appendix 121) codes for a protein involved in lantibiotic biosynthesis.

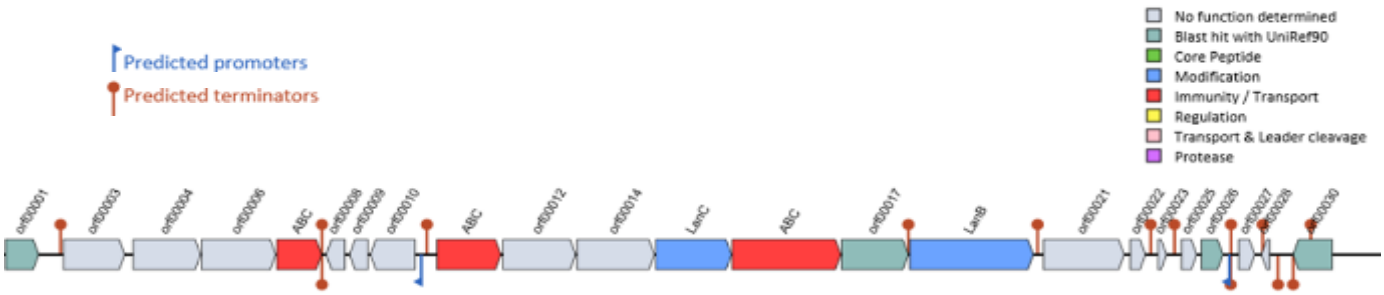


Figure 6.31 Predicted bacteriocin DpG BGC in the whole genome shotgun sequence of *D. pigrum* ATCC 51524 as predicted by BAGEL. Information about the genes present in this BGC are present in appendix 121.

Bacteriocin DpG BGC was also present in the genome of *D. pigrum* (NCBI accession MUYF0100003.1) however, antiSMASH did not predict it in this study.

The amino acid sequence labelled as HMPREF9703_01020 in the antiSMASH results for *D. pigrum* ATCC 51524, was 96.43% similar to an amino acid sequence (labelled as BWX42_03205) in the genome of *D. pigrum* (NCBI accession MUYF0100003.1) (Appendix 122).

6.4.3.7 Polyketide DpA

Polyketide DpA BGC was predicted in the whole genome shotgun sequence of *D. pigrum* ATCC 51524 (Figure 6.32). The amino acid sequence (labelled as HMPREF9703_00336) was 97.45% similar to an amino acid sequence (labelled as orf_1832) in the genome of *D. pigrum* (NCBI accession MUYF0100003.1) as predicted by PRISM (Appendix 124).

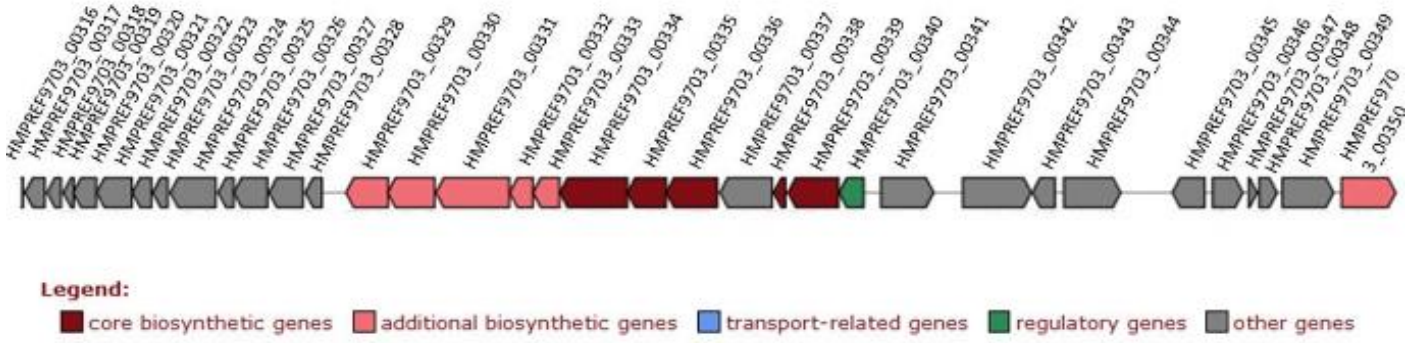


Figure 6.32 Polyketide DpA BGC in the whole genome shotgun sequence of *D. pigrum* ATCC 51524 as predicted by antiSMASH. Information about the genes present in this BGC are present in appendix 123.

In addition, a gene involved in the production of polyketide DpA was also predicted by PRISM in the whole genome shotgun sequence of *D. pigrum* ATCC 51524 (Figure 6.33). The amino acid sequence (labelled as orf_338 in PRISM) was 100% identical to an amino acid sequence (labelled as orf_1832 in PRISM) in the genome of *D. pigrum* (NCBI accession MUYF0100003.1).

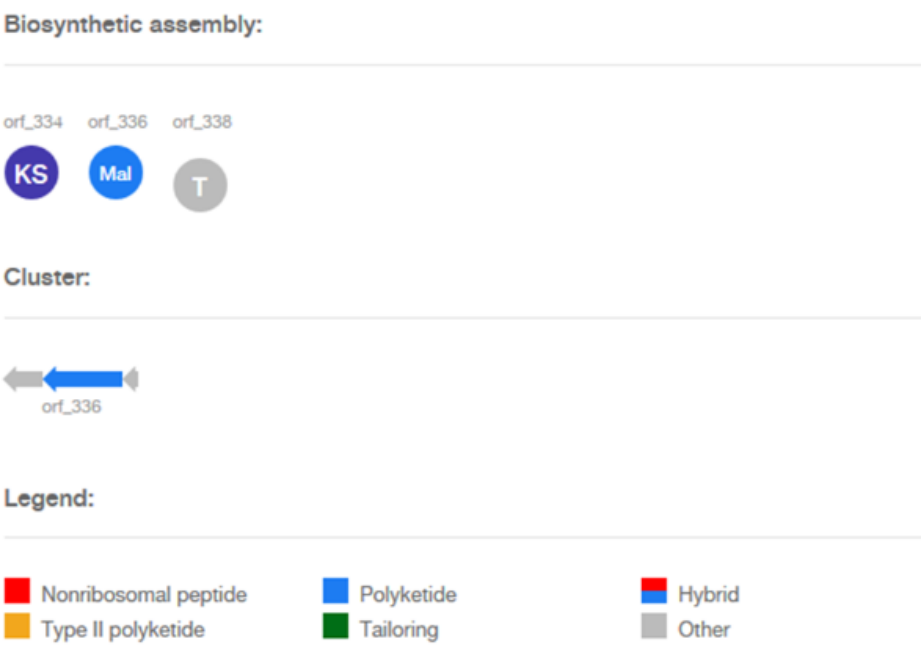


Figure 6.33 Polyketide DpA BGC in the whole genome shotgun sequence of *D. pigrum* ATCC 51524 as predicted by PRISM. Domains in biosynthetic assembly are T; thiolation, Mal; Acyltransferase and KS; ketosynthase domain. Information about the genes present in this BGC are present in appendix 125.

6.5 Discussion

The human microbiome of the upper airway is diverse and resides in an ecological niche that is poor in nutrients (Krismer *et al*., 2017). This suggests that colonizing bacteria compete strongly for their survival, using a variety of mechanisms to inhibit colonization of competitors (Krismer *et al*., 2017, Brugger *et al*., 2020).

This study has shown the presence of BGCs that have the potential to encode proteins involved in antimicrobial production in the anterior nares of carriers and non-carriers of *S. aureus*. In the carrier group, all the BGCs belonged to *S. aureus*. In the non-carrier group, most BGCs belonged to *Actinobacteria* (mainly *Corynebacterium singular*, *Corynebacterium simulans* and to a lesser extent, *Lawsonella clevelandensis*) and the Firmicute *Dolosigranulum pigrum.*

Bioinformatics analysis in this chapter showed that bacteriocins and a polyketide are likely to be produced by *Dolosigranulum pigrum* residing in the anterior nares of *Staphylococcus aureus* non-carriers. In addition, bacteriocins, siderophores, NRP and Type 1 polyketide are likely to be produced by *Corynebacterium* spp.. *Corynebacterium simulans* was predicted to produce NRP CsinA. Due to the presence of a gene that can code for the production of a protein involved in colicin biosynthesis, NRP

CsinA is likely to be a colicin-like compound, possibly secreted by a type II secretion system. Type II secretion system is used by bacteria to move substrates across their cell membrane and provide self-resistance to the producer bacterium (Naskar *et al*., 2021). Colicin is a plasmid- encoded bacteriocin produced by some strains of *E. coli,* and this is the first study demonstrating the possibility of *Corynebacterium* spp. to produce a compound related to colicin (Hahn-Löbmann et al., 2019).

Moreover, some non-carriers, including subject 1 and subject 2, showed the co-presence of potential antimicrobial-producing *D. pigrum* and *Corynebacterium* spp. The research presented in this chapter cannot predict if one of these species or both species is contributing to the exclusion of *S. aureus* in the anterior nares. Recent metagenomic studies have indicated that the joint presence of *Corynebacterium* spp. and *Dolosigranulum pigrum* was negatively associated with *S. aureus* in the anterior nares (Escapa *et al*., 2018, Brugger *et al*., 2020). In addition to the exclusion of *S. aureus*, a previous report has demonstrated the ability of the nasal bacterium *Corynebacterium accolens* to inhibit the growth of *Streptococcus pneumoniae* in the anterior nares by releasing a triacylglycerol lipase (Bomar *et al*., 2016).

Interestingly, subject 31 showed the co-existence of *Lawsonella clevelandensis* and *D. pigrum*, both of which carried BGCs with genes possibly coding to produce antibiotic-like compounds. *L. clevelandensis* is a common bacterium of the anterior nares (Escapa *et al*., 2018) and was previously described as a novel genus and species within suborder Corynebacterineae (phylum Actinobacteria) in 2015 (Nicholson *et al*., 2015). Like this thesis, a recent study identified a BGC encoding proteins involved in the biosynthesis of T1PKS in the genome of *L. clevelandensis* (Leung *et al*., 2021). However, their findings do not show if this bacterium has the potential to competitively exclude *S. aureus* in the anterior nares (Leung *et al*., 2021). This study therefore contributes to the existing literature on the presence of *L. clevelandensis* in the anterior nares, suggesting its capability to exclude *S. aureus*.

D. pigrum was one of the most interesting potential excluders of *S. aureus* because its BGCs were present in most of the individuals in the non-carrier group, 5 individuals to be exact. In order to study the potential of the various *D. pigrum* BGCs that were present in non-carriers of *S. aureus* to inhibit that bacterium, the predicted BGCs in this study should also be confirmed to be predicted in the whole genome shotgun sequence of *D. pigrum* ATCC 51524 as this strain is commercially available. Our findings showed that the whole genome shotgun sequence of *D. pigrum* ATCC

51524 is predicted to carry genes involved in the production of six different bacteriocins and one polyketide.

D. pigrum, a commensal bacterium of the upper respiratory tracts, has been recently described as beneficial in the context of decreasing *Streptococcus pneumoniae* infection, offering protection against *S. pneumoniae* lung colonization (Raya Tonetti *et al*., 2021). In addition, the immunomodulatory effect of *D. pigrum* has recently been shown to reduce the severity of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-Cov-2) by inducing a differential cytokine profile (Islam *et al*., 2021). Recent animal model studies revealed that after *S. aureus* injection of *Galleria mellonella* larvae, *D. pigrum* AMBR11 could decrease mortality in this animal model (De Boeck *et al.*, 2021).

While our investigation was in progress in 2018, another group was studying the presence of BGCs predicted to carry genes that encode antibiotic-like compounds in the whole genome shotgun sequence of *D. pigrum* ATCC 51524 (Brugger *et al*., 2020). Their findings were not published until the year 2020. They hypothesized a negative association between *D. pigrum* and *S. aureus* in the nasal passage microbiome of adults (Brugger *et al*., 2020). The accessory genome of 11 *D. pigrum* strains, including *D. pigrum* ATCC 51524, revealed a diverse range of

BGCs predicted to carry genes involved in the production of candidate antibiotics (Brugger *et al*., 2020). Putative type II lanthipeptide BGCs were predicted in the genome of 4 of the 11 *D. pigrum* strains. Using antiSMASH and ClusterFinder, they managed to identify two lanthipeptides BGCs in the genome of *D. pigrum* ATCC 51524. In this thesis, the whole genome shotgun sequence of *D. pigrum* ATCC 51524 was predicted to have BGCs carrying genes involved in the production of 6 bacteriocins in addition to one polyketide.

Some *D. pigrum* bacteriocin BGCs (bacteriocin DpD, DpE, DpD and DpG BGCs) were composed of genes that code for ABC transporters. These transporters can be involved in the transport of these bacteriocins out of the bacterial cell (Zeng and Charkowski, 2021). They can also contribute to self-resistance of *D. pigrum* to the produced antibiotic (Méndez and Salas, 2001)

As predicted by antiSMASH, bacteriocin DpB BGC and bacteriocin DpC BGC belonging to *D. pigrum* (accession number MUYF01000003.1), were not like any BGC in the antiSMASH platform. This finding suggests uniqueness of these BGCs as they lack similarity to any of the BGCs in the curated MIBiG dataset of antiSMASH.

DpA BGC has the potential to carry genes that code for proteins involved in the production of a molecule related to bovicin or a novel molecule. Bovicin is a lantibiotic produced by *Streptococcus bovis* HC5 that can inhibit hyper ammonia-producing bacteria (Xavier and Russell, 2009). In addition, this bacteriocin shows a broad spectrum of activity (Mantovani *et al*., 2002). Another study demonstrated the ability of bovicin to reduce *S. aureus* adhesion to polystyrene surfaces (Pimentel-Filho *et al*., 2014). No study has reported the isolation of a compound related to bovicin from the nasal bacterium *D. pigrum*. It would be worthwhile to attempt to isolate this compound from *D. pigrum* ATCC 51524 and test its ability to eradicate *S. aureus* nasal carriage.

Moreover, DpA BGC has the potential to carry genes that code for the production of a compound related to thermophilin, or a novel compound. Thermophilin is a broad spectrum bacteriocin isolated from *Streptococcus thermophilus.* It can inhibit a range of bacteria, including *Streptococcus mutans* and *Listeria monocytogenes* (Renye *et al*., 2021).

Bioinformatics analysis also revealed that bacteriocin DpD BGC could code for the production of a class 1 bacteriocin. This class of bacteriocins are subdivided into lantibiotics which include linear peptide nisin and the globular peptide mersacidin (Yang *et al*., 2014). Moreover, analysis also

predicted a gene that can code for a lantibiotic mersacidin modifying enzyme. Mersacidin is a small lantibiotic capable of inhibiting MRSA colonizing the nasal epithelia of a mouse model by complexing to lipid II , inhibiting bacterial cell wall synthesis (Kruszewska et al., 2004). Moreover, since this BGC has a gene that codes for iron-dependent-repressor IdeR, it could be regulated by the presence of iron (Marcos-Torres *et al*., 2021). The results also suggest that bacteriocin DpD BGC, has the potential to encode for the production of a compound related to haloduracin. Haloduracin is a two-peptide lantibiotic produced by *Bacillus halodurans* (Lawton *et al*., 2007). This lantibiotic was previously discovered by genome mining and its combined use with chloramphenicol has been shown to inhibit *S. aureus*, *E. faecium* and *E. faecalis* (Danesh *et al*., 2016, Lawton *et al*., 2007).

Transcription of bacterial BGCs is controlled in a pathway-specific manner by global regulators including transcriptional repressors (Covington *et al*., 2021). Bacteriocin DpE BGC has a gene that codes for a transcriptional repressor YvoA which can control transcription of bacteriocin DpE. As an attempt to activate this possibly cryptic BGC, deleting this repressor will derepress expression of this BGC and initiate the production of messenger RNA. This method of repressor inactivation by gene deletion has been employed previously for induction of several metabolites from otherwise-silent BGCs (Sidda et al., 2014). Moreover, bacteriocin DpE BGC, carries

genes that can code for the production of a propionicin-like bacteriocin. Propionicin is a bacteriocin isolated from *Propionibacterium* spp., including *Propionibacterium freudenreichii* (Brede *et al*., 2004). A propionicin-like compound has not yet been isolated from a bacterium residing in the anterior nares.

D. pigrum polyketide DpA BGC has a gene that code for the production of a β-ketoacyl ACP synthase. This enzyme can catalyze the clasein condensation between acetyl CoA and malonyl ACP (Sachdeva and Reynolds, 2008). It is also a key enzyme that initiates fatty acid biosynthesis ACP (Sachdeva and Reynolds, 2008). The predictions in this chapter cannot recognize whether the product of this BGC is a polyketide or a fatty acid, since the production of both share a similar enzymatic domain structure (Jenke-Kodama *et al*., 2005). Moreover, bioinformatics analysis predicted the product of this BGC to be related to xantholipin, a polycyclic antibiotic produced by *Streptomyces* species with strong antimicrobial effects against MRSA (Wu *et al*., 2017).

The negative association between *D. pigrum* and *S. aureus* in the anterior nares can offer the potential for development of novel therapeutic approaches. In the next chapter the capacity of *D. pigrum* ATCC 51524 to

inhibit *S. aureus* will be discussed, since genome data suggests this strain has the BGCs identified in the metagenome of non *S. aureus* carriers.

Chapter 7
Inhibition of *Staphylococcus aureus* JE2 by *Dolosigranulum pigrum* ATCC 51524

7.0 Inhibition of *Staphylococcus aureus* JE2 by *Dolosigranulum pigrum* ATCC 51524

7.1 Introduction

7.1.1 Initial isolation and identification of *Dolosigranulum pigrum*

Dolosigranulum was first isolated in 1988 at the Microbiology Department of Colchester General Hospital (Aguirre *et al*., 1993). It was isolated from spinal cord tissue of a newly diagnosed case of multiple sclerosis. A bacterium isolated in this study was catalase negative, Gram positive cocci 1.0-1.5 µm in size arranged in clumps. At the time, the phenotypic profile of this bacterium was not compatible with any bacteria on the analytical profile index (API) database but was closest to *Gemella haemolysan* (Aguirre *et al.*, 1993).

In 1991, *Dolosigranulum* was isolated again from a patient with neurotropic cornea at the RAF Institute of Pathology and Tropical Medicine, Halton, Aylesbury (Hall *et al*., 2001). At this point, the bacterium was determined to be phylogenetically closest to *Aerococcus* and *Globicatella* but was classified as a new genus, *Dolosigranulum* and named *Dolosigranulum pigrum* gen.nov.,sp nov. (Aguirre *et al*., 1993).

7.1.2 Phenotypic characteristics of *Dolosigranulum pigrum*

Dolosigranulum pigrum is a Gram positive, facultatively anaerobic bacterium from the phylum Firmicutes. The cells of this bacterium mostly appear in pairs, tetrads, and groups (Aguirre *et al*., 1993, Hall *et al*., 2001). Belonging to the family of Carnobacteriaceae, it is a non-spore forming, non-motile bacterium (Lecuyer *et al*., 2007).

7.1.3 Colonization of the anterior nares by *D. pigrum*

The anterior nares is a complex ecosystem characterized by the presence of a diverse microbial community, with *S. aureus* being one of the colonizers (Liu *et al*., 2015). *D. pigrum* is a colonizer of the human nasal cavity, functioning as a commensal bacterium (De Boeck *et al*., 2017). However, its association with other nasal bacteria, including *S. aureus*, is yet to be explored. *D. pigrum* has been suggested to limit the colonization of *S. aureus* in the anterior nares (Liu *et al*., 2015), with a recent study proposing the inhibition of *S. aureus* growth by *D. pigrum in vitro* (Brugger *et al*., 2020). In agreement with these studies, previous bioinformatics analysis in Chapter 6 also suggests the capability of *D. pigrum* to exclude *S. aureus* in the anterior nares through the production of bacteriocins and a polyketide. Despite this, more research is needed to explore the potential mechanisms of *S. aureus* inhibition by *D. pigrum.*

7.2 Aims and objectives of this chapter

The aim of this research was to evaluate whether the commensal bacterium of the anterior nares, *Dolosigranulum pigrum,* was able to inhibit *S. aureus* JE2 and seek insight into the underlying inhibitory mechanism. The objectives were:

- To assess the inhibitory activity of *Dolosigranulum pigrum* ATCC 51524 grown in different media against *Staphylococcus aureus* JE2 using agar well diffusion assays.
- To increase the ability of *Dolosigranulum pigrum* ATCC 51524 to produce antibiotics by isolating rifampicin- and streptomycin-resistant *D. pigrum* mutants and evaluate their antimicrobial activity in comparison to that of the wild-type using agar well diffusion assays.

7.3 Materials and methods

7.3.1 Inhibition of *Staphylococcus aureus* JE2 by colonies of *Dolosigranulum pigrum* ATCC 51524

Bacterial stock of *D. pigrum* ATCC 51524 was prepared by streaking glycerol stock onto Todd Hewitt (TH) agar (Sigma, UK). The plates were incubated at 37°C for 48 hours until pinpoint *D. pigrum* ATCC 51524 colonies were formed. A standard *S. aureus* JE2 culture was prepared as described in Chapter 2 section 2.1.1, except the culture was adjusted to OD_{600nm} of 0.1. A lawn of this culture was prepared onto TH agar as described in Chapter 2 section 2.2.1.

One colony of *D. pigrum* ATCC 51524 was gently dotted and spread onto these plates using a 1 µl sterile plastic loop. The plates were incubated for 18 hours at 37°C. If inhibition was present around *D. pigrum* ATCC 51524 colonies after incubation, it might indicate the production of antimicrobial compounds. Three biological replicates were used for this assay originating from three agar plates that had the *D. pigrum* colonies.

7.3.1.1 Small scale preparation of *Dolosigranulum pigrum* ATCC 51524 bacteria and supernatant

This assay was done to identify if the antimicrobial activity of *D. pigrum* ATCC 51524 colonies was due to the bacteria or the supernatant.

The medium used for initial screening of antimicrobial activity by *D. pigrum* ATCC 51524 was TH broth (VWR, USA). Five *D. pigrum* colonies were inoculated in 10mL TH broth in a 50 mL tube. This bacterial culture was incubated at 37°C for 22 hours in a shaking incubator at 200 rpm. After 22 hours, the absorbance (OD_{600nm}) was adjusted to 0.01 and the culture was incubated in a 200 rpm shaking incubator at 37°C. After 16, 22, 24 and 30 hours incubation, 200µl of culture was collected. First, 100 µl of bacterial cultures were centrifuge for 5 minutes at 4193 xg and then were filtered and collected using 0.22 µm pore size syringe filters (Fisher Scientific). These filtered supernatants and the remaining 100 µl bacteria were tested for inhibitory activity against *S. aureus* JE2 using agar well diffusion assays as described in Chapter 2 section 2.2.1. The agar well diffusion assay plates were incubated at 37°C.

7.3.1.2 Screening for inhibitory activity of the supernatant of *D. pigrum* grown in 10 different media using agar well diffusion assays

Five *D. pigrum* ATCC 51524 colonies were inoculated in 10mL media in a 50 mL tube. The 10 media used in this assay were ; 1) TH broth, 2) TH broth with 200 µM 2, 2'-dipyridyl, 3) TH broth supplemented with 0.5% yeast, 4) TH broth supplemented with 0.5% yeast and 200 µM 2, 2'-dipyridyl, 5) Nutrient broth (Sigma, UK), 6) Nutrient broth with 200 µM 2, 2'-dipyridyl, 7) TS broth (Sigma, UK), 8) TS broth with 200 µM 2, 2'-dipyridyl, 9) M9 minimal medium (Sigma, UK), 10) M9 minimal medium with 200 µM 2, 2'-dipyridyl. The iron chelator 2, 2'-dipyridyl was added to some of the growth media to initiate iron starvation in the bacterial culture and enhance antibiotic production by *D. pigrum.*

D. pigrum ATCC 51524 bacterial cultures were incubated at 37°C for 22 hours in a shaking incubator at 200 rpm. After 22 hours, the absorbance (OD600nm) was measured and was adjusted to 0.01. 200µl of culture was collected after 16, 22, 24 and 30 hours incubation. Filtered supernatants were prepared as described in section 7.3.1.1 of this chapter. These filtered supernatants and the remaining 100 µl of bacterial suspension were tested for inhibitory activity against *S. aureus* JE2 using agar well diffusion assays as described in Chapter 2 section 2.2.1. The agar well diffusion assay plates were incubated at 37°C.

7.3.1.3 Inhibitory activity of highly concentrated *D. pigrum* ATCC 51524 supernatant against *S. aureus* JE2.

The aim of this assay was to identify the optimum time for antimicrobial activity production in supernatants of *D. pigrum* ATCC 51524 grown in 200mL Todd Hewitt broth prepared in 1L bottles when concentrated by 20 times.

Five *D. pigrum* ATCC 51524 colonies were inoculated in TH broth 10mL in a 50 mL tube. The tubes were incubated at 37°C in a shaking incubator at 200 rpm for 22 hours. After 22 hours, the *D. pigrum* ATCC 51524 inoculum was used to adjust the absorbance of two 1 liter bottles containing 200mL Todd Hewitt broth to a starting of OD_{600nm} of 0.01. At time points 0, 2, 4, 6, 16, 18, 20, 22, 24, 26, 28, 30, 40, 42, 44, 46, 48, 50, 52 and 54 hours, 2 mL of bacterial culture was collected. These cultures were centrifuged for 15 minutes at 4193 xg. The top layer of supernatants were collected and filtered using 0.22 µm pore size syringe filters (Fisher Scientific). These filtered supernatants were concentrated twenty times (20X) using a SpeedVac vacuum concentrator set at 5.1 vacuum pressure for 2 hours. The concentrated supernatants were filtered at 4193 xg for 2 minutes using 0.22 µm centrifuge filters and tested for antimicrobial activity against *S. aureus* JE2 using agar well diffusion assays following Chapter

2 section 2.2.1. Due to the thick consistency, the 20X concentrated supernatants were pipetted into the wells and were left undisturbed for 2 hours at room temperature to allow maximum absorption into the agar wells before incubation for 18 hours at 37°C. This assay was done using three biological replicates.

7.3.1.4 Treatment of *Dolosigranulum pigrum* ATCC 51524 bacteria with catalase

200 mL of 30-hour *D. pigrum* ATCC 51524 was prepared in a 1 Liter bottle to collect the bacterial pellet and a large volume of supernatant. This time point was chosen because *D. pigrum* ATCC 51524 bacteria showed highest inhibitory activity after 30-hours incubation. The bacterial culture was prepared as described in section 7.3.1.3 of this Chapter. The bacterial pellets were collected after 30 hours of incubation and the supernatant was collected and filtered.

The pellets were resuspended in 5mL phosphate-buffered saline (PBS). The *D. pigrum* ATC 51524 resuspended pellet (1 mL) was disrupted for 5 minutes by sonication using an ultrasonic disintegrator UP200Ht, 200 watts 26 kHz. The 5-minute sonication process was performed by repeating 5 sessions of sonication for 1 minute with resting on ice for 1 minute in-between sessions. The sonicated *D. pigrum* ATCC 51524 lysate

(50 µl) was treated with 50 µl 1mg/ml catalase for 2 hours at room temperature. After 2 hours, the catalase-treated *D. pigrum* lysate (100 µl) was tested for inhibitory activity against *S. aureus* JE2 using agar well diffusion assays following Chapter 2, part 2.2.1. Another sample of sonicated *D. pigrum* was tested without treatment with catalase. This untreated control was tested using potassium phosphate instead of catalase. If the antimicrobial activity of the enzyme treated lysate did not form margins of inhibition, it suggests that growth inhibition was due to the action of the inhibitory compound hydrogen peroxide (Mélançon and Grenier, 2003). If the catalase treated lysate produced antimicrobial activity, it suggests that the antimicrobial is not affected by catalase, which might suggest that the *D. pigrum*-associated antimicrobial is causing inhibition of *S. aureus* JE2.

7.3.2 Activation of cryptic biosynthetic gene clusters using ribosomal engineering

7.3.2.1 Determination of MICs

The minimum inhibitory concentration of rifampicin and streptomycin against *D. pigrum* ATCC 51524 were determined using broth microdilution assay as described in Chapter 2 section 2.2.2. A range of concentrations of streptomycin and rifampicin were used in this assay (1-64 ug/ml).

7.3.2.2 Isolation of rifampicin-resistant and streptomycin-resistant *Dolosigranulum pigrum* ATCC 51524 mutants.

Five colonies of *D. pigrum* ATCC 51524 were inoculated in 10 mL TH broth in a 50 mL tube. The inoculum was incubated for 22 hours at 37°C while shaking at 200 rpm. TH broth agar plates containing either 10 ug/ml rifampicin, 10 ug/ml streptomycin, 50 ug/ml rifampicin and 50 ug/ml streptomycin were used. 100 ul of the 22- hour *D. pigrum* ATCC 51524 culture was spread onto the agar plates using an L-shaped cell spreader. The plates were incubated at 37°C for 72 hours and the presence of resistant colonies was observed. Once colonies were seen, cultures of the resistant *D. pigrum* ATCC 51524 colonies were grown in 10 mL TH broth and were used to prepare a bacterial stock.

7.3.2.3 Initial screening of rifampicin and streptomycin resistant *Dolosigranulum pigrum* to inhibit *S. aureus* JE2.

D. pigrum rifampicin or streptomycin resistant colonies isolated in part 7.3.2.2 were assessed for inhibitory activity against *S. aureus* JE2. Each mutant *D. pigrum* colony was inoculated in 10 mL TH broth in a 50 mL tube. The supernatants in this assay were collected after 16, 22, 24, 30, 40, 46, 48 and 54 hours incubation. The supernatants were prepared and

tested for inhibitory activity against *S. aureus* JE2 following section 7.3.1.3 of this chapter.

7.4 Results

7.4.1 Initial detection of *Staphylococcus aureus* JE2 inhibition by *Dolosigranulum pigrum* ATCC 51524

The formation of clear zones around *D. pigrum* ATCC 51524 colonies indicated the inhibition of *S. aureus* growth (Figure 7.1).

Figure 7.1 Inhibition *of S. aureus* JE2 by *D. pigrum* ATCC 51524 bacterial colonies.

Six single *D. pigrum* ATCC 51524 colonies were tested using duplicate colonies originating from three biological replicates. Three biological replicates from three different agar plates (agar plate A, C and D) with *D. pigrum* ATCC 51524 colonies were used. WTA; wild-type *Dolosigranulum pigrum* ATCC 51524 colony originating from plate A, WTC; wild-type *Dolosigranulum pigrum* ATCC 51524 colony originating from plate C, WTD; wild-type *Dolosigranulum pigrum* ATCC 51524 colony originating from plate D.

7.4.2 Screening for inhibitory activity of *Dolosigranulum pigrum* ATCC 51524 grown in 10 different media using agar well diffusion assays

D. pigrum ATCC 51524 did not grow in nutrient broth ± 200 µM 2, 2'-dipyridyl, TS broth ± 200 µM 2, 2'-dipyridyl and M9 minimal medium ± 200 µM 2, 2'-dipyridyl. *D. pigrum* ATCC 51524 was able to grow in TH broth ± 200 µM 2, 2'-dipyridyl and TH broth supplemented with 0.5% yeast ± 200 µM 2, 2'-dipyridyl, but the supernatants did not inhibit *S. aureus* JE2 using agar well diffusion assays.

D. pigrum ATCC 51524 grown in TH broth ± 200 µM 2, 2'-dipyridyl and TH broth supplemented with 0.5% yeast ± 200 µM 2, 2'-dipyridyl was able to inhibit *S. aureus* JE2. None of the supernatants collected were able to inhibit *S. aureus* JE2 (Table 7.1).

Table 7.1 Inhibitory activity of *D. pigrum* ATCC 51524 bacteria and supernatant grown in 4 different media.

Medium used to grow *D. pigrum* ATCC 51524	Tested sample	Margins of inhibition (mm)			
		16h	22h	24h	30h
THB	**Bacteria**	6 (5-7)	9 (7-10)	8 (8)	5 (3-6)
	Supernatant	0 (0)	0 (0)	0 (0)	0 (0)
THB + 2,2 dipyridyl	**Bacteria**	4 (4)	6 (4-6)	6 (5-8)	4 (3-5)
	Supernatant	0 (0)	0 (0)	0 (0)	0 (0)
THBY	**Bacteria**	7 (7)	7 (6-9)	6 (6-7)	7 (6-8)
	Supernatant	0 (0)	0 (0)	0 (0)	0 (0)
THBY + 2,2 dipyridyl	**Bacteria**	12 (12-13)	10 (7-10)	12 (12)	11 (10-12)
	Supernatant	0 (0)	0 (0)	0 (0)	0 (0)

This table represents the inhibitory activity of *D. pigrum* ATCC 51524 bacteria and supernatant when grown in different media. Experiments were done using triplicates for three independent biological repeats. Values are represented as median(range) and were measured in millimeters (mm). h; hour, THB; Todd Hewitt broth, THB+2,2 dipyridyl; Todd Hewitt broth with 200 µM 2, 2'-dipyridyl, THBY; Todd Hewitt broth supplemented with 0.5% yeast, THBY + 2,2' dipyridyl; Todd Hewitt broth supplemented with 0.5% yeast and 200 µM 2, 2'-dipyridyl. Other media were not included in this table because *D. pigrum* ATCC 51524 did not grow in them.

7.4.3 Inhibitory activity of *D. pigrum* ATCC 51524 supernatant and bacteria grown in 1 Liter Todd Hewitt broth against *S. aureus* JE2

Results of this assay showed that *D. pigrum* ATCC 51524 bacteria, grown in Todd Hewitt broth, harvested at 22-46 hours, was able to inhibit the growth of *S. aureus* JE2 (Figure 7.2).

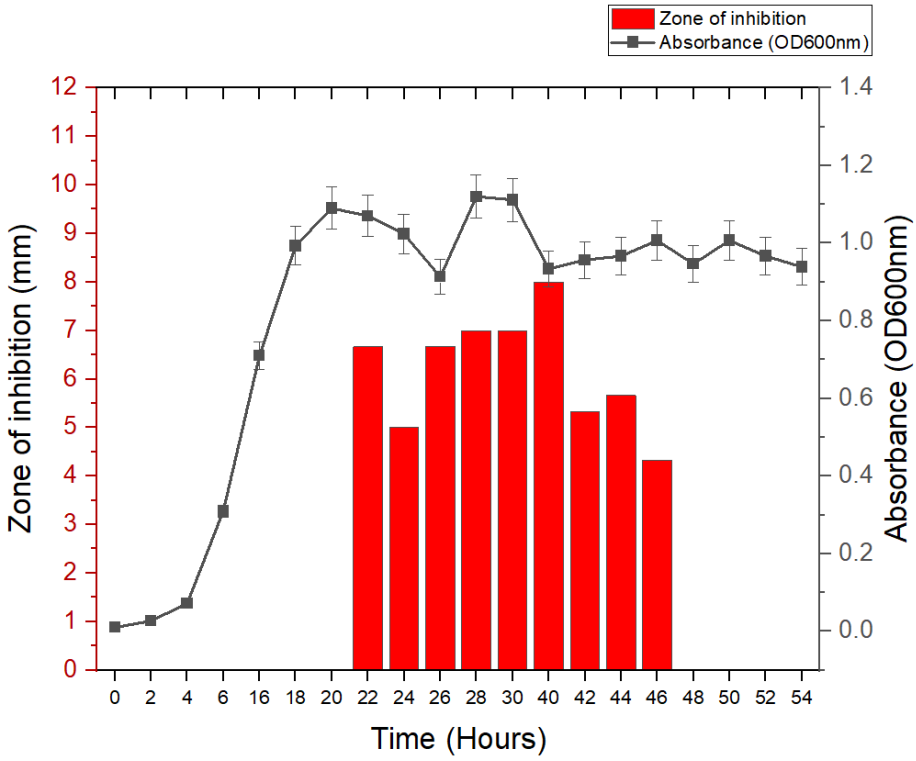


Figure 7.2 Inhibition of *S. aureus* JE2 by the bacterial culture of *D. pigrum* ATCC 51524 grown for 0,2,4,6,16,18,20,22, 24, 26, 28, 30, 40, 42, 44, 46, 48, 50, 52, and 54 hours.

The non-concentrated *D. pigrum* ATCC 51524 supernatants were unable to inhibit the growth of *S. aureus* JE2. Therefore, the supernatants were concentrated by 20X using a vacuum concentrator. The concentrated supernatant harvested at 26, 30 and 42 hours were able to inhibit *S. aureus* (Figure 7.3).

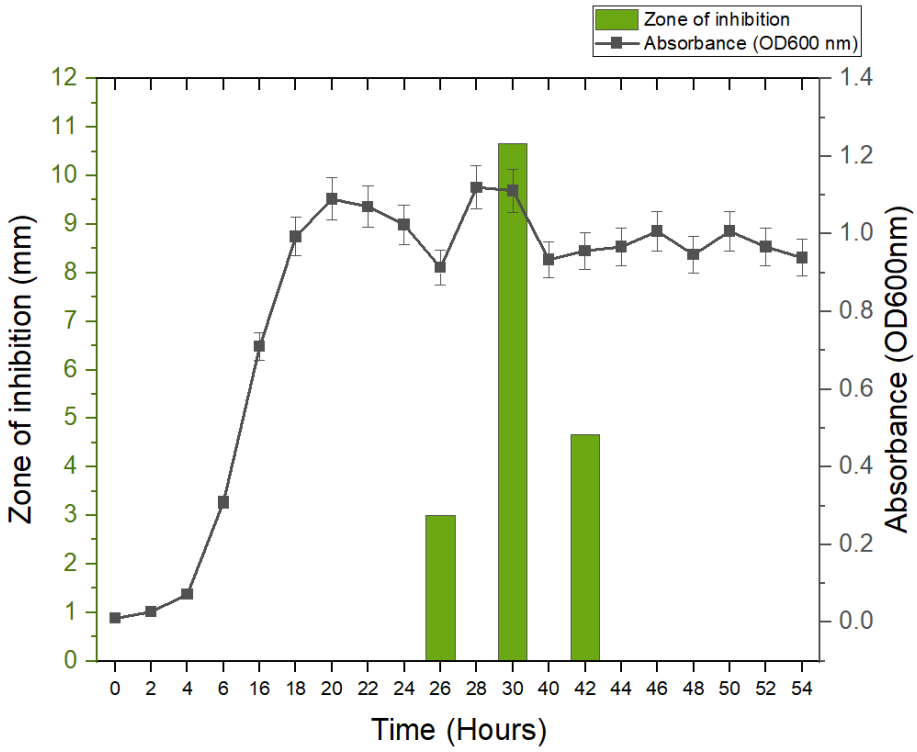


Figure 7.3 Inhibition of *S. aureus* JE2 by 20X concentrated supernatants of *D. pigrum* ATCC 51524 grown for 0,2,4,6,16,18,20,22, 24, 26, 28, 30, 40, 42, 44, 46, 48, 50, 52 and 54 hours.

7.4.4. Antimicrobial effect of catalase-treated *D. pigrum* ATCC 51524 against *S. aureus* JE2

Catalase (1 mg/ml) was used to exclude any antimicrobial activity due to hydrogen peroxide. Complete inactivation of antimicrobial activity of sonicated *D. pigrum* ATCC 51524 lysate was observed after treatment with 1 mg/ml catalase (Table 7.2). This suggests that hydrogen peroxide may have been the cause of antimicrobial activity exhibited by *D. pigrum* ATCC 51524 bacteria.

Table 7.2. Treatment of sonicated *D. pigrum* ATCC 51524 lysate with catalase.

Treatment	Content of well	Margins of inhibition (mm)
1	*D. pigrum* without treatment with catalase (1 mg/ml)	6 (6-7)
2	*D. pigrum* after treatment with catalase (1 mg/ml)	0 (0)
3	Control 1	0 (0)
4	Control 2	0 (0)

Control 1; Mixture of 50 µl 1mg/mL catalase + 50µl PBS, Control 2; 50µl potassium phosphate+ 50µl PBS, mm; millimeters. Values represent one experiment with three biological replicates and are expressed as median (range)

7.4.5 Isolation of rifampicin-resistant and streptomycin-resistant *Dolosigranulum pigrum* ATCC 51524 mutants.

After ruling out inhibition of *S. aureus* JE2 by the non-concentrated supernatant of *D. pigrum* ATCC 51524, ribosomal engineering was used to enhance antimicrobial production.

The MIC of rifampicin and streptomycin against *D. pigrum* ATCC 51524 was 1 µg/ml and 2 µg/ml respectively. 26 *D. pigrum* ATCC 51524 colonies that were resistant to either 50 µg/ml rifampicin, 50 µg/ml streptomycin, 10 µg/ml rifampicin or 10 µg/ml streptomycin were isolated.

Nine *D. pigrum* ATCC 51524 colonies were resistant to 50 µg/ml rifampicin, while two colonies were resistant to 50 µg/ml streptomycin. In addition, 13 *D. pigrum* ATCC 51524 colonies were resistant to 10 µg/ml rifampicin while two *D. pigrum* ATCC 51524 colonies were resistant to 10 µg/ml streptomycin. Bacterial stocks were prepared for each mutant strain.

7.4.6 Small scale screening of rifampicin and streptomycin resistant *Dolosigranulum pigrum* for production of antimicrobials against *S. aureus* JE2.

Out of the 26 mutants isolated in part 7.3.2.2 of this chapter, 19 were screened for inhibitory activity against *S. aureus* JE2 using agar well

diffusion assays. Eight out of the 19 mutants were able to inhibit S*. aureus* JE2.

Most of the mutant's supernatants harvested at 22 hours produced inhibitory activity (Table 7.3). The mutant DP8rif10 produced inhibitory activity at 22, 24, 30, 40 and 46 hours. Highest inhibitory activity by the supernatant of *D. pigrum* DP8rif10 was observed at 30 hours.

Table 7.3. Inhibitory activity of the supernatants of mutant *D. pigrum* collected at different time points.

Name of mutated *D. pigrum*	Antibiotic used for selection	Concentration (µg/ml) used for selection	Margins of inhibition formed by supernatant (mm) collected at different timepoints							
			16h	**22h**	**24h**	**30h**	**40h**	**46h**	**48h**	**54h**
DP4rif50	rifampicin	50	0(0)	3(3-4)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
DP5rif50	rifampicin	50	0(0)	6(5-6)	3(4-3)	0(0)	0(0)	0(0)	0(0)	0(0)
DP6rif50	rifampicin	50	0(0)	6(5-6)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
DP8rif50	rifampicin	50	0(0)	4(4-6)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
DP5rif10	rifampicin	10	0(0)	5(3-6)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
DP8rif10	rifampicin	10	0(0)	7(6-8)	6(3-9)	10(8-12)	4(4-5)	5(4-6)	0(0)	0(0)
DP1strept50	streptomycin	50	0(0)	4(3-4)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
DP2strept10	streptomycin	10	0(0)	5(5-7)	0(0)	0(0)	0(0)	0(0)	0(0)	0(0)
Control	-	-	0(0)	0(0)	0(0)	9(6-12)	0(0)	0(0)	0(0)	0(0)

This table represents the inhibitory activity of the supernatants of *D. pigrum* mutants. Experiments were done using three biological replicates on three different days. Values represent one experiment with three biological replicates and are expressed as median (range). h; hour, mm; millimeter, DP; *Dolosigranulum pigrum*, rif; rifampicin, strept; streptomycin, control; the supernatant of *D. pigrum* ATCC 51524 harvested at 30 hours and concentrated 20 times using a vacuum concentrator, -; not applicable to the control.

Catalase (1 mg/ml) was used to treat the 30-hour supernatant of *D. pigrum* DP8rif10 to exclude hydrogen peroxide as the cause of antimicrobial activity as described in part 7.3.1.5 of this chapter. The supernatant was still able to inhibit *S. aureus* JE2 even after treatment with catalase (1 mg/ml), ruling out the inhibitory effect of hydrogen peroxide (Table 7.4).

Table 7.4 Treatment of *D. pigrum* DP8rif10 30-hour supernatant with catalase.

Treatment	Content of well	Margins of inhibition (mm)
1	Non-treated DP8rif10 supernatant	10(9-10)
2	Treated DP8rif10 supernatant	10 (10-12)
3	Control 1	0 (0)
4	Control 2	0 (0)

Treated supernatant; *D. pigrum* DP8rif10 30-hour supernatant treated with 1 mg/ml catalase, Non-treated supernatant; *D. pigrum* DP8rif10 30-hour supernatant without catalase treatment. Control 1; Mixture of 50 µl 1mg/mL catalase + 50µl Todd Hewitt broth, Control 2; 50µl potassium phosphate+ 50µl Todd Hewitt broth, mm; millimeters. Values represent one experiment with three biological replicates and are expressed as median (range).

7.5 Discussion

Dolosigranulum pigrum has emerged in a few studies as a protective bacterium against colonization by *S. aureus* in the anterior nares (Bosch *et al*., 2016, Brugger *et al*., 2020). This thesis, along with other studies demonstrating the inverse relationship between *S. aureus* and *D. pigrum* (Brugger *et al*., 2020), signify that the microbial community in the anterior nares has the potential to produce new secondary metabolites with antimicrobial properties. In the previous bioinformatics chapter, Chapter 6, a possible negative association between *D. pigrum* and *S. aureus* in the anterior nares was identified. Consistent with these results, analysis of 11 *D. pigrum* genomes by Brugger *et al.* (2020) has also shown that these genomes contain BGCs that carry genes potentially encoding the production of putative bacteriocins including lanthipeptides, as well as bactericidal proteins. It has been proposed by Brugger *et al.* (2020) that *D. pigrum* might use many mechanisms to inhibit the growth of *S. aureus*. Interestingly, they found that the genome of *D. pigrum* strains that contained both a lanthipeptide BGC and a bacteriocin BGC strongly inhibited *S. aureus* growth *in vitro* (Brugger *et al*., 2020).

In this study, production of hydrogen peroxide was one mechanism that inhibited *S. aureus* and could explain the negative association between *D. pigrum* ATCC 51524 and *S. aureus*. This study cannot confirm that *S. aureus* was killed by the supernatant of *D. pigrum* ATCC 51524 because

it was only active when highly concentrated. This could be due to the low amount of antimicrobial compound in the supernatant or more likely due to poorly expressed BGCs (Tanaka *et al*., 2013) in the whole genome shotgun sequence of *D. pigrum* ATCC 51524. Lack of secretion, by which the antimicrobials from *D. pigrum* are released by a secretory system, which bacteria employ to move substrates across their cell membrane to the external environment, is another reason for inactivity (Naskar *et al*., 2021). The growth conditions used in this study might have not promoted the expression of the secretory-system-related genes found in the BGCs.

Results from small scale screening revealed that adding the iron chelator 2,2’ dipyridyl, did not enhance the production of antimicrobials produced by *D. pigrum* ATCC 51524.This finding indicates that iron deprivation while *D. pigrum* was growing in broth does not affect the antimicrobial production of this bacterium suggesting that *D. pigrum* antimicrobial activity against *S. aureus* is not linked to iron availability (Nguyen *et al*., 2015).

The addition of 0.5 % yeast extract as a nitrogen source played an important role in growth of *D. pigrum* ATCC 51524, but not for antimicrobial production. Even though other studies revealed that addition of yeast extract influenced antibiotic production by other bacteria such as *Bacillus licheniformis*, adding yeast to growth medium did not show a similar result

(Anthony *et al*., 2009). This is the first study to assess the antimicrobial activity of *D. pigrum* ATCC 51524 when grown in different medium. Small scale screening results from this chapter are therefore very useful to optimize culture conditions of *D. pigrum* ATCC 51524 to produce antimicrobial compounds.

An attempt to activate potentially poorly expressed secondary metabolite BGCs was implemented with the use of ribosomal engineering . One interesting mutant in this study, *D. pigrum* DP8rif10, was able to inhibit *S. aureus* JE2 without the need to concentrate the supernatant. This approach might have activated some cryptic secondary metabolite BGCs to produce increased amounts of antimicrobial(s) (Tanaka *et al*., 2013) or more interestingly, might have led to the production of novel antimicrobials (Hosaka *et al*., 2009). Further work is required to confirm if the *rpoB* mutation is responsible for enhanced antibiotic production.

In summary, this study supports the hypothesis that *D. pigrum* can inhibit the growth of *S. aureus.* Furthermore, this Chapter shows improved antimicrobial production by *D. pigrum* ATCC 51524 through the selection of rifampicin resistance and thus, the possibility of extracting the compound(s) responsible for antimicrobial activity could be possible in the future. To date, no antimicrobial secondary metabolites produced by *D.*

pigrum have been isolated. Brugger *et al.* (2020) have tried to identify *S. aureus* secondary metabolites using bioassay-guided fractionation, but their approach failed. It will therefore be interesting to isolate and extract the active antimicrobial compound(s) in the supernatant of *D. pigrum* DP8rif10 using solvents and fractionation. The work described in this chapter will help future studies aimed at investigating the therapeutic potential of *D. pigrum* as a producer of novel antimicrobials or as a probiotic.

Chapter 8.0

General discussion and future directions

8.0 General discussion and future directions

8.1 General discussion

The emergence of antibiotic resistant bacteria has caused a serious burden on healthcare settings (Founou *et al*., 2017). According to a recent thorough review, antimicrobial resistance is a prominent cause of death worldwide, with the highest burdens in low-resource settings (Murray *et al*., 2022).

Diseases caused by antibiotic resistant bacteria are estimated to cause 10 million deaths each year by 2050 (O'Neil, 2014). This estimate was based on the Review on Antimicrobial Resistance commissioned by the UK Government (O'Neil, 2014). However, it has been scrutinized, as the data was based on selected pathogens, including *E. coli*, *K. pneumoniae* and *S. aureus*. They were derived from the European Antimicrobial Resistance Surveillance network (EARS-Net), which is not a population-based surveillance network as it only records invasive infections diagnosed in hospitals (de Kraker *et al*., 2016). Moreover, they criticised their use of crude mortality proportions or adjusted odds ratios (ORs) for patients with resistant infections versus patients with susceptible infections. These ORs are not easily transformed to attributable mortality proportions since they were expressed due to a specific antimicrobial resistance profile of a particular pathogen, in addition to the effect of therapy timing, and host-

and pathogen-related factors (de Kraker *et al*., 2016). Even though this estimate has been criticised, there is a general agreement that spread of antimicrobial resistance is a global threat that needs to be urgently addressed (Guest *et al*., 2020).

In addition to the rise in infections caused by antibiotic resistant bacteria, there is a decline in the global pipeline of antibiotics in development (Theuretzbacher *et al*., 2020). Even though a few antibiotics are currently in Phase III clinical trials, few of them are active against Gram negative bacteria (PEW,2021). To approach this problem, the focal point of this research was aimed to discover new antibiotics from two under-explored sources; bacteria from of the anterior nares and water samples. The studies presented in this thesis focused on 1) the inhibitory activity of five NW *Pseudomonas* sp. strains isolated from fresh water against the ESKAPE pathogens, 2) the enhancement of antibiotic production by *Pseudomonas* sp. strain NW27 using ribosomal engineering approaches, 3) the extraction of antimicrobial compounds in the supernatant of *Pseudomonas* sp. strain NW27-A1, 4) the use of IMG/M database to identify if the metagenome of *S. aureus* non-carriers carry genes with the potential to code for the production of antimicrobials involved in the exclusion of *S. aureus* in the anterior nares and 5) the enhancement of antimicrobial activity against *S. aureus* produced by *D. pigrum* ATCC 51524 using ribosomal engineering.

Many soil-dwelling bacteria are already used commercially, whereas research on antibiotic-producing bacteria in water is less extensive (Tawiah *et al*., 2012). Antibiotic-producing *Pseudomonas* species have been previously isolated from soil and plants (Mitsutomi *et al*., 2017, Rio *et al*., 1972). Cepacidine A, is one of the novel antifungal agents isolated from the plant pathogen *Pseudomonas cepacian* (Lee et al., 1994). However, novel antibiotics produced by *Pseudomonas* species isolated from water environments are not isolated yet. Moreover, novel antibiotics have been discovered in previously unexplored human microbial habitats, such as the anterior nares. There have been discoveries of novel antibiotics, such as lugdunin isolated from *S. lugdunensis* in this untapped environment, emphasizing the importance of these habitats as unexplored sources of novel natural products (Zipperer *et al*., 2016).

In Chapter 3, the antimicrobial activity of NW *Pseudomonas* sp. strains isolated from water were determined. In agreement with other studies, this research demonstrated the ability of one strain many compounds (OSMAC) approach to affect secondary metabolite production by microbial strains (Romano *et al.*, 2018). Many NW *Pseudomonas* sp. strains supernatants prepared using different conditions were able to elicit antimicrobial activity against indicator bacteria. Some of the supernatants of *Pseudomonas* sp. strains NW7, NW9, NW10 and NW27 were inhibitory to more than one indicator bacterium. Moreover, some BGCs were

common in the genomes of *Pseudomonas* sp. strain NW7 and NW9, including 2,4-DAPG BGC and sessilin A BGC. 2,4-DAPG works by disrupting the cell membrane permeability, triggering a reactive oxygen burst, leading to interruption of cell haemostasis (Troppens *et al*., 2013). Moreover, researcher have found that this compound causes the loss of mitochondrial membrane potential in eukaryotic cell, demonstrating its possible toxicity (Troppens *et al*., 2013). However, development of resistance to this compound is not know yet.

The supernatants of *Pseudomonas* sp. strains NW16 were only able to inhibit *K. pneumoniae,* demonstrating the narrow antimicrobial activity of the compound(s) in these supernatants. Using different culture conditions, it was shown that the supernatant of *Pseudomonas* sp. strain NW27 grown in LB broth for 48 hours at 25°C was producing surfactant(s) and was able to inhibit the growth of *S. aureus*, *E. faecium* and *A. baumannii*. This was an interesting finding because the supernatant was active against *A. baumannii*, one of the top pathogens in the global priority list of antibiotic-resistant bacteria (WHO, 2017). Moreover, *A. baumannii* was recently categorised as an urgent threat by the CDC (CDC, 2019), emphasizing the need for options to treat *A. baumannii*-associated infections.

Moreover, antiSMASH computationally predicted many BGCs in the whole genome shotgun sequence of *Pseudomonas* sp. strain NW27. Some of these were likely to encode for secondary metabolites that were related to pyoverdine, lankacidin C, arylpolyene, fengycin, viscosin, bananamide and syringomycin. In addition, the metabolic product of 6 other BGCs were too dissimilar to known BGCs for the antiSMASH to suggest similar products. Since genomic data suggested the presence of unique BGCs, antibiotic-producing potential of this bacterial strain was investigated. Its activity was further enhanced using ribosomal engineering techniques. This study was conducted as previous studies demonstrated the enhanced antimicrobial activity of rifampicin-resistant bacterial mutants (Ochi *et al*., 2017, Xie *et al*., 2016). In this thesis an interesting rifampicin-resistant mutant was isolated, named *Pseudomonas* sp. strain NW27-A1, which demonstrated enhanced antimicrobial activity against *S. aureus*, *E. faecium* and *A. baumannii*. Like other studies, rifampicin-selected mutations could have activated the weakly expressed BGCs in the genome of *Pseudomonas* sp. strain NW27 (Beck *et al*., 2021). It could also suggest the production of compound(s) at a higher concentration than the wild type of *Pseudomonas* sp. strain NW27 strain. The results at this point were promising enough to proceed with extraction methods, where the antimicrobial compounds were extracted from the supernatant of *Pseudomonas* sp. strain NW27-A1. Even though acetonitrile extracts were successful in inhibiting *S. aureus* and *E. faecium*, the extraction method

was not successful at extracting the antibacterial compound(s) responsible for inhibiting *A. baumannii*. In addition, multiple fractions were actively reducing cellular metabolism of *E. faecium*.

As described in Chapter 6, a bioinformatics-based technique was used to explore the anterior nares to discover other antibiotic-producing bacteria. Several bacteria including *D. pigrum* and *Corynebacterium* spp., had BGCs that have the potential to encode proteins involved in the production of secondary metabolites with antimicrobial properties. Due to minimal information available in existing literature at the time, the major finding in this part of the research was the ability of *D. pigrum* to potentially produce antimicrobials that exclude *S. aureus* in the anterior nares. At the time of this research, the whole genome shotgun sequence of *D. pigrum* (ATCC 51524) was publicly available and findings in this study showed the presence of BGCs that have the potential to encode for six different bacteriocins and one polyketide. It is unknown whether *S. aureus* can become resistant to any of these compounds in the future. Resistance development investigations can be performed in the future to determine whether or not these compounds are likely to lead to the development of resistance in *S. aureus*. Moreover, these results demonstrate that computational approaches to discover BGCs in metagenomes could potentially be used to identify novel molecules involved in microbial exclusion (Wang *et al*., 2022a).

In Chapter 7, *D. pigrum* ATCC 51524 was tested to see whether it could inhibit *S. aureus* JE2 *in vitro*. After growing *D. pigrum* ATCC 51524 in different culture media, it was clear that the supernatants collected were not able to inhibit *S. aureus*. Some studies have revealed that the addition of yeast extract enhanced antimicrobial production, but this thesis did not find this to occur with *D. pigrum* (Anthony *et al*., 2009, Yan *et al*., 2018).

According to bioinformatics analysis suggested the presence of a gene that codes for an iron-dependent repressor IdeR in the bacteriocin DpD BGC. IdeR functions in repressing the expression of iron uptake genes when intracellular iron levels are sufficient, acting as a repressor of siderophore biosynthesis (Marcos-Torres *et al*., 2021) . Bacteriocin DpD produced by *D. pigrum*, could be regulated by the presence of iron. However, as suggested in this thesis, iron level is not associated with antibiotic production by *D. pigrum*. *D. pigrum* is not well studied and this research advanced our knowledge about the influence of different growth conditions on *D. pigrum* ATCC 51524.

Secretion of secondary metabolites by *D. pigrum* ATCC 51524 is another possibility for the lack of antibacterial activity. To date, no study has been carried out to investigate the antibiotic secretion system in *D. pigrum*:

however, this study has found that ABC transporters can be involved in secretion of bacteriocin DpD, bacteriocin DpE and bacteriocin DpG. The expression of the genes that code for these transporters can be influenced by the presence of certain substances and/or certain growth conditions, both of which may have not been present in our study.

The work described in Chapter 7 showed that the production of hydrogen peroxide by *D. pigrum* could be associated with the competitive exclusion of *S. aureus* JE2 in the anterior nares. This finding builds on other research, which reported the elimination of *S. aureus* colonization of the oral cavity through hydrogen peroxide production by viridans group streptococci (Uehara *et al*., 2001). However, findings in this thesis do not necessarily imply that hydrogen peroxide is the sole mechanism of *S. aureus* exclusion.

Since *D. pigrum* ATCC 51524 did not possess appreciable antimicrobial activity, enhancement of antimicrobial production was implemented. The supernatant of *D. pigrum* DP8rif10, one of the most interesting *D. pigrum* ATCC 51524 mutants in this study, was shown to inhibit *S. aureus* JE2 without the need to concentrate the supernatant. In addition, results showed that hydrogen peroxide was not the cause of antimicrobial activity in the supernatant of *D. pigrum* DP8rif10. This mutant is therefore

interesting, and it is worth investigating the mechanism of action of antimicrobial compounds responsible for inhibitory activity (Desbois *et al*., 2013).

8.2 Future directions

Since the supernatant of *Pseudomonas* sp. strain NW27-A1 was inhibiting *A. baumannii*, it is possible that other compounds targeting Gram negative bacteria can be produced if more culture conditions are screened in the future. Moreover, to confirm the suggestion that some BGCs were activated and are responsible for the enhanced antimicrobial activity in *Pseudomonas* sp. strain NW27-A1, transcriptome analysis and proteomics analysis can be performed in the future (Beck *et al*., 2021). Differential expression between BGCs of *Pseudomonas* sp. strain NW27 and *Pseudomonas* sp. strain NW27-A1 can be done by which the expression of core genes in the BGCs can be observed. In addition, proteomics analysis can be done by which whole protein is isolated, sequenced and mapped to the genome. Since the antibacterial potential of this mutant was evident in this study, it is worth investigating the full profile of the activated gene clusters using real time RT-PCR (Zhang et al., 2015). Moreover, it is likely that the genes involved in the secretory system of *Pseudomonas* sp. strain NW27-A1 were upregulated, which led to the secretion of the antimicrobials. Future work is required to

understand the transcriptional organization and regulation of antibiotic export complex in *Pseudomonas* sp. strain NW27 and NW27-A1 (Martín *et al*., 2005). Future work is also required to determine the mutation in the *rpoB* gene of *Pseudomonas* NW27-A1.

Since the acetonitrile extract prepared from the supernatant of *Pseudomonas* sp. strain NW27-A1 was able to inhibit *S. aureus* and *E. faecium*, future work is required to determine the mode of action of the extract (Schäfer and Wenzel, 2020).

A critical factor of success of an antibiotic in development is its minimal cytotoxic effect in humans (Copaescu *et al*., 2021). Antibacterials interact with targets that are specific to bacteria and have a strong affinity for bacterial targets rather than eukaryotic ones, causing selective toxocity (Dalhoff, 2021). Since some *Pseudomonas* spp. produce toxic metabolites, one limitation of this study was the lack of information regarding the cytotoxic activity of the extracts. (Michalska and Wolf, 2015, Würtele *et al*., 2001). It could be of benefit to obtain information about the toxicity of the extracts by nephrotoxicity and MTT cytotoxicity assays (Abdelraouf *et al*., 2012).

This thesis supports previous findings in the literature that optimization of incubation conditions can show a positive influence on antibiotic production by microbes (Ripa *et al*., 2009). It is therefore worthwhile to investigate more *D. pigrum* ATCC 51524 incubation conditions as a part of future work. Furthermore, a synthetic medium can be composed using average nutrient concentrations found in human anterior nares (Krismer *et al*., 2014, Cole *et al*., 2021). Using this synthetic medium as one of the incubation media to grow *D. pigrum* ATCC 51524 might optimise production of antimicrobials by this bacterium. Moreover, this study demonstrated some *D. pigrum* BGCs, including bacteriocin DpD, bacteriocin DpE and bacteriocin DpG BGCs, with predicted promoter regions. Given suitable conditions, sigma factors can bind to promoters of genes in these BGCs, increasing the transcription of the biosynthetic genes (Nazir *et al*., 2018). Although many different sigma factor binding sites are known in other bacteria such as *Bacillus subtilis*, less information regarding sigma factors in *D. pigrum* is available in the literature (De Jong *et al*., 2012). Based on the whole genome shotgun sequence of *D. pigrum* ATCC 51524, structures in promoter regions of bacteriocin DpD, DpE and DpG BGCs can be further studied in the future (Nazir *et al*., 2018).

Even though a few studies have investigated the eradication of *S. aureus* in the anterior nares by *D. pigrum* there is still intriguing evidence from this

thesis which suggests that the *D. pigrum* mutant (DP8rif10) has the potential to inhibit growth of *S. aureus* (Bosch *et al*., 2016, Escapa *et al*., 2018, Brugger *et al*., 2020),. As a part of future work in advancing our knowledge on this topic, it would be interesting to extract the antimicrobial compounds produced by the mutated *D. pigrum* DP8rif10. Future work is also required to confirm the mutation in the *rpoB* gene in *D. pigrum* DP8rif10. Moreover, it would be of benefit to investigate if these extracts exhibit bactericidal or bacteriostatic activity through time-kill kinetics assays (Foerster *et al*., 2016).

S. aureus nasal decolonisation procedures can alter the nasal microbiome, which can have a diverse impact on the hosts health (Baede *et al*., 2022). This research did not provide information on the nasal microbiome's impact following treatment with compounds purified from *D. pigrum* DP8rif10. It would therefore be essential to investigate this possibility by applying the purified antimicrobial compound(s) in cotton rat nose model and performing RNA-based 16S species-level metabarcoding to assess microbial diversity before and after treatment (Desbois *et al*., 2013). Using the cotton rat nose as a model for *S. aureus* colonisation, it is also worth investigating the *in vivo* effectiveness of these extracts for reducing nasal burden of *S. aureus* (Desbois *et al*., 2013). A better understanding of *D. pigrum* could lead to its use in the future as a novel probiotic for the anterior nares.

References

Abdelraouf, K., Braggs, K. H., Yin, T., Truong, L. D., Hu, M. & Tam, V. H. (2012). Characterization of polymyxin B-induced nephrotoxicity: implications for dosing regimen design. *Antimicrobial agents and chemotherapy,* 56**,** 4625-4629.

Acquaro Junior, V. R., Rodrigues, J. P. & Moraes, L. A. B. (2019). Solid phase microextraction as a powerful alternative for screening of secondary metabolites in actinomycetes. *Journal of Mass Spectrometry,* 54**,** 823-833.

Aedo, S. & Tomasz, A. (2016). Role of the Stringent Stress Response in the Antibiotic Resistance Phenotype of Methicillin-Resistant *Staphylococcus aureus*. *Antimicrob Agents Chemother,* 60**,** 2311-7.

Aghapour, Z., Gholizadeh, P., Ganbarov, K., Bialvaei, A. Z., Mahmood, S. S., Tanomand, A., Yousefi, M., Asgharzadeh, M., Yousefi, B. & Kafil, H. S. (2019). Molecular mechanisms related to colistin resistance in *Enterobacteriaceae*. *Infect Drug Resist,* 12**,** 965-975.

Aguirre, M., Morrison, D., Cookson, B. D., Gay, F. W. & Collins, M. D. (1993). Phenotypic and phylogenetic characterization of some Gemella-like organisms from human infections: description of *Dolosigranulum pigrum* gen. nov., sp. nov. *J Appl Bacteriol,* 75**,** 608-12.

Ahmed, M. O. & Baptiste, K. E. (2017). Vancomycin-Resistant Enterococci: A Review of Antimicrobial Resistance Mechanisms and Perspectives of Human and Animal Health. *Microbial Drug Resistance,* 24**,** 590-606.

Aldred, K. J., Kerns, R. J. & Osheroff, N. (2014). Mechanism of quinolone action and resistance. *Biochemistry,* 53**,** 1565-74.

Alsohim, A. S., Taylor, T. B., Barrett, G. A., Gallie, J., Zhang, X. X., Altamirano-Junqueira, A. E., Johnson, L. J., Rainey, P. B. & Jackson, R. W. (2014). The biosurfactant viscosin produced by *Pseudomonas fluorescens* SBW25 aids spreading motility and plant growth promotion. *Environ Microbiol,* 16**,** 2267-81.

Almario, J., Bruto, M., Vacheron, J., Prigent-Combaret, C., Moënne-Loccoz, Y. and Muller, D., (2017). Distribution of 2,4-Diacetylphloroglucinol Biosynthetic Genes among the *Pseudomonas* spp. Reveals Unexpected Polyphyletism. *Frontiers in Microbiology*, 8.

Álvarez-Marín, R., Lepe, J. A., Gasch-Blasi, O., Rodríguez-Martínez, J. M., Calvo-Montes, J., Lara-Contreras, R., Martín-Gandul, C., Tubau-Quintano, F., Cano-García, M. E., Rodríguez-López, F., Rodríguez-Baño, J., Pujol-Rojo, M., Torre-Cisneros, J., Martínez-Martínez, L., Pascual-Hernández, Á. & Jiménez-Mejías, M. E. (2021). Clinical characteristics and outcome of bacteraemia caused by *Enterobacter cloacae* and *Klebsiella aerogenes*: more similarities than differences. *Journal of Global Antimicrobial Resistance,* 25**,** 351-358.

Andersen, P. S., Pedersen, J. K., Fode, P., Skov, R. L., Fowler, V. G., Jr., Stegger, M. & Christensen, K. (2012). Influence of host genetics and environment on nasal carriage of *Staphylococcus aureus* in danish middle-aged and elderly twins. *J Infect Dis,* 206**,** 1178-84.

Annavajhala, M. K., Gomez-Simmonds, A. & Uhlemann, A.-C. (2019). Multidrug-Resistant *Enterobacter cloacae* Complex Emerging as a Global, Diversifying Threat. *Frontiers in Microbiology,* 10.

Anthony, T., Rajesh, T., Kayalvizhi, N. & Gunasekaran, P. (2009). Influence of medium components and fermentation conditions on the production of

bacteriocin(s) by *Bacillus licheniformis* AnBa9. *Bioresour Technol,* 100**,** 872-7.

Antonoplis, A., Zang, X., Wegner, T., Wender, P. A. & Cegelski, L. (2019). Vancomycin-Arginine Conjugate Inhibits Growth Of Carbapenem-Resistant *E. Coli* And Targets Cell-Wall Synthesis. *Acs Chem Biol,* 14**,** 2065-2070.

Aoki, H. & Okuhara, M. (1980). Natural beta-lactam antibiotics. *Annu Rev Microbiol,* 34**,** 159-81.

Aqib, A. I. & Alsayeqh, A. F. (2022). Vancomycin drug resistance, an emerging threat to animal and public health. *Frontiers in Veterinary Science,* 9.

Arbune, M., Gurau, G., Niculet, E., Iancu, A. V., Lupasteanu, G., Fotea, S., Vasile, M. C. & Tatu, A. L. (2021). Prevalence of Antibiotic Resistance of ESKAPE Pathogens Over Five Years in an Infectious Diseases Hospital from South-East of Romania. *Infect Drug Resist,* 14**,** 2369-2378.

Arias, C. A. & Murray, B. E. (2012). The rise of the *Enterococcus*: beyond vancomycin resistance. *Nature Reviews Microbiology,* 10**,** 266-278.

Artsimovitch, I., Patlan, V., Sekine, S.-I., Vassylyeva, M. N., Hosaka, T., Ochi, K., Yokoyama, S. & Vassylyev, D. G. (2004). Structural basis for transcription regulation by alarmone ppGpp. *Cell,* 117**,** 299-310.

Baede, V. O., Barray, A., Tavakol, M., Lina, G., Vos, M. C. & Rasigade, J.-P. (2022). Nasal microbiome disruption and recovery after mupirocin treatment in *Staphylococcus aureus* carriers and noncarriers. *Scientific Reports,* 12**,** 19738.

Bai, K., Yan, H., Chen, X., Lyu, Q., Jiang, N., Li, J. & Luo, L. (2021). The Role of RelA and SpoT on ppGpp Production, Stress Response, Growth Regulation, and Pathogenicity in *Xanthomonas campestris* pv. campestris. *Microbiol Spectr,* 9**,** e0205721.

Baral, B., Akhgari, A. & Metsa-Ketela, M. (2018). Activation of microbial secondary metabolic pathways: Avenues and challenges. *Synth Syst Biotechnol,* 3**,** 163-178.

Baranowska, I. & Wojciechowska, I. (2012). Development of SPE/HPLC-DAD to Determine Residues of Selected Disinfectant Agents in Surface Water. *Polish Journal of Environmental Studies,* 21**,** 269-277.

Barbhaiya, H. and Rao, K., (1985). Production of pyoverdine, the fluorescent pigment of *Pseudomonas aeruginosa* PAO1. *FEMS Microbiology Letters*, 27(2), pp.233-235.

Barna, J. C. & Williams, D. H. (1984). The structure and mode of action of glycopeptide antibiotics of the vancomycin group. *Annu Rev Microbiol,* 38**,** 339-57.

Baron, S. S. & Rowe, J. J. (1981). Antibiotic action of pyocyanin. *Antimicrob Agents Chemother,* 20**,** 814-20.

Bartal, C., Rolston, K. V. I. & Nesher, L. (2022). Carbapenem-resistant *Acinetobacter baumannii*: Colonization, Infection and Current Treatment Options. *Infectious Diseases and Therapy,* 11**,** 683-694.

Beck, C., Gren, T., Ortiz-López, F. J., Jørgensen, T. S., Carretero-Molina, D., Martín Serrano, J., Tormo, J. R., Oves-Costales, D., Kontou, E. E., Mohite, O. S., Mingyar, E., Stegmann, E., Genilloud, O. & Weber, T. (2021). Activation and Identification of a Griseusin Cluster in *Streptomyces* sp. CA-256286 by Employing Transcriptional Regulators and Multi-Omics Methods. *Molecules (Basel, Switzerland),* 26**,** 6580.

Bernier, S. P. & Surette, M. G. (2013). Concentration-dependent activity of antibiotics in natural environments. *Front Microbiol,* 4**,** 20.

Bernstein, L. S. (1982). Combination of trimethoprim with sulfonamides other than sulfamethoxazole. *Rev Infect Dis,* 4**,** 411-8.

Beveridge, T. J. (2001). Use of the Gram stain in microbiology. *Biotechnic & Histochemistry,* 76**,** 111-118.

Bhalodi, A. A., Van Engelen, T. S. R., Virk, H. S. & Wiersinga, W. J. (2019). Impact of antimicrobial therapy on the gut microbiome. *Journal of Antimicrobial Chemotherapy,* 74**,** i6-i15.

Biesbroek, G., Tsivtsivadze, E., Sanders, E. A., Montijn, R., Veenhoven, R. H., Keijser, B. J. & Bogaert, D. (2014). Early respiratory microbiota composition determines bacterial succession patterns and respiratory health in children. *Am J Respir Crit Care Med,* 190**,** 1283-92.

Biniarz, P. & Lukaszewicz, M. (2017). Direct quantification of lipopeptide biosurfactants in biological samples via HPLC and UPLC-MS requires sample modification with an organic solvent. *Appl Microbiol Biotechnol,* 101**,** 4747-4759.

Blahová, J., Králiková, K., Krcméry, V. & Jezek, P. (2000). Low-Frequency transduction of imipenem resistance and high-frequency transduction of ceftazidime and aztreonam resistance by the bacteriophage AP-151 isolated from a *Pseudomonas aeruginosa* strain. *J Chemother,* 12**,** 482-6.

Blair, J. M. A., Webber, M. A., Baylay, A. J., Ogbolu, D. O. & Piddock, L. J. V. (2015). Molecular mechanisms of antibiotic resistance. *Nature Reviews Microbiology,* 13**,** 42-51.

Blanchard, C., Brooks, L., Beckley, A., Colquhoun, J., Dewhurst, S. & Dunman, P. M. (2016). Neomycin Sulfate Improves the Antimicrobial Activity of Mupirocin-Based Antibacterial Ointments. *Antimicrob Agents Chemother,* 60**,** 862-72.

Blin, K., Kim, H., Medema, M. and Weber, T., (2017). Recent development of antiSMASH and other computational approaches to mine secondary metabolite biosynthetic gene clusters. *Briefings in Bioinformatics*, 20(4), pp.1103-1113.

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S., Medema, M. and Weber, T., (2019). antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research*, 47(W1), pp.W81-W87.

Bosch, F. & Rosich, L. (2008). The contributions of Paul Ehrlich to pharmacology: a tribute on the occasion of the centenary of his Nobel Prize. *Pharmacology,* 82**,** 171-9.

Bode, H. B., Bethe, B., Hofs, R. & Zeeck, A. (2002). Big effects from small changes: possible ways to explore nature's chemical diversity. *Chembiochem,* 3**,** 619-27.

Bode, L. G., Kluytmans, J. A., Wertheim, H. F., Bogaers, D., Vandenbroucke-Grauls, C. M., Roosendaal, R., Troelstra, A., Box, A. T., Voss, A., Van Der Tweel, I., Van Belkum, A., Verbrugh, H. A. & Vos, M. C. (2010). Preventing surgical-site infections in nasal carriers of *Staphylococcus aureus*. *N Engl J Med*, 362, 9-17.

Bodour, A. and Miller-Maier, R. (1998). Application of a modified drop-collapse technique for surfactant quantitation and screening of biosurfactant-producing microorganisms. *Journal of Microbiological Methods*, 32(3), pp.273-280.

Bomar, L., Brugger, S. D., Yost, B. H., Davies, S. S. & Lemon, K. P. (2016). *Corynebacterium accolens* Releases Antipneumococcal Free Fatty Acids from Human Nostril and Skin Surface Triacylglycerols. *mBio,* 7**,** e01725-15.

Bosch, A., Levin, E., Van Houten, M. A., Hasrat, R., Kalkman, G., Biesbroek, G., De Steenhuijsen Piters, W. A. A., De Groot, P. C. M., Pernet, P., Keijser, B. J. F., Sanders, E. A. M. & Bogaert, D. (2016). Development of Upper Respiratory Tract Microbiota in Infancy is Affected by Mode of Delivery. *EBioMedicine,* 9**,** 336-345.

Bose, J. L. (2016). Chemical and UV Mutagenesis. *Methods Mol Biol,* 1373**,** 111-5.

Bourne, C. R. (2014). Utility of the Biosynthetic Folate Pathway for Targets in Antimicrobial Discovery. *Antibiotics (Basel),* 3**,** 1-28.

Boyaci, H., Chen, J., Lilic, M., Palka, M., Mooney, R. A., Landick, R., Darst, S. A. & Campbell, E. A. (2018). Fidaxomicin jams *Mycobacterium tuberculosis* RNA polymerase motions needed for initiation via RbpA contacts. *Elife,* 7.

Brandel, J., Humbert, N., Elhabiri, M., Schalk, I., Mislin, G. and Albrecht-Gary, A., (2012). Pyochelin, a siderophore of *Pseudomonas aeruginosa*: Physicochemical characterization of the iron(iii), copper(ii) and zinc(ii) complexes. *Dalton Transactions*, 41(9), p.2820.

Brede, D. A., Faye, T., Johnsborg, O., Odegård, I., Nes, I. F. & Holo, H. (2004). Molecular and genetic characterization of propionicin F, a bacteriocin from *Propionibacterium freudenreichii*. *Applied and environmental microbiology,* 70**,** 7303-7310.

Breijyeh, Z., Jubeh, B. & Karaman, R. (2020). Resistance of Gram-Negative Bacteria to Current Antibacterial Agents and Approaches to Resolve It. *Molecules,* 25.

Brodersen, D. E., Clemons, W. M., Jr., Carter, A. P., Morgan-Warren, R. J., Wimberly, B. T. & Ramakrishnan, V. (2000). The Structural Basis for the Action of the Antibiotics Tetracycline, Pactamycin, and Hygromycin B on the 30S Ribosomal Subunit. *Cell,* 103**,** 1143-1154.

Brouwers, R., Vass, H., Dawson, A., Squires, T., Tavaddod, S. & Allen, R. J. (2020). Stability of β-lactam antibiotics in bacterial growth media. *PLoS One,* 15**,** e0236198.

Brown, S., Santa Maria, J. P., Jr. & Walker, S. (2013). Wall teichoic acids of gram-positive bacteria. *Annu Rev Microbiol,* 67**,** 313-36.

Brugger, S., Baumberger, C., Jost, M., Jenni, W., Brugger, U. and Mühlemann, K., (2012). Automated Counting of Bacterial Colony Forming Units on Agar Plates. *PLoS ONE*, 7(3), p.e33695.

Brugger, S. D., Eslami, S. M., Pettigrew, M. M., Escapa, I. F., Henke, M. T., Kong, Y. & Lemon, K. P. (2020). *Dolosigranulum pigrum* Cooperation and Competition in Human Nasal Microbiota. *mSphere,* 5.

Büchler, A. C., Ragozzino, S., Wicki, M., Spaniol, V., Jäger, S., Seth-Smith, H. M. B., Goldenberger, D., Hinic, V., Egli, A., Frei, R. & Widmer, A. F. (2022). Patients exposed to vancomycin-resistant enterococci during in-hospital outbreaks in a low endemic setting: a proposal for risk-based screening. *Antimicrobial Resistance & Infection Control,* 11**,** 60.

Buijs, Y., Isbrandt, T., Zhang, S. D., Larsen, T. O. & Gram, L. (2020). The Antibiotic Andrimid Produced by *Vibrio coralliilyticus* Increases Expression of Biosynthetic Gene Clusters and Antibiotic Production in *Photobacterium galatheae*. *Front Microbiol,* 11**,** 622055.

Burgess, R. R. (2009). Chapter 4 Preparing a Purification Summary Table. *In:* Burgess, R. R. & Deutscher, M. P. (eds.) *Methods in Enzymology.* Academic Press.

Butler, M. S. & Paterson, D. L. (2020). Antibiotics in the clinical pipeline in October 2019. *The Journal of Antibiotics,* 73**,** 329-364.

Cai, X.-C., Xi, H., Liang, L., Liu, J.-D., Liu, C.-H., Xue, Y.-R. & Yu, X.-Y. (2017). Rifampicin-Resistance Mutations in the *rpoB* Gene in *Bacillus velezensis* CC09 have Pleiotropic Effects. *Frontiers in microbiology,* 8**,** 178-178.

Caltabiano, A. M., Foley, J. P. & Striegel, A. M. (2018). Aqueous size-exclusion chromatography of polyelectrolytes on reversed-phase and hydrophilic interaction chromatography columns. *J Chromatogr A,* 1532**,** 161-174.

Camarinha-Silva, A., Jauregui, R., Chaves-Moreno, D., Oxley, A. P., Schaumburg, F., Becker, K., Wos-Oxley, M. L. & Pieper, D. H. (2014). Comparing the anterior nare bacterial community of two discrete human populations using Illumina amplicon sequencing. *Environ Microbiol,* 16**,** 2939-52.

Carter, A. P., Clemons, W. M., Brodersen, D. E., Morgan-Warren, R. J., Wimberly, B. T. & Ramakrishnan, V. (2000). Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature,* 407**,** 340-348.

Carter, E. L., Jager, L., Gardner, L., Hall, C. C., Willis, S. & Green, J. M. (2007). *Escherichia coli abg* genes enable uptake and cleavage of the folate catabolite p-aminobenzoyl-glutamate. *J Bacteriol,* 189**,** 3329-34.

Carvalho, T. (2023). A new class of antibiotics delivers promising trial results against tuberculosis. *Nat Med,* 29**,** 4-5.

Cavallo, G. & Martinetto, P. (1981). [The mechanism of action of aminoglycosides]. *G Batteriol Virol Immunol,* 74**,** 335-46.

CDC (2000). From the Centers for Disease Control and Prevention. Achievements in public health, 1900-1999: changes in the public health system. *JAMA,* 283**,** 735-8.

CDC (2019) Antibiotic Resistance Threats in the United States, 2019. Available at: https://www.cdc.gov/drugresistance/pdf/threats-report/2019-ar-threats-report-508.pdf [Accessed 22 April 2020].

Chakraborty, A. B. & Berger, S. J. (2005). Optimization of reversed-phase peptide liquid chromatography ultraviolet mass spectrometry analyses using an automated blending methodology. *J Biomol Tech,* 16**,** 327-35.

Champagne-Jorgensen, K., Jose, T. A., Stanisz, A. M., Mian, M. F., Hynes, A. P. & Bienenstock, J. (2021). Bacterial membrane vesicles and phages in blood after consumption of *lacticaseibacillus rhamnosu*s JB-1. *Gut Microbes,* 13**,** 1993583.

Chanioti, S., Liadakis, G. & Tzia, C. (2014). Solid–Liquid Extraction.

Chao, Y. P. & Liao, J. C. (1994). Metabolic responses to substrate futile cycling in *Escherichia coli*. *J Biol Chem,* 269**,** 5122-6.

Chapelle, C., Gaborit, B., Dumont, R., Dinh, A. & Vallée, M. (2021). Treatment of UTIs Due to *Klebsiella pneumoniae* Carbapenemase-Producers: How to Use New Antibiotic Drugs? A Narrative Review. *Antibiotics (Basel),* 10.

Chen, J.-C., Wu, H.-L., Wang, T., Dong, M.-Y., Chen, Y. & Yu, R.-Q. (2022). High-Performance Liquid Chromatography–Diode Array Detection Combined with Chemometrics for Simultaneous Quantitative Analysis of

Five Active Constituents in a Chinese Medicine Formula Wen-Qing-Yin. *Chemosensors,* 10**,** 238.

Cequier-Sánchez, E., Rodríguez, C., Ravelo, A. G. & Zárate, R. (2008). Dichloromethane as a solvent for lipid extraction and assessment of lipid classes and fatty acids from samples of different natures. *J Agric Food Chem,* 56**,** 4297-303.

Chellat, M. F. & Riedl, R. (2017). Pseudouridimycin: The First Nucleoside Analogue That Selectively Inhibits Bacterial RNA Polymerase. *Angew Chem Int Ed Engl,* 56**,** 13184-13186.

Chen, Y., Mehok, A. R., Mant, C. T. & Hodges, R. S. (2004). Optimum concentration of trifluoroacetic acid for reversed-phase liquid chromatography of peptides revisited. *J Chromatogr A,* 1043**,** 9-18.

Chiș, A. A., Rus, L. L., Morgovan, C., Arseniu, A. M., Frum, A., Vonica-Țincu, A. L., Gligor, F. G., Mureșan, M. L. & Dobrea, C. M. (2022). Microbial Resistance to Antibiotics and Effective Antibiotherapy. *Biomedicines,* 10**,** 1121.

Cimermancic, P., Medema, M. H., Claesen, J., Kurita, K., Wieland Brown, L. C., Mavrommatis, K., Pati, A., Godfrey, P. A., Koehrsen, M., Clardy, J., Birren, B. W., Takano, E., Sali, A., Linington, R. G. & Fischbach, M. A. (2014). Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell,* 158**,** 412-421.

Clardy, J., Fischbach, M. A. & Currie, C. R. (2009). The natural history of antibiotics. *Curr Biol,* 19**,** R437-41.

Cole, A. L., Sundar, M., Lopez, A., Forsman, A., Yooseph, S. & Cole, A. M. (2021). Identification of Nasal Gammaproteobacteria with Potent Activity against *Staphylococcus aureus*: Novel Insights into the "Noncarrier" State. *mSphere,* 6.

Comroe, J. H., Jr. (1978). Pay dirt: the story of streptomycin. Part I. From Waksman to Waksman. *Am Rev Respir Dis,* 117**,** 773-81.

Cong, Y., Yang, S. & Rao, X. (2020). Vancomycin resistant *Staphylococcus aureus* infections: A review of case updating and clinical features. *Journal of Advanced Research,* 21**,** 169-176.

Conly, J. M. & Johnston, L. B. (2004). Coming full circle: From antibiotics to probiotics and prebiotics. *Can J Infect Dis Med Microbiol,* 15**,** 161-3.

Copaescu, A., Choshi, P., Pedretti, S., Mouhtouris, E., Peter, J. & Trubiano, J. A. (2021). Dose Dependent Antimicrobial Cellular Cytotoxicity—Implications for ex vivo Diagnostics. *Frontiers in Pharmacology,* 12.

Costa, D. M., Johani, K., Melo, D. S., Lopes, L. K. O., Lopes Lima, L. K. O., Tipple, A. F. V., Hu, H. & Vickery, K. (2019). Biofilm contamination of high-touched surfaces in intensive care units: epidemiology and potential impacts. *Letters in Applied Microbiology,* 68**,** 269-276.

Covington, B. C., Xu, F. & Seyedsayamdost, M. R. (2021). A Natural Product Chemist's Guide to Unlocking Silent Biosynthetic Gene Clusters. *Annu Rev Biochem,* 90**,** 763-788.

Cox, G. & Wright, G. D. (2013). Intrinsic antibiotic resistance: mechanisms, origins, challenges and solutions. *Int J Med Microbiol,* 303**,** 287-92.

Coyne, S., Courvalin, P. & Périchon, B. (2011). Efflux-mediated antibiotic resistance in *Acinetobacter* spp. *Antimicrob Agents Chemother,* 55**,** 947-53.

Crellin, E., Mansfield, K. E., Leyrat, C., Nitsch, D., Douglas, I. J., Root, A., Williamson, E., Smeeth, L. & Tomlinson, L. A. (2018). Trimethoprim use for urinary tract infection and risk of adverse outcomes in older patients: cohort study. *BMJ,* 360**,** k341.

Cummins, P. M., Rochfort, K. D. & O'connor, B. F. (2017). Ion-Exchange Chromatography: Basic Principles and Application. *Methods Mol Biol,* 1485**,** 209-223.

Dadashi, M., Hajikhani, B., Darban-Sarokhalil, D., Van Belkum, A. & Goudarzi, M. (2020). Mupirocin resistance in *Staphylococcus aureus*: A systematic review and meta-analysis. *J Glob Antimicrob Resist,* 20**,** 238-247.

D'aes, J., Kieu, N. P., Léclère, V., Tokarski, C., Olorunleke, F. E., De Maeyer, K., Jacques, P., Höfte, M. & Ongena, M. 2014. To settle or to move? The interplay between two classes of cyclic lipopeptides in the biocontrol strain Pseudomonas CMR12a. *Environmental Microbiology,* 16**,** 2282-2300.

Dalhoff, A. (2021). Selective toxicity of antibacterial agents-still a valid concept or do we miss chances and ignore risks? *Infection,* 49**,** 29-56.

Dame, J. A., Beylis, N., Nuttall, J. & Eley, B. (2020). *Pseudomonas aeruginosa* bloodstream infection at a tertiary referral hospital for children. *BMC Infectious Diseases,* 20**,** 729.

Danesh, A., Ljungh, Å., Mattiasson, B. & Mamo, G. (2016). Synergistic effect of haloduracin and chloramphenicol against clinically important Gram-positive bacteria. *Biotechnology reports (Amsterdam, Netherlands),* 13**,** 37-41.

Das Ghatak, P., Mathew-Steiner, S. S., Pandey, P., Roy, S. & Sen, C. K. (2018). A surfactant polymer dressing potentiates antimicrobial efficacy in biofilm disruption. *Sci Rep,* 8**,** 873.

Daugelavicius, R., Bakiene, E. & Bamford, D. H. (2000). Stages of polymyxin B interaction with the *Escherichia coli* cell envelope. *Antimicrob Agents Chemother,* 44**,** 2969-78.

Davin-Regli, A. & Pagès, J. M. (2015). *Enterobacter aerogenes* and *Enterobacter cloacae*; versatile bacterial pathogens confronting antibiotic treatment. *Front Microbiol,* 6**,** 392.

De Boeck, I., Wittouck, S., Martens, K., Spacova, I., Cauwenberghs, E., Allonsius, C. N., Jorissen, J., Wuyts, S., Van Beeck, W., Dillen, J., Bron, P. A., Steelant, B., Hellings, P. W., Vanderveken, O. M. & Lebeer, S. (2021). The nasal mutualist *Dolosigranulum pigrum* AMBR11 supports homeostasis via multiple mechanisms. *iScience,* 24**,** 102978.

De Boeck, I., Wittouck, S., Wuyts, S., Oerlemans, E. F. M., Van Den Broek, M. F. L., Vandenheuvel, D., Vanderveken, O. & Lebeer, S. (2017). Comparing the Healthy Nose and Nasopharynx Microbiota Reveals Continuity As Well As Niche-Specificity. *Front Microbiol,* 8**,** 2372.

De Jong, A., Pietersma, H., Cordes, M., Kuipers, O. P. & Kok, J. (2012). PePPER: a webserver for prediction of prokaryote promoter elements and regulons. *BMC Genomics,* 13**,** 299.

De Kraker, M. E., Stewardson, A. J. & Harbarth, S. (2016). Will 10 Million People Die a Year due to Antimicrobial Resistance by 2050? *PLoS Med,* 13**,** e1002184.

De Niederhäusern, S., Bondi, M., Messi, P., Iseppi, R., Sabia, C., Manicardi, G. & Anacarso, I. (2011). Vancomycin-resistance transferability from VanA enterococci to *Staphylococcus aureus*. *Curr Microbiol,* 62**,** 1363-7.

Denissen, J., Reyneke, B., Waso-Reyneke, M., Havenga, B., Barnard, T., Khan, S. & Khan, W. (2022). Prevalence of ESKAPE pathogens in the environment: Antibiotic resistance status, community-acquired infection and risk to human health. *International Journal of Hygiene and Environmental Health,* 244**,** 114006.

Desbois, A. P., Sattar, A., Graham, S., Warn, P. A. & Coote, P. J. (2013). MRSA decolonization of cotton rat nares by a combination treatment comprising lysostaphin and the antimicrobial peptide ranalexin. *J Antimicrob Chemother,* 68**,** 2569-75.

Detweiler, C. S. (2020). Infection-based chemical screens uncover host-pathogen interactions. *Curr Opin Microbiol,* 54**,** 43-50.

Denton, M., Wilcox, M. H., Parnell, P., Green, D., Keer, V., Hawkey, P. M., Evans, I. & Murphy, P. (2004). Role of environmental cleaning in controlling an outbreak of *Acinetobacter baumannii* on a neurosurgical intensive care unit. *J Hosp Infect,* 56**,** 106-10.

Dhanda, G., Sarkar, P., Samaddar, S. & Haldar, J. (2019). Battle against Vancomycin-Resistant Bacteria: Recent Developments in Chemical Strategies. *J Med Chem,* 62**,** 3184-3205.

Dinu, V., Lu, Y., Weston, N., Lithgo, R., Coupe, H., Channell, G., Adams, G. G., Torcello Gómez, A., Sabater, C., Mackie, A., Parmenter, C., Fisk, I., Phillips-Jones, M. K. & Harding, S. E. 2020. The antibiotic vancomycin induces complexation and aggregation of gastrointestinal and submaxillary mucins. *Sci Rep,* 10**,** 960.

Di Tella, D., Tamburro, M., Guerrizio, G., Fanelli, I., Sammarco, M. L. & Ripabelli, G. (2019). Molecular Epidemiological Insights into Colistin-Resistant and Carbapenemases-Producing Clinical *Klebsiella pneumoniae* Isolates. *Infect Drug Resist,* 12**,** 3783-3795.

Dixon, R. A. & Chopra, I. (1986). Polymyxin B and polymyxin B nonapeptide alter cytoplasmic membrane permeability in *Escherichia coli*. *J Antimicrob Chemother,* 18**,** 557-63.

Dößelmann, B., Willmann, M., Steglich, M., Bunk, B., Nübel, U., Peter, S. & Neher, R. A. (2017). Rapid and Consistent Evolution of Colistin Resistance in Extensively Drug-Resistant *Pseudomonas aeruginosa* during Morbidostat Culture. *Antimicrobial Agents and Chemotherapy,* 61**,** e00043-17.

Donia, M. S., Cimermancic, P., Schulze, C. J., Wieland Brown, L. C., Martin, J., Mitreva, M., Clardy, J., Linington, R. G. & Fischbach, M. A. (2014). A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell,* 158**,** 1402-1414.

Donlan, R. M. (2002). Biofilms: microbial life on surfaces. *Emerg Infect Dis,* 8**,** 881-90.

Dumas, Z., Ross-Gillespie, A. and Kümmerli, R., (2013). Switching between apparently redundant iron-uptake mechanisms benefits bacteria in

changeable environments. *Proceedings of the Royal Society B: Biological Sciences*, 280(1764), p.20131055.

Edwards, D. I. (1986). Reduction of nitroimidazoles in vitro and DNA damage. *Biochem Pharmacol,* 35**,** 53-8.

Eheth, J., Djimeli, C., Nana, P., Arfao, A., Ewoti, O., Moungang, L., Bricheux, G., Sime-Ngando, T. and Nola, M., (2019). Less effect of wells physicochemical properties on the antimicrobial susceptibility *Pseudomonas aeruginosa* isolated in equatorial region of Central Africa. *Applied Water Science*, 9(2).

El Aila, N. A., Al Laham, N. A. & Ayesh, B. M. (2017). Nasal carriage of methicillin resistant *Staphylococcus aureus* among health care workers at Al Shifa hospital in Gaza Strip. *BMC Infect Dis,* 17**,** 28.

Eliopoulos, G. M. & Gold, H. S. (2001). Vancomycin-Resistant Enterococci: Mechanisms and Clinical Observations. *Clinical Infectious Diseases,* 33**,** 210-219.

El-Sayed Ahmed, M. A. E.-G., Zhong, L.-L., Shen, C., Yang, Y., Doi, Y. & Tian, G.-B. (2020). Colistin and its role in the Era of antibiotic resistance: an extended review (2000–2019). *Emerging Microbes & Infections,* 9**,** 868-885.

Epand, R. F., Savage, P. B. & Epand, R. M. (2007). Bacterial lipid composition and the antimicrobial efficacy of cationic steroid compounds (Ceragenins). *Biochim Biophys Acta,* 1768**,** 2500-9.

Epand, R. M., Walker, C., Epand, R. F. & Magarvey, N. A. (2016). Molecular mechanisms of membrane targeting antibiotics. *Biochimica et Biophysica Acta (BBA) - Biomembranes,* 1858**,** 980-987.

Escapa, I. F., Chen, T., Huang, Y., Gajare, P., Dewhirst, F. E. & Lemon, K. P. (2018). New Insights into Human Nostril Microbiome from the Expanded Human Oral Microbiome Database (eHOMD): a Resource for the Microbiome of the Human Aerodigestive Tract. *mSystems,* 3.

Fàbrega, A., Madurga, S., Giralt, E. & Vila, J. (2009). Mechanism of action of and resistance to quinolones. *Microb Biotechnol,* 2**,** 40-61.

Fagerquist, C. K., Lightfield, A. R. & Lehotay, S. J. (2005). Confirmatory and quantitative analysis of beta-lactam antibiotics in bovine kidney tissue by dispersive solid-phase extraction and liquid chromatography-tandem mass spectrometry. *Anal Chem,* 77**,** 1473-82.

Falk, N. A. 2019. Surfactants as Antimicrobials: A Brief Overview of Microbial Interfacial Chemistry and Surfactant Antimicrobial Activity. *J Surfactants Deterg,* 22**,** 1119-1127.

Fernandes De Oliveira, L. M., Steindorff, M., Darisipudi, M. N., Mrochen, D. M., Trube, P., Broker, B. M., Bronstrup, M., Tegge, W. & Holtfreter, S. (2021). Discovery of *Staphylococcus aureus* Adhesion Inhibitors by Automated Imaging and Their Characterization in a Mouse Model of Persistent Nasal Colonization. *Microorganisms,* 9.

Ferrarini, E., De Roo, V., Geudens, N., Martins, J. C. & Höfte, M. (2022). Altering in vivo membrane sterol composition affects the activity of the cyclic

lipopeptides tolaasin and sessilin against Pythium. *Biochimica et Biophysica Acta (BBA) - Biomembranes,* 1864**,** 184008.

Ferrer-Espada, R., Shahrour, H., Pitts, B., Stewart, P. S., Sánchez-Gómez, S. & Martínez-De-Tejada, G. (2019). A permeability-increasing drug synergizes with bacterial efflux pump inhibitors and restores susceptibility to antibiotics in multi-drug resistant *Pseudomonas aeruginosa* strains. *Scientific Reports,* 9**,** 3452.

Fishovitz, J., Hermoso, J. A., Chang, M. & Mobashery, S. (2014). Penicillin-binding protein 2a of methicillin-resistant *Staphylococcus aureus*. *IUBMB Life,* 66**,** 572-7.

Foerster, S., Unemo, M., Hathaway, L. J., Low, N. & Althaus, C. L. (2016). Time-kill curve analysis and pharmacodynamic modelling for in vitro evaluation of antimicrobials against *Neisseria gonorrhoeae*. *BMC Microbiology,* 16**,** 216.

Fontana, R., Canepari, P., Lleò, M. M. & Satta, G. (1990). Mechanisms of resistance of enterococci to beta-lactam antibiotics. *Eur J Clin Microbiol Infect Dis,* 9**,** 103-5.

Founou, R., Founou, L. and Essack, S., (2017). Clinical and economic impact of antibiotic resistance in developing countries: A systematic review and meta-analysis. *PLOS ONE*, 12(12), p.e0189621.

Franciolli, M., Bille, J., Glauser, M. P. & Moreillon, P. (1991). Beta-lactam resistance mechanisms of methicillin-resistant *Staphylococcus aureus*. *J Infect Dis,* 163**,** 514-23.

Frangipani, E., Bonchi, C., Minandri, F., Imperi, F. & Visca, P. (2014). Pyochelin potentiates the inhibitory activity of gallium on *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother,* 58**,** 5572-5.

Frank, D. N., Feazel, L. M., Bessesen, M. T., Price, C. S., Janoff, E. N. & Pace, N. R. (2010). The human nasal microbiota and *Staphylococcus aureus* carriage. *PLoS One,* 5**,** e10598.

Fridman, O., Goldberg, A., Ronin, I., Shoresh, N. & Balaban, N. Q. (2014). Optimization of lag time underlies antibiotic tolerance in evolved bacterial populations. *Nature,* 513**,** 418-21.

Fuller, A. T., Mellows, G., Woolford, M., Banks, G. T., Barrow, K. D. & Chain, E. B. (1971). Pseudomonic Acid: an Antibiotic produced by *Pseudomonas fluorescens*. *Nature,* 234**,** 416-417.

Fusco, G., & Minelli, A. (2019). Genetics and Cytogenetics of Reproduction. *In:* Minelli, A. & Fusco, G. (eds.) *The Biology of Reproduction.* Cambridge: Cambridge University Press.

Gaca, A. O., Kajfasz, J. K., Miller, J. H., Liu, K., Wang, J. D., Abranches, J. & Lemos, J. A. (2013). Basal Levels of (p)ppGpp in *Enterococcus faecalis*: the Magic beyond the Stringent Response. *mBio,* 4**,** e00646-13.

Gao, W., Howden, B. P. & Stinear, T. P. (2018). Evolution of virulence in *Enterococcus faecium*, a hospital-adapted opportunistic pathogen. *Current Opinion in Microbiology,* 41**,** 76-82.

Garcia L. (2010). Broth Microdilution MIC Test, p 25-41. *In Clinical Microbiology Procedures Handbook, 3rd Edition*. ASM Press, Washington, DC. doi: 10.1128/9781555817435.ch5.2

Gardete, S. & Tomasz, A. (2014). Mechanisms of vancomycin resistance in *Staphylococcus aureus*. *J Clin Invest,* 124**,** 2836-40.

Geng, H. & Belas, R. (2010). Expression of tropodithietic acid biosynthesis is controlled by a novel autoinducer. *J Bacteriol,* 192**,** 4377-87.

Geudens, N. & Martins, J. C. (2018). Cyclic Lipodepsipeptides From *Pseudomonas* spp. – Biological Swiss-Army Knives. *Frontiers in Microbiology,* 9.

Ghebremedhin, B., Layer, F., König, W. & König, B. (2008). Genetic classification and distinguishing of *Staphylococcus* species based on different partial gap, 16S rRNA, *hsp60*, *rpoB*, *sodA*, and *tuf* gene sequences. *J Clin Microbiol,* 46**,** 1019-25.

Ghosh, M., Miller, P. A., Mollmann, U., Claypool, W. D., Schroeder, V. A., Wolter, W. R., Suckow, M., Yu, H., Li, S., Huang, W., Zajicek, J. & Miller, M. J. (2017). Targeted Antibiotic Delivery: Selective Siderophore Conjugation with Daptomycin Confers Potent Activity against Multidrug Resistant *Acinetobacter baumannii* Both in Vitro and in Vivo. *J Med Chem,* 60**,** 4577-4583.

Gilbert, P., Collier, P. J. & Brown, M. R. (1990). Influence of growth rate on susceptibility to antimicrobial agents: biofilms, cell cycle, dormancy, and stringent response. *Antimicrob Agents Chemother,* 34**,** 1865-8.

Girlich, D., Poirel, L., Szczepanowski, R., Schlüter, A. & Nordmann, P. (2012). Carbapenem-hydrolyzing GES-5-encoding gene on different plasmid types recovered from a bacterial community in a sewage treatment plant. *Appl Environ Microbiol,* 78**,** 1292-5.

Glatter, K. A. & Finkelman, P. (2021). History of the Plague: An Ancient Pandemic for the Age of COVID-19. *Am J Med,* 134**,** 176-181.

Glauert, A. M. & Thornley, M. J. (1969). The topography of the bacterial cell wall. *Annu Rev Microbiol,* 23**,** 159-98.

Glushakova, A. M., Kachalkin, A. V., Prokof'eva, T. V. & Lysak, L. V. (2022). *Enterobacteriaceae* in soils and atmospheric dust aerosol accumulations of Moscow city. *Current Research in Microbial Sciences,* 3**,** 100124.

Gogry, F. A., Siddiqui, M. T., Sultan, I. & Haq, Q. M. R. (2021). Current Update on Intrinsic and Acquired Colistin Resistance Mechanisms in Bacteria. *Front Med (Lausanne),* 8**,** 677720.

González-Jaramillo, L. M., Aranda, F. J., Teruel, J. A., Villegas-Escobar, V. & Ortiz, A. 2017. Antimycotic activity of fengycin C biosurfactant and its interaction with phosphatidylcholine model membranes. *Colloids Surf B Biointerfaces,* 156**,** 114-122.

Gottesman, T., Fedorowsky, R., Yerushalmi, R., Lellouche, J. & Nutman, A. (2021). An outbreak of carbapenem-resistant *Acinetobacter baumannii* in a COVID-19 dedicated hospital. *Infection Prevention in Practice,* 3**,** 100113.

Goudarzi, H., Mirsamadi, E. S., Ghalavand, Z., Hakemi Vala, M., Mirjalali, H. & Hashemi, A. (2019). Rapid detection and molecular survey of blaVIM, blaIMP and blaNDM genes among clinical isolates of *Acinetobacter baumannii* using new multiplex real-time PCR and melting curve analysis. *BMC Microbiol,* 19**,** 122.

Gregory, T. V., Ellis, K., Valeriani, R., Khan, F., Wu, X., Murin, L., Alibayov, B., Vidal, A. G. J., Zhao, T. & Vidal, J. E. (2021). MoWa: A Disinfectant for Hospital Surfaces Contaminated With Methicillin-Resistant *Staphylococcus aureus* (MRSA) and Other Nosocomial Pathogens. *Front Cell Infect Microbiol,* 11**,** 676638.

Guest, J. F., Keating, T., Gould, D. & Wigglesworth, N. (2020). Modelling the annual NHS costs and outcomes attributable to healthcare-associated infections in England. *BMJ Open,* 10**,** e033367.

Guo, S., Wang, Y., Bilal, M., Hu, H., Wang, W. & Zhang, X. (2020). Microbial Synthesis of Antibacterial Phenazine-1,6-dicarboxylic Acid and the Role of PhzG in *Pseudomonas chlororaphis* GP72AN. *J Agric Food Chem,* 68**,** 2373-2380.

Gupta, A., Ampofo, K., Rubenstein, D. & Saiman, L. (2003). Extended Spectrum β Lactamase-producing *Klebsiella pneumoniae* Infections: a Review of the Literature. *Journal of Perinatology,* 23**,** 439-443.

Hadjithomas, M., Chen, I. A., Chu, K., Huang, J., Ratner, A., Palaniappan, K., Andersen, E., Markowitz, V., Kyrpides, N. C. & Ivanova, N. N. (2017). IMG-ABC: new features for bacterial secondary metabolism analysis and targeted biosynthetic gene cluster discovery in thousands of microbial genomes. *Nucleic Acids Res,* 45**,** D560-d565.

Hall, G. S., Gordon, S., Schroeder, S., Smith, K., Anthony, K. & Procop, G. W. (2001). Case of synovitis potentially caused by *Dolosigranulum pigrum*. *J Clin Microbiol,* 39**,** 1202-3.

Hamamoto, H., Panthee, S., Hashimoto, K., Tsuchida, T. & Sekimizu, K. (2021). An efficient method to screen for the soil bacteria producing therapeutically effective antibiotics. *The Journal of Antibiotics,* 74**,** 850-855.

Hamidian, M. & Nigro, S. J. (2019). Emergence, molecular mechanisms and global spread of carbapenem-resistant *Acinetobacter baumannii*. *Microb Genom,* 5.

Hammad, S. F., Abdallah, I. A., Bedair, A. & Mansour, F. R. (2021). Homogeneous liquid-liquid extraction as an alternative sample preparation technique for biomedical analysis. *J Sep Sci*.

Hamzaoui, Z., Ocampo-Sosa, A., Maamar, E., Fernandez Martinez, M., Ferjani, S., Hammami, S., Harbaoui, S., Genel, N., Arlet, G., Saidani, M., Slim, A., Boutiba-Ben Boubaker, I. & Martinez-Martinez, L. (2018). An Outbreak of NDM-1-Producing *Klebsiella pneumoniae*, Associated with OmpK35 and OmpK36 Porin Loss in Tunisia. *Microb Drug Resist,* 24**,** 1137-1147.

Hahn-Löbmann, S., Stephan, A., Schulz, S., Schneider, T., Shaverskyi, A., Tusé, D., Giritch, A. & Gleba, Y. (2019). Colicins and Salmocins – New Classes of Plant-Made Non-antibiotic Food Antibacterials. *Frontiers in Plant Science,* 10.

Haque, M., Sartelli, M., Mckimm, J. & Abu Bakar, M. (2018). Health care-associated infections - an overview. *Infect Drug Resist,* 11**,** 2321-2333.

Hennion, M. C. (1999). Solid-phase extraction: method development, sorbents, and coupling with liquid chromatography. *J Chromatogr A,* 856**,** 3-54.

Héritier, C., Poirel, L., Fournier, P. E., Claverie, J. M., Raoult, D. & Nordmann, P. (2005). Characterization of the naturally occurring oxacillinase of *Acinetobacter baumannii*. *Antimicrob Agents Chemother,* 49**,** 4174-9.

Hider, R. C. & Kong, X. 2010. Chemistry and biology of siderophores. *Natural product reports,* 27**,** 637-657.

Hiller, N. L., Ahmed, A., Powell, E., Martin, D. P., Eutsey, R., Earl, J., Janto, B., Boissy, R. J., Hogg, J., Barbadora, K., Sampath, R., Lonergan, S., Post, J. C., Hu, F. Z. & Ehrlich, G. D. (2010). Generation of genic diversity among *Streptococcus pneumoniae* strains via horizontal gene transfer during a chronic polyclonal pediatric infection. *PLoS Pathog,* 6**,** e1001108.

Ho, M. X., Hudson, B. P., Das, K., Arnold, E. & Ebright, R. H. (2009). Structures of RNA polymerase-antibiotic complexes. *Curr Opin Struct Biol,* 19**,** 715-23.

Horan, T., Gaynes, R. & Mayhall, C. (2004). Hospital epidemiology and infection control.

Hormaeche, E. & Edwards, P. R. (1960). A Proposed Genus Enterobacter. *International Journal of Systematic and Evolutionary Microbiology,* 10**,** 71-74.

Hosaka, T., Xu, J. & Ochi, K. (2006). Increased expression of ribosome recycling factor is responsible for the enhanced protein synthesis during the late growth phase in an antibiotic-overproducing *Streptomyces coelicolor* ribosomal rpsL mutant. *Mol Microbiol,* 61**,** 883-97.

Hosaka, T., Ohnishi-Kameyama, M., Muramatsu, H., Murakami, K., Tsurumi, Y., Kodani, S., Yoshida, M., Fujie, A. & Ochi, K. (2009). Antibacterial discovery in actinomycetes strains with mutations in RNA polymerase or ribosomal protein S12. *Nat Biotechnol,* 27**,** 462-4.

Hou, J. P. & Poole, J. W. (1971). Beta-lactam antibiotics: their physicochemical properties and biological activities in relation to structure. *J Pharm Sci,* 60**,** 503-32.

Howard, A., O'donoghue, M., Feeney, A. & Sleator, R. D. (2012). *Acinetobacter baumannii*: an emerging opportunistic pathogen. *Virulence,* 3**,** 243-50.

Howe, A. & Sofou, S. (2021). Daptomycin-Induced Lipid Phases on Model Lipid Bilayers: Effect of Lipid Type and of Lipid Leaflet Order on Membrane Permeability. *The Journal of Physical Chemistry B,* 125**,** 5775-5785.

Hrenovic, J., Durn, G., Goic-Barisic, I. & Kovacic, A. (2014). Occurrence of an environmental *Acinetobacter baumannii* strain similar to a clinical isolate in paleosol from Croatia. *Appl Environ Microbiol,* 80**,** 2860-6.

Huddleston, J. R. (2014). Horizontal gene transfer in the human gastrointestinal tract: potential spread of antibiotic resistance genes. *Infect Drug Resist,* 7**,** 167-76.

Huffman, B. A., Poltash, M. L. & Hughey, C. A. (2012). Effect of polar protic and polar aprotic solvents on negative-ion electrospray ionization and chromatographic separation of small acidic molecules. *Anal Chem,* 84**,** 9942-50.

Hunter, C. A. (2004).Quantifying intermolecular interactions: guidelines for the molecular recognition toolbox. *Angew Chem Int Ed Engl,* 43**,** 5310-24.

Huovinen, P., Sundström, L., Swedberg, G. & Sköld, O. (1995). Trimethoprim and sulfonamide resistance. *Antimicrob Agents Chemother,* 39**,** 279-89.

Hutchins, C. F., Moore, G., Thompson, K. A., Webb, J. & Walker, J. T. (2017). Contamination of hospital tap water: the survival and persistence of *Pseudomonas aeruginosa* on conventional and 'antimicrobial' outlet fittings. *J Hosp Infect,* 97**,** 156-161.

Hutchings, M. I., Truman, A. W. & Wilkinson, B. (2019). Antibiotics: past, present and future. *Curr Opin Microbiol,* 51**,** 72-80.

Ian, F., Woods, D., Lewis, M., Gan, Q. & Nancarrow, P. (2012). The Importance of Acetonitrile in the Pharmaceutical Industry and Opportunities for its Recovery from Waste. *Organic Process Research & Development,* 16**,** 612-624.

Iglewski, B. H. (1996). *Pseudomonas*. In: Baron, S. (ed.) Medical Microbiology. 4th ed. Galveston (TX).

Islam, M. A., Albarracin, L., Melnikov, V., Andrade, B. G. N., Cuadrat, R. R. C., Kitazawa, H. & Villena, J. (2021). *Dolosigranulum pigrum* Modulates

Immunity against SARS-CoV-2 in Respiratory Epithelial Cells. *Pathogens,* 10.

Isnansetyo, A., (2003). Antibacterial activity of 2,4-diacetylphloroglucinol produced by *Pseudomonas* sp. AMSN isolated from a marine alga, against vancomycin-resistant *Staphylococcus aureus*. *International Journal of Antimicrobial Agents*, 22(5), pp.545-547.

Jackson, N., Czaplewski, L. & Piddock, L. J. V. (2018). Discovery and development of new antibacterial drugs: learning from experience? *J Antimicrob Chemother,* 73**,** 1452-1459.

Jain, M., Sharma, A., Sen, M. K., Rani, V., Gaind, R. & Suri, J. C. (2019). Phenotypic and molecular characterization of *Acinetobacter baumannii* isolates causing lower respiratory infections among ICU patients. *Microb Pathog,* 128**,** 75-81.

Jenke-Kodama, H., Sandmann, A., Müller, R. & Dittmann, E. (2005). Evolutionary Implications of Bacterial Polyketide Synthases. *Molecular Biology and Evolution,* 22**,** 2027-2039.

Jeon, M., Huh, K., Ko, J.-H., Cho, S. Y., Huh, H. J., Lee, N. Y., Kang, C.-I., Chung, D. R. & Peck, K. R. (2021). Difference in the Clinical Outcome of Bloodstream Infections Caused by *Klebsiella aerogenes* and *Enterobacter cloacae* Complex. *Open Forum Infectious Diseases,* 8.

Jian, Y., Zhao, L., Zhao, N., Lv, H.-Y., Liu, Y., He, L., Liu, Q. & LI, M. (2021). Increasing prevalence of hypervirulent ST5 methicillin susceptible *Staphylococcus aureus* subtype poses a serious clinical threat. *Emerging Microbes & Infections,* 10**,** 109-122.

Jiang, S. C. & Paul, J. H. (1998). Gene transfer by transduction in the marine environment. *Appl Environ Microbiol,* 64**,** 2780-7.

Jiménez-Guerra, G., Heras-Cañas, V., Gutiérrez-Soto, M., Del Pilar Aznarte-Padial, M., Expósito-Ruiz, M., Navarro-Marí, J. M. & Gutiérrez-Fernández, J. (2018). Urinary tract infection by *Acinetobacter baumannii* and *Pseudomonas aeruginosa*: evolution of antimicrobial resistance and therapeutic alternatives. *J Med Microbiol,* 67**,** 790-797.

Johnston, C., Martin, B., Fichant, G., Polard, P. & Claverys, J. P. (2014). Bacterial transformation: distribution, shared mechanisms and divergent control. *Nat Rev Microbiol,* 12**,** 181-96.

Julian, W. T., Vasilchenko, A. V., Shpindyuk, D. D., Poshvina, D. V. & Vasilchenko, A. S. (2020). Bacterial-Derived Plant Protection Metabolite 2,4-Diacetylphloroglucinol: Effects on Bacterial Cells at Inhibitory and Subinhibitory Concentrations. *Biomolecules,* 11.

Jung, D., Rozek, A., Okon, M. & Hancock, R. E. (2004a). Structural transitions as determinants of the action of the calcium-dependent antibiotic daptomycin. *Chem Biol,* 11**,** 949-57.

Khan, H. A., Baig, F. K. & Mehboob, R. (2017). Nosocomial infections: Epidemiology, prevention, control and surveillance. *Asian Pacific Journal of Tropical Biomedicine,* 7**,** 478-482.

Khedkar, S., Smyshlyaev, G., Letunic, I., Maistrenko, Oleksandr M., Coelho, Luis P., Orakov, A., Forslund, S. K., Hildebrand, F., Luetge, M., Schmidt,

T. S. B., Barabas, O. & Bork, P. (2022). Landscape of mobile genetic elements and their antibiotic resistance cargo in prokaryotic genomes. *Nucleic Acids Research,* 50**,** 3155-3168.

Kislik, V. (2012). Chapter 13. Advances in Development of Solvents for Liquid–Liquid Extraction. *Solvent Extraction Classical and Novel Approaches***,** 451-481.

Kizny Gordon, A. E., Mathers, A. J., Cheong, E. Y. L., Gottlieb, T., Kotay, S., Walker, A. S., Peto, T. E. A., Crook, D. W. & Stoesser, N. (2017). The Hospital Water Environment as a Reservoir for Carbapenem-Resistant Organisms Causing Hospital-Acquired Infections-A Systematic Review of the Literature. *Clin Infect Dis,* 64**,** 1435-1444.

Kluytmans, J., Van Belkum, A. & Verbrugh, H. (1997). Nasal carriage of Staphylococcus aureus: epidemiology, underlying mechanisms, and associated risks. *Clin Microbiol Rev,* 10**,** 505-20.

Kokai-Kun, J. F., Walsh, S. M., Chanturiya, T. & Mond, J. J. (2003). Lysostaphin cream eradicates *Staphylococcus aureus* nasal colonization in a cotton rat model. *Antimicrob Agents Chemother,* 47**,** 1589-97.

Kotsogianni, I., Wood, T. M., Alexander, F. M., Cochrane, S. A. & Martin, N. I. (2021). Binding Studies Reveal Phospholipid Specificity and Its Role in the Calcium-Dependent Mechanism of Action of Daptomycin. *ACS Infectious Diseases,* 7**,** 2612-2619.

Krismer, B., Liebeke, M., Janek, D., Nega, M., Rautenberg, M., Hornig, G., Unger, C., Weidenmaier, C., Lalk, M. & PescheL, A. (2014). Nutrient

limitation governs *Staphylococcus aureus* metabolism and niche adaptation in the human nose. *PLoS Pathog,* 10**,** e1003862.

Krismer, B., Weidenmaier, C., Zipperer, A. & Peschel, A. (2017). The commensal lifestyle of *Staphylococcus aureus* and its interactions with the nasal microbiota. *Nature Reviews Microbiology,* 15**,** 675-687.

Kruszewska, D., Sahl, H.-G., Bierbaum, G., Pag, U., Hynes, S. O. & Ljungh, Å. (2004). Mersacidin eradicates methicillin-resistant *Staphylococcus aureus* (MRSA) in a mouse rhinitis model. *Journal of Antimicrobial Chemotherapy,* 54**,** 648-653.

Kugeler, K. J., Staples, J. E., Hinckley, A. F., Gage, K. L. & Mead, P. S. (2015). Epidemiology of human plague in the United States, 1900-2012. *Emerg Infect Dis,* 21**,** 16-22.

Kumar, S. N., Nambisan, B., Sundaresan, A., Mohandas, C. & Anto, R. J. (2014). Isolation and identification of antimicrobial secondary metabolites from *Bacillus cereus* associated with a rhabditid entomopathogenic nematode. *Annals of Microbiology,* 64**,** 209-218.

Kumpitsch, C., Koskinen, K., Schopf, V. & Moissl-Eichinger, C. (2019). The microbiome of the upper respiratory tract in health and disease. *BMC Biol,* 17**,** 87.

Kunder, D. A., Cain, C. L., O'shea, K., Cole, S. D. & Rankin, S. C. (2015). Genotypic relatedness and antimicrobial resistance of *Staphylococcus schleiferi* in clinical samples from dogs in different geographic regions of the United States. *Veterinary Dermatology,* 26**,** 406-e94.

Lawton, E. M., Cotter, P. D., Hill, C. & Ross, R. P. (2007). Identification of a novel two-peptide lantibiotic, haloduracin, produced by the alkaliphile *Bacillus halodurans* C-125. *FEMS Microbiol Lett,* 267**,** 64-71.

Lazzaro, M., Feldman, M. F. & García Véscovi, E. (2017). A Transcriptional Regulatory Mechanism Finely Tunes the Firing of Type VI Secretion System in Response to Bacterial Enemies. *mBio,* 8.

Leclercq, R., Derlot, E., Duval, J. & Courvalin, P. (1988). Plasmid-mediated resistance to vancomycin and teicoplanin in *Enterococcus faecium*. *New England Journal of Medicine,* 319**,** 157-161.

Lecuyer, H., Audibert, J., Bobigny, A., Eckert, C., Janniere-Nartey, C., Buu-Hoi, A., Mainardi, J. L. & Podglajen, I. (2007). *Dolosigranulum pigrum* causing nosocomial pneumonia and septicemia. *J Clin Microbiol,* 45**,** 3474-5.

Leung, M. H. Y., Tong, X., Boifot, K. O., Bezdan, D., Butler, D. J., Danko, D. C., Gohli, J., Green, D. C., Hernandez, M. T., Kelly, F. J., Levy, S., Mason-Buck, G., Nieto-Caballero, M., Syndercombe-Court, D., Udekwu, K., Young, B. G., Mason, C. E., Dybwad, M. & Lee, P. K. H. (2021). Characterization of the public transit air microbiome and resistome reveals geographical specificity. *Microbiome,* 9**,** 112.

Lee, C. H., Kim, S., Hyun, B., Suh, J. W., Yon, C., Kim, C., Lim, Y. & Kim, C. (1994). Cepacidine A, a novel antifungal antibiotic produced by *Pseudomonas cepacia*. I. Taxonomy, production, isolation and biological activity. *J Antibiot (Tokyo),* 47**,** 1402-5.

Lee, V. T., J; Hassim, M.H; Chemmangattuvalappil , N. G. (2020). Design of Solvent Mixtures for Selective Extraction by Quantifying Thermodynamic and Sustainability Aspects. *Springer*

Lerminiaux, N. A. & Cameron, A. D. (2019). Horizontal transfer of antibiotic resistance genes in clinical environments. *Canadian journal of microbiology,* 65**,** 34-44.

Levison, M. E. & Levison, J. H. (2009). Pharmacokinetics and pharmacodynamics of antibacterial agents. *Infect Dis Clin North Am,* 23**,** 791-815, vii.

Leviton, I. (1999). Inhibitors of Protein Synthesis. *Cancer Investigation,* 17**,** 87-92.

Lewis, K. (2013). Platforms for antibiotic discovery. *Nat Rev Drug Discov,* 12**,** 371-87.

Li, B., Qiu, Y., Shi, H. & Yin, H. (2016). The importance of lag time extension in determining bacterial resistance to antibiotics. *Analyst,* 141**,** 3059-67.

Li, X., Hernandez, V., Rock, F. L., Choi, W., Mak, Y. S. L., Mohan, M., Mao, W., Zhou, Y., Easom, E. E., Plattner, J. J., Zou, W., Pérez-Herrán, E., Giordano, I., Mendoza-Losana, A., Alemparte, C., Rullas, J., Angulo-Barturen, I., Crouch, S., Ortega, F., Barros, D. & Alley, M. R. K. (2017). Discovery of a Potent and Specific *M. tuberculosis* Leucyl-tRNA Synthetase Inhibitor: (S)-3-(Aminomethyl)-4-chloro-7-(2-hydroxyethoxy)benzo[c][1,2]oxaborol-1(3H)-ol (GSK656). *J Med Chem,* 60**,** 8011-8026.

Li, H., Luo, Y. F., Williams, B. J., Blackwell, T. S. & Xie, C. M. (2012a). Structure and function of OprD protein in *Pseudomonas aeruginosa*: from antibiotic resistance to novel therapies. *Int J Med Microbiol,* 302**,** 63-8.

Lim, H. S., Yeu, J. E., Hong, S. P. & Kang, M. S. (2018). Characterization of Antibacterial Cell-Free Supernatant from Oral Care Probiotic Weissella cibaria, CMU. *Molecules,* 23.

Lima, W. G., Silva Alves, G. C., Sanches, C., Antunes Fernandes, S. O. & De Paiva, M. C. (2019). Carbapenem-resistant *Acinetobacter baumannii* in patients with burn injury: A systematic review and meta-analysis. *Burns,* 45**,** 1495-1508.

Lin, L., Zheng, Q., Wei, T., Zhang, Z., Zhao, C., Zhong, H., Xu, Q., Lin, J. and Guo, L., (2020). Isolation and Characterization of Fengycins Produced by *Bacillus amyloliquefaciens* JFL21 and Its Broad-Spectrum Antimicrobial Potential Against Multidrug-Resistant Foodborne Pathogens. *Frontiers in Microbiology*, 11.

Lister, P. D., Wolter, D. J. & Hanson, N. D. (2009). Antibacterial-resistant *Pseudomonas aeruginosa*: clinical impact and complex regulation of chromosomally encoded resistance mechanisms. *Clinical microbiology reviews,* 22**,** 582-610.

Liu, C. M., Price, L. B., Hungate, B. A., Abraham, A. G., Larsen, L. A., Christensen, K., Stegger, M., Skov, R. & Andersen, P. S. (2015). *Staphylococcus aureus* and the ecology of the nasal microbiome. *Sci Adv,* 1**,** e1400216.

Liu, Y., Link, H., Liu, L., Du, G., Chen, J. & Sauer, U. (2016). A dynamic pathway analysis approach reveals a limiting futile cycle in N-acetylglucosamine overproducing *Bacillus subtilis*. *Nat Commun,* 7**,** 11933.

Liu, Y., Dai, C., Zhou, Y., Qiao, J., Tang, B., Yu, W., Zhang, R., Liu, Y. and Lu, S., (2021). Pyoverdines Are Essential for the Antibacterial Activity of *Pseudomonas chlororaphis* YL-1 under Low-Iron Conditions. *Applied and Environmental Microbiology*, 87(7).

Liu, Y.-Y., Wang, Y., Walsh, T. R., Yi, L.-X., Zhang, R., Spencer, J., Doi, Y., Tian, G., Dong, B. & Huang, X. (2016). Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. *The Lancet infectious diseases,* 16**,** 161-168.

Livermore, D. M. 1990. Antibiotic uptake and transport by bacteria. *Scand J Infect Dis Suppl,* 74**,** 15-22.

Li, X., Xing, J., Li, B., Wang, P. & Liu, J. (2012b). Use of tuf as a target for sequence-based identification of Gram-positive cocci of the genus *Enterococcus*, *Streptococcus*, coagulase-negative *Staphylococcus*, and *Lactococcus*. *Annals of Clinical Microbiology and Antimicrobials,* 11**,** 31.

Lobritz, M. A., Andrews, I. W., Braff, D., Porter, C. B. M., Gutierrez, A., Furuta, Y., Cortes, L. B. G., Ferrante, T., Bening, S. C., Wong, F., Gruber, C., Bakerlee, C. W., Lambert, G., Walker, G. C., Dwyer, D. J. & Collins, J. J. (2022). Increased energy demand from anabolic-catabolic processes drives β-lactam antibiotic lethality. *Cell Chemical Biology,* 29**,** 276-286.e4.

Lohman, T. M. (1992). Escherichia coli DNA helicases: mechanisms of DNA unwinding. *Mol Microbiol,* 6**,** 5-14.

Lorme, F., Maataoui, N., Rondinaud, E., Esposito-Farèse, M., Clermont, O., Ruppe, E., Arlet, G., Genel, N., Matheron, S., Andremont, A. & Armand-Lefevre, L. (2018). Acquisition of plasmid-mediated cephalosporinase producing *Enterobacteriaceae* after a travel to the tropics. *PLoS One,* 13**,** e0206909.

Mac, S., Fitzpatrick, T., Johnstone, J. & Sander, B. (2019). Vancomycin-resistant enterococci (VRE) screening and isolation in the general medicine ward:

a cost-effectiveness analysis. *Antimicrobial Resistance & Infection Control,* 8**,** 168.

Maffioli, S. I., Sosio, M., Ebright, R. H. & Donadio, S. (2019). Discovery, properties, and biosynthesis of pseudouridimycin, an antibacterial nucleoside-analog inhibitor of bacterial RNA polymerase. *J Ind Microbiol Biotechnol,* 46**,** 335-343.

Malik, M., Zhao, X. & Drlica, K. (2006). Lethal fragmentation of bacterial chromosomes mediated by DNA gyrase and quinolones. *Molecular microbiology,* 61**,** 810-25.

Malinga, N. Z. Z., Shobo, C. O., Molechan, C., Amoako, D. G., Zishiri, O. T. & Bester, L. A. (2022). Molecular Surveillance and Dissemination of *Klebsiella pneumoniae* on Frequently Encountered Surfaces in South African Public Hospitals. *Microb Drug Resist,* 28**,** 306-316.

Mallet, A. I. (1998). Solid-phase extraction: principles and practice E. M Thurman and M. S MillS: vol 147 in series chemical analysis, ed. J. D. Winefordner, John Wiley and Sons, New York, 1998 ISBN 0471 61422X *Rapid Communications in Mass Spectrometry,* 12**,** 988-988.

Mantovani, H. C., Hu, H., Worobo, R. W. & Russell, J. B. (2002). Bovicin HC5, a bacteriocin from *Streptococcus bovis* HC5. *Microbiology,* 148**,** 3347-3352.

Marambio-Jones, C. & Hoek, E. M. V. (2010). A review of the antibacterial effects of silver nanomaterials and potential implications for human health and the environment. *Journal of Nanoparticle Research,* 12**,** 1531-1551.

Maranata, G. J., Surya, N. O. & Hasanah, A. N. (2021). Optimising factors affecting solid phase extraction performances of molecular imprinted polymer as recent sample preparation technique. *Heliyon,* 7**,** e05934.

Marcos-Torres, F. J., Maurer, D., Juniar, L. & Griese, J. J. (2021). The bacterial iron sensor IdeR recognizes its DNA targets by indirect readout. *Nucleic Acids Res,* 49**,** 10120-10135.

Markowitz, V. M., Chen, I. M., Palaniappan, K., Chu, K., Szeto, E., Pillay, M., Ratner, A., Huang, J., Woyke, T., Huntemann, M., Anderson, I., Billis, K., Varghese, N., Mavromatis, K., Pati, A., Ivanova, N. N. & Kyrpides, N. C. (2014). IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res,* 42**,** D560-7.

Martín, J. F., Casqueiro, J. & Liras, P. 2005. Secretion systems for secondary metabolites: how producer cells send out messages of intercellular communication. *Current Opinion in Microbiology,* 8**,** 282-293.

Martin H, C., Ibáñez, R., Nothias, L.-F., Boya P, C. A., Reinert, L. K., Rollins-Smith, L. A., Dorrestein, P. C. & Gutiérrez, M. (2019). Viscosin-like lipopeptides from frog skin bacteria inhibit *Aspergillus fumigatus* and *Batrachochytrium dendrobatidis* detected by imaging mass spectrometry and molecular networking. *Scientific Reports,* 9**,** 3019.

Martis, N., Leroy, S. & Blanc, V. (2014). Colistin in multi-drug resistant *Pseudomonas aeruginosa* blood-stream infections: a narrative review for the clinician. *J Infect,* 69**,** 1-12.

Marturano, J. E. & Lowery, T. J. (2019). ESKAPE Pathogens in Bloodstream Infections Are Associated With Higher Cost and Mortality but Can Be Predicted Using Diagnoses Upon Admission. *Open Forum Infect Dis,* 6**,** ofz503.

Mašlaňová, I., Stříbná, S., Doškař, J. & Pantůček, R. (2016). Efficient plasmid transduction to *Staphylococcus aureus* strains insensitive to the lytic action of transducing phage. *FEMS Microbiol Lett,* 363.

Matthijs, S., Vander Wauven, C., Cornu, B., Ye, L., Cornelis, P., Thomas, C. and Ongena, M., (2014). Antimicrobial properties of *Pseudomonas* strains producing the antibiotic mupirocin. *Research in Microbiology*, 165(8), pp.695-704.

Mcdermott, P. F., Walker, R. D. & White, D. G. (2003). Antimicrobials: modes of action and mechanisms of resistance. *Int J Toxicol,* 22**,** 135-43.

Mcdonagh, J. (1912). Some toxic effects of Salvarsan. *British Medical Journal,* 1**,** 272.

Mélançon, D. & Grenier, D. (2003). Production and properties of bacteriocin-like inhibitory substances from the swine pathogen *Streptococcus suis* serotype 2. *Appl Environ Microbiol,* 69**,** 4482-8.

Méndez, C. & Salas, J. A. (2001). The role of ABC transporters in antibiotic-producing organisms: drug secretion and resistance mechanisms. *Res Microbiol,* 152**,** 341-50.

Mentasti, M., Prime, K., Sands, K., Khan, S. & Wootton, M. (2020). Rapid detection of OXA-23-like, OXA-24-like, and OXA-58-like carbapenemases from *Acinetobacter* species by real-time PCR. *Journal of Hospital Infection,* 105**,** 741-746.

Michalska, M. & Wolf, P. 2015. *Pseudomonas* Exotoxin A: optimized by evolution for effective killing. *Frontiers in Microbiology,* 6.

Miethke, M., Pieroni, M., Weber, T., Brönstrup, M., Hammann, P., Halby, L., Arimondo, P. B., Glaser, P., Aigle, B., Bode, H. B., Moreira, R., Li, Y., Luzhetskyy, A., Medema, M. H., Pernodet, J.-L., Stadler, M., Tormo, J. R., Genilloud, O., Truman, A. W., Weissman, K. J., Takano, E., Sabatini, S., Stegmann, E., Brötz-Oesterhelt, H., Wohlleben, W., Seemann, M., Empting, M., Hirsch, A. K. H., Loretz, B., Lehr, C.-M., Titz, A., Herrmann, J., Jaeger, T., Alt, S., Hesterkamp, T., Winterhalter, M., Schiefer, A., Pfarr, K., Hoerauf, A., Graz, H., Graz, M., Lindvall, M., Ramurthy, S., Karlén, A., Van Dongen, M., Petkovic, H., Keller, A., Peyrane, F., Donadio, S., Fraisse, L., Piddock, L. J. V., Gilbert, I. H., Moser, H. E. & Müller, R. (2021). Towards the sustainable discovery and development of new antibiotics. *Nature Reviews Chemistry,* 5**,** 726-749.

Mihelič, M., Vlahoviček-Kahlina, K., Renko, M., Mesnage, S., Doberšek, A., Taler-Verčič, A., Jakas, A. & Turk, D. (2017). The mechanism behind the selection of two different cleavage sites in NAG-NAM polymers. *IUCrJ,* 4**,** 185-198.

Milani, R. V., Wilt, J. K., Entwisle, J., Hand, J., Cazabon, P. & Bohan, J. G. (2019). Reducing inappropriate outpatient antibiotic prescribing: normative comparison using unblinded provider reports. *BMJ Open Qual,* 8**,** e000351.

Miller, S. I. (2016). Antibiotic Resistance and Regulation of the Gram-Negative Bacterial Outer Membrane Barrier by Host Innate Immune Molecules. *mBio,* 7**,** e01541-16.

Mishra, M., Panda, S., Barik, S., Sarkar, A., Singh, D. V. & Mohapatra, H. (2020). Antibiotic Resistance Profile, Outer Membrane Proteins, Virulence Factors and Genome Sequence Analysis Reveal Clinical Isolates of

Enterobacter Are Potential Pathogens Compared to Environmental Isolates. *Frontiers in Cellular and Infection Microbiology,* 10.

Mishra, N. N., Bayer, A. S., Tran, T. T., Shamoo, Y., Mileykovskaya, E., Dowhan, W., Guan, Z. & Arias, C. A. (2012). Daptomycin Resistance in Enterococci Is Associated with Distinct Alterations of Cell Membrane Phospholipid Content. *PLOS ONE,* 7**,** e43958.

Mitsutomi, S., Sekimizu, K. & Kaito, C. (2017). Isolation of antibiotic-producing *Pseudomonas* species with low-temperature cultivation of temperate soil. *Drug Discov Ther,* 11**,** 267-275.

Mohd Asri, N. A., Ahmad, S., Mohamud, R., Mohd Hanafi, N., Mohd Zaidi, N. F., Irekeola, A. A., Shueb, R. H., Yee, L. C., Mohd Noor, N., Mustafa, F. H., Yean, C. Y. & Yusof, N. Y. (2021). Global Prevalence of Nosocomial Multidrug-Resistant *Klebsiella pneumoniae*: A Systematic Review and Meta-Analysis. *Antibiotics (Basel),* 10.

Mollema, F. P., Richardus, J. H., Behrendt, M., Vaessen, N., Lodder, W., Hendriks, W., Verbrugh, H. A. & Vos, M. C. (2010). Transmission of methicillin-resistant *Staphylococcus aureus* to household contacts. *J Clin Microbiol,* 48**,** 202-7.

Moradali, M. F., Ghods, S. & Rehm, B. H. (2017). *Pseudomonas aeruginosa* Lifestyle: A Paradigm for Adaptation, Survival, and Persistence. *Front Cell Infect Microbiol,* 7**,** 39.

Moreno, S. N., Mason, R. P., Muniz, R. P., Cruz, F. S. & Docampo, R. (1983). Generation of free radicals from metronidazole and other nitroimidazoles by Tritrichomonas foetus. *Journal of Biological Chemistry,* 258**,** 4051-4054.

Mosaei, H., Zenkin, N., Lovett, S. T. & Hinton, D. (2020). Inhibition of RNA Polymerase by Rifampicin and Rifamycin-Like Molecules. *EcoSal Plus,* 9.

Mott, M. L. & Berger, J. M. (2007). DNA replication initiation: mechanisms and regulation in bacteria. *Nat Rev Microbiol,* 5**,** 343-54.

Murray, B. E. (2000). Vancomycin-Resistant Enterococcal Infections. *New England Journal of Medicine,* 342**,** 710-721.

Murray, J., Muruko, T., Gill, C. I. R., Kearney, M. P., Farren, D., Scott, M. G., Mcmullan, G. & Ternan, N. G. (2017). Evaluation of bactericidal and anti-biofilm properties of a novel surface-active organosilane biocide against healthcare associated pathogens and *Pseudomonas aeruginosa* biolfilm. *PLoS One,* 12**,** e0182624.

Murray, C. J. L., Ikuta, K. S., Sharara, F., Swetschinski, L., Robles Aguilar, G., Gray, A., Han, C., Bisignano, C., Rao, P., Wool, E., Johnson, S. C., Browne, A. J., Chipeta, M. G., Fell, F., Hackett, S., Haines-Woodhouse, G., Kashef Hamadani, B. H., Kumaran, E. A. P., Mcmanigal, B., Agarwal, R., Akech, S., Albertson, S., Amuasi, J., Andrews, J., Aravkin, A., Ashley, E., Bailey, F., Baker, S., Basnyat, B., Bekker, A., Bender, R., Bethou, A., Bielicki, J., Boonkasidecha, S., Bukosia, J., Carvalheiro, C., Castañeda-Orjuela, C., Chansamouth, V., Chaurasia, S., Chiurchiù, S., Chowdhury, F., Cook, A. J., Cooper, B., Cressey, T. R., Criollo-Mora, E., Cunningham, M., Darboe, S., Day, N. P. J., De Luca, M., Dokova, K., Dramowski, A., Dunachie, S. J., Eckmanns, T., Eibach, D., Emami, A., Feasey, N., Fisher-Pearson, N., Forrest, K., Garrett, D., Gastmeier, P., Giref, A. Z.,

Greer, R. C., Gupta, V., Haller, S., Haselbeck, A., Hay, S. I., Holm, M., Hopkins, S., Iregbu, K. C., Jacobs, J., Jarovsky, D., Javanmardi, F., Khorana, M., Kissoon, N., Kobeissi, E., Kostyanev, T., Krapp, F., Krumkamp, R., Kumar, A., Kyu, H. H., Lim, C., Limmathurotsakul, D., Loftus, M. J., Lunn, M., Ma, J., Mturi, N., Munera-Huertas, T., Musicha, P., Mussi-Pinhata, M. M., Nakamura, T., Nanavati, R., Nangia, S., Newton, P., Ngoun, C., Novotney, A., Nwakanma, D., Obiero, C. W., Olivas-Martinez, A., Olliaro, P., Ooko, E., et al. (2022). Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet,* 399**,** 629-655.

Münch, D., Müller, A., Schneider, T., Kohl, B., Wenzel, M., Bandow, J. E., Maffioli, S., Sosio, M., Donadio, S., Wimmer, R. & Sahl, H. G. (2014). The lantibiotic NAI-107 binds to bactoprenol-bound cell wall precursors and impairs membrane functions. *J Biol Chem,* 289**,** 12063-12076.

Naskar, S., Hohl, M., Tassinari, M. & Low, H. H. (2021). The structure and mechanism of the bacterial type II secretion system. *Molecular Microbiology,* 115**,** 412-424.

Nazir, F., Ibrahim, M., Zaman, G., Hussain, A., Yar, A. M. & Bo, Z. (2018). Genetic Diversity and Functional Analysis of Sigma Factors in *Enterobacter cloacae* Complex Resourced From Various Niche. *Evol Bioinform Online,* 14**,** 1176934318754878.

Neve, R., Carrillo, B. and Phelan, V., (2021). Impact of Artificial Sputum Medium Formulation on *Pseudomonas aeruginosa* Secondary Metabolite Production. *Journal of Bacteriology*, 203(21).

Ndlovu, T., Rautenbach, M., Vosloo, J. A., Khan, S. & Khan, W. (2017). Characterisation and antimicrobial activity of biosurfactant extracts produced by *Bacillus amyloliquefaciens* and *Pseudomonas aeruginosa* isolated from a wastewater treatment plant. *AMB Express,* 7**,** 108.

Nguyen, A. T., Jones, J. W., Ruge, M. A., Kane, M. A. & Oglesby-Sherrouse, A. G. (2015). Iron Depletion Enhances Production of Antimicrobials by *Pseudomonas aeruginosa*. *J Bacteriol,* 197**,** 2265-75.

Nicholson, A. C., Bell, M., Humrighouse, B. W. & Mcquiston, J. R. (2015). Complete Genome Sequences for Two Strains of a Novel Fastidious, Partially Acid-Fast, Gram-Positive *Corynebacterineae* Bacterium, Derived from Human Clinical Samples. *Genome Announc,* 3.

Nielsen, K. F., Mansson, M., Rank, C., Frisvad, J. C. & Larsen, T. O. (2011). Dereplication of microbial natural products by LC-DAD-TOFMS. *J Nat Prod,* 74**,** 2338-48.

Nielsen, T., Sørensen, D., Tobiasen, C., Andersen, J., Christophersen, C., Givskov, M. and Sørensen, J., (2002). Antibiotic and Biosurfactant Properties of Cyclic Lipopeptides Produced by Fluorescent *Pseudomonas* spp. from the Sugar Beet Rhizosphere. *Applied and Environmental Microbiology*, 68(7), pp.3416-3423.

Ni, R. T., Onishi, M., Mizusawa, M., Kitagawa, R., Kishino, T., Matsubara, F., Tsuchiya, T., Kuroda, T. & Ogawa, W. (2020). The role of RND-type efflux pumps in multidrug-resistant mutants of *Klebsiella pneumoniae*. *Scientific Reports,* 10**,** 10876.

Nikaido, H. & Rosenberg, E. Y. 1981. Effect on solute size on diffusion rates through the transmembrane pores of the outer membrane of *Escherichia coli*. *Journal of General Physiology,* 77**,** 121-135.

Nikaido, H. (2003). Molecular basis of bacterial outer membrane permeability revisited. *Microbiol Mol Biol Rev,* 67**,** 593-656.

Nouwen, J. L., Fieren, M. W., Snijders, S., Verbrugh, H. A. & Van Belkum, A. (2005). Persistent (not intermittent) nasal carriage of *Staphylococcus aureus* is the determinant of CPD-related infections. *Kidney Int,* 67**,** 1084-92.

Oakley, A. J. (2019). A structural view of bacterial DNA replication. *Protein Sci,* 28**,** 990-1004.

Ochi, K. (2007). From microbial differentiation to ribosome engineering. *Biosci Biotechnol Biochem,* 71**,** 1373-86.

Ochi, K., Tanaka, Y. & Tojo, S. (2014). Activating the expression of bacterial cryptic genes by rpoB mutations in RNA polymerase or by rare earth elements. *Journal of Industrial Microbiology and Biotechnology,* 41**,** 403-414.

Ochi, K. (2017). Insights into microbial cryptic gene activation and strain improvement: principle, application and technical aspects. *The Journal of Antibiotics,* 70**,** 25-40.

Oelschlaeger, T. A. & Tall, B. D. (1997). Invasion of cultured human epithelial cells by *Klebsiella pneumoniae* isolated from the urinary tract. *Infect Immun,* 65**,** 2950-8

Ofek, I., Cohen, S., Rahmani, R., Kabha, K., Tamarkin, D., Herzig, Y. & Rubinstein, E. (1994). Antibacterial synergism of polymyxin B nonapeptide and hydrophobic antibiotics in experimental gram-negative infections in mice. *Antimicrob Agents Chemother,* 38**,** 374-7.

Oh, J., Conlan, S., Polley, E. C., Segre, J. A. & Kong, H. H. (2012). Shifts in human skin and nares microbiota of healthy children and adults. *Genome Med,* 4**,** 77.

Oliveira, D. M. P. D., Forde, B. M., Kidd, T. J., Harris, P. N. A., Schembri, M. A., Beatson, S. A., Paterson, D. L. & Walker, M. J. (2020). Antimicrobial Resistance in ESKAPE Pathogens. *Clinical Microbiology Reviews,* 33**,** e00181-19.

Oliveira, J. & Reygaert, W. C. (2022). Gram Negative Bacteria. *StatPearls.* Treasure Island (FL): StatPearls Publishing

Olsen, K., Falch, B. M., Danielsen, K., Johannessen, M., Ericson Sollid, J. U., Thune, I., Grimnes, G., Jorde, R., Simonsen, G. S. & Furberg, A. S. (2012). *Staphylococcus aureus* nasal carriage is associated with serum 25-hydroxyvitamin D levels, gender and smoking status. The Tromso Staph and Skin Study. *Eur J Clin Microbiol Infect Dis,* 31**,** 465-73.

Olsson, L. M., Boulund, F., Nilsson, S., Khan, M. T., Gummesson, A., Fagerberg, L., Engstrand, L., Perkins, R., Uhlén, M., Bergström, G., Tremaroli, V. & Bäckhed, F. (2022). Dynamics of the normal gut microbiota: A longitudinal

one-year population study in Sweden. *Cell Host & Microbe,* 30**,** 726-739.e3.

Omoboye, O. O., Geudens, N., Duban, M., Chevalier, M., Flahaut, C., Martins, J. C., Leclère, V., Oni, F. E. & Höfte, M. (2019a). Pseudomonas sp. COW3 Produces New Bananamide-Type Cyclic Lipopeptides with Antimicrobial Activity against *Pythium myriotylum* and *Pyricularia oryzae*. *Molecules,* 24.

Omoboye, O., Oni, F., Batool, H., Yimer, H., De Mot, R. and Höfte, M., (2019b). *Pseudomonas* Cyclic Lipopeptides Suppress the Rice Blast Fungus *Magnaporthe oryzae* by Induced Resistance and Direct Antagonism. *Frontiers in Plant Science*, 10.

O'Neill, J. & Grande-Bretagne (2014). Antimicrobial Resistance: Tackling a Crisis for the Health and Wealth of Nations, Review on Antimicrobial Resistance.

Ono, S., Suzuki, S., Ito, D., Tagawa, S., Shiina, T. & Masuda, S. (2020). Plastidial (p)ppGpp Synthesis by the Ca2+-Dependent RelA–SpoT Homolog Regulates the Adaptation of Chloroplast Gene Expression to Darkness in Arabidopsis. *Plant and Cell Physiology,* 61**,** 2077-2086.

Ötles, S. & Kartal, C. (2016). Solid-Phase Extraction (SPE): Principles and Applications in Food Samples. *Acta Sci Pol Technol Aliment,* 15**,** 5-15.

Ovung, A. & Bhattacharyya, J. (2021). Sulfonamide drugs: structure, antibacterial property, toxicity, and biophysical interactions. *Biophys Rev,* 13**,** 259-272.

Özen, A. I. & Ussery, D. W. (2012). Defining the *Pseudomonas* genus: where do we draw the line with Azotobacter? *Microb Ecol,* 63**,** 239-48.

Pajerski, W., Ochonska, D., Brzychczy-Wloch, M., Indyka, P., Jarosz, M., Golda-Cepa, M., Sojka, Z. & Kotarba, A. (2019). Attachment efficiency of gold nanoparticles by Gram-positive and Gram-negative bacterial strains governed by surface charges. *Journal of Nanoparticle Research,* 21**,** 186.

Page, B., Page, M. & Noel, C. (1993). A new fluorometric assay for cytotoxicity measurements in-vitro. *Int J Oncol,* 3**,** 473-6.

Palaniappan, K., Chen, I. A., Chu, K., Ratner, A., Seshadri, R., Kyrpides, N. C., Ivanova, N. N. & Mouncey, N. J. (2020). IMG-ABC v.5.0: an update to the IMG/Atlas of Biosynthetic Gene Clusters Knowledgebase. *Nucleic Acids Res,* 48**,** D422-D430.

Pandey, N. & Cascella, M. (2022). Beta Lactam Antibiotics. *StatPearls.* Treasure Island (FL): StatPearls Publishing

Paterson, D. L. & Bonomo, R. A. (2005). Extended-spectrum beta-lactamases: a clinical update. *Clin Microbiol Rev,* 18**,** 657-86.

Peek, M. E., Bhatnagar, A., Mccarty, N. A. & Zughaier, S. M. (2012). Pyoverdine, the Major Siderophore in *Pseudomonas aeruginosa*, Evades NGAL Recognition. *Interdiscip Perspect Infect Dis,* 2012**,** 843509.

Perl, T. M., Cullen, J. J., Wenzel, R. P., Zimmerman, M. B., Pfaller, M. A., Sheppard, D., Twombley, J., French, P. P., Herwaldt, L. A., (2002) Mupirocin & The Risk of *Staphylococcus aureus* Study Intranasal mupirocin to prevent postoperative *Staphylococcus aureus* infections. *N Engl J Med,* 346**,** 1871-7.

Peters, N. K., Dixon, D. M., Holland, S. M. & Fauci, A. S. (2008). The Research Agenda of the National Institute of Allergy and Infectious Diseases for Antimicrobial Resistance. *The Journal of Infectious Diseases,* 197**,** 1087-1093.

Pettit, R. K., Weber, C. A. & Pettit, G. R. (2009). Application of a high throughput Alamar blue biofilm susceptibility assay to *Staphylococcus aureus* biofilms. *Ann Clin Microbiol Antimicrob,* 8**,** 28.

PEW (2021) Antibiotics Currently in Global Clinical Development [Online]. Available online at: https://www.pewtrusts.org/en/research-and-analysis/data-visualizations/2014/antibiotics-currently-in-clinical-development (Accessed 8 November 2021).

Peypoux, F., Bonmatin, J. M. & Wallach, J. (1999). Recent trends in the biochemistry of surfactin. *Applied Microbiology and Biotechnology,* 51**,** 553-563.

Piddock, L. J. (2006). Multidrug-resistance efflux pumps? not just for resistance. *Nature Reviews Microbiology,* 4**,** 629-636.

Piddock, L. J. (2015). Teixobactin, the first of a new class of antibiotics discovered by iChip technology? *J Antimicrob Chemother,* 70**,** 2679-80.

Pilla, G., Mcvicker, G. & Tang, C. M. (2017). Genetic plasticity of the *Shigella* virulence plasmid is mediated by intra- and inter-molecular events between insertion sequences. *PLoS Genet,* 13**,** e1007014.

Pimentel-Filho, N., Martins, M., Nogueira, G., Mantovani, H. & Vanetti, M. (2014). Bovicin HC5 and nisin reduce *Staphylococcus aureus* adhesion to polystyrene and change the hydrophobicity profile and Gibbs free energy of adhesion. *International journal of food microbiology,* 190C**,** 1-8.

Pinto, G. A., Durão, F. O., Fiúza, A. M. A., Guimarães, M. M. B. L. & Madureira, C. M. N. (2004). Design optimisation study of solvent extraction: chemical reaction, mass transfer and mixer–settler hydrodynamics. *Hydrometallurgy,* 74**,** 131-147.

Pogliano, J., Pogliano, N. & Silverman, J. A. (2012). Daptomycin-mediated reorganization of membrane architecture causes mislocalization of essential cell division proteins. *J Bacteriol,* 194**,** 4494-504.

Poirel, L., Naas, T. & Nordmann, P. (2010). Diversity, epidemiology, and genetics of class D beta-lactamases. *Antimicrob Agents Chemother,* 54**,** 24-38.

Poirel, L., Jayol, A. & Nordmann, P. (2017). Polymyxins: antibacterial activity, susceptibility testing, and resistance mechanisms encoded by plasmids or chromosomes. *Clinical microbiology reviews,* 30**,** 557-596.

Poole, K. (2011). *Pseudomonas aeruginosa*: resistance to the max. *Front Microbiol,* 2**,** 65.

Pourahmad Jaktaji, R. & Jazayeri, N. 2013. Expression of *acrA* and *acrB* Genes in *Esherichia coli* Mutants with or without marR or acrR Mutations. *Iran J Basic Med Sci,* 16**,** 1254-8.

Pratiwi, R., Hidayat, I., Hanafi, M. and Mangunwardoyo, W., (2017). Antibacterial compound produced by *Pseudomonas aeruginosa* strain UICC B-40, an endophytic bacterium isolated from *Neesia altissima*. *Journal of Microbiology*, 55(4), pp.289-295.

Prozorov, A. A. (2005). The Bacterial Cell Cycle: DNA Replication, Nucleoid Segregation, and Cell Division. *Microbiology,* 74**,** 375-387.

Quaiyum, S., Fortunato, H., Gonzaga, LJ. and Okabe, S. (2018). Antimicrobial Activity in the Marine Cheilostome Bryozoan Cryptosula zavjalovensis Kubanin, 1976. *Journal of Antimicrobial Agents*, 04(03).

Qureshi, K. A., Bholay, A. D., Rai, P. K., Mohammed, H. A., Khan, R. A., Azam, F., Jaremko, M., Emwas, A. H., Stefanowicz, P., Waliczek, M., Kijewska, M., Ragab, E. A., Rehan, M., Elhassan, G. O., Anwar, M. J. & Prajapati, D. K. (2021). Isolation, characterization, anti-MRSA evaluation, and in-silico multi-target anti-microbial validations of actinomycin X2 and actinomycin D produced by novel *Streptomyces smyrnaeus* UKAQ_23. *Sci Rep,* 11**,** 14539.

Radajewski, S., Webster, G., Reay, D. S., Morris, S. A., Ineson, P., Nedwell, D. B., Prosser, J. I. & Murrell, J. C. (2002). Identification of active methylotroph populations in an acidic forest soil by stable-isotope probingccThe GenBank accession numbers for the sequences reported in this paper are AY080911–AY080961. *Microbiology,* 148**,** 2331-2342.

Raetz, C. R. & Whitfield, C. (2002). Lipopolysaccharide endotoxins. *Annu Rev Biochem,* 71**,** 635-700.

Ramirez, D. & Giron, M. (2022). Enterobacter Infections. *StatPearls.* Treasure Island (FL): StatPearls Publishing

Rasigade, J.-P. & Vandenesch, F. (2014). *Staphylococcus aureus*: A pathogen with still unresolved issues. *Infection, Genetics and Evolution,* 21**,** 510-514.

Raya Tonetti, F., Tomokiyo, M., Ortiz Moyano, R., Quilodran-Vega, S., Yamamuro, H., Kanmani, P., Melnikov, V., Kurata, S., Kitazawa, H. & Villena, J. (2021). The Respiratory Commensal Bacterium *Dolosigranulum pigrum* 040417 Improves the Innate Immune Response to Streptococcus pneumoniae. *Microorganisms,* 9.

Redfern, J., Wallace, J., Van Belkum, A., Jaillard, M., Whittard, E., Ragupathy, R., Verran, J., Kelly, P. & Enright, M. C. (2021). Biofilm associated genotypes of multiple antibiotic resistant *Pseudomonas aeruginosa*. *BMC Genomics,* 22**,** 572.

Reen, F. J., Barret, M., Fargier, E., O'muinneacháin, M. & O'gara, F. (2013). Molecular evolution of LysR-type transcriptional regulation in *Pseudomonas aeruginosa*. *Mol Phylogenet Evol,* 66**,** 1041-9.

Reen, F. J., Romano, S., Dobson, A. D. & O'gara, F. 2015. The Sound of Silence: Activating Silent Biosynthetic Gene Clusters in Marine Microorganisms. *Mar Drugs,* 13**,** 4754-83.

Rehman, T. U., Aslam, R., Aqib, A. I., Mohsin, M., Manzoor, A., Shoaib, M., Naseer, M. A., Hasan, A., Sattar, H., Fakhar, E. A. K. M., Muzammil, I. & Yao, W. (2020). Phylogeny of hospital acquired MRSA, and its comparative phenotypic clinico-epidemiology with vancomycin resistant S. aureus (VRSA). *Microb Pathog,* 149**,** 104537.

Renye, J. A. & Steinberg, D. H. (2021). Thermophilin 110 inhibits growth and biofilm formation of *Streptococcus mutans. Biotechnology Reports,* 31**,** e00647.

Reygaert, W. C. (2018). An overview of the antimicrobial resistance mechanisms of bacteria. *AIMS Microbiol,* 4**,** 482-501.

Rieusset, L., Rey, M., Muller, D., Vacheron, J., Gerin, F., Dubost, A., Comte, G. and Prigent-Combaret, C., (2020). Secondary metabolites from plant-associated *Pseudomonas* are overproduced in biofilm. *Microbial Biotechnology*, 13(5), pp.1562-1580

Rio, L. A. D., Olivares, J., Blesa, M. C. & Mayor, F. (1972). Antibiotics from *Pseudomonas reptilivora* Taxonomic Classification and Optimal Conditions of Fermentation for Antibiotic Production. *Antimicrobial Agents and Chemotherapy,* 2**,** 186-188.

Ripa, F. A., Nikkon, F., Zaman, S. & Khondkar, P. (2009). Optimal Conditions for Antimicrobial Metabolites Production from a New *Streptomyces* sp. RUPA-08PR Isolated from Bangladeshi Soil. *Mycobiology,* 37**,** 211-214.

Riquelme, S., Liimatta, K., Wong Fok Lung, T., Fields, B., Ahn, D., Chen, D., Lozano, C., Sáenz, Y., Uhlemann, A., Kahl, B., Britto, C., DiMango, E. and Prince, A., (2020). *Pseudomonas aeruginosa* Utilizes Host-Derived Itaconate to Redirect Its Metabolism to Promote Biofilm Formation. *Cell Metabolism*, 31(6), pp.1091-1106.e6.

Rohde, M., Fischetti, V. A., Novick, R. P., Ferretti, J. J., Portnoy, D. A., Braunstein, M. & Rood, J. I. (2019). The Gram-Positive Bacterial Cell Wall. *Microbiology Spectrum,* 7**,** 7.3.10.

Romano, S., Jackson, S. A., Patry, S. & Dobson, A. D. W. (2018). Extending the "One Strain Many Compounds" (OSMAC) Principle to Marine Microorganisms. *Marine drugs,* 16**,** 244.

Rubin, R. P. (2007). A brief history of great discoveries in pharmacology: in celebration of the centennial anniversary of the founding of the American Society of Pharmacology and Experimental Therapeutics. *Pharmacol Rev,* 59**,** 289-359.

Rutledge, P. J. & Challis, G. L. (2015). Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat Rev Microbiol,* 13**,** 509-23.

Sachdeva, S. & Reynolds, K. A. (2008). *Mycobacterium tuberculosis* β-Ketoacyl Acyl Carrier Protein Synthase III (mtFabH) Assay: Principles and Method. *In:* CHAMPNEY, W. S. (ed.) *New Antibiotic Targets.* Totowa, NJ: Humana Press.

Saini, A. K. & Kumar, V. (2021). Chapter 2 - Ribosome structure. *In:* KUMAR, V. (ed.) *Emerging Concepts in Ribosome Structure, Biogenesis, and Function.* Academic Press.

Sajeed Ali, S. & Vidhale, N. (2011). Evaluation of Siderophore produced by different Clinical isolate Pseudomonas aeruginosa. *International Journal of Microbiology Research,* 03**,** 131-135.

Sakamoto, Y., Yamauchi, Y., Jo, T., Michihata, N., Hasegawa, W., Takeshima, H., Matsui, H., Fushimi, K., Yasunaga, H. & Nagase, T. (2021). In-hospital mortality associated with community-acquired pneumonia due to methicillin-resistant *Staphylococcus aureus*: a matched-pair cohort study. *BMC Pulmonary Medicine,* 21**,** 345.

Sanchez, L. A., Sierra, M. G., Sineriz, F. & Delgado, O. (2013). Andrimid production at low temperature by a psychrotolerant *Serratia proteamaculans* strain. *World J Microbiol Biotechnol,* 29**,** 1773-81.

Santajit, S. & Indrawattana, N. (2016). Mechanisms of Antimicrobial Resistance in ESKAPE Pathogens. *Biomed Res Int,* 2016**,** 2475067.

Santamaria, G., Liao, C., Lindberg, C., Chen, Y., Wang, Z., Rhee, K., Pinto, F. R., Yan, J. & Xavier, J. B. (2022). Evolution and regulation of microbial secondary metabolism. *eLife,* 11**,** e76119.

Sasirekha, B. & Srividya, S. (2016). Siderophore production by *Pseudomonas aeruginosa* FP6, a biocontrol strain for Rhizoctonia solani and Colletotrichum gloeosporioides causing diseases in chilli. *Agriculture and Natural Resources,* 50**,** 250-256.

Schäfer, A. B. & Wenzel, M. (2020). A How-To Guide for Mode of Action Analysis of Antimicrobial Peptides. *Front Cell Infect Microbiol,* 10**,** 540898.

Schenck, L. P., Surette, M. G. & Bowdish, D. M. (2016). Composition and immunological significance of the upper respiratory tract microbiota. *FEBS Lett,* 590**,** 3705-3720.

Scherer, K., Wiedemann, I., Ciobanasu, C., Sahl, H.-G. & Kubitscheck, U. (2013). Aggregates of nisin with various bactoprenol-containing cell wall precursors differ in size and membrane permeation capacity. *Biochimica et Biophysica Acta (BBA) - Biomembranes,* 1828**,** 2628-2636.

Schiessl, K., Janssen, E., Kraemer, S., McNeill, K. and Ackermann, M., (2017). Magnitude and Mechanism of Siderophore-Mediated Competition at Low Iron Solubility in the *Pseudomonas aeruginosa* Pyochelin System. *Frontiers in Microbiology*, 8.

Schwarz, J., Hubmann, G., Rosenthal, K. & Lutz, S. (2021). Triaging of Culture Conditions for Enhanced Secondary Metabolite Diversity from Different Bacteria. *Biomolecules,* 11.

Seltmann, G. & Holst, O. (2002). Periplasmic Space and Rigid Layer. *In:* Seltmann, G. & Holst, O. (eds.) *The Bacterial Cell Wall.* Berlin, Heidelberg: Springer Berlin Heidelberg.

Severinov, K. (2000). RNA polymerase structure-function: insights into points of transcriptional regulation. *Curr Opin Microbiol,* 3**,** 118-25.

Severinov, K., Mustaev, A., Severinova, E., Kozlov, M., Darst, S. A. & Goldfarb, A. (1995). The beta subunit Rif-cluster I is only angstroms away from the active center of *Escherichia coli* RNA polymerase. *J Biol Chem,* 270**,** 29428-32.

Shaalan, R. A. & Belal, T. S. 2010. HPLC-DAD stability indicating determination of nitrofurazone and lidocaine hydrochloride in their combined topical dosage form. *J Chromatogr Sci,* 48**,** 647-53.

Shalaby, M.-A. W., Dokla, E. M. E., Serya, R. A. T. & Abouzid, K. A. M. (2020). Penicillin binding protein 2a: An overview and a medicinal chemistry perspective. *European Journal of Medicinal Chemistry,* 199**,** 112312.

Shariati, A., Dadashi, M., Moghadam, M. T., Van Belkum, A., Yaslianifard, S. & Darban-Sarokhalil, D. (2020). Global prevalence and distribution of vancomycin resistant, vancomycin intermediate and heterogeneously vancomycin intermediate *Staphylococcus aureus* clinical isolates: a systematic review and meta-analysis. *Scientific Reports,* 10**,** 12689.

Shilts, M. H., Rosas-Salazar, C., Tovchigrechko, A., Larkin, E. K., Torralba, M., Akopov, A., Halpin, R., Peebles, R. S., Moore, M. L., Anderson, L. J., Nelson, K. E., Hartert, T. V. & Das, S. R. (2016). Minimally Invasive Sampling Method Identifies Differences in Taxonomic Richness of Nasal Microbiomes in Young Infants Associated with Mode of Delivery. *Microb Ecol,* 71**,** 233-42.

Sharma, R., Jamwal, V., Singh, V. P., Wazir, P., Awasthi, P., Singh, D., Vishwakarma, R. A., Gandhi, S. G. & ChaubeY, A. (2017). Revelation and cloning of valinomycin synthetase genes in *Streptomyces lavendulae* ACR-DA1 and their expression analysis under different fermentation and elicitation conditions. *J Biotechnol,* 253**,** 40-47.

Sheng, H., Huang, J., Han, Z., Liu, M., Lü, Z., Zhang, Q., Zhang, J., Yang, J., Cui, S. & Yang, B. (2020). Genes and Proteomes Associated with Increased Mutation Frequency and Multidrug Resistance of Naturally Occurring Mismatch Repair-Deficient Salmonella Hypermutators. *Frontiers in Microbiology,* 11.

Shentu, X., Liu, N., Tang, G., Tanaka, Y., Ochi, K., Xu, J. & Yu, X. (2016). Improved antibiotic production and silent gene activation in Streptomyces diastatochromogenes by ribosome engineering. *J Antibiot (Tokyo),* 69**,** 406-10.

Shima, J., Hesketh, A., Okamoto, S., Kawamoto, S. & Ochi, K. (1996). Induction of actinorhodin production by rpsL (encoding ribosomal protein S12)

mutations that confer streptomycin resistance in *Streptomyces lividans* and *Streptomyces coelicolor* A3(2). *J Bacteriol,* 178**,** 7276-84.

Shi, L., Mcmillan, J. R., Yu, D., Chen, X., Tucker, C. J., Wasserman, E., Mohler, C. & Chen, Z. 2021. Effect of Surfactant Concentration and Hydrophobicity on the Ordering of Water at a Silica Surface. *Langmuir,* 37**,** 10806-10817.

Shostak, K., Bonner, C., Sproule, A., Thapa, I., Shields, S. W. J., Blackwell, B., Vierula, J., Overy, D. & Subramaniam, R. (2020). Activation of biosynthetic gene clusters by the global transcriptional regulator TRI6 in *Fusarium graminearum*. *Molecular Microbiology,* 114**,** 664-680.

Sidda, J. D., Song, L., Poon, V., Al-Bassam, M., Lazos, O., Buttner, M. J., Challis, G. L. & CorrE, C. (2014). Discovery of a family of γ-aminobutyrate ureas via rational derepression of a silent bacterial gene cluster. *Chemical Science,* 5**,** 86-89.

Silhavy, T. J., Kahne, D. & Walker, S. (2010). The bacterial cell envelope. *Cold Spring Harb Perspect Biol,* 2**,** a000414.

Simionato, A. S., Navarro, M. O. P., De Jesus, M. L. A., Barazetti, A. R., Da Silva, C. S., Simões, G. C., Balbi-Peña, M. I., De Mello, J. C. P., Panagio, L. A., De Almeida, R. S. C., Andrade, G. & De Oliveira, A. G. (2017). The Effect of Phenazine-1-Carboxylic Acid on Mycelial Growth of Botrytis cinerea Produced by *Pseudomonas aeruginosa* LV Strain. *Frontiers in Microbiology,* 8.

Skinnider, M., Dejong, C., Rees, P., Johnston, C., Li, H., Webster, A., Wyatt, M. and Magarvey, N., (2015). Genomes to natural products PRediction Informatics for Secondary Metabolomes (PRISM). *Nucleic Acids Research*, p.gkv1012.

Skinnider, M., Merwin, N., Johnston, C. and Magarvey, N., (2017). PRISM 3: expanded prediction of natural product chemical structures from microbial genomes. *Nucleic Acids Research*, 45(W1), pp.W49-W54.

Skold, O. (2000). Sulfonamide resistance: mechanisms and trends. *Drug Resist Updat,* 3**,** 155-160.

Slack, P. (1989). The black death past and present. 2. Some historical problems. *Trans R Soc Trop Med Hyg,* 83**,** 461-3.

Sorensen, K. N., Kim, K. H. & Takemoto, J. Y. (1996). In vitro antifungal and fungicidal activities and erythrocyte toxicities of cyclic lipodepsinonapeptides produced by *Pseudomonas syringae* pv. syringae. *Antimicrob Agents Chemother,* 40**,** 2710-3.

Spapen, H., Jacobs, R., Van Gorp, V., Troubleyn, J. & Honoré, P. M. (2011). Renal and neurological side effects of colistin in critically ill patients. *Ann Intensive Care,* 1**,** 14.

Springer, B., Kidan, Y. G., Prammananan, T., Ellrott, K., Böttger, E. C. & Sander, P. (2001). Mechanisms of streptomycin resistance: selection of mutations in the 16S rRNA gene conferring resistance. *Antimicrob Agents Chemother,* 45**,** 2877-84.

Stokes, J. M., Lopatkin, A. J., Lobritz, M. A. & Collins, J. J. (2019). Bacterial Metabolism and Antibiotic Efficacy. *Cell Metab,* 30**,** 251-259.

Stokes, J. M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N. M., Macnair, C. R., French, S., Carfrae, L. A., Bloom-Ackermann, Z., Tran, V. M., Chiappino-Pepe, A., Badran, A. H., Andrews, I. W., Chory, E. J., Church, G. M., Brown, E. D., Jaakkola, T. S., Barzilay, R. & Collins, J. J. (2020). A Deep Learning Approach to Antibiotic Discovery. *Cell,* 180**,** 688-702.e13.

Storm, D. R. (1974). Mechanism of bacitracin action: a specific lipid-peptide interaction. *Ann N Y Acad Sci,* 235**,** 387-98.

Sugarman, B. & Epps, L. R. (1982). Effect of estrogens on bacterial adherence to HeLa cells. *Infect Immun,* 35**,** 633-8.

Sugawara, E., Kojima, S. & Nikaido, H. (2016). *Klebsiella pneumoniae* Major Porins OmpK35 and OmpK36 Allow More Efficient Diffusion of β-Lactams than Their *Escherichia coli* Homologs OmpF and OmpC. *J Bacteriol,* 198**,** 3200-3208.

Sulek, K. (1968). [Nobel prize for Gerhard Domagk in 1939 for discovery of the antibacterial activity of prontosil]. *Wiad Lek,* 21**,** 1089.

Sur, S., Romo, T. and Grossfield, A., (2018). Selectivity and Mechanism of Fengycin, an Antimicrobial Lipopeptide, from Molecular Dynamics. *The Journal of Physical Chemistry B*, 122(8), pp.2219-2226.

Suresh, P., Varathraju, G., Shanmugaiah, V., Almaary, K. S., Elbadawi, Y. B. & Mubarak, A. (2021). Partial purification and characterization of 2, 4-diacetylphloroglucinol producing Pseudomonas fluorescens VSMKU3054 against bacterial wilt disease of tomato. *Saudi J Biol Sci,* 28**,** 2155-2167.

Sutherland, R., Boon, R. J., Griffin, K. E., Masters, P. J., Slocombe, B. & White, A. R. (1985). Antibacterial activity of mupirocin (pseudomonic acid), a new antibiotic for topical use. *Antimicrob Agents Chemother,* 27**,** 495-8.

Svetlov, M. S., Koller, T. O., Meydan, S., Shankar, V., Klepacki, D., Polacek, N., Guydosh, N. R., Vázquez-Laslop, N., Wilson, D. N. & Mankin, A. S. (2021). Context-specific action of macrolide antibiotics on the eukaryotic ribosome. *Nature Communications,* 12**,** 2803.

Szita, G., Tabajdi, V., Fábián, A., Biró, G., Reichart, O. and Körmöczy, P., (1998). A novel, selective synthetic acetamide containing culture medium for isolating *Pseudomonas aeruginosa* from milk. *International Journal of Food Microbiology*, 43(1-2), pp.123-127.

Tally, F. P. & Debruin, M. F. (2000). Development of daptomycin for gram-positive infections. *J Antimicrob Chemother,* 46**,** 523-6.

Tally, F. P., Zeckel, M., Wasilewski, M. M., Carini, C., Berman, C. L., Drusano, G. L. & Oleson, F. B., JR. (1999). Daptomycin: a novel agent for Gram-positive infections. *Expert Opin Investig Drugs,* 8**,** 1223-38.

Tamehiro, N., Hosaka, T., Xu, J., Hu, H., Otake, N. & Ochi, K. (2003). Innovative approach for improvement of an antibiotic-overproducing industrial strain of *Streptomyces albus*. *Appl Environ Microbiol,* 69**,** 6412-7.

Tanaka, Y., Kasahara, K., Hirose, Y., Murakami, K., Kugimiya, R. & OchI, K. (2013). Activation and products of the cryptic secondary metabolite biosynthetic gene clusters by rifampin resistance (*rpoB*) mutations in actinomycetes. *J Bacteriol,* 195**,** 2959-70.

Tang, Y. Q. & Weng, N. (2013). Salting-out assisted liquid–liquid extraction for bioanalysis. *Bioanalysis,* 5**,** 1583-1598.

Tan, S. Y. & Tatsumura, Y. (2015). Alexander Fleming (1881-1955): Discoverer of penicillin. *Singapore Med J,* 56**,** 366-7.

Tawiah, A. A., Gbedema, S. Y., Adu, F., Boamah, V. E. & Annan, K. (2012). Antibiotic producing microorganisms from River Wiwi, Lake Bosomtwe and the Gulf of Guinea at Doakor Sea Beach, Ghana. *BMC Microbiology,* 12**,** 234.

Telke, A. A., Olaitan, A. O., Morand, S. & Rolain, J.-M. (2017). soxRS induces colistin hetero-resistance in *Enterobacter asburiae* and *Enterobacter cloacae* by regulating the acrAB-tolC efflux pump. *Journal of Antimicrobial Chemotherapy,* 72**,** 2715-2721.

Thakur, P., Saini, N. K., Thakur, V. K., Gupta, V. K., Saini, R. V. & Saini, A. K. (2021). Rhamnolipid the Glycolipid Biosurfactant: Emerging trends and promising strategies in the field of biotechnology and biomedicine. *Microbial Cell Factories,* 20**,** 1.

Thavasi, R., Sharma, S. and Jayalakshmi, S., (2013). Evaluation of Screening Methods for the Isolation of Biosurfactant Producing Marine Bacteria. *Journal of Petroleum & Environmental Biotechnology*, 04(02).

Theophel, K., Schacht, V. J., Schlüter, M., Schnell, S., Stingu, C.-S., Schaumann, R. & Bunge, M. (2014). The importance of growth kinetic analysis in determining bacterial susceptibility against antibiotics and silver nanoparticles. *Frontiers in Microbiology,* 5.

Theuretzbacher, U., Outterson, K., Engel, A. & Karlen, A. (2020). The global preclinical antibacterial pipeline. *Nat Rev Microbiol,* 18**,** 275-285.

Thomas, C. M. & Nielsen, K. M. (2005). Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria. *Nature Reviews Microbiology,* 3**,** 711-721.

Thomashow, L. S., Weller, D. M., Bonsall, R. F. & Pierson, L. S. (1990). Production of the antibiotic phenazine-1-carboxylic Acid by fluorescent *pseudomonas* species in the rhizosphere of wheat. *Appl Environ Microbiol,* 56**,** 908-12.

Thöming, J., Tomasch, J., Preusse, M., Koska, M., Grahl, N., Pohl, S., Willger, S., Kaever, V., Müsken, M. and Häussler, S., (2020). Parallel evolutionary paths to produce more than one *Pseudomonas aeruginosa* biofilm phenotype. *npj Biofilms and Microbiomes*, 6(1).

Thrum, H. (1977). [Classes of antibiotics and their mechanisms of action]. *Z Gesamte Inn Med,* 32**,** Suppl: 209-14.

Tindall, B. J., Sutton, G. & GarritY, G. M. (2017). *Enterobacter aerogenes* Hormaeche and Edwards 1960 (Approved Lists 1980) and *Klebsiella mobilis* Bascomb et al. 1971 (Approved Lists 1980) share the same nomenclatural type (ATCC 13048) on the Approved Lists and are homotypic synonyms, with consequences for the name *Klebsiella mobilis* Bascomb et al. 1971 (Approved Lists 1980). *International Journal of Systematic and Evolutionary Microbiology,* 67**,** 502-504.

Trimble, M. J., Mlynárčik, P., Kolář, M. & Hancock, R. E. (2016). Polymyxin: alternative mechanisms of action and resistance. *Cold Spring Harbor perspectives in medicine,* 6**,** a025288.

Troppens, D. M., Dmitriev, R. I., Papkovsky, D. B., O'gara, F. & Morrissey, J. P. (2013). Genome-wide investigation of cellular targets and mode of action of the antifungal bacterial metabolite 2,4-diacetylphloroglucinol in *Saccharomyces cerevisiae*. *FEMS Yeast Res,* 13**,** 322-34.

Tiwari, V., Meena, K. And Tiwari, M., (2018). Differential anti-microbial secondary metabolites in different ESKAPE pathogens explain their adaptation in the hospital setup. *Infection, Genetics and Evolution*, 66, pp.57-65.

Tsai, Y. K., Liou, C. H., Lin, J. C., Fung, C. P., Chang, F. Y. & Siu, L. K. (2020). Effects of different resistance mechanisms on antimicrobial resistance in *Acinetobacter baumannii*: a strategic system for screening and activity testing of new antibiotics. *Int J Antimicrob Agents,* 55**,** 105918.

Tshepelevitsh, S., Hernits, K., Jenco, J., Hawkins, J. M., Muteki, K., Solich, P. & Leito, I. (2017). Systematic Optimization of Liquid-Liquid Extraction for Isolation of Unidentified Components. *ACS Omega,* 2**,** 7772-7776.

Tsubery, H., Ofek, I., Cohen, S. & Fridkin, M. (2000). Structure-function studies of polymyxin B nonapeptide: implications to sensitization of gram-negative bacteria. *J Med Chem,* 43**,** 3085-92.

Tugrul, T. and Cansunar, E., (2005). Detecting Surfactant-producing Microorganisms by the Drop-collapse Test. *World Journal of Microbiology and Biotechnology*, 21(6-7), pp.851-853.

Tumbarello, M., De Pascale, G., Trecarichi, E. M., Spanu, T., Antonicelli, F., Maviglia, R., Pennisi, M. A., Bello, G. & Antonelli, M. (2013). Clinical

outcomes of *Pseudomonas aeruginosa* pneumonia in intensive care unit patients. *Intensive Care Med,* 39**,** 682-92.

Uddin, T. M., Chakraborty, A. J., Khusro, A., Zidan, B. M. R. M., Mitra, S., Emran, T. B., Dhama, K., Ripon, M. K. H., Gajdács, M., Sahibzada, M. U. K., Hossain, M. J. & Koirala, N. (2021). Antibiotic resistance in microbes: History, mechanisms, therapeutic strategies and future prospects. *Journal of Infection and Public Health,* 14**,** 1750-1766.

Uehara, Y., Kikuchi, K., Nakamura, T., Nakama, H., Agematsu, K., Kawakami, Y., Maruchi, N. & Totsuka, K. (2001). H(2)O(2) produced by viridans group streptococci may contribute to inhibition of methicillin-resistant Staphylococcus aureus colonization of oral cavities in newborns. *Clin Infect Dis,* 32**,** 1408-13.

Uppalapati, S. R., Sett, A. & Pathania, R. (2020). The Outer Membrane Proteins OmpA, CarO, and OprD of *Acinetobacter baumannii* Confer a Two-Pronged Defense in Facilitating Its Success as a Potent Human Pathogen. *Frontiers in Microbiology,* 11.

Van Bambeke, F., Balzi, E. & Tulkens, P. M. (2000). Antibiotic efflux pumps. *Biochemical pharmacology,* 60**,** 457-470.

Van Bambeke, F., Mingeot-Leclercq, M.-P., Glupczynski, Y. & Tulkens, P. M. (2017). 137 - Mechanisms of Action. *In:* Cohen, J., Powderly, W. G. & Opal, S. M. (eds.) *Infectious Diseases (Fourth Edition).* Elsevier.

Van Hal, S. J., Willems, R. J. L., Gouliouris, T., Ballard, S. A., Coque, T. M., Hammerum, A. M., Hegstad, K., Westh, H. T., Howden, B. P., Malhotra-

Kumar, S., Werner, G., Yanagihara, K., Earl, A. M., Raven, K. E., Corander, J., Bowden, R., Pinholt, M., Loens, K., Xavier, B. B., Matheeussen, V., Goossens, H. & Enterococcal, G. (2021). The global dissemination of hospital clones of *Enterococcus faecium*. *Genome Medicine,* 13**,** 52.

Van Heel, A., de Jong, A., Montalbán-López, M., Kok, J. and Kuipers, O., (2013). BAGEL3: automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. *Nucleic Acids Research*, 41(W1), pp.W448-W453.

Van Heel, A., de Jong, A., Song, C., Viel, J., Kok, J. and Kuipers, O., (2018). BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Research*, 46(W1), pp.W278-W281.

Van Hoek, A. H., Mevius, D., Guerra, B., Mullany, P., Roberts, A. P. & Aarts, H. J. (2011). Acquired antibiotic resistance genes: an overview. *Front Microbiol,* 2**,** 203.

Vassilev, V., Lavermicocca, P., Di Giorgio, D. & Iacobellis, N. S. (1996). Production of syringomycins and syringopeptins by *Pseudomonas syringae* pv. *atrofaciens*. *Plant Pathology,* 45**,** 316-322.

Veerabadhran, M., Chakraborty, S., Mitra, S., Karmakar, S. & Mukherjee, J. (2018). Effects of flask configuration on biofilm growth and metabolites of intertidal Cyanobacteria isolated from a mangrove forest. *J Appl Microbiol,* 125**,** 190-202.

Velkov, T., Roberts, K. D., Nation, R. L., Thompson, P. E. & Li, J. (2013). Pharmacology of polymyxins: new insights into an 'old' class of antibiotics. *Future Microbiol,* 8**,** 711-24.

Venugopal, A. A. & Johnson, S. (2012). Fidaxomicin: a novel macrocyclic antibiotic approved for treatment of Clostridium difficile infection. *Clin Infect Dis,* 54**,** 568-74.

Venter, H., Henningsen, Michael L. & Neville, S. (2017). Antimicrobial resistance in healthcare, agriculture and the environment: The biochemistry behind the headlines. *Essays In Biochemistry,* 61**,** 1-10.

Viljanen, P. & Vaara, M. (1984). Susceptibility of gram-negative bacteria to polymyxin B nonapeptide. *Antimicrob Agents Chemother,* 25**,** 701-5.

Vincent, J. L., Rello, J., Marshall, J., Silva, E., Anzueto, A., Martin, C. D., Moreno, R., Lipman, J., Gomersall, C., Sakr, Y. & Reinhart, K. (2009). International study of the prevalence and outcomes of infection in intensive care units. *Jama,* 302**,** 2323-9.

Vlassov, V. V., Laktionov, P. P. & Rykova, E. Y. (2007). Extracellular nucleic acids. *Bioessays,* 29**,** 654-67.

Von Eiff, C., Becker, K., Machka, K., Stammer, H. & Peters, G. (2001). Nasal carriage as a source of *Staphylococcus aureus* bacteremia. Study Group. *N Engl J Med,* 344**,** 11-6.

Waksman, S. A., Schatz, A. & Reynolds, D. M. (2010). Production of antibiotic substances by actinomycetes. *Ann N Y Acad Sci,* 1213**,** 112-24.

Walker, V. & Mills, G. A. (2002). Solid-phase extraction in clinical biochemistry. *Ann Clin Biochem,* 39**,** 464-77.

Wang, X. & Kasperski, K. (2018). Quantitative determination of multiple surfactants in water using HPLC-MS/MS. *Analytical Methods,* 10**,** 2512-2525.

Wang, C., Jiang, C. & Armstrong, D. W. (2008). Considerations on HILIC and polar organic solvent-based separations: use of cyclodextrin and macrocyclic glycopetide stationary phases. *J Sep Sci,* 31**,** 1980-90.

Wang, Z., Koirala, B., Hernandez, Y., Zimmerman, M., Park, S., Perlin, D. S. & Brady, S. F. (2022a). A naturally inspired antibiotic to target multidrug-resistant pathogens. *Nature,* 601**,** 606-611.

Wang, Q., Li, X., Tang, W., Guan, X., Xiong, Z., Zhu, Y., Gong, J. & Hu, B. (2022b). Differential Gene Sets Profiling in Gram-Negative and Gram-Positive Sepsis. *Front Cell Infect Microbiol,* 12**,** 801232.

Wang, X., Shen, Y., Huang, W. & Zhou, Y. (2019). Characterisation of community-acquired *Staphylococcus aureus* causing skin and soft tissue infections in a children's hospital in Shanghai, China. *Epidemiology and infection,* 147**,** e323-e323.

Ward, J. B. (1981). Teichoic and teichuronic acids: biosynthesis, assembly, and location. *Microbiol Rev,* 45**,** 211-43.

Wertheim, H., Verveer, J., Boelens, H., van Belkum, A., Verbrugh, H. and Vos, M., (2005). Effect of Mupirocin Treatment on Nasal, Pharyngeal, and Perineal Carriage of *Staphylococcus aureus* in Healthy Adults. *Antimicrobial Agents and Chemotherapy*, 49(4), pp.1465-1467.

Werner, G., Strommenger, B. & Witte, W. (2008). Acquired vancomycin resistance in clinically relevant pathogens. *Future Microbiol,* 3**,** 547-62.

Wieland, K., Chhatwal, P. & Vonberg, R.-P. (2018). Nosocomial outbreaks caused by *Acinetobacter baumannii* and *Pseudomonas aeruginosa*: Results of a systematic review. *American Journal of Infection Control,* 46**,** 643-648.

Wolfender, J. L. (2009). HPLC in natural product analysis: the detection issue. *Planta Med,* 75**,** 719-34.

Woodford, N. & Ellington, M. J. (2007). The emergence of antibiotic resistance by mutation. *Clinical Microbiology and Infection,* 13**,** 5-18.

World Health Organization WHO (2015) Global action plan on antimicrobial resistance [Online] Available online at: https://www.who.int/publications/i/item/9789241509763 (Accessed 19 September 2021).

World Health Organization WHO (2017) Global Priority List of Antibiotic-Resistant Bacteria to Guide Research, Discovery, and Development of New Antibiotics. Available at: https://www.who.int/medicines/publications/WHO-PPL-Short_Summary_25Feb-ET_NM_WHO.pdf?ua=1 [Accessed 1 March 2020].

World Health Organization WHO (2021) Antibacterial agents in clinical and preclinical development: an overview and analysis. Available online at: https://www.who.int/publications/i/item/9789240047655 (Accessed 10 February, 2023).

Wróbel, A., Arciszewska, K., Maliszewski, D. & Drozdowska, D. (2020). Trimethoprim and other nonclassical antifolates an excellent template for searching modifications of dihydrofolate reductase enzyme inhibitors. *The Journal of Antibiotics,* 73**,** 5-27.

Wu, L., Wang, Z., Guan, Y., Huang, X., Shi, H., Liu, Y. and Zhang, X., (2020). The (p)ppGpp-mediated stringent response regulatory system globally inhibits primary metabolism and activates secondary metabolism in *Pseudomonas protegens* H78. *Applied Microbiology and Biotechnology*, 104(7), pp.3061-3079.

Wu, S., Huang, T., Xie, D., Wo, J., Wang, X., Deng, Z. & Lin, S. 2017. Xantholipin B produced by the stnR inactivation mutant *Streptomyces flocculus* CGMCC 4.1223 WJN-1. *The Journal of Antibiotics,* 70**,** 90-95

Wu, M. & Li, X. (2015). Chapter 87 - *Klebsiella pneumoniae* and *Pseudomonas aeruginosa*. *In:* Tang, Y.-W., Sussman, M., Liu, D., Poxton, I. & Schwartzman, J. (eds.) *Molecular Medical Microbiology (Second Edition).* Boston: Academic Press.

Würtele, M., Wolf, E., Pederson, K. J., Buchwald, G., Ahmadian, M. R., Barbieri, J. T. & Wittinghofer, A. (2001). How the Pseudomonas aeruginosa ExoS toxin downregulates Rac. *Nat Struct Biol,* 8**,** 23-6.

Xavier, B. M. & Russell, J. B. (2009). The ability of non-bacteriocin producing *Streptococcus bovis* strains to bind and transfer bovicin HC5 to other sensitive bacteria. *Anaerobe,* 15**,** 168-72.

Xie, Y., Liu, Z., Zhang, G., Mo, X., Ding, X., Xia, L. & Hu, S. (2016). A rifampicin-resistant (*rpoB*) mutation in *Pseudomonas protegens* Pf-5 strain leads to improved antifungal activity and elevated production of secondary metabolites. *Research in Microbiology,* 167**,** 625-629.

Xu, C., Bilya, S. R. & Xu, W. (2019). adeABC efflux gene in *Acinetobacter baumannii*. *New Microbes New Infect,* 30**,** 100549.

Yadav, A. K., Espaillat, A. & Cava, F. (2018). Bacterial Strategies to Preserve Cell Wall Integrity Against Environmental Threats. *Front Microbiol,* 9**,** 2064.

Yajko, D. M., Madej, J. J., Lancaster, M. V., Sanders, C. A., Cawthon, V. L., Gee, B., Babst, A. & Hadley, W. K. (1995). Colorimetric method for determining MICs of antimicrobial agents for *Mycobacterium tuberculosis*. *J Clin Microbiol,* 33**,** 2324-7.

Yang, K. M., Kim, J. S., Kim, H. S., Kim, Y. Y., Oh, J. K., Jung, H. W., Park, D. S. & Bae, K. H. (2021). *Lactobacillus reuteri* AN417 cell-free culture supernatant as a novel antibacterial agent targeting oral pathogenic bacteria. *Sci Rep,* 11**,** 1631.

Yang, S. C., Lin, C. H., Sung, C. T. & Fang, J. Y. (2014). Antibacterial activities of bacteriocins: application in foods and pharmaceuticals. *Front Microbiol,* 5**,** 241.

Yang, X., Xie, L., Li, Y. & Wei, C. (2009). More than 9,000,000 unique genes in human gut bacterial community: estimating gene numbers inside a human body. *PLoS One,* 4**,** e6074.

Yarlagadda, V., Manjunath, G. B., Sarkar, P., Akkapeddi, P., Paramanandham, K., Shome, B. R., Ravikumar, R. & Haldar, J. (2016). Glycopeptide Antibiotic To Overcome the Intrinsic Resistance of Gram-Negative Bacteria. *ACS Infect Dis,* 2**,** 132-9.

Zanger, P., Nurjadi, D., Gaile, M., Gabrysch, S. & Kremsner, P. G. (2012). Hormonal Contraceptive Use and Persistent *Staphylococcus aureus* Nasal Carriage. *Clinical Infectious Diseases,* 55**,** 1625-1632.

Zeng, X. & Lin, J. (2013). Beta-lactamase induction and cell wall metabolism in Gram-negative bacteria. *Frontiers in microbiology,* 4**,** 128.

Zeng, Y. & Charkowski, A. O. (2021). The Role of ATP-Binding Cassette Transporters in Bacterial Phytopathogenesis. *Phytopathology®,* 111**,** 600-610.

Zhang, A., Wan, L., Wu, C., Fang, Y., Han, G., Li, H., Zhang, Z. & Wang, H. (2013). Simultaneous determination of 14 phenolic compounds in grape canes by HPLC-DAD-UV using wavelength switching detection. *Molecules,* 18**,** 14241-57.

Zhang, A., Wan, L., Wu, C., Fang, Y., Han, G., Li, H., Zhang, Z. & Wang, H. 2013. Simultaneous determination of 14 phenolic compounds in grape canes by HPLC-DAD-UV using wavelength switching detection. *Molecules,* 18**,** 14241-57.

Zhang, Y., Huang, H., Xu, S., Wang, B., Ju, J., Tan, H. & Li, W. (2015). Activation and enhancement of Fredericamycin A production in deepsea-derived *Streptomyces somaliensis* SCSIO ZH66 by using ribosome engineering and response surface methodology. *Microb Cell Fact,* 14**,** 64.

Zhen, X., Lundborg, C. S., Zhang, M., Sun, X., Li, Y., Hu, X., Gu, S., Gu, Y., Wei, J. & Dong, H. (2020). Clinical and economic impact of methicillin-resistant *Staphylococcus aureus*: a multicentre study in China. *Scientific Reports,* 10**,** 3900.

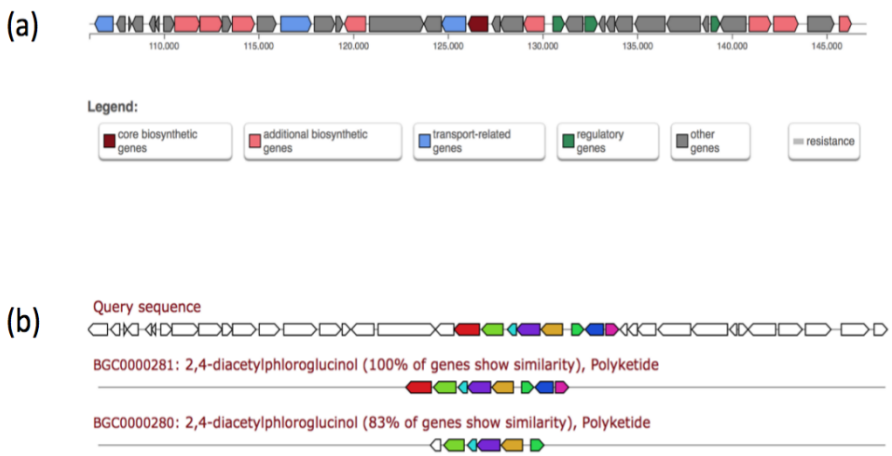
Zhou, L., De Jong, A., Yi, Y. & Kuipers, O. P. (2021). Identification, Isolation, and Characterization of Medipeptins, Antimicrobial Peptides From *Pseudomonas mediterranea* EDOX. *Frontiers in Microbiology,* 12.

Zhu, K. & Rock, C. O. (2008). RhlA converts beta-hydroxyacyl-acyl carrier protein intermediates in fatty acid synthesis to the beta-hydroxydecanoyl-beta-hydroxydecanoate component of rhamnolipids in *Pseudomonas aeruginosa*. *J Bacteriol,* 190**,** 3147-54.

Zhu, S., Duan, Y. & Huang, Y. (2019). The Application of Ribosome Engineering to Natural Product Discovery and Yield Improvement in *Streptomyces*. *Antibiotics (Basel),* 8.

Zipperer, A., Konnerth, M. C., Laux, C., Berscheid, A., Janek, D., Weidenmaier, C., Burian, M., Schilling, N. A., Slavetinsky, C., Marschal, M., Willmann, M., Kalbacher, H., Schittek, B., Brotz-Oesterhelt, H., Grond, S., Peschel, A. & Krismer, B. (2016). Human commensals producing a novel antibiotic impair pathogen colonization. *Nature,* 535**,** 511-6.

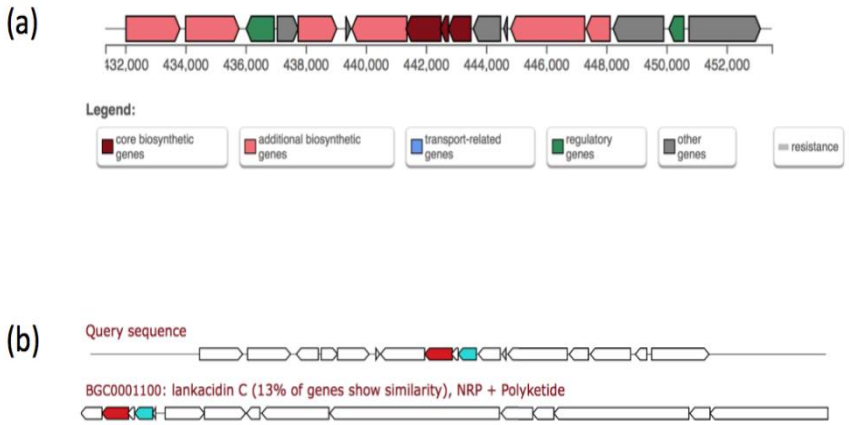
Appendices



Appendix 1: Type 3 polyketide synthase (T3PKS) BGC located from 106060-147109 in the genome of *Pseudomonas* sp. strain NW7.

(a) Diagrammatical representation of T3PKS BGC in the genome of *Pseudomonas* sp. strain NW7 as predicted by antiSMASH.

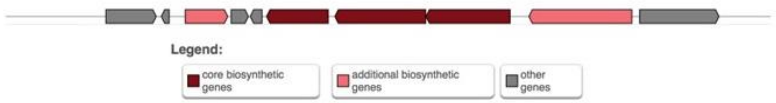
(b) Comparison of known BGCs with the T3PKS BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative T3PKS BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.



Appendix 2: Redox-cofactor BGC located from 431338-453500 in the genome of *Pseudomonas* sp. strain NW7.

(a) Diagrammatical representation of redox-cofactor BGC in the genome of *Pseudomonas* sp. strain NW7 as predicted by antiSMASH.

(b) Comparison of known BGCs with the redox-cofactor BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative redox-cofactor BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.



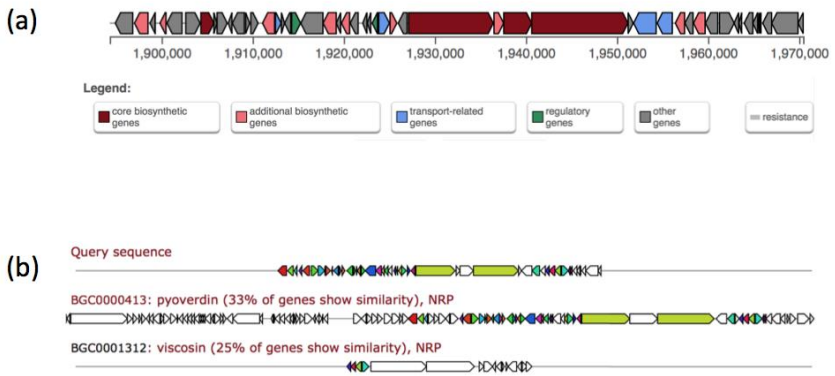
Appendix 3: NAGGN BGC located from 1678447-1693123 in the genome of *Pseudomonas* sp. strain NW7.

Diagrammatical representation of NAGGN BGC in the genome of *Pseudomonas* sp. strain NW7 as predicted by antiSMASH.

Appendix 4: Nonribosomal peptide synthetase (NRPS) BGC located from 1812447-1863999 in the genome of *Pseudomonas* sp. strain NW7.

(a) Diagrammatical representation of NRPS BGC in the genome of *Pseudomonas* sp. strain NW7 as predicted by antiSMASH.

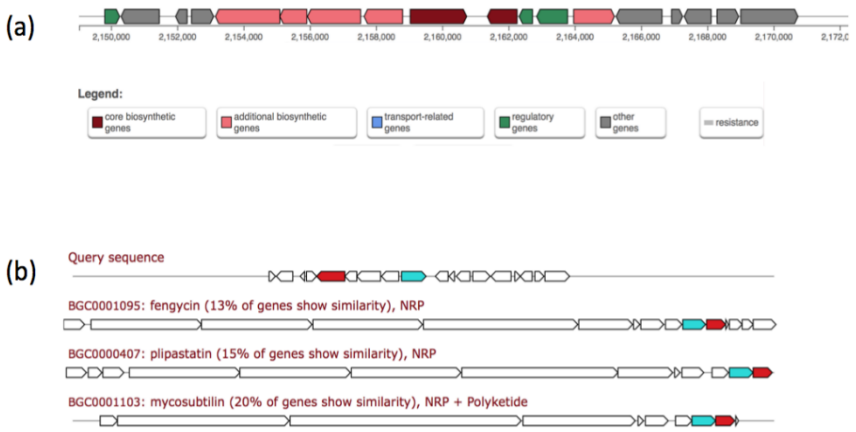
(b) Comparison of known BGCs with the NRPS BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative NRPS BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.



Appendix 5: Ranthipeptide/NRPS BGC located from 1894315-1970440 in the genome of *Pseudomonas* sp. strain NW7.

(a) Diagrammatical representation of ranthipeptide/NRPS BGC in the genome of *Pseudomonas* sp. strain NW7 as predicted by antiSMASH.

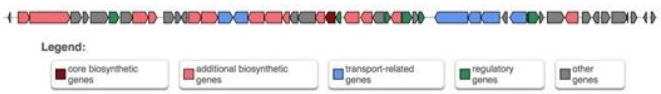
(b) Comparison of known BGCs with the ranthipeptide/NRPS BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative ranthipeptide/NRPS BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.



Appendix 6: Betalactone BGC located from 2149033-2172258 in the genome of *Pseudomonas* sp. strain NW7.

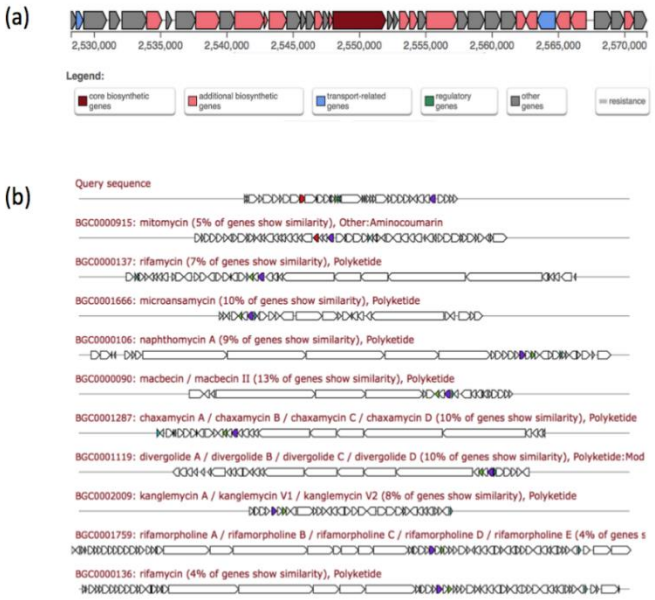
(a) Diagrammatical representation of betalactone BGC in the genome of *Pseudomonas* sp. strain NW7 as predicted by antiSMASH.

(b) Comparison of known BGCs with the betalactone BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative betalactone BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.



Appendix 7: Acyl amino acids BGC located from 2442147-2503076 in the genome of *Pseudomonas* sp. strain NW7.

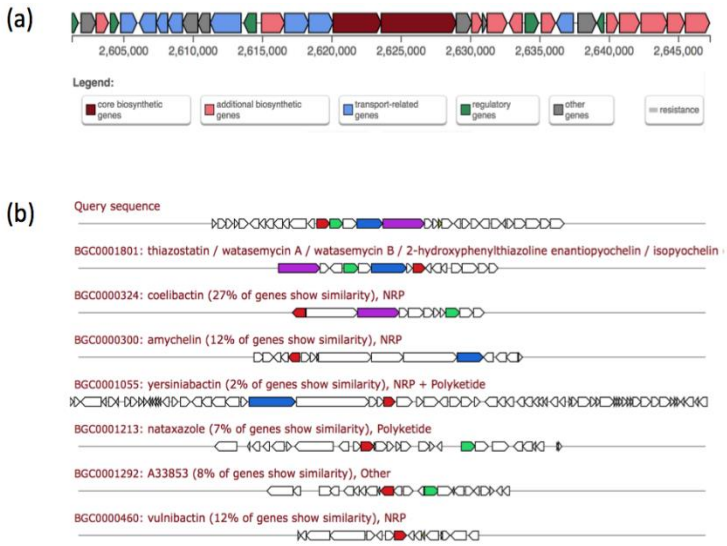
Diagrammatical representation of acyl amino acids BGC in the genome of *Pseudomonas* sp. strain NW7 as predicted by antiSMASH. Figure was adapted from antiSMASH version 6.0.1.



Appendix 8: Nonribosomal peptide synthetase (NRPS) BGC located from 2528263-2571683 in the genome of *Pseudomonas* sp. strain NW7.

(a) Diagrammatical representation of NRPS BGC in the genome of *Pseudomonas* sp. strain NW7 as predicted by antiSMASH.

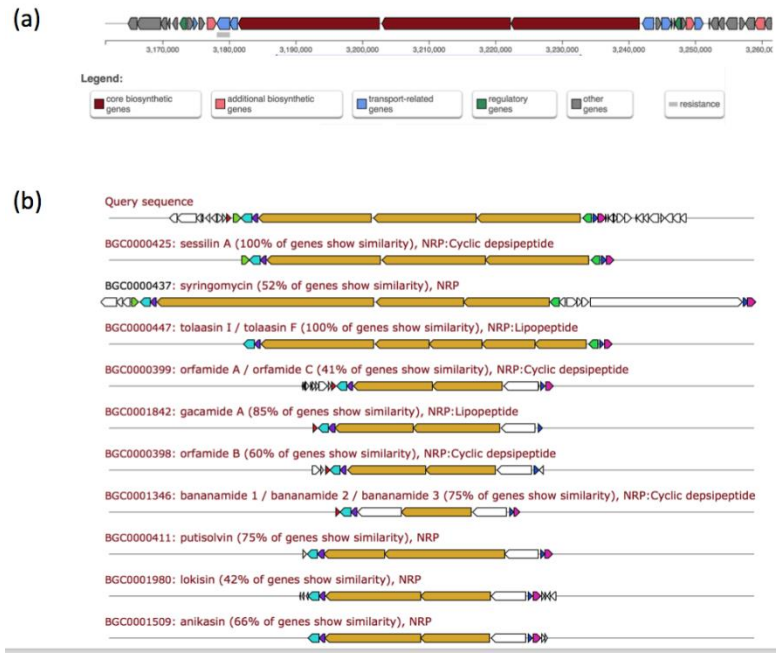
(b) Comparison of known BGCs with the NRPS BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative NRPS BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.



Appendix 9: Nonribosomal peptide synthetase (NRPS)-type BGC located from 2601269-2647278 in the genome of *Pseudomonas* sp. strain NW7.

(a) Diagrammatical representation of NRPS BGC in the genome of *Pseudomonas* sp. strain NW7 as predicted by antiSMASH.

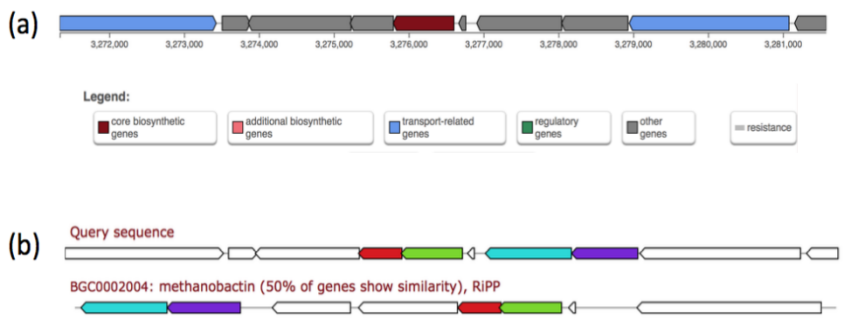
(b) Comparison of known BGCs with the NRPS BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative NRPS BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.



Appendix 10: Nonribosomal peptide synthetase (NRPS) BGC located from 3161399-3261614 in the genome of *Pseudomonas* sp. strain NW7.

(a) Diagrammatical representation of NRPS BGC in the genome of *Pseudomonas* sp. strain NW7 as predicted by antiSMASH.

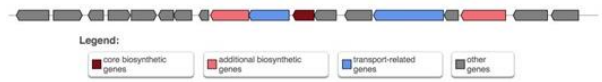
(b) Comparison of known BGCs with the NRPS BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative NRPS BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.



Appendix 11: Unspecified ribosomally synthesised and post-translationally modified peptide product (RiPP)-like-BGC located from 3271339-3281579 in the genome of *Pseudomonas* sp. strain NW7.

(a) Diagrammatical representation of RiPP-like-BGC in the genome of *Pseudomonas* sp. strain NW7 as predicted by antiSMASH.

(b) Comparison of known BGCs with the RiPP-like-BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative RiPP-like-BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.



Appendix 12: CDPS BGC located from 5012236-5032985 in the genome of *Pseudomonas* sp. strain NW7.

Diagrammatical representation of CDPS BGC in the genome of *Pseudomonas* sp. strain NW7 as predicted by antiSMASH. Figure was adapted from antiSMASH version 6.0.1.

Appendix 13: RiPP-like BGC located from 5072019-5082989 in the genome of *Pseudomonas* sp. strain NW7.

Diagrammatical representation of RiPP-like BGC in the genome of *Pseudomonas* sp. strain NW7 as predicted by antiSMASH. Figure was adapted from antiSMASH version 6.0.1.

Appendix 14: RiPP-like BGC located from 5113312-5122147 in the genome of *Pseudomonas* sp. strain NW7.

Diagrammatical representation of RiPP-like BGC in the genome of *Pseudomonas* sp. strain NW7 as predicted by antiSMASH. Figure was adapted from antiSMASH version 6.0.1.

Appendix 15: RRE-containing BGC located from 5686261-5706545 in the genome of *Pseudomonas* sp. strain NW7.

Diagrammatical representation of RRE-containing BGC in the genome of *Pseudomonas* sp. strain NW7 as predicted by antiSMASH. Figure was adapted from antiSMASH version 6.0.1.

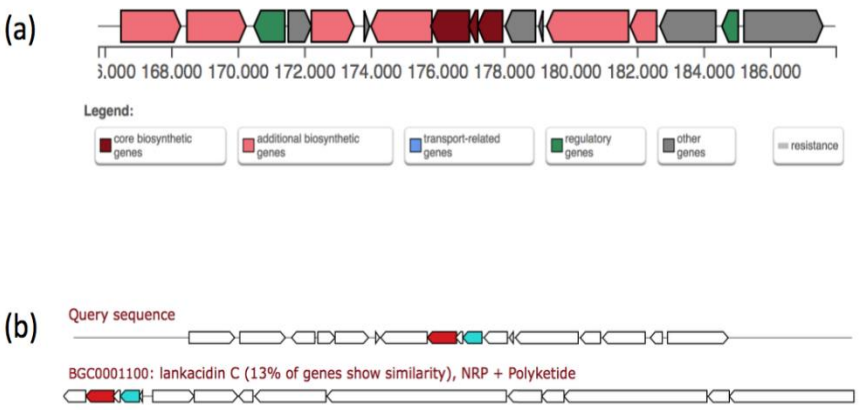
Appendix 16: Arylpolyene-type BGC located from 6053423-6097040 in the genome of *Pseudomonas* sp. strain NW7.

(a) Diagrammatical representation of arylpolyene BGC in the genome of *Pseudomonas* sp. strain NW7 as predicted by antiSMASH.

(b) Comparison of known BGCs with the arylpolyene BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative arylpolyene BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.

Appendix 17: RiPP-like BGC located from 6624809-6635654 in the genome of *Pseudomonas* sp. strain NW7.

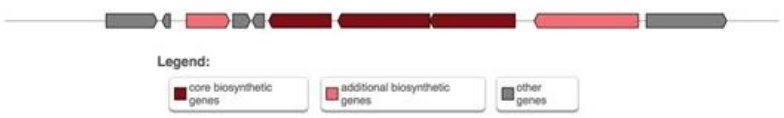
Diagrammatical representation of RiPP-like BGC in the genome of *Pseudomonas* sp. strain NW7 as predicted by antiSMASH. Figure was adapted from antiSMASH version 6.0.1.



Appendix 18: Redox-cofactor BGC located from 165807-187969 in the genome of *Pseudomonas* sp. strain NW9.

(a) Diagrammatical representation of redox-cofactor BGC in the genome of *Pseudomonas* sp. strain NW9 as predicted by antiSMASH.

(b) Comparison of known BGCs with the redox-cofactor BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative redox-cofactor BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.



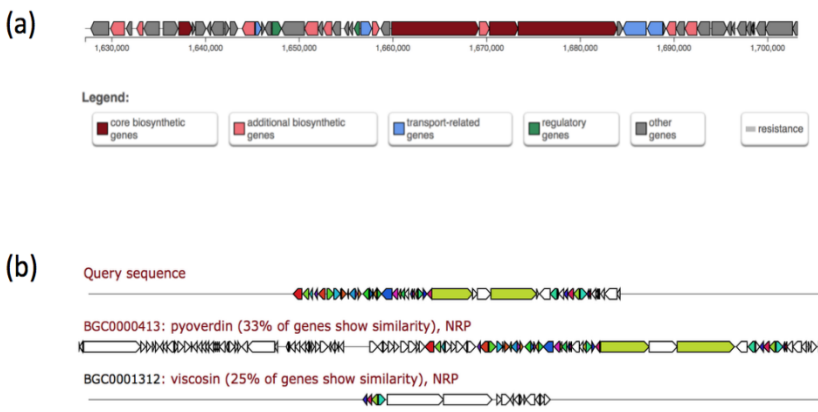
Appendix 19: NAGGN BGC located from 1411111-1425787 in the genome of *Pseudomonas* sp. strain NW9.

Diagrammatical representation of NAGGN BGC in the genome of *Pseudomonas* sp. strain NW9 as predicted by antiSMASH. Figure was adapted from antiSMASH version 6.0.1.

Appendix 20: Nonribosomal peptide synthetase (NRPS) BGC located from 1545308-1596860 in the genome of *Pseudomonas* sp. strain NW9.

(a) Diagrammatical representation of NRPS BGC in the genome of *Pseudomonas* sp. strain NW9 as predicted by antiSMASH.

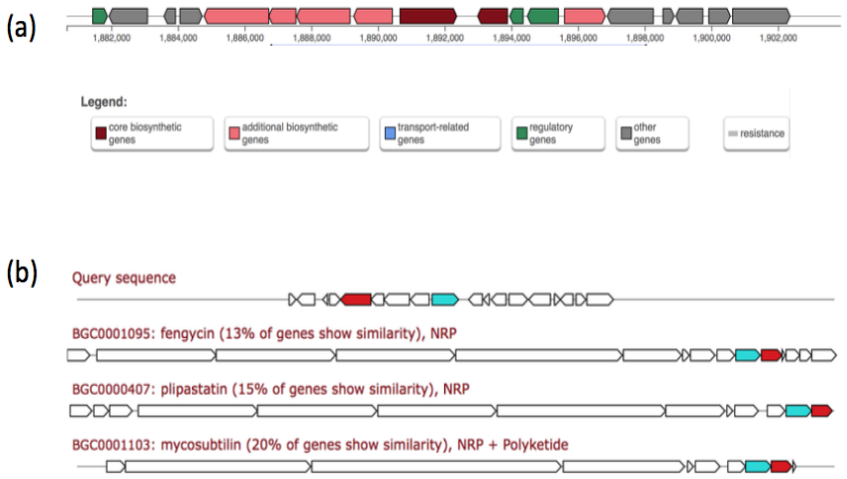
(b) Comparison of known BGCs with the NRPS BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative NRPS BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.



Appendix 21: Ranthipeptide/NRPS BGC located from 1627182-1703255 in the genome of *Pseudomonas* sp. strain NW9.

(a) Diagrammatical representation of ranthipeptide/NRPS BGC in the genome of *Pseudomonas* sp. strain NW9 as predicted by antiSMASH.

(b) Comparison of known BGCs with the ranthipeptide/NRPS BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative ranthipeptide/NRPS BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.



Appendix 22: Betalactone BGC located from 1880661-1903886 in the genome of *Pseudomonas* sp. strain NW9.

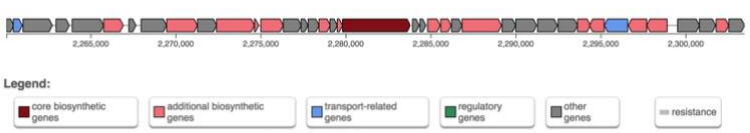
(a) Diagrammatical representation of betalactone BGC in the genome of *Pseudomonas* sp. strain NW9 as predicted by antiSMASH.

(b) Comparison of known BGCs with the betalactone BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative betalactone BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.

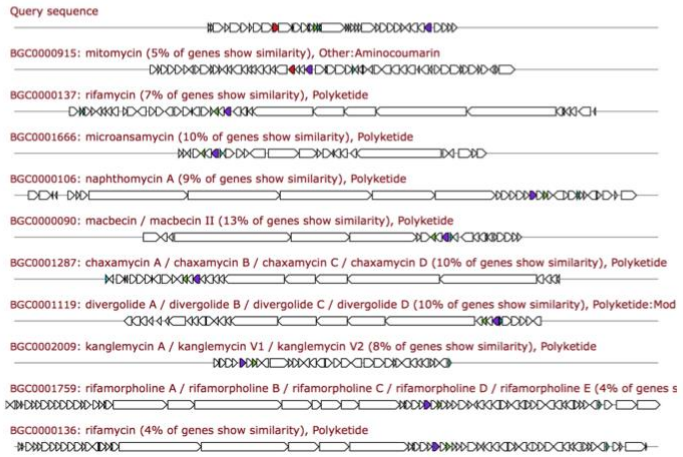
Appendix 23: Acyl amino acids BGC located from 2173764-2234693 in the genome of *Pseudomonas* sp. strain NW9.

Diagrammatical representation of acyl amino acids BGC in the genome of *Pseudomonas* sp. strain NW9 as predicted by antiSMASH. Figure was adapted from antiSMASH version 6.0.1

(a)



(b)



Appendix 24: Nonribosomal peptide synthetase (NRPS) BGC located from 2260050-2303470 in the genome of *Pseudomonas* sp. strain NW9.

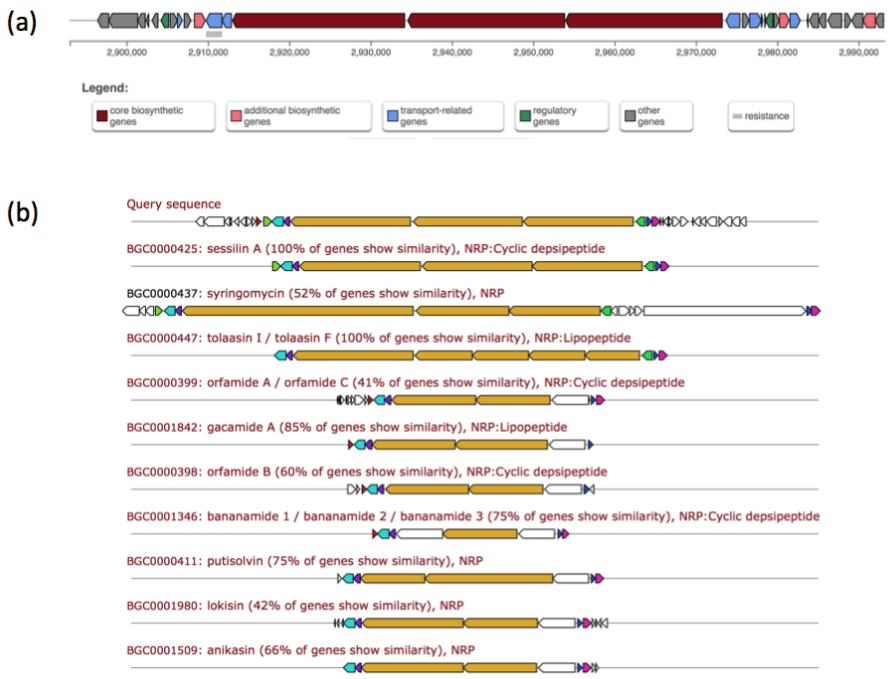
(a) Diagrammatical representation of NRPS BGC in the genome of *Pseudomonas* sp. strain NW9 as predicted by antiSMASH.

(b) Comparison of known BGCs with the NRPS BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative NRPS BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.

Appendix 25: Nonribosomal peptide synthetase (NRPS) BGC located from 2333056-2379065 in the genome of *Pseudomonas* sp. strain NW9.

(a) Diagrammatical representation of NRPS BGC in the genome of *Pseudomonas* sp. strain NW9 as predicted by antiSMASH.

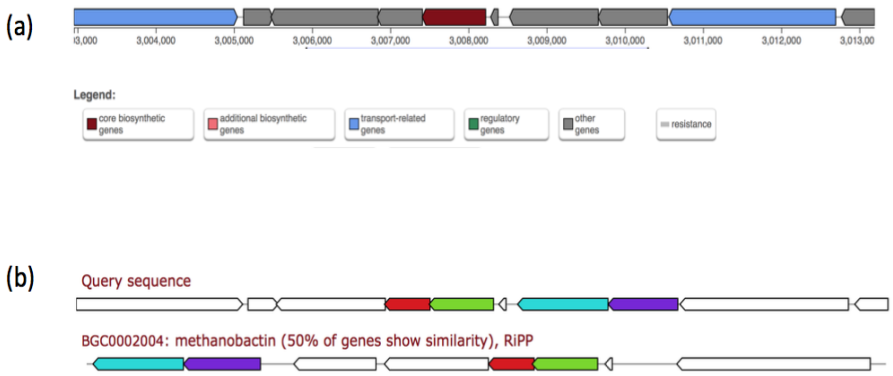
(b) Comparison of known BGCs with the NRPS BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative NRPS BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.



Appendix 26: Nonribosomal peptide synthetase (NRPS) BGC located from 2893015-2993230 in the genome of *Pseudomonas* sp. strain NW9.

(a) Diagrammatical representation of NRPS BGC in the genome of *Pseudomonas* sp. strain NW9 as predicted by antiSMASH.

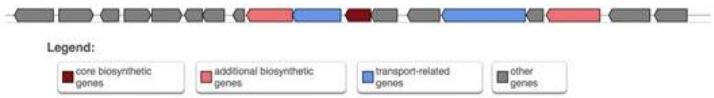
(b) Comparison of known BGCs with the NRPS BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative NRPS BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.



Appendix 27: Unspecified ribosomally synthesised and post-translationally modified peptide product (RiPP)-like BGC located from 3002955-3013195 in the genome of *Pseudomonas* sp. strain NW9.

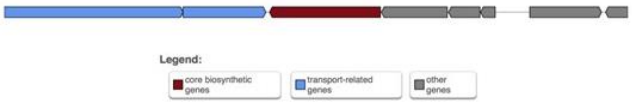
(a) Diagrammatical representation of RiPP-like BGC in the genome of *Pseudomonas* sp. strain NW9 as predicted by antiSMASH.

(b) Comparison of known BGCs with the RiPP-like BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative RiPP-like BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.



Appendix 28: CDPS BGC located from 4743475-4764224 in the genome of *Pseudomonas* sp. strain NW9.

Diagrammatical representation of CDPS BGC in the genome of *Pseudomonas* sp. strain NW9 as predicted by antiSMASH. Figure was adapted from antiSMASH version 6.0.1.

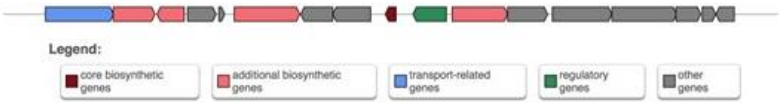


Appendix 29: RiPP-like BGC located from 4803258-4814228 in the genome of *Pseudomonas* sp. strain NW9.

Diagrammatical representation of RiPP-like BGC in the genome of *Pseudomonas* sp. strain NW9 as predicted by antiSMASH. Figure was adapted from antiSMASH version 6.0.1.

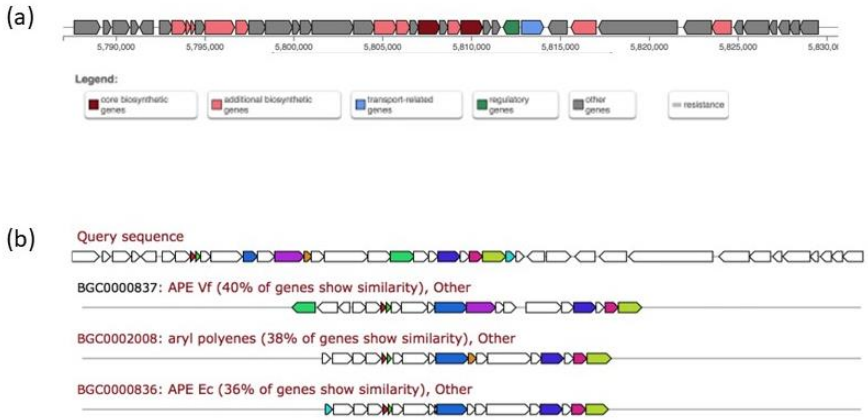
Appendix 30: RiPP-like BGC located from 4844551-4853386 in the genome of *Pseudomonas* sp. strain NW9.

Diagrammatical representation of RiPP-like BGC in the genome of *Pseudomonas* sp. strain NW9 as predicted by antiSMASH. Figure was adapted from antiSMASH version 6.0.1.



Appendix 31: RRE-containing BGC located from 5419211-5439495 in the genome of *Pseudomonas* sp. strain NW9.

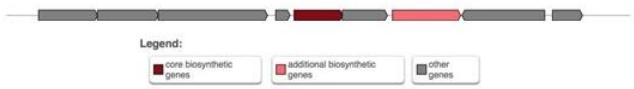
Diagrammatical representation of RRE-containing BGC in the genome of *Pseudomonas* sp. strain NW9 as predicted by antiSMASH. Figure was adapted from antiSMASH version 6.0.1.



Appendix 32: Arylpolyene BGC located from 5787030-5830647 in the genome of *Pseudomonas* sp. strain NW9.

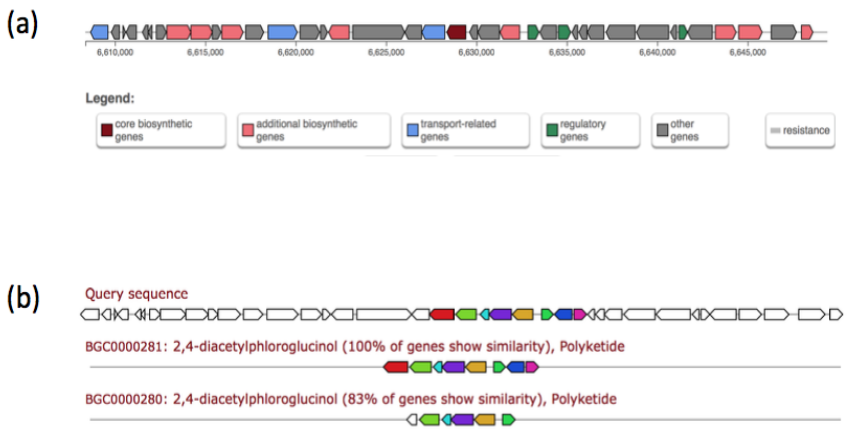
(a) Diagrammatical representation of arylpolyene BGC in the genome of *Pseudomonas* sp. strain NW9 as predicted by antiSMASH.

(b) Comparison of known BGCs with the arylpolyene BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative arylpolyene BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.



Appendix 33: RiPP-like BGC located from 6358136-6368981 in the genome of *Pseudomonas* sp. strain NW9.

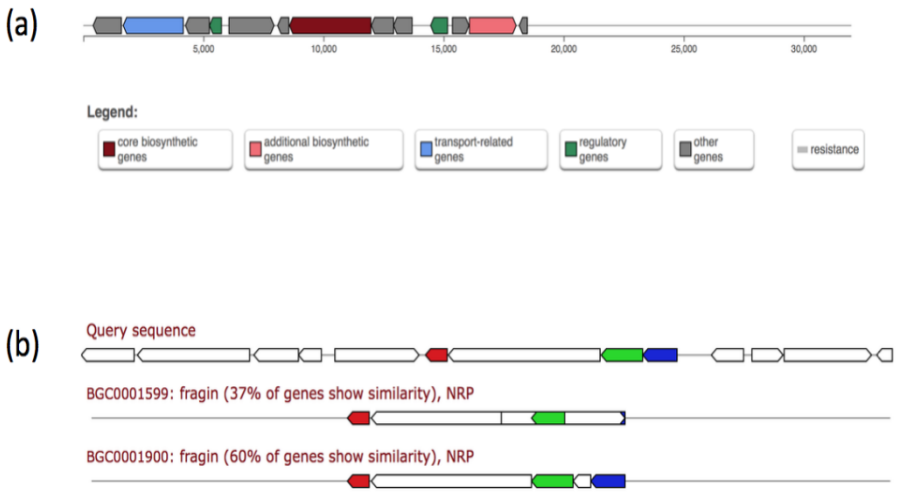
Diagrammatical representation of RiPP-like BGC in the genome of *Pseudomonas* sp. strain NW9 as predicted by antiSMASH. Figure was adapted from antiSMASH version 6.0.1.



Appendix 34: Type III polyketide synthase (T3PKS) BGC located from 6608365-6649414 in the genome of *Pseudomonas* sp. strain NW9.

(a) Diagrammatical representation of T3PKS BGC in the genome of *Pseudomonas* sp. strain NW9 as predicted by antiSMASH.

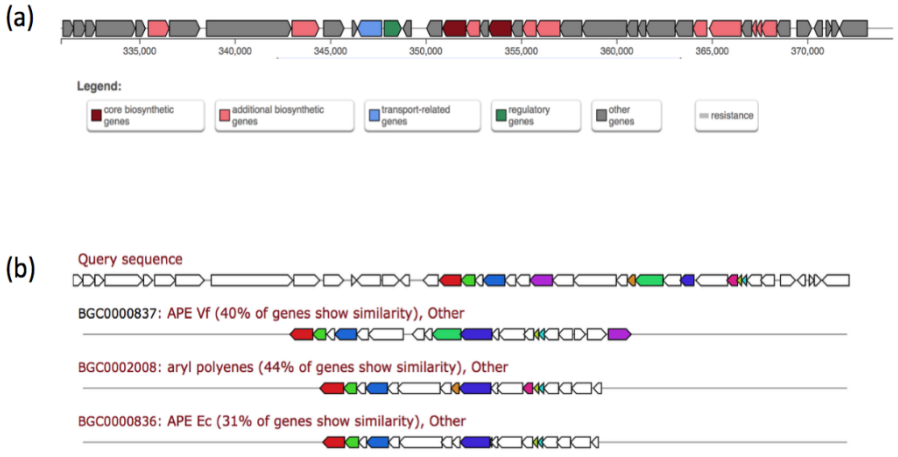
(b) Comparison of known BGCs with the T3PKS BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative T3PKS BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.



Appendix 35: Nonribosomal peptide synthetase (NRPS) BGC located from 1-31967 in the genome of *Pseudomonas* sp. strain NW10.

(a) Diagrammatical representation of NRPS BGC in the genome of *Pseudomonas* sp. strain NW10 as predicted by antiSMASH.

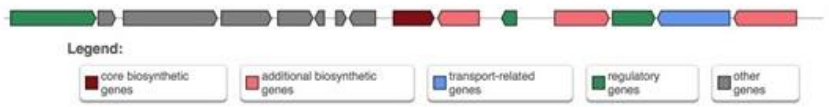
(b) Comparison of known BGCs with the NRPS BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative NRPS BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.



Appendix 36: Arylpolyene BGC located from 330883-374487 in the genome of *Pseudomonas* sp. strain NW10.

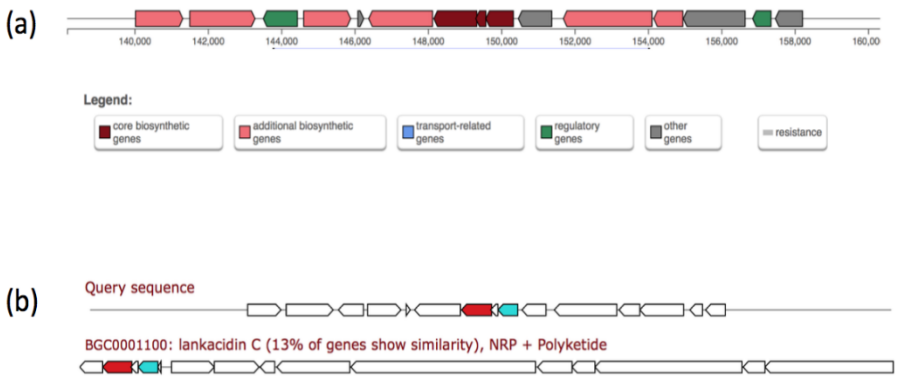
(a) Diagrammatical representation of arylpolyene BGC in the genome of *Pseudomonas* sp. strain NW10 as predicted by antiSMASH.

(b) Comparison of known BGCs with the arylpolyene BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative arylpolyene BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.



Appendix 37: PpyS-KS BGC located from 170210-191259 in the genome of *Pseudomonas* sp. strain NW9.

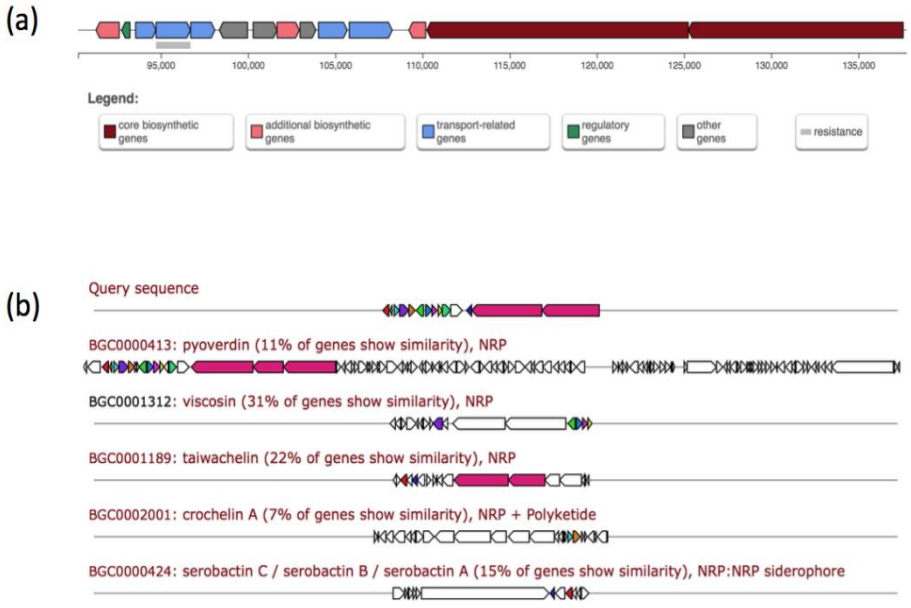
Diagrammatical representation of PpyS-KS BGC in the genome of *Pseudomonas* sp. strain NW10 as predicted by antiSMASH. Figure was adapted from antiSMASH version 6.0.1.



Appendix 38: Redox-cofactor BGC located from 138161-160329 in the genome of *Pseudomonas* sp. strain NW10.

(a) Diagrammatical representation of redox-cofactor BGC in the genome of *Pseudomonas* sp. strain NW10 as predicted by antiSMASH.

(b) Comparison of known BGCs with the redox-cofactor BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative redox-cofactor BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.



Appendix 39: Nonribosomal peptide synthetase (NRPS) BGC located from 90254-137776 in the genome of *Pseudomonas* sp. strain NW10.

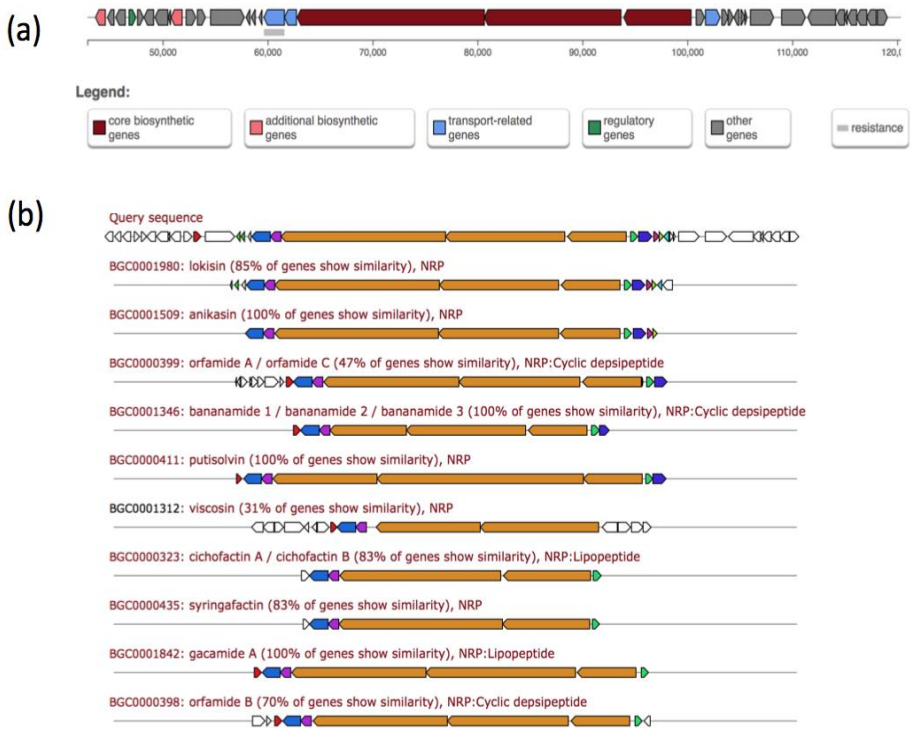
(a) Diagrammatical representation of NRPS BGC in the genome of *Pseudomonas* sp. strain NW10 as predicted by antiSMASH.

(b) Comparison of known BGCs with the NRPS BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative NRPS BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region.

Appendix 40: Nonribosomal peptide synthetase (NRPS) BGC located from 14775-67773 in the genome of *Pseudomonas* sp. strain NW10.

(a) Diagrammatical representation of NRPS BGC in the genome of *Pseudomonas* sp. strain NW10 as predicted by antiSMASH.

(b) Comparison of known BGCs with the NRPS BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative NRPS BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.



Appendix 41: Nonribosomal peptide synthetase (NRPS) BGC located from 42820-120355 in the genome of *Pseudomonas* sp. strain NW10.

(a) Diagrammatical representation of NRPS BGC in the genome of *Pseudomonas* sp. strain NW10 as predicted by antiSMASH.

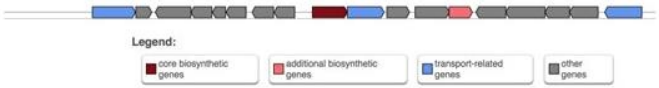
(b) Comparison of known BGCs with the NRPS BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative NRPS BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.

Appendix 42: NAGGN BGC located from 9254-24131 in the genome of *Pseudomonas* sp. strain NW10.

Diagrammatical representation of NAGGN BGC in the genome of *Pseudomonas* sp. strain NW10 as predicted by antiSMASH. Figure was adapted from antiSMASH version 6.0.1.

Appendix 43: Hsterlactone/ Butyrlactone BGC located from 15650-36204 in the genome of *Pseudomonas* sp. strain NW10.

Diagrammatical representation of hsterlactone/butyrlactone BGC in the genome of *Pseudomonas* sp. strain NW10 as predicted by antiSMASH. Figure was adapted from antiSMASH version 6.0.1.

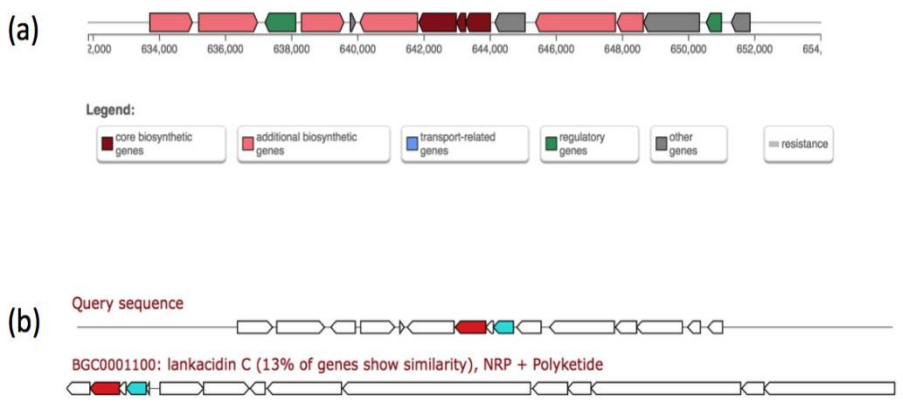


Appendix 44: RRE-containing BGC located from 37131-58261 in the genome of *Pseudomonas* sp. strain NW10.

Diagrammatical representation of RRE-containing BGC in the genome of *Pseudomonas* sp. strain NW10 as predicted by antiSMASH. Figure was adapted from antiSMASH version 6.0.1.

Appendix 45: RiPP-like BGC located from 1-6422 in the genome of *Pseudomonas* sp. strain NW10.

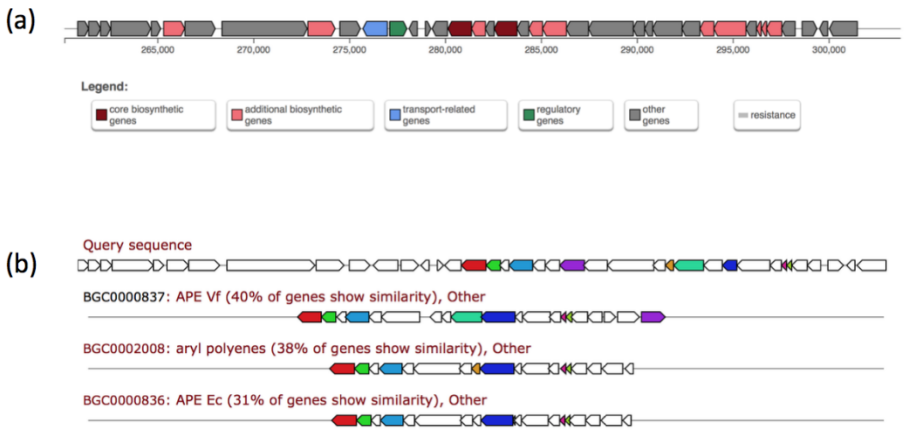
Diagrammatical representation of RiPP-like BGC in the genome of *Pseudomonas* sp. strain NW10 as predicted by antiSMASH. Figure was adapted from antiSMASH version 6.0.1.



Appendix 46: Redox-cofactor BGC located from 631855-654023 in the genome of *Pseudomonas* sp. strain NW16.

(a) Diagrammatical representation of redox-cofactor BGC in the genome of *Pseudomonas* sp. strain NW16 as predicted by antiSMASH.

(b) Comparison of known BGCs with the redox-cofactor BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative redox-cofactor BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.



Appendix 47: Arylpolyene BGC located from 260151-303755 in the genome of *Pseudomonas* sp. strain NW16.

(a) Diagrammatical representation of arylpolyene BGC in the genome of *Pseudomonas* sp. strain NW16 as predicted by antiSMASH.

(b) Comparison of known BGCs with the arylpolyene BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative arylpolyene BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.

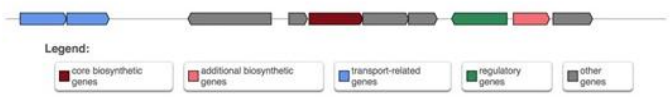
Appendix 48: NAGGN BGC located from 199596-214471 in the genome of *Pseudomonas* sp. strain NW16.

Diagrammatical representation of NAGGN BGC in the genome of *Pseudomonas* sp. strain NW16 as predicted by antiSMASH. Figure was adapted from antiSMASH version 6.0.1.

Appendix 49: Non-ribosomal peptide synthetase (NRPS) BGC located from 355091-408089 in the genome of *Pseudomonas* sp. strain NW16.

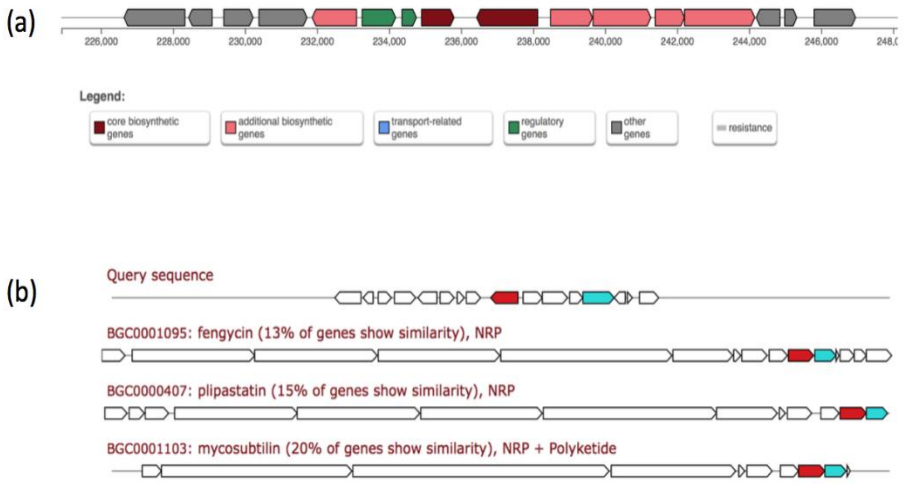
(a) Diagrammatical representation of NRPS BGC in the genome of *Pseudomonas* sp. strain NW16 as predicted by antiSMASH.

(b) Comparison of known BGCs with the NRPS BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative NRPS BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version



Appendix 50: RiPP-like BGC located from 55395-66282 in the genome of *Pseudomonas* sp. strain NW16.

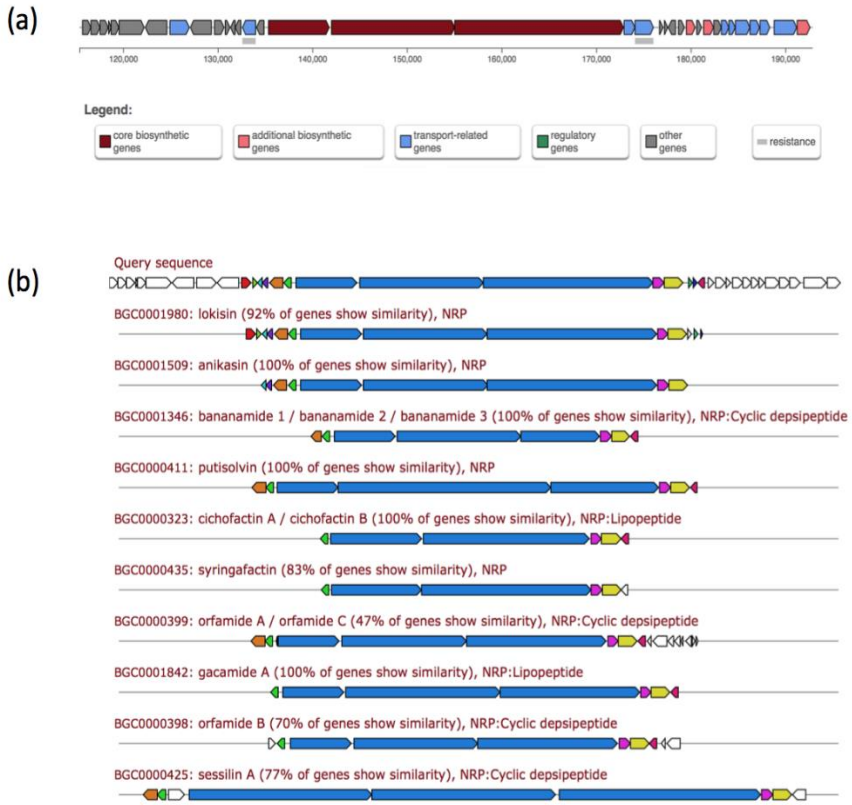
Diagrammatical representation of RiPP-like BGC in the genome of *Pseudomonas* sp. strain NW16 as predicted by antiSMASH. Figure was adapted from antiSMASH version 6.0.1.



Appendix 51: Betalactone BGC located from 224905-248128 in the genome of *Pseudomonas* sp. strain NW16.

(a) Diagrammatical representation of betalactone BGC in the genome of *Pseudomonas* sp. strain NW16 as predicted by antiSMASH.

(b) Comparison of known BGCs with the betalactone BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative betalactone BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.



Appendix 52: Non-ribosomal peptide synthetase (NRPS) BGC located from 115369-192888 in the genome of *Pseudomonas* sp. strain NW16.

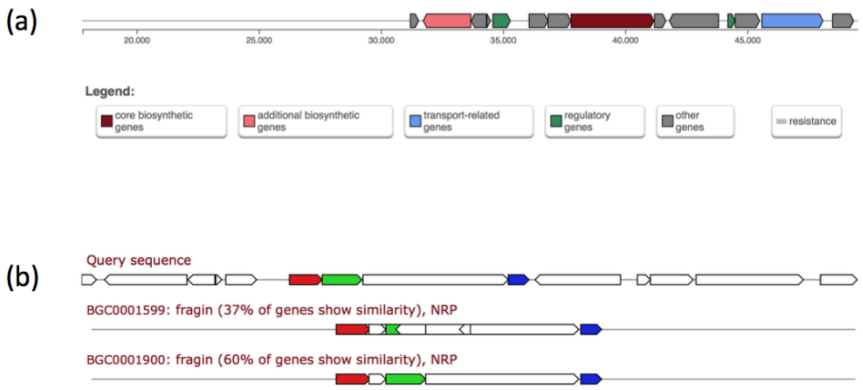
(a) Diagrammatical representation of NRPS BGC in the genome of *Pseudomonas* sp. strain NW16 as predicted by antiSMASH.

(b) Comparison of known BGCs with the NRPS BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative NRPS BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.

Appendix 53: Nonribosomal peptide synthetase (NRPS) BGC located from 1-47523 in the genome of *Pseudomonas* sp. strain NW16.

(a) Diagrammatical representation of NRPS BGC in the genome of *Pseudomonas* sp. strain NW16 as predicted by antiSMASH.

(b) Comparison of known BGCs with the NRPS BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative NRPS BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.



Appendix 54: Nonribosomal peptide synthetase (NRPS) BGC located from 17778-49528 in the genome of *Pseudomonas* sp. strain NW16.

(a) Diagrammatical representation of NRPS BGC in the genome of *Pseudomonas* sp. strain NW16 as predicted by antiSMASH.

(b) Comparison of known BGCs with the NRPS BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative NRPS BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.

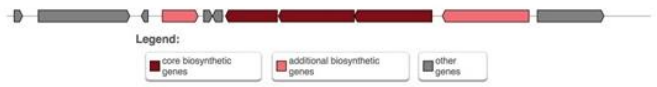
Appendix 55: Siderophore BGC located from 273564-284764 in the genome of *Pseudomonas* sp. strain NW27.

Diagrammatical representation of siderophore BGC in the genome of *Pseudomonas* sp. strain NW27 as predicted by antiSMASH. Figure was adapted from antiSMASH version 6.0.1.

Appendix 56: Nonribosomal peptide synthetase (NRPS) BGC located from 372448-42514 in the genome of *Pseudomonas* sp. strain NW27.

(A) Diagrammatical representation of NRPS BGC in the genome of *Pseudomonas* sp. strain NW27 as predicted by antiSMASH.

(B) Comparison of known BGCs with the NRPS BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative NRPS BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.



Appendix 57: NAGGN BGC located from 494680-509398 in the genome of *Pseudomonas* sp. strain NW27.

Diagrammatical representation of NAGGN BGC in the genome of *Pseudomonas* sp. strain NW27 as predicted by antiSMASH. Figure was adapted from antiSMASH version 6.0.1.

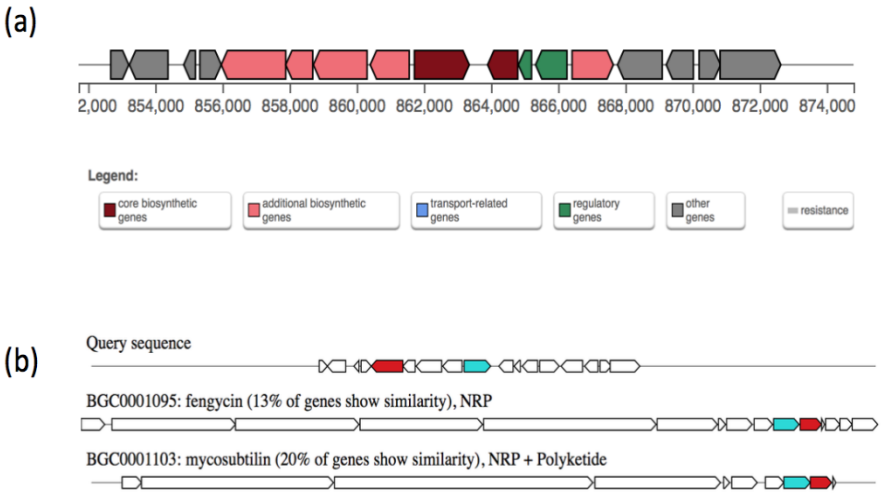
Appendix 58: Siderophore BGC located from 644664-656514 in the genome of *Pseudomonas* sp. strain NW27.

Diagrammatical representation of siderophore BGC in the genome of *Pseudomonas* sp. strain NW27 as predicted by antiSMASH. Figure was adapted from antiSMASH version 6.0.1.

Appendix 59: Nonribosomal peptide synthetase (NRPS) BGC located from 700777-745389 in the genome of *Pseudomonas* sp. strain NW27.

(a) Diagrammatical representation of NRPS BGC in the genome of *Pseudomonas* sp. strain NW27 as predicted by antiSMASH.

(b) Comparison of known BGCs with the NRPS BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative NRPS BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted



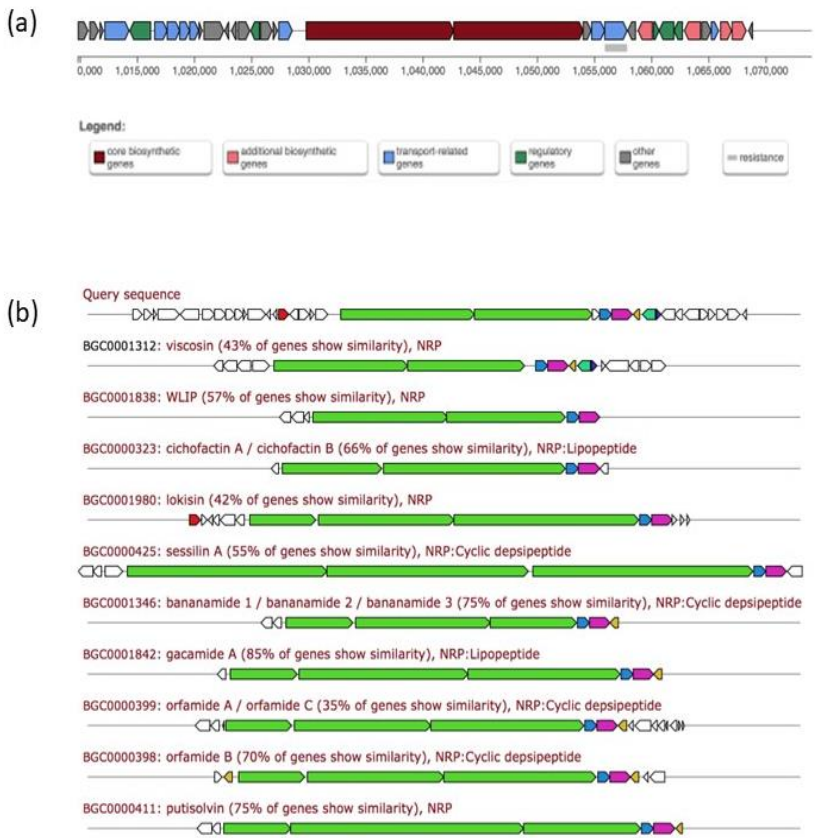
Appendix 60: Betalactone BGC located from 851716 - 874797 in the genome of *Pseudomonas* sp. strain NW27.

(a) Diagrammatical representation of betalactone BGC in the genome of *Pseudomonas* sp. strain NW27 as predicted by antiSMASH.

(b) Comparison of known BGCs with the betalactone BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative betalactone BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.

Appendix 61: RiPP-like BGC located from 919356-929788 in the genome of *Pseudomonas* sp. strain NW27.

Diagrammatical representation of RiPP-like BGC in the genome of *Pseudomonas* sp. strain NW27 as predicted by antiSMASH. Figure was adapted from antiSMASH version 6.0.1.

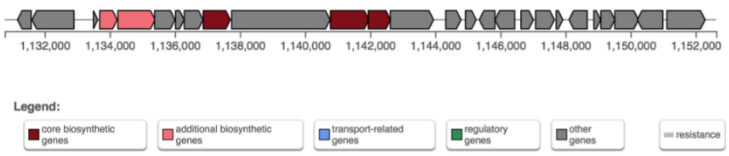


Appendix 62: Nonribosomal peptide synthetase (NRPS) BGC located from 1009807 - 1074048 in the genome of *Pseudomonas* sp. strain NW27.

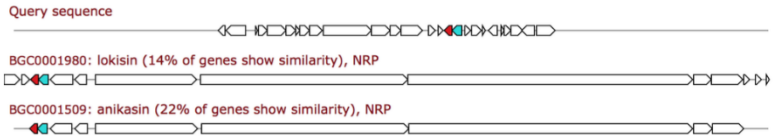
(a) Diagrammatical representation of NRPS BGC in the genome of *Pseudomonas* sp. strain NW27 as predicted by antiSMASH.

(b) Comparison of known BGCs with the NRPS BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative NRPS BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.

(a)



(b)



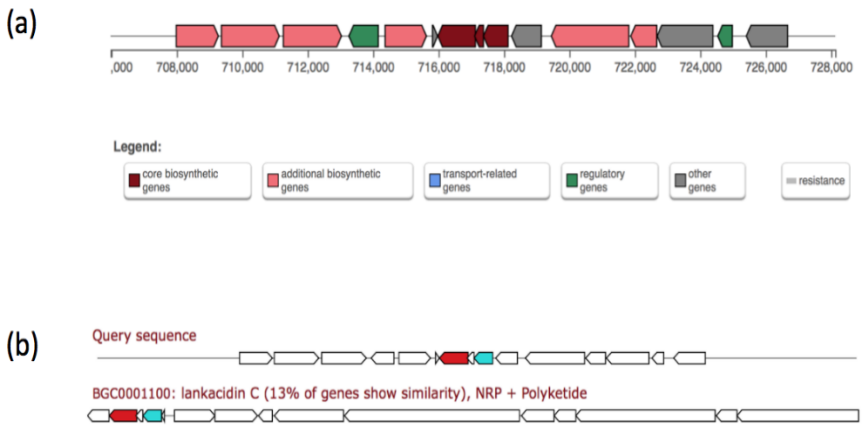
Appendix 63: Thioamidite-type/RiPP-like BGC located from 1130778 - 1152626 in the genome of *Pseudomonas* sp. strain NW27.

(a) Diagrammatical representation of thioamidite-type/RiPP-like BGC in the genome of *Pseudomonas* sp. strain NW27 as predicted by antiSMASH.

(b) Comparison of known BGCs with the thioamidite-type/RiPP-like BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative thioamidite-type/RiPP-like BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.

Appendix 64: RiPP-like BGC located from 192722-203567 in the genome of *Pseudomonas* sp. strain NW27.

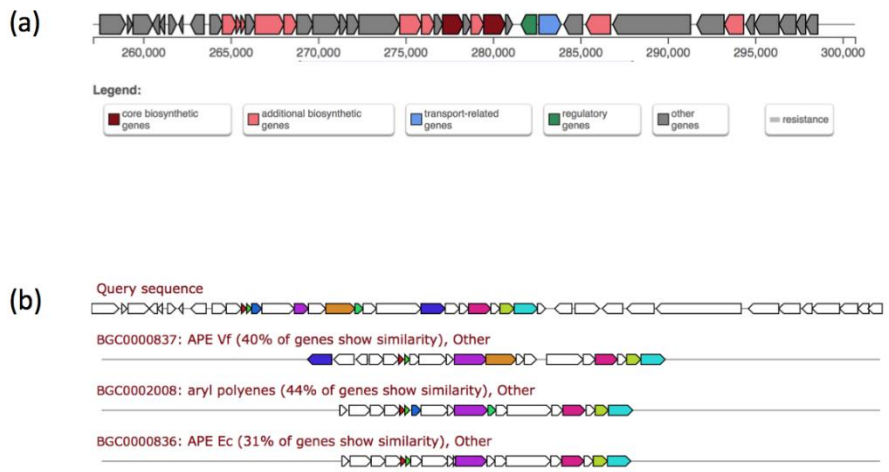
Diagrammatical representation of RiPP-like BGC in the genome of *Pseudomonas* sp. strain NW27 as predicted by antiSMASH. Figure was adapted from antiSMASH version 6.0.1.



Appendix 65: Redox-cofactor BGC located from 705977-728124 in the genome of *Pseudomonas* sp. strain NW27.

(a) Diagrammatical representation of redox-cofactor BGC in the genome of *Pseudomonas* sp. strain NW27 as predicted by antiSMASH.

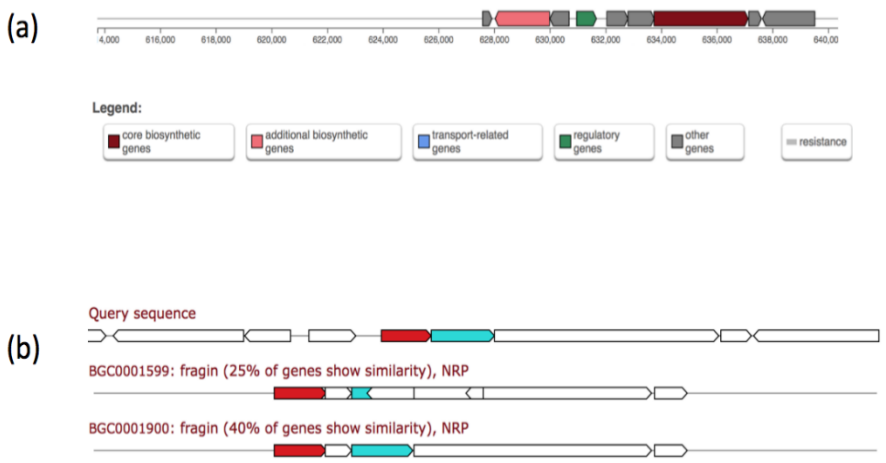
(b) Comparison of known BGCs with the redox-cofactor BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative redox-cofactor BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.



Appendix 66: Arylpolyene BGC located from 257124-300699 in the genome of *Pseudomonas* sp. strain NW27.

(a) Diagrammatical representation of arylpolyene BGC in the genome of *Pseudomonas* sp. strain NW27 as predicted by antiSMASH.

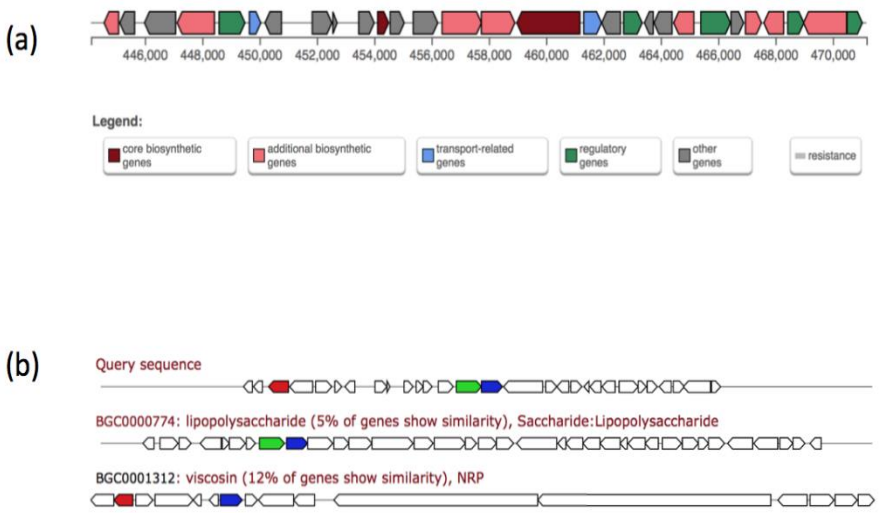
(b) Comparison of known BGCs with the arylpolyene BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative arylpolyene BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.



Appendix 67: Nonribosomal peptide synthetase (NRPS)-like BGC located from 613749-640358 in the genome of *Pseudomonas* sp. strain NW27.

(a) Diagrammatical representation of NRPS-like BGC in the genome of *Pseudomonas* sp. strain NW27 as predicted by antiSMASH.

(b) Comparison of known BGCs with the NRPS-like BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative NRPS-like BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.



Appendix 68: Thiopeptide BGC located from 444121-471178 in the genome of *Pseudomonas* sp. strain NW27.

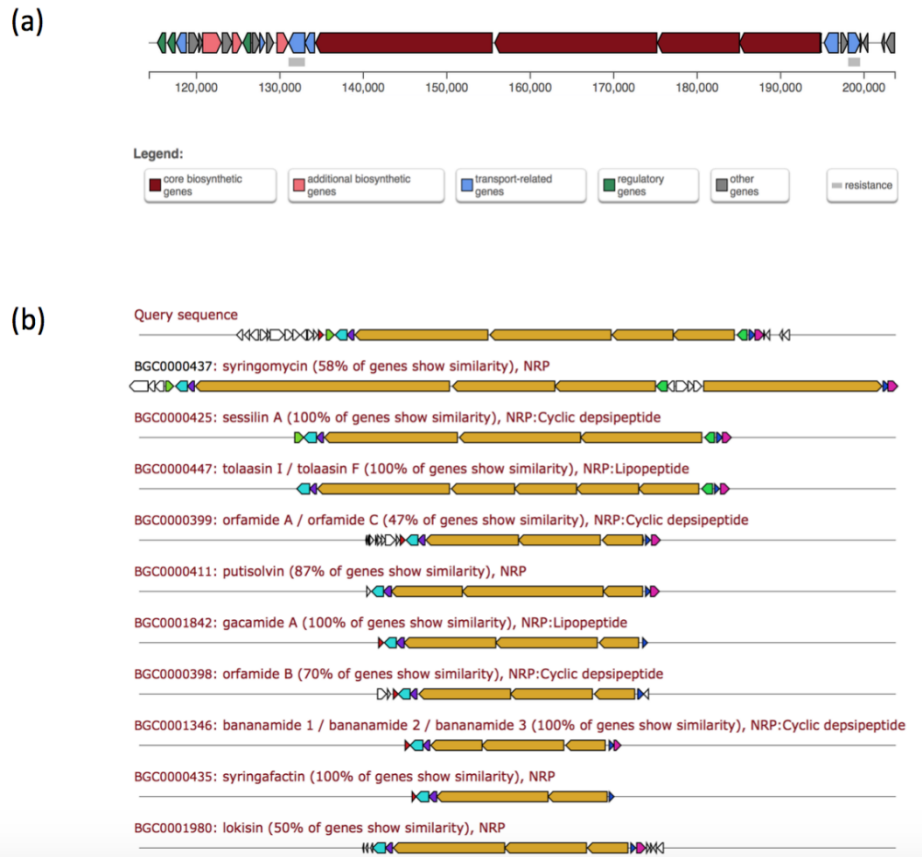
(a) Diagrammatical representation of thiopeptide BGC in the genome of *Pseudomonas* sp. strain NW27 as predicted by antiSMASH.

(b) Comparison of known BGCs with the thiopeptide BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative thiopeptide BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.

Appendix 69: Non-ribosomal peptide synthetase (NRPS) BGC located from 126517-192409 in the genome of *Pseudomonas* sp. strain NW27.

(a) Diagrammatical representation of NRPS BGC in the genome of *Pseudomonas* sp. strain NW27 as predicted by antiSMASH.

(b) Comparison of known BGCs with the NRPS BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative NRPS BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.



Appendix 70: Nonribosomal peptide synthetase (NRPS) BGC located from 114344-203719 in the genome of *Pseudomonas* sp. strain NW27.

(A) Diagrammatical representation of NRPS BGC in the genome of *Pseudomonas* sp. strain NW27 as predicted by antiSMASH.

(B) Comparison of known BGCs with the NRPS BGC. The figure represents a set of BGCs with genes that encode for proteins with a predicted function. These genes are also predicted in the query BGC. The query sequence is the putative NRPS BGC. The percentage represents the percentage of genes in the BGCs that have a significant BLAST hit to genes within the current region. These figures were adapted from antiSMASH version 6.0.1.

Appendix 71: RiPP-like BGC located from 3144-14019 in the genome of *Pseudomonas* sp. strain NW27.

Diagrammatical representation of RiPP-like BGC in the genome of *Pseudomonas* sp. strain NW27 as predicted by antiSMASH. Figure was adapted from antiSMASH version 6.0.1.

Appendix 72: Tentative *S. aureus* genes identified in the metagenome of five individuals.

Individual	Tentative *S. aureus* gene	Protein product	Nucleotide sequence	length (bp)	Gene ID in IMG/M
5	***fmhA***	**FemA/FemB family glycyltransferase**	GTGCCTACCATTTGCAGTGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTAATTTTTATGGCATTACTGGTGATTTTTCAGAGTCAGCAGAAGACTATGGCGTACAACAATTCAAAAAAGGATTTAACGCACACGTAGAAGAATATATTGGTGATTTTATTAAACCAATCAAACCTATACTTTACAAATTACAAACA	204	SRS018369_Baylor_scaffold_180__gene_30
11	***clfB***	**Clumping factor B**	GCAAATATTAATATTGCGGATGAAATGTTTGATAATAAAATTACTTATAACTATAGTTCGCCAATTGCAGGAATTGATAAGCCAAATGGCGCGAACATTTCTTCTCAAATTATTGGTGTAGATACAGCTTCAGGTCAAAACACATACAAGCAAACGGTATTTGTTAACCCTAAGCAACGAGTTTTAGGTAATACGTGGGTGTATATTAAAGGTTACCAAGATAAAATCGAAGAAAGTAGCGGTAAAGTAA	414	C171444__gene_3501
	fnbA	**Fibronectin binding protein A**	AGATATACATTTACTGATTACATCAAAGATAAAGTAAATGTTACAGCTAATTTAGAAATAAACTTATTCATAGATCCTAAAACTGTACAAAGCAATGGACAACAAACAATAACTTCAAAATTAAATGGCAAAGAAACATCAGGAACTATGCAAATAACATATAAAGATGGTGTTAAAAATCAATATACAAATGTAAATGGTTCGATTGAAACGTTTGACAAAGAAAAAAATAAATTTACACACGTAGCATATATTAAGCCTATAAATGGAAACAATTCAGATAGTGTTACTGTAACGGGTATGTTGACACAAGGAAGTAATGAGAATGGAACACAACCAAATGTTAAAATATATGAGTATGTTGGAGTTGAAAATGGG	378	C170634__gene_3113
	sodA	**Superoxide dismutase**	ACATTATTTGGATCAGGTTGGACTTGGTTAGTTGTTAATGATGGAAAATTAGAAATTGTGACAACGCCAAACCAAGATAATCCATTAACAGAAGGCAAAACACCAATCTTACTATTTGATGTTTGGGAGCATGCCTACTATCTGAAATATCAAAATAAACGTCCAGACTATATGACTGCATTTTGGAATATTGTTAACTGGAAAAAAGTTGATGAATTATACCAAGCAGCAAAATAA	237	C167625__gene_1733
13	***arcC***	**Carbamate kinase**	ATGTCTAAAATCGTCGTAGCTTTGGGTGGTAACGCTTTAGGACAATCGCCTAAAGAACAATTAGATTTATTAAAAAGTACATCTAAATCTCTAGTAAGTTTAATCGATAAAGGATACGAAATTGTAATTAGTCATGGTAATGGTCCACAAGTTGGAAGTATTAACTTAGGTTTAAATTATGCGGCGGAACACAAACAAGGTCCTCCTTTTCCTTTCCCTGAATGTGGTGCTATGAGTCAAGCTTATATTGGCTATCAGATGCAAGAAAGTTTACAAAACGAACTTCATTCAATGGGCATAGATAAACAAGTTGTTACGCTAGTAACACAAGTTCAAGTTGCAAGCGATGATTCTGCTTTCAATAACCCTACTAAACCAATTGGATTATTTTATACAAAAGAGCAGGCTGATAAATTTACGAAAGAAAAAGGTTATACTTTTGTAGAAGATTCTGGACGCGGCTATCGTCGCGTTGTCCCTTCCCCACAACCTATAAGTATAGTTGAGTTAGATAGTATAGAAACATTAATCACTCATGGAACACTAGTTATCGCTGCCGGCGGTGGTGGAATTCCAG	930	>C349066__gene_13204

			TAATTAAGGAAAATGAAGTTTATACAGGTGTTGATGCAGTTATTGATAAAGAT AAAACGAGTGCTTTATTAGCAGCACATTTACAATCTGATCAATTAATCATATT AACTGCTGTGGACCATG TTTACATTAACTATGGAAAAGAAAACCAAAGAGGTCTCGATGAAGTGTCTGT GGATGAAATGAAAAAACATATCTCTGATGGTCAATTTGCTAAAGGAAGTATG CTCCCAAAAGTTGAAGCTGCACTTCAATTTCTTGAAAAAAATACTAAAGGCA GTGTTTTGATTACATCTCTAGCAGGATTAGGGGACGCTTTAGACGGTAAAAT AGGAACATTAATTAAGAATTGA		
28	***sodA***	**Superoxide dismutase**	ATGGCTTTTGAATTACCAAAATTACCATACGCATTTGATGCATTAGAACCACA TTTTGACAAAGAAACTATGGAAATTCATCATGACAGACATCATAACACTTATG TTACGAAATTAAATGCTGCAGTAGAAGGTACAGATTTAGAATCTAAATCTATT GAAGAAATTGTTGCTAATTTAGACAGTGTACCAGCTAACATCCAAACTGCTG TACGTAATAATGGCGGTGGACATTTAAACCATTCATTATTCTGGAGTTACTTT CACCAAACTCAGAAGAAAAAGGTACTGTAGTAGAAAAAATTAAAGAACAATG GGGTTCTTTAGAAGAATTTAAAAAAGAATTTGCTGACAAAGCAGCTGCACGC TTTGGTTCAGGTTGGGCTTGGTTAGTCGTAAACAATGGCCAGTTAGAAATTG TGACTACACCAAACCAAGATAATCCATTAACTGAGGGTAAAACACCTATTTTA GGTTTAGACGTATGGGAACACGCTTATTACCTAAAATATCAAAACAAACGCC CTGACTACATTGGCGCATTTTGGAATGTAGTTAACTGGGAAAAAGTTGACGA ATTATATAATGCAACAAAA	597	C253285__ gene_1278 1
28	***sodA***	**Superoxide dismutase**	ATGGCATTTAAATTACCAAATTTACCATATGCATATGATGCATTGGAACCATA TATAGATCAAAGAACAATGGAGTTTCATCACGACAAACATCACAATACGTAC GTGACGAAATTAAACGCAACAGTTGAAGGAACAGAGTTAGAGCATCAATCAC TAGCGGATATGATTGCTAACTTAGACAAGGTACCGGAAGCGATGAGGATGT CAGTCCGTAATAATGGCGGTGGTCATTTTAACCATTCATTATTCTGGGAAATA CTATCACCTAATTCTGAAGAAAAAGGTGGCGTAATAGATGACATCAAAGCGC AGTGGGGCACTTTAGATGAATTTAAAAATGAATTTGCAAATAAAGCAACAACA TTATTTGGATCAGGTTGGACTTGGTTAGTTGTTAATGATGGCAAATTAGAAAT TGTGACAACGCCAAACCAAGATAATCCATTAACAGAAGGCAAAACACCAATC TTACTATTTGATGTTTGGGAGCATGCCTACTATCTGAAATATCAAAATAAACG TCCAGACTATATGACTGCATTTTGGAATATTGTTAACTGGAAAAAAGTTGATG AATTATACCAAGCAGCAAAATAA	600	>SRS0244 24_LANL_s caffold_124 8__gene_2 700
28	***clfB***	**Clumping factor B**	AGCGATTCGGACTCAGATAGCGACTCCGATTCAAGAGTTACACCACCAAATA ATGAACAGAAAGCACCATCAAATCCTAAAGGTGAAGTAAACCATTCTAATAA GGTATCAAAACAACACAAAACTGATGCTTTACCAGAAACAGGAGATAAGAGC GAAAACACAAATGCAACTTTATTTGGTGCAATGATGGCATTATTAGGATCATT ACTATTGTTTAGAAAACGCAAGCAAGATCATAAAGAAAAAGCGTAA	255	>C235965_ _gene_342 0
28	***clfB***	**Clumping factor B**	TTGAAAAAAAGAATTGATTATTTGTCGAATAAGCAGAATAAGTATTCGATTAG ACGTTTTACAGTAGGTACCACATCAGTAATAGTAGGGGCAACTATACTATTT GGGATAGGCAATCATCAAGCACAAGCTTCAGAACAATCGAACGATACAACA CAATCTTCGAAAAATAATGCAAGTGCAGATTCCGAAAAAAACAATATGATAGA AACACCTCAATTAAATACAACGGCTAATGATACATCTGATA TTAGTGCAAACACAAACAGTGCGAATGTAGATAGCACAGCAAAACCAATGTC TACACAAACGAGCAATACCACTACAACAGAGCCAGCTTCAACAAATGAAACA	1812	>C255459_ _gene_151 92

			CCTCAACCGACGGCAATTAAAGATCAAGCAACTGCTGCAAAAATGCAAGATC AAACTGTTCCTCAAGAAGCAAATTCTCAAGTAGATAATAAAACAACGAATGAT GCTAATAGCATAGCAACAAACAGTGAGCTTAAAAATCCTCA AACATTAGATTTACCACAATCATCACCACAAACGATTTCCAATGCGCAAGGA ACTAGTAAACCGAGTGTTAGAACGAGAGCTGTACGTAGTCTTGCAGTTGCTG AACCTGTAGTAAATGCTGCTGATGCTAAAGGTACAAATGTAAATGGTCAAGT TACGGCAAGTGATTTCAAGTTAGAAAAGACTACATTTGACCCTAACCAAAGT GGTAACACATTTATGGCGGCAAAATTTACAGTGACTGGACAA GTGAAAGCAGGGGATTATTTTACAGCGAAGTTACCAGATAGTGTAAATGGTA ATGGAGATGTGGATTACTCTAATTCAAATAATACGATGCCAATTGCAGATATT AAAAGTACGAATGGCGATGTTGTAGCTAAAGCAACATATGATATCTTGACTA AGACGTATACATTTGTCTTTACAGATTATGTAAATAATAAAGAAAATATTAACG GACAATTTTCATTACCTTTATTTACAGACCGAGCAAAGG CACCTAAATCAGGAACATATGATGCGAATATTAATATTGCGGATGAAATGTTT AATAATAAAATTACTTATAACTATAGTTCGCCAATTGCAGGAATTGATAAACC AAATGGCGCGAACATTTCTTCTCAAATTATTGGTGTAGATACAGCTTCAGGT CAAAACACATACAAGCAAACAGTATTTGTTAACCCTAAGCAACGAGTTTTAG GTAATACGTGGGTGTATATTAAAGGTTACCAAGATAAAATCGAAGAAAGTAG CGGTAAAGTAAGTGCTACAGATACAAAACTGAGAATTTTTGAAGTGAATGAT ACATCTAAATTATCAGATAGCTACTATGCAGACCCA AATGATTCGAATCTTAAAGAAGTAACGAATGAGTTTAAGGATAAAATCACTTA TAAATACCAAAATGTAGCAAGTATTAATTTTGGCGATATTACTAAAACGTATG TTGTATTAGTGGAAGGTCACTATGATAATACTGGTAAAAACTTGAAAACACAG GTTATTCAAGAAAATATTGACCCAGCGACAGGTAAAGACTACAGTATTTTCG GTTGGAATAATGAGAATGTTGTACGTTATGGTGGTGGAAGTGCTGATGGTGA TTCAGCAGTAAATCCGAAAGACCCAACTCCAGGGCCACCGGTTGCCCCAGA ACCAAGTCCAGAGCCAGACCCAGAGCCAACGCCAGATCCAGAACCAAGTCC AGACCCGGATCCGGATTCAGATTCAGATAGTGAGTCAGATTCAGACAGCGA CTCAGGTTCAGATAGCGATTCGGATTCAGACAGCGATTCGGATTCA		
28	***fnbA***	**Fibronectin binding protein A**	ATCGAACTAGTGGATGAATTACCTAAAGAACATGGTCAAGCTCAAGGTCCAA TCGAAGAGATTACTGAAAATTACCAGCATATTTCTCATTCTGGTTTAGGAACT GAAAATGGTCACGGGAATTATGGCGTGATTGAAGAAATTGAAGAAAATAGTC ATGTTGATATTAAGAGTGAATTAGGTTATGAAGGTGGCCAAAATAGCGGTAA CCAGTCATTCGAGGAAGACACAGAAGAAGACAAACCTAAATATGAACAAGGT GGCAATATCGTAGATATCGATTTCGACAGTGTGCCACAAATTCATGGTCAAA ATAATGGTGATCAGTCATTCGAGGAAGACACAGAAAAAGACAAACCTAAGTA TGAACAAGGTGGTAATATCATTGATATCGACTTCGATAGTGTGCCACATATT CATGGATTCAATAAGCATAATGAAATTATTGAAGAAGATACAAACAAAGATAA ACCTAATTATCAATTTGGTGGACACAACAGTGTTGATTTTGAAGAAGATACAC TTCCACAAGTAAGTGGTCATAATGAAGGTCAACAAACGATTGAAGAAGATAC AACACCTCCAATCGTGCCACCAACGCCACCGACACCAGAAGTACCAATCGT GCCACCAACACCACCGACACCAGAA	651	>C247629_ _gene_903 0
28	***fnbA***	**Fibronectin binding protein A**	TTTGAAGAATCTACACATGAAAATTCAAAACATCACGCTGATGTTGTTGAATA TGAAGAGGATACAAACCCAGGTGGTGGTCAGGTTACTACTGAGTCTAACTTA GTTGAATTTGACGAAGAGTCTACAAAAGGTATTGTAACTGGCGCAGTGAGC	834	>SRS0244 24_LANL_s caffold_858

			GATCATACAACAGTTGAAGATACGAAAGAATATACAACTGAAAGTAATCTTAT CGAACTAGTGGATGAATTACCTGAAGAACATGGTCAAGCACAAGGGCCAAT CGAGGAAATCACTGAAAACAATCATCATATTTCTCATTCTGGTTTAGGAACTG AAAATGGTCACGGTAATTATGGCGTGATTGAAGAAATCGAAGAAAATAGCCA CGTTGATATTAAGAGTGAATTAGGTTACGAAGGTGGCCAAAATAGCGGTAAC CAGTCATTCGAGGAAGACACAGAAGAAGATAAACCTAAATATGAACAAGGTG GCAATATCGTAGATATCGATTTCGATAGTGTACCTCAAATTCATGGTCAAAAT AAAGGTGATCAGTCATTCGAAGAAGATACAGAGAAAGACAAGCCTAAATATG AACATGGCGGTAACATCATTGATATCGACTTCGACAGCGTGCCACATATTCA CGGATTCAATAAGCACACTGAAATTATTGAAGAAGATACAAACAAAGACAAA CCTAATTATCAATTCGGTGGACACAACATTGTTGATTTTGAAGAAGATACACT TCCAAAAGTAAGCGGCCAAAATGAAGGTCAACAAACGATTGAAGAAGATACA ACGCCGCCAACACCACCAACACCACCAACACCAGAAGTACCAAGTGAGCCG		__gene_1418
28	***isdA***	**Iron regulated surface determinant protein A**	TCTATCATTTTAGGTTCCCTTGTATACATAGGCGCAGACAGCCAACAAGTCA ATGCGGCAACAGAAGCTACGAACGCAACTAATAATCAAAGCACACAAGTTTC TCAAGCAACATCACAACCAATTAATTTCCAAGTGCAAAAAGATGGCTCTTCA GAGAAGTCACACATGGATGACTATATGCAACACCCTGGTAAAGTGATTAAAC AAAATAATAAATATTATTTCCAAACCGTGTTAAACAATGCATCATTCTGGAAA GAATACAAATTTTACAATGCAAACAATCAAGAATTAGCAACAACTGTTGTTAA CGATAATAAAAAAGCGGATACTAGAACAATCAATGTTGCAGTTGAACCTGGA TATAAGAGCTTAACTACTAAAGTACATATTGTCGTGCCACAAATTAATTACAA TCATAGATATACTACGCATTTGGAATTTGAAAAAGCAATTCCTACATTAGCTG ACGCAGCAAAACCAAACAATGTTAAACCGGTTCAACCAAAACCAGCTCAACC TAAAACACCTACTGAGCAAACTAAACCAGTTCAACCTAAAGTTGAAAAAGTTA AACCTACTGTAACTACAACAAGCAAAGTTGAAGACAATCACTCTACTAAAGTT GTAAGTACTGACACAACAAAAGATCAAACTAAAACACAAACTGCTCATACAG TTAAAACAGCACAAACTG CT	702	>C248343__gene_9437
28	***isdA***	**Iron regulated surface determinant protein A**	ACAGCACAAACTGCTCAAGAACAAAATAAAGTTCAAACACCTGTTAAAGATG TTGCAACAGCGAAATCTGAAAGCAACAATCAAGCTGTAAGTTATAATAAATCA CAACAAACTAACAAAGTTACAAAACATAACGAAACGCCTAAACAAGCATCTA AAGCTAAAGAATTACCAAAAACTGGTTTAACTTCAGTTGATAACTTTATTAGC ACAGTTGCCTTCGCAACACTTGCCCTTTTAGGTTCATTATCTTTATTACTTTTC AAAAGAAAAGAATCTAAATAA	285	>C252359__gene_12017
28	***sceD***	**Transglycosylase**	ATGAAGAAAACATTACTCGCATCATCATTAGCAGTAGGTTTAGGAATCGTAG CAGGAAATGCAGGTCACGAAGCCCATGCAAGTGAAGCGGACTTAAATAAAG CATCTTTAGCGCAAATGGCGCAATCAAATGATCAAACATTAAATCAAAAACCA ATTGAAGCTGGGGCTTATAATTATACATTTGACTATGAAGGGTTTACTTATCA CTTTGAATCAGATGGTACACACTTTGCTTGGAATTACCATGCAACAGGTGCT AATGGAGCAAACATGAGTGCACAAGCACCTGCAACTAATAATGTTGAACCAT CAGCTGTTCAAGCTAATCAAGTACAATCACAAGAAGTTGAAGCACCACAAAA TGCTCAAACTCAACAACCACAAGCATCAACATCAAACAATTCACAAGTTACT GCAACACCAACTGAATCAAAAGCATCAGAAGGTTCATCAGTAAATGTGAATG CTCATCTAAAACAAATTGCTCAACGTGAATCAGGTGGCAATATTCATGCTGT	696	>C253293__gene_12784

			AAATCCAACATCAGGTGCAGCTGGTAAGTATCAATTCTTACAATCAACTTGG GATTCAGTAGCACCTGCTAAATATAAAGGTGTATCACCAGCAAATGCTCCTG AAAGTGTTCAAGATGCCGCAGCAGTAAAATTATATAACACTGGTGGCGCTGG ACATTGGGTTACTGCATAA		
28	***oatA***	**O-acetyltransferase**	ATGGATACAAAAGACTTTAAACGTTTAGAAAAAATGTATTCCCCGCGATACTT ACCCGGATTAGATGGATTGAGGGCATTCGCAGTTATAGGAATCATTATTTAT CACTTGAATGCACAATGGTTAAGTGGGGGCTTTTTAGGAGTAGATACATTCT TCGTTATTTCAGGTTATTTAATAACAAGTTTGTTGATAAGTGAGTATTATCGG ACGCAAAAAATCGATTTGCTAGAGTTTTGGAAGCGACGAT TGAAACGACTCATTCCGGCAGTGTTGTTTTTAATTTGTGTCGTGCTTACGTTC ACACTGATATTTAAACCGGAATTGATTATACAAATGAAACGAGATGCTATTGC AGCTATATTCTATGTTTCAAACTGGTGGTACATCTCACAGAATGTAGATTATT TTAACCAATTTGCTATTGAACCACTAAAACATTTATGGTCTTTAGCCATTGAA GAACAATTTTACTTGCTTTTCCCATTGGTTATCACGTTCTTATTACATAGATTT AAACCGAGAAATATTATTCAAACGCTATTTATTGTATCGTTGATTTCTTTAGG ACTTATGATAGTGATTCATTTCATCACTGGAGATAATTCACGTGTGTATTTTG GGACAGATACACGACTGCAAACTTTATTGCTTGGTTGTATATTAGCATTTATT TGGCCTCCGTTTGCTTTGAAAAAAGATATTTCTAAAAAGATTGTCGTATCATT AGATATTATAGGGATATCTGGTTTTGCGGTTCTAATGACTTTGTTCTTTATAG TTGGAGACCAAGATCAATGGATCTATAATGGAGGATTTTACATTATATCATTT GCAACTTTATTCATTATTGCAATTGCGGTACATCCTTCTAGTTTATTTGCTAAA TTTTTAAGTATGAAACCTTTACTAATTATAGGTAAACGATCATATAGCTTATAT TTATGGCATTATCCTATCATTGTTTTTGTGAACAGTTATTACGTACAAGGACA AATACCGGTATACGTTTATATTATAGAAATTTTGTTAACAGCGTTAATGGCTG AAATTTCGTATCGCTTTATTGAAACACCTATACGTAAAAAAGGATTTAAAGCT TTTGCATTTTTACCTAAAAAGAAGGGGCAATTTGCTAGAACAGTGTTAG TTATCCTATTATTGGTTCCGTCTATCGTTGTGCTCAGTGGACAGTTTGATGCA CTTGGCAAACAACATGAAGCTGAGAAGAAAGAGAAGAAGACGGAATTTAAAA CAACGAAGAAAAAAGTCGTTAAAAAAGATAAGCAAGAGGATAAGCAGACAG CGAATAGCAAAGAGGATATTAAAAAGTCATCACCACTATTAATTGGTGACTC GGTCATGGTGGATATTGGTAATGTCTTTACTAAGAAAATACCAAATGCACAA ATTGATGGTAAAGTTGGACGGCAACTCGTTGATGCTACACCAATTGTGAAAT CGCAATATAAAGACTATGCTAAAAAAGGTCAAAAAGTTGTAGTAGAGCTTGG TACAAATGGGGCATTTACGAAAGATCAATTAAATGAACTATTGGATAGTTTTG GAAAAGCAGACATATATTTAGTTTCTATTAGAGTACCTAGAGATTATGAAGGT AGAATAAATAAATTAATTTATGAGGCAGCTGAAAAGCGCTCTAATGTACATCT AGTCGATTGGTATAAAGCTTCTGCAGGTCATCCGGAGTACTTTGCATATGAC GGTATTCACTTAGAATATGCAGGTAGTAAAGCGCTGACTGATTTGATTGTAA AAACGATGGAAACACATGCTACAAATAAGAAATAA	1812	>SRS024424_LANL_scaffold_1242__gene_2634
30	***fnbA***	**Fibronectin binding protein A**	ATGGCTTATGATGGCTTATTTACAAAGAAAATGGTTGAGTCTTTACAATTTTT AACAACTGGACGCGTTCACAAAATCAATCAACCTGATAATGACACGATACTA ATGGTTGTACGTCAAAATAGACAAAACCATCAATTGTTATTGTCAATCCATCC AAACTTTTCAAGATTACAATTGACTACAAAAAAATATGATAATCCATTTAATCC ACCCATGTTTGCGCGTGTTTTTAGAAAACACTTAGAAGGTGGTATTATCGAA TCGATTAAGCAAATTGGTAATGATCGTCGTATTGAAATCGATATAAAGAGTAA	1698	SRS011105_Baylor_scaffold_899__gene_1274

			AGATGAAATTGGCGATACTATTTACCGTACTGTCATCCTGGAGATTATGGGT AAACATAGTAACTTAATTTTAGTAGATGAAAATCGCAAAATAATTGAAGGATT TAAACACTTAACACCAAATACGAATCACTATCGTACAGTAATGCCAGGTTTTA ATTATGAAGCACCACCTACTCAGCACAAAATGAATCCGTATGATATTACAGG TGCAGAGGTGTTGAAATATATCGATTTTAACACAGGTAATATTGCTAAACAAT TATTGAATCAGTTTGAAGGATTTAGCCCTTTAATTACGAATGAAATCGTTAAT CGTCGTCAATTTATGACTTCATCAACATTACCAGAAGCATTCGATGAAGTAAT GGCAGAAACCAAGTTACCACCTACTCCTATTTTTCATAAAAATCATGAAACAG GTAAAGAGGATTTCTATTTTATAAAGTTAAATCAATTTAATGATGATACAGTTA CGTACGATTCATTAAATGATTTACTTGATCGTTTTTATGATGCGCGTGGCGAA CGTGAACGCGTTAAACAACGTGCGAATGATTTAGTTCGATTTGTTCAACAGC AGTTGCACAAATATCAAAATAAATTAGCGAAGTTGATTGAAGAATATGAGCA GTCTAAAAATAAAGATACTGAACAGTTATATGGTGAATTGATCACTGCTAATA TATATCGAATTAAGCAAGGCGATAAAGAAGTAACGGCATTGAATTATTATAC GAATGAAGAAGTTGTCATTCCTTTAAATCCTACAAAATCCCCATCTGCAAATG CTCAATATTATTATAAACAATATAATCGTATGAAAACGAGAGAACGTGAATTA CAACATCAAATTCAATTGACGAAAGACAATATAGATTATTTTTCAACAATCGA ACAACAATTACATCATATTTCTGTCCATGACATTGATGAAATTAGAGATGAAT TAGCAGAACAAGGCTTTATGAAACAGCGTAAAAATCAAACTAAGAAAAAGAA AGAGCAGATTCAATTACAACATTATGTATCAACTGATGGCGACGATATATATG TTGGCAAGAATAACAAGCAAAATGATTATTTAACAAATAAAAAAGCTAAAAAA ACTCACACATGGTTACACACAAAAGATATTCCGGGTTCACATGTCGTTATATT TAATGATGCACCAAGTGATACGACAATCAAGGAAGCGGCTATGTTAGCAGG ATATTTTTCAAAAGCTGGTAATTCTGGACAAATACCTGTTGATTACACGTTAA TTACAAATGTGCACAAAC CATCAGGTGCAAAGCCTGGGTTTGTAACATATGACAATCAAAAAACTTTGTA TGCTACACCTGATTATGAACATATTCAAAAAATGAAACAATCATAA		

Appendix 73: Information about the genes in the BGC that have the potential to be involved in antimicrobial production

Individual	IMG Metagenome ID	BGC	Gene count in BGC	Gene ID in IMG/ABC	Nucleotide sequence	Amino acid sequence
1	7000000567	16140 2349	4	C176097 __gene_1 1238	ATGGAACAAAAACAAAGCGTGGCAATTGA GACAGAACAGTTAGAGTTGGTTCAAGCAG TATCTGATGAGAGGTTAGAAGTGTTGGTAG GTGGGAAAGGCCCCGGACTGTTAAAGACA CTTACAAAAGACTGTCCTGGTACAGTTTCC GATGTTTGTGTAAATTTAGCAGTTGTTAGC ACCTGTAAAAAATGTGAGTAA	MEQKQSVAIETEQLELVQAVSDERLEVLV GGKGPGLLKTLTKDCPGTVSDVCVNLAVV STCKKCE
1	7000000567	16140 2358	3	SRS0159 37_WUG C_scaffol d_1130__ gene_709	AATATGTTAATTCTAAAGTATGCATCATCAG AATTGAATACGAAAAATGATATAGACAAAA GATACGATATTTATATGTCACATCTGATTAG CCAATTAAAAGATAAGTGGGAGTGCGGAA TTCCTTATAAGAATTCTCCAGGTATGATGTT GGGGCTCTCAGGAATAGGGTTTGGATTAT TAAGTTTAATGAATGAAGACCTACCGTTTA TATTATTACTTGAATAA	NMLILKYASSELNTKNDIDKRYDIYMSHLIS QLKDKWECGIPYKNSPGMMLGLSGIGFGL LSLMNEDLPFILLLE
1	7000000567	16140 2355	4	C177077 __gene_1 2362	ATAGTAGAGCCGATCACTGAACTTTACGCA GCCCTTATCTCGGTGTCCATCAAGATGAAT CTGGTTGATTTCACGGACCACCGTGCGGA ATCACACAAGGTGTTCATCGAGGACGCGC AGTGGGATCAGGTCTTCAGGGATACTCAG ATGCACCGCATCGCGGAGTACCCAAACAC TAGTGACCCGTTGCGCGAATGCGGGCAAC GACTCATCATCGTTGGTGCCGATGACGAC GATGTCCCCACGCTCAACAGTGAAGACAT TCTGGGCTACCTGCGTGCTCACCTACCGG GTTACATGGTTCCGGCTTCCGTCAATGTAC TACCGGAATTGCCGTTAACCTCAAATGGCA AGGTAGACCGCAAGGCCCTAGCGCAACTC TGCTTAGAACCGGTGGGAAGCCCCAACAA CCGGATCGATCCTCCACGCAACGAAACGG AAGAGCAGATAGCGACTATCTGGCGCGAT GTCCTCGACACCACGGAGGTTGGGCGCAA TGATGACTTCTATGCGCTTGGGGGTGACT CACTGCTCATGGCCGAAACGGTAACCCGG CTCCGCCAAGAAATACCTGGCCTACAGCA GCACACGTGGGACGCCCTTATGCGTGGCG TACTCAAGGTGCCAACCATCGCAGGCATTT	IVEPITELYAALISVSIKMNLVDFTDHRAESH KVFIEDAQWDQVFRDTQMHRIAEYPNTSD PLRECGQRLIIVGADDDDVPTLNSEDILGY LRAHLPGYMVPASVNVLPELPLTSNGKVD RKALAQLCLEPVGSPNNRIDPPRNETEEQI ATIWRDVLDTTEVGRNDDFYALGGDSLLM AETVTRLRQEIPGLQQHTWDALMRGVLKV PTIAGISALAQAAGSSCQPEALKAVNSANH TSPELTALSTVASGSPTGSSNLHVYRLPKD ATFCRVMIHAGTGRLKDYEFLMPELLQRQ PEIAHVGFTAGDADRFLDYTTRTLIRDLAQ SYAQELDELDMESYQLVGYCIGGMLALET AKALTELGRDVRQVTCISTHQCPHRVTNE LLCELAYGCIFNADLSAMGANFDLKTLAAA LEHTLDGINRNISDEELCTLEGPYADIGEFF QKMAVLSPRARRKLIYRSIREFDTDSESTR GMLDILYDVFRHSLLGTIDYVPDVYFGDVV VLQPTEGVPGFYPSLGGDIDWPATVLGNL QIHAVAGSHATCLLQENVPSLLPFFTEREQ RNG

					CCGCGCTAGCACAGGCTGCAGGAAGTAGC TGTCAACCCGAAGCATTAAAAGCCGTTAAT TCTGCTAATCATACCTCCCCTGAACTGACT GCTCTGTCTACTGTAGCCAGTGGCTCTCC AACTGGGTCTTCAAATCTGCATGTATACCG ATTGCCGAAGGACGCAACGTTCTGCCGCG TGATGATCCACGCTGGCACTGGCCGCCTG AAGGACTACGAGTTCTTGATGCCCGAACT GCTTCAACGCCAGCCAGAGATTGCCCACG TTGGTTTTACCGCTGGTGACGCAGATCGAT TCCTGGACTACACAACCCGTACGCTGATC AGGGATCTGGCCCAGAGCTACGCCCAGGA ACTTGATGAGTTGGATATGGAATCTTATCA GTTGGTCGGCTACTGCATTGGCGGTATGT TAGCGCTGGAGACCGCGAAGGCGCTCAC CGAACTCGGCCGCGACGTGCGCCAGGTA ACTTGCATCAGCACCCACCAGTGCCCGCA CCGGGTCACTAATGAACTGCTGTGTGAGC TCGCCTATGGGTGCATTTTCAACGCTGACC TAAGTGCAATGGGGGCGAACTTCGACCTC AAAACCCTCGCAGCAGCCCTGGAGCACAC CCTTGATGGCATCAACCGCAATATCAGTGA CGAAGAACTGTGTACCCTTGAAGGCCCAT ACGCAGATATCGGGGAATTCTTCCAAAAGA TGGCGGTGTTGAGTCCCAGGGCGCGCCG CAAACTCATCTACCGAAGCATCCGTGAATT CGACACAGACTCCGAGTCCACTCGCGGGA TGCTGGACATCCTATACGATGTCTTCCGGC ATTCGTTACTCGGAACCATCGACTACGTCC CCGATGTCTATTTCGGTGATGTAGTCGTAC TGCAGCCTACCGAAGGTGTACCTGGTTTTT ACCCGAGCCTAGGCGGAGACATCGATTGG CCAGCGACCGTGCTCGGCAACCTGCAGAT ACATGCCGTGGCCGGTTCTCATGCCACCT GCCTGCTACAGGAGAACGTACCCTCACTA CTTCCGTTCTTCACGGAGAGGGAACAACG AAATGGCTAA	
1	7000000567	16140 2357	2	C177313 __gene_1 2852	ATGCGAGATACTTTTAAATTTGTTAAACAG CAAGATATTATGGATTGTGGTGTGGCTTGT ATTCAGATGATTTTAAAACACTATGACTCTG ATATGCCTGCCCATAAACTCCGGCACATGA CAGGTACTGATATTGACGGAACTTCTGCCC TAGGATTGAAGTCCACATTAGAAGAACTAC AATTTGAATGTCTAGCTGTTCAAGCAGATA ATAGTGTGTGGACAGATCCTGACATGAACT	MRDTFKFVKQQDIMDCGVACIQMILKHYD SDMPAHKLRHMTGTDIDGTSALGLKSTLE ELQFECLAVQADNSVWTDPDMNYPAIAHV LLEDNRQHYIIIYGFKNNKLLIADPAEGKYK MSPEEFTTIWTNILLIPKPNKQYQPVVEKIG GLTSFLPALFKSKKIMFFTIIASFLATSLSIVS SYYFQGIIDRVIPQEDLSLLNIMSIGLLGVYL LRVLFDYIRSQLLIILGQSMSSHIMLDYFKH

					ATCCTGCTATTGCCCATGTCTTACTAGAAG ACAATCGTCAACATTACATTATAATTTATGG ATTCAAGAATAATAAACTTTTAATAGCTGAC CCAGCTGAAGGAAAATATAAGATGAGTCC CGAAGAATTTACTACTATTTGGACAAATATT CTCCTTATTCCTAAACCTAACAAACAGTAT CAGCCTGTTGTCGAGAAGATTGGTGGATT AACCTCGTTTTTACCGGCTTTATTTAAGTCT AAAAAGATAATGTTTTTTACTATAATAGCTT CATTTTTAGCTACATCACTGAGTATTGTTAG TTCATATTATTTTCAAGGAATTATTGATAGA GTTATCCCTCAAGAGGATCTATCTCTACTC AACATTATGTCTATCGGATTATTAGGTGTG TATTTACTTCGAGTATTGTTTGACTATATTA GAAGTCAGTTGTTAATTATTTTAGGTCAGA GCATGAGTAGTCATATTATGCTAGACTATT TTAAGCATGTCTTATATCTTCCTATGCAGTT TTTTAACACACGAAAAAACGGAGACATTAT CTCACGATTTTTAGATGCGAATAAGATTATT CGAGCACTTGCTAGTTCAGCTCTTGCTCTA TTCCTGGATATAACGATGGTTGCTATTGTA GGAACATTTTTATTTATTCAGAATAGAATCT TATTTGTAATTACACTTCTATCACTTCCTAT ATACTTTGTCACTGTATTATTTTTTGTGAAG CGTTATAATAAGGCGAACGAAGAAGAAATG AGTTCCTCAGCAATACTAAATTCAAGTATTA TTGAAAGCTTGGATGGC ATGGAGACAATTAAGTCATATAATAGCGAA CAAAATGTATATGAAACCGTTCAAAATCAG TTCTTAAATCTAATGAAAAAATCCTATAAAA CCGTTACATTAGATAATATGCAACGTTCAA TTAAACAATCTGTCCAGTTACTCACTAGTG CAGGGGTTTTATGGGCAGGGTCATATTATA TTCTTCATAATTCAATGAGTATTGGTCAATT GATTACCTATAATGCATTATTAGTATTTTTC ACTAACCCACTAGAAAATATTATTAACTTAC AAGCAGAACTACAGACCGCTGAAATAGCA AGTAAACGTATGAATGAAGTTCTTGCCATT GAATCAGAGTACAATCTCACTGATAAACAG GCCGCTATCACTTTTGATGGAGGGATACG AATAAATCATTTGACATTTTCTTACAACTTA AAAGAGTCTACTCTAAAAAATATTACATGTC ATATTCCATTCAATCAGACGGTTGCTCTAG TCGGAATGAGTGGCTC TGGAAAATCTACTCTAGCCAAATTGTTGCT	VLYLPMQFFNTRKNGDIISRFLDANKIIRAL ASSALALFLDITMVAIVGTFLFIQNRILFVITL LSLPIYFVTVLFFVKRYNKANEEEMSSSAIL NSSIIESLDGMETIKSYNSEQNVYETVQNQ FLNLMKKSYKTVTLDNMQRSIKQSVQLLTS AGVLWAGSYYILHNSMSIGQLITYNALLVF FTNPLENIINLQAELQTAEIASKRMNEVLAI ESEYNLTDKQAAITFDGGIRINHLTFSYNLK ESTLKNITCHIPFNQTVALVGMSGSGKSTL AKLLLRLHEVAEGTIQYDNILINKIPHNYLRD HVTYLPQESFFFKGTIIENLLFGLSHQPSEL EIITACEQAKVRHVIDSLPLGFNTPLEEGAA NLSGGQKQRLAIARALLRDTNIYIFDEATSS VDTVTEQKIIQSISELKNKLIIHITHHLPIAKQ CDQILVMHDGQLVEEGTHEQLLANSGTYH TLWQSTFP

					ACGACTTCATGAGGTAGCTGAAGGGACTA TTCAGTATGATAATATACTGATTAATAAAAT TCCACACAACTATTTACGAGACCATGTGAC ATACCTTCCACAAGAATCTTTTTTCTTCAAA GGAACAATTATTGAAAATTTATTGTTTGGGT TATCTCACCAACCTAGTGAATTGGAAATTA TAACTGCTTGTGAACAAGCAAAGGTTCGAC ATGTGATTGACAGCTTGCCATTAGGATTTA ACACACCTCTAGAAGAAGGAGCAGCTAAT CTATCTGGTGGTCAAAAACAACGTCTAGCT ATTGCTCGAGCGTTACTTAGAGATACGAAT ATTTATATTTTTGATGAAGCGACAAGTAGT GTAGATACAGTCACAGAACAAAAAATTATT CAATCGATTTCTGAACTAAAAAATAAATTAA TTATTCACATAACTCACCACTTGCCTATTG CTAAACAATGTGACCAAATACTTGTAATGC ATGATGGTCAACTTGTAGAAGAAGGGACT CATGAGCAATTGTTAGCCAATTCAGGAACG TACCACACATTATGGCAATCAACTTTCCCA TAA	
2	7000000273	161380740	5	C180579__gene_8656	TTGCTAGCACTCGCTCTTCTCCTCGCCGCT GCTTTACTCGCTAGTCTACTTTTTGGTGCC CGGCCTATTGCGTTCGCTGATGCCGTGCA AGCATTAATTAATCTACCTTCATTGCTAGAC GAACCTCAGTCTGGTTCCGCGGACGCTCG CGTTATCGCAGATCTCCGCTGGCCCCGCA CCCTCATCGGCTTATTTGTCGGTGCCGCA CTCGGTGTCGCAGGTGCATTAATCCAAGG CCATACCCGAAACCCACTGGCTGACCCAG GACTATTGGGTGTTTCAGCGGGCGCAGCA TTCGCTGTCGTCTTAGGCTTTTTCGCGTTT GGACTCACTTCTGCCTTGAGCACATCCGT CGTGGCTTTTTGTGGTGCAGTTCTTGCCAC TGCATTAGTTTTCGGTCTTGCCTCGCTGGG TGGGGGACAGATCAATCCACTCACGTTAAT CCTTGGC GGCGCTGCGCTAACAGCGGTGCTACATTC CATGACCACTGCGCTCACACTCATCGACG ACAACAGTCTGGACCGGATGCGCTTCTGG AGCGTTGGCGCATTATCTGGCAGGGATCT TTCAATTTTTTGGGGCACCGCGCCGTTTAT CGCGGTCGGACTCATAGTTGCACTTGCAA CCGCGCCTACGTTGAATCTGCTGAATCTTG GCGACGACGCAGCTAGTGCGCTTGGCGTC	LLALALLLAAALLASLLFGARPIAFADAVQA LINLPSLLDEPQSGSADAR VIADLRWPRTLIGLFVGAALGVAGALIQGH TRNPLADPGLLGVSAGAAFA VVLGFFAFGLTSALSTSVVAFCGAVLATAL VFGLASLGGGQINPLTLILG GAALTAVLHSMTTALTLIDDNSLDRMRFW SVGALSGRDLSIFWGTAPFIA VGLIVALATAPTLNLLNLGDDAASALGVNT YRARLAGMVVIALLAG

					AATACTTACCGGGCTCGGTTGGCCGGCAT GGTGGTCATCGCGTTACTTGCGGGA	
2	7000000273	161380741	5	C180623__gene_8877	GTGCACAATGGAGGACAAACTGTAATTATA CTTCATTTTGAGGAAGGTACTTCTGTTGTC TATAAACCAAGAGATTTAAAAATGGATGTA ACTTTTCAAAGTACAATTAAATGGTTCAATG AAGTGACCAAAAGTCATTTATATAGTTTAAA AATCATAAATCATACTGAATACGGTTGGGT AGAGTATATTCCACACGAAGAGTGTAAGGA TTATAGTGATTTCAAAAACTATTATACTGAA TTGGGACAATTATTATTTTTGTTTTATCTAC TTCGTGGGAATGATATACATTATGAAAATA TTATTGCAAAAGGAAAACATCCTGTATTGA TAGACTTGGAGACATTATTTCATAACAATA CTTCAAATACGAGTGGTATAGATACAGCTG CTGATAGAGTAAATGAACTATTAGAGAATT CTGTCAGGACAGTAGGAATTTTACCTAATT TAGTTTGGGCACAAAATGGTAAAAACGGTG TAGATATTAGTGCCATATCAACTAGTGAGA ATAAGGAAATTCCTATAGAACAAGCCAGTA TTACTAATGTAAACAAAGATAATATGAAAGT AGAATATAAGACATCTACGTTAGCATCTCA AAAAAATAACCCATATATAATAGGAGAAGA AATTAGTCTTACAAGTTATCATAAGTATTTG AAAAAAGGGTTTATTGAATCATACACTAAA ATTAAAAAACATAAATAAAAAAGAAATAATAA ATCAAGTTGAAAATTACAAAGAGATTTATG CAAGACAAATTTTAAGACCTACACAATATTA TACAACATTAAT CCAGATATCATTGCATCCTGATTTTTTGAG AAGTGCTATTGATAGGGAAATGTTGTTTTC CAAGCTATGGATATACTTTGATGAAAATAA TTCATTTAGAAAAGTTTCGGAAATAGAGTTT ACTTCTTTGTTAAAGAATGATATTCCGTGG TTAATATCAAATGTCAGCAAAAACAATATAA CGACTAAGGATGGTTCAGAAATTGAATCGA TATTCAAGCACTCTAGTATTGCTCTCGTTA AAGAAAAGATAAATATTTTAGGGGACAAAG ATTTGACGTTACAGGTGGAATTGATTGAAA CAGCTTTGAATTATGATTCTGAATATAATAA AGCGGAGTCCCAGCGAGAAAATGATAGAA AAATTATAGAGATAAATGATGACAAATTAAA TAAAAATCATTTAGACCAGCAACTTTTAGAA ATATCTACTAATATAGGTGATTATTTGATTA ATCAGTCATTTATTGGAATGAACGGGGATG	VHNGGQTVIILHFEEGTSVVYKPRDLKMDV TFQSTIKWFNEVTKSHLYSLKIINHTEYGW VEYIPHEECKDYSDFKNYYTELGQLLFLFY LLRGNDIHYENIIAKGKHPVLIDLETLFHNN TSNTSGIDTAADRVNELLENSVRTVGILP NLVWAQNGKNGVDISAISTSENKEIPIEQA SITNVNKDNMKVEYKTSTLASQKNNPYIIG EEISLTSYHKYLKKGFIESYTKIKNINKKEIIN QVENYKEIYARQILRPTQYYTTLIQISLHPD FLRSAIDREMLFSKLWIYFDENNSFR KVSEIEFTSLLKNDIPWLISNVSKNNITTKD GSEIESIFKHSSIALVKEKINILGDKDLTLQV ELIETALNYDSEYNKAESQRENDRKIIEIND DKLNKNHLDQQLLEISTNIGDYLINQSFIGM NGDVSWIDMNVIGEKANDWNMVPTSMDL YSGLSGIMIYFIFLYKETKQNKYLIMVKRCY KSIINYIKNVRKRTNINSEVMFGGFSGETPII YALTILEEELGGIFDLDELEKIRSWIFKECKK NISVGNEHDIIIGSSGVIAILLRYYDLTSNDAI LEVCQQYAEQIIDNYIEMDNNSIAWIGIASR NALGGFAHGVSGIVWALSKLYSYLPDERYI EVIEKALRYEDYLYSEDDKNWVDRRETEE GIEYNNLSSNMPVAWCHGASGILLSRASL KKHNLPLSEKRKNKIDEDIEIAVRTTLKNGF GYSHCLCHGDLGNMLILKYASSELNTKNDI DKRYDIYMSHLISQLKDKWECGIPYKNSPG MMLGLSGIGFGLLSLMNEDLPFILLLE

					TTAGTTGGATTGATATGAATGTAATTGGAG AAAAAGCTAATGATTGGAACATGGTGCCTA CTAGTATGGATTTATACAGTGGATTATCAG GAATAATGATTTATTTTATATTTCTGTATAA AGAAACTAAACAAAACAAATATTTAATCATG GTAAAAAGATGTTATAAAAGTATTATAAATT ATATAAAAAATGTTAGAAAAAGAACAAATAT TAATTCAGAGGTTATGTTTGGTGGTTTTTCT GGTGAAACACCTATTATATACGCACTTACG ATACTAGAAGAAGAATTAGGAGGAATATTT GACTTAGATGAATTAGAAAAAATAAGATCA TGGATTTTCAAAGAATGTAAGAAAAATATAT CAGTTGGAAATGAGCATGATATTATAATTG GTTCATCAGGGGTAATAGCTATTTTATTGA GATATTACGATTTGACATCTAATGATGCTAT TTTAGAAGTGTGTCAACAGTATGCAGAACA AATCATTGATAACTATATAGAGATGGATAA CAACAGCATTGCTTGGATAGGTATTGCTTC GAGAAATGCATTGGGAGG TTTTGCGCATGGCGTTTCTGGAATTGTTTG GGCATTATCGAAACTGTATTCTTACTTACC AGATGAACGATATATAGAAGTTATTGAAAA AGCACTAAGATATGAAGATTATTTATATAGT GAAGATGATAAAAACTGGGTTGATAGACGT GAAACAGAAGAAGGTATTGAATATAATAAC TTGAGTTCCAACATGCCCGTTGCATGGTGT CATGGAGCAAGTGGAATTCTATTAAGTAGA GCGTCATTA AAAAAACATAATCTACCTTTATCAGAAAAAA GAAAAAATAAAATTGATGAAGACATTGAAA TTGCTGTTAGAACCACTTTAAAAAATGGAT TTGGATATAGTCACTGTTTGTGTCATGGAG ATTTGGGAAATATGTTAATTCTAAAGTATGC ATCATCAGAATTGAATACGAAAAATGATAT AGACAAAAGATACGATATTTATATGTCACA TCTGATTAGCCAATTAAAAGATAAGTGGGA GTGCGGAATTCCTTATAAGAATTCTCCAGG TATGATGTTGGGGCTCTCAGGAATAGGGT TTGGATTATTAAGTTTAATGAATGAAGACCT ACCGTTTATATTATTACTTGAATAA	
8	7000000413	161390751	6	SRS019986_Baylor_scaffold_1888__gene_2397	ATGGATGTCACAGCGCTGATTAACGACCTT GAATCCCGCGGTATAGCGCTATGGGTGAA CGGCGATCGACTCAACTACCGTTCACCCA AGGGTTCCTTGCGCGAGGAAGACCTCGCG GCCTTAAGGTCAAATAAGGAAAAGGTTTTG	MDVTALINDLESRGIALWVNGDRLNYRSP KGSLREEDLAALRSNKEKVLAWLREREAV PHDEQARFAPFPMTDIQRAYATGQNEGY DLGGTGCHSYAEIRTERLDRSRLEQAWHE LIQRHDMLSAVVVPPDSLQVVKSRSLPVL

					GCATGGTTGCGTGAGCGGGAGGCCGTTC	QAVDLAGHNPDVPDAEYLRHRAKLENRSY
					CGCATGATGAGCAAGCACGGTTCGCCCCC	PLGTWPLHEFQLLQFDECSILQFSVDM
					TTCCCCATGACCGATATTCAGCGCGCATAC	IIADFVSVRVMVEELLTLYAGNVLPELEDTT
					GCAACCGGTCAGAACGAAGGCTATGATCT	FRDIITSRNHHSQSAAGFAARTNAKKYWS
					GGGTGGCACCGGGTGCCACAGCTACGCC	EIIPSLSGKPLLPTLTSADRTSEMPVRFTRR
					GAAATACGAACGGAACGCCTTGATCGCAG	TWRCSPAAWSKLTDAASTHGVTPSATLLT
					CCGCCTTGAACAGGCCTGGCACGAACTCA	AYADVLRRWSSTSDFCVNVTSMNRDSAIA
					TCCAGCGCCATGACATGCTCTCTGCAGTG	GINRIIGDFTEMTLHACHPHTGTFSERVHA
					GTCGTTCCACCGGATTCGCTGCAGGTGGT	TQEQLSEELSHAAYSGVDVLRDIARTTGQ
					TAAATCTCGCAGTTTGCCGGTGCTACAAGC	PAVIPVVFTSALGADTPHNNGPAYNLVSG
					TGTAGACCTCGCGGGCCACAACCCAGATG	VSRTPQVWIDCQAFQDGGSCNVNWDVRE
					TCCCGGACGCCGAGTATCTTCGTCACCGC	DVFEPALIDDMWESFTDLLDRLVDDGSAW
					GCAAAATTGGAAAACCGTAGCTACCCGTT	QETDSVHLPDKTIAIRNRIHKTHVQQTTRC
					GGGCACCTGGCCACTGCACGAATTCCAGC	LHDGFWDNVQQHPHQPALVCGGKTYSYQ
					TTTTGCAATTTGACGAATGCAGCATCCTAC	HLAGYVGALQHELSDVGPGDYIAIVLGNG
					AGTTTTCGGTGGACATGATCATCGCCGACT	VWQIAAAVAVVSTGAAYVPIDHEQPAIRQR
					TTGTAAGCGTGAGGGTGATGGTCGAGGAA	SMIEACRPANVITNSHFSEENTDISNINVDT
					CTTCTAACTCTCTACGCAGGTAACGTTCTA	LSPIQYSGTIASPVSPTETAYIIFTSGSTGIP
					CCGGAACTAGAGGACACCACGTTTCGGGA	KGVVVTHSAAMNTIDSVNNLLGRNKRRTV
					TATAATCACCTCCCGTAACCACCACAGCCA	LGVSKLSFDLSVYDIFGTFASGGTLVLPLD
					GAGTGCCGCTGGCTTCGCTGCACGCACCA	EESRNPSKWIDFLVDNNVDTWNSVPALFQ
					ATGCAAAAAAGTACTGGTCCGAGATCATCC	MLVREVEVTRHPNILSLDLVMLSGDRIPGT
					CCTCGTTGTCCGGCAAACCCTTACTGCCG	LPAHAAPHFPNAELISLGGATEGGIWSIFH
					ACGCTGACCTCAGCCGATCGCACATCCGA	PMTCHTNETSIPYGTALPNQGMWVLDEA
					AATGCCGGTGCGTTTCACCCGGCGTACCT	CNECPDWVRGQIHISGESLATGYLNDPTS
					GGCGATGCTCGCCAGCAGCATGGTCAAAG	TAEKFFFSEKHGTRMYATGDIGSYRPDGVI
					CTCACCGACGCAGCAAGCACTCACGGCGT	EFHGRRDNQLKINGYRVETGEIEGVLESN
					CACTCCTTCGGCGACGCTACTTACCGCCT	DFVERAIVLAQETSDPIKLHAFVTDAQSNK
					ACGCCGATGTACTGCGCCGTTGGTCCTCC	DELKDAGQIRNSELRTMLEQRWTPADTSL
					ACAAGTGACTTTTGCGTCAATGTCACTTCG	DTGIFATWMRLGNEAAMAALLAAFQQAGV
					ATGAACCGAGATTCTGCCATTGCCGGAAT	FLVAGKHHTLTEITAAIHPSEEYRELITRWL
					CAACCGTATTATCGGAGACTTTACCGAGAT	NILTGEGLATKDDEGWTVSQQTLDFFVFG
					GACCCTACATGCTTGCCATCCGCACACTG	EAWDQFGNMEAEIN
					GAACCTTTAGTGAGCGCGTGCACGCAACC	NSKELFNYQRHAAEALLSQLRGEISPTEVF
					CAGGAGCAGCTATCCGAAGAGCTATCACA	FPEGDTHNARTIYGENRISKAMNAAAAEA
					CGCTGCGTATTCCGGGGTTGACGTGTTGC	VIGIAEHHADHPVRILEVGAGIGATTEKIVS
					GCGATATAGCCCGCACTACCGGGCAGCCC	RLPENVIEYRFTDISTFFLHKAQKMFAHCG
					GCGGTAATCCCGGTAGTATTCACCAGCGC	AMTYGLFDMNSDCTSQDVEFGGYDIILCA
					GTTGGGTGCAGACACACCGCACAACAATG	NVLHNSVNIEESFTRLKQLRRPGGVIVIVEP
					GTCCGGCCTACAACCTTGTCTCCGGCGTA	ITELYAALISVSIKMNLVDFTDHRAESHKVFI
					AGCCGAACGCCGCAGGTATGGATTGACTG	EDAQWDQVFRDTQMHRIAEYPNTSDPLR
					TCAGGCATTCCAGGACGGAGGCTCTTGCA	ECGQRLIIVGADDDDVPTLNSEDILGYLRA
					ACGTCAACTGGGACGTGCGAGAGGACGTC	HLPGYMVPASVNVLPELPLTSNGKVDRKA
					TTTGAACCGGCGTTGATCGACGACATGTG	LAQLCLEPVGSPNNRIDPPRNETEEQIATI
					GGAGTCATTTACCGACCTACTCGATCGCCT	WRDVLDTTEVGRNDDFYALGGDSLLMAE

					GGTCGACGACGGTTCGGCGTGGCAGGAA ACTGATTCAGTCCACTTGCCCGATAAGACG ATCGCCATCCGTAACCGCATTCACAAAACC CACGTACAGCAGACCACCCGATGCCTCCA CGACGGGTTTTGGGATAACGTCCAGCAGC ACCCACACCAGCCAGCGCTGGTGTGTGGC GGAAAAACCTACAGCTACCAACATCTGGCT GGCTACGTCGGGGCATTACAGCATGAGCT CTCTGATGTCGGTCCCGGAGATTACATTG CCATCGTCCTAGGTAACGGAGTATGGCAA ATAGCTGCCGCTGTTGCCGTGGTATCAAC TGGTGCAGCCTACGTGCCGATCGATCACG AGCAACCCGCAATCCGCCAACGTTCAATG ATAGAAGCGTGTCGTCCTGCCAACGTCAT CACTAATTCTCATTTCTCCGAAGAAAATAC CGATATCTCCAACATCAACGTCGATACTCT CAGCCCGATACAGTACAGTGGCACCATAG CCTCACCGGTTTCCCCCACCGAAACCGCC TACATCATCTTCACCTCAGGCAGCACCGG AATTCCAAAAGGTGTTGTCGTTACGCACTC GGCTG CAATGAACACCATAGATAGCGTGAACAATC TCCTCGGCCGAAATAAAAGACGCACGGTA TTGGGGGTATCGAAGTTATCCTTTGACTTA TCCGTGTATGACATCTTCGGTACCTTCGCA AGCGGTGGCACTTTGGTCCTGCCGTTAGA CGAGGAATCGCGAAACCCCAGCAAATGGA TCGATTTCCTTGTTGATAACAACGTCGATA CCTGGAACTCTGTTCCTGCACTGTTCCAAA TGCTTGTGCGAGAGGTAGAGGTGACCCGT CACCCGAACATCCTCAGCCTAGACCTGGT TATGCTGTCCGGAGACAGGATCCCGGGCA CTCTCCCAGCGCATGCTGCACCCCACTTC CCTAATGCTGAACTCATCAGCCTCGGCGG CGCCACGGAGGGCGGAATATGGTCGATTT TCCACCCGATGACCTGTCACACAAATGAAA CTAGCATACCATATGGTACCGCGTTACCCA ACCAAGGAATGTGGGTGCTCGACGAGGCC TGCAATGAGTGTCCGGACTGGGTGCGAGG GCAAATCCATATTTCCGGGGAAAGCCTAG CCACCGGTTACCTCAACGATCCAACCTCC ACAGCAGAGAAATTTTTCTTTTCCGAAAAG CATGGCACACGTATGTACGCTACCGGGGA TATAGGAAGCTACCGCCCAGACGGCGTAA TCGAGTTTCACGGCCGTCGGGACAATCAA	TVTRLRQEIPGLQQHTWDALMRGVLKVPT IAGISALAQAAGSSCQPEALKAVNSANHTS PELTALSTVASGSPTGSSNLHVYRLPKDAT FCRVMIHAGTGRLKDYEFLMPELLQRQPEI AHVGFTAGDADRFLDYTTRTLIRDLAQSYA QELDELDMESYQLVGYCIGGMLALETAKA LTELGRDVRQVTCISTHQCPHRVTNELLC ELAYGCIFNADLSAMGANFDLKTLAAALEH TLDGINRNISDEELCTLEGPYADIGEFFQK MAVLSPRARRKLIYRSIREFDTDSESTRGM LDILYDVFRHSLLGTIDYVPDVYFGDVVVL QPTEGVTGFYPSLGGDIDWPATVLGNLQI HAVAGSHATCLLQENVPSLLPFFTEREQR NG

					CTAAAAATCAACGGCTATCGAGTGGAAACC GGCGAGATAGAAGGAGTCCTCGAATCCAA CGACTTCGTCGAACGCGCTATAGTACTCG CCCAGGAAACCAGTGACCCAATTAAACTAC ATGCTTTCGTCACTGATGCACAGAGCAACA AAGATGAGCTCAAGGATGCTGGGCAAATC AGAAACTCGGAGCTCCGTACAATGCTTGA GCAGCGCTGGACACCTGCAGATACGAGCC TTGACACGGGAATATTTGCCACATGGATGC GGCTGGGAAACGAAGCCGCGATGGCTGC TTTGCTCGCGGCTTTCCAACAAGCCGGTG TTTTCCTCGTTGCAGGAAAACACCACACTC TTACGGAAATCACCGCGGCGATCCATCCC TCCGAAGAGTATCGCGAACTCATAACCCG TTGGCTCAATATTCTGACTGGAGAAGGCTT AGCCACTAAGGATGATGAGGGTTGGACTG TCAGCCAACAGACCCTGGACTTCTTCGTGT TCGGAGAGGCCTGGGATCAGTTCGGCAAC ATGGAAGCGGAAATCAATAACAGTAAGGA ATTATTCAATTACCAACGGCACGCAGCTGA GGCGCTTCTTTCCCAGTTGCGCGGAGAGA TCAGTCCCACCGAGGTGTTCTTCCCGGAG GGAGATACCCATAACGCCCGCACCATTTA CGGCGAAAACCGCATTAGCAAGGCGATGA ATGCTGCGGCCGCAGAGGCAGTGATCGG CATTGCCGAGCACCACGCTGATCATCCGG TAAGGATCCTCGAGGTCGGCGCTGGCATA GGTGCTACGACGGAAAAAATTGTTAGCCG ACTCCCCGAGAACGTAATCGAATACCGCTT CACCGATATTTCGACATTCTTCCTACATAA AGCCCAAAAAATGTTCGCGCATTGTGGTG CCATGACCTATGGCCTGTTTGATATGAACT CAGACTGCACATCGCAAGATGTAGAGTTC GGCGGTTACGACATCATTTTGTGCGCCAA CGTACTGCATAATTCAGTAAATATCGAGGA GTCATTCACCCGCTTAAAGCAGCTGCGCC GCCCCGGGGGCGTGATCGTCATAGTAGAG CCGATCACTGAACTTTACGCAGCCCTTATC TCGGTGTCCATCAAGATGAATCTGGTTGAT TTCACGGACCACCGTGCGGAATCACACAA GGTGTTCATCGAGGACGCGCAGTGGGATC AGGTCTTCAGGGATACCCAGATGCACCGC ATCGCGGAGTACCCAAACACTAGTGACCC GTTGCGCGAATGCGGGCAACGACTCATCA TCGTTGGTGCCGATGACGACGATGTCCCC	

					ACGCTCAACAGTGAAGACATTCTGGGCTA CCTGCGTGCTCACCTACCGGGTTACATGG TTCCGGCTTCCGTCAATGTACTACCGGAAT TGCCGTTAACCTCAAATGGCAAGGTAGAC CGCAAGGCCCTA GCGCAACTCTGCTTAGAACCGGTGGGAAG CCCCAACAACCGGATCGATCCTCCACGCA ACGAAACGGAAGAGCAGATAGCGACTATC TGGCGCGATGTCCTCGACACCACGGAGGT TGGGCGCAATGATGACTTCTATGCGCTTG GGGGTGACTCACTGCTCATGGCCGAAACG GTAACCCGGCTCCGCCAAGAAATACCTGG CCTACAGCAGCACACGTGGGACGCCCTTA TGCGTGGCGTACTCAAGGTGCCAACCATC GCAGGCATTTCCGCGCTAGCACAGGCTGC AGGAAGTAGCTGTCAACCCGAAGCATTAA AAGCCGTTAATTCTGCTAATCATACCTCCC CTGAACTGACTGCTCTGTCTACTGTAGCCA GTGGCTCTCCAACTGGGTCTTCAAATCTGC ATGTATACCGATTGCCGAAGGACGCAACG TTCTGCCGCGTGATGATCCACGCTGGCAC TGGCCGCCTGAAGGACTACGAGTTCTTGA TGCCCGAACTGCTTCAACGCCAGCCAGAG ATTGCCCACGTTGGTTTTACCGCTGGTGAC GCAGATCGATTCCTGGACTACACAACCCG TACGCTGATCAGGGATCTGGCCCAGAGCT ACGCCCAGGAACTTGATGAGTTGGATATG GAATCTTATCAGTTGGTCGGCTACTGCATT GGCGGTATGTTAGCGCTGGAGACCGCGAA GGCGCTCACCGAACTCGGCCGCGACGTG CGCCAGGTAACTTGCATCAGCACCCACCA GTGCCCGCACCGGGTCACTAATGAACTGC TGTGTGAGCTCGCCTATGGGTGCATTTTCA ACGCTGACCTAAGTGCAATGGGGGCGAAC TTCG ACCTCAAAACCCTCGCAGCAGCCCTGGAG CACACCCTTGATGGCATCAACCGCAATATC AGTGACGAAGAACTGTGTACCCTTGAAGG CCCATACGCAGATATCGGGGAATTCTTCCA AAAGATGGCGGTGTTGAGTCCCAGGGCGC GCCGCAAACTCATCTACCGAAGCATCCGT GAATTCGACACAGACTCCGAGTCCACTCG CGGGATGCTGGACATCCTATACGATGTCTT CCGGCATTCGTTACTCGGAACCATCGACT ACGTCCCCGATGTCTATTTCGGTGATGTAG	

					TCGTACTGCAGCCTACCGAAGGTGTAACT GGTTTTTACCCGAGCCTAGGCGGAGACAT CGATTGGCCAGCGACCGTGCTCGGCAACC TGCAGATACATGCCGTGGCCGGTTCTCAT GCCACCTGCCTGCTACAGGAGAACGTACC CTCACTACTTCCGTTCTTCACGGAGAGGG AACAACGAAATGGCTAA	
12	7000000529	16140 0179	3	C158735 __gene_8 686	ATGAAATTGATAGATAAGTTTAAAAATGGTT TATATTCTTTTGAAAGAGATATTCAGCATGA TGAGACTAATAATTATAAATCAGAAAACAG GTTACAGTATTGGAAAAAATTTCTAGGTGT TAACGAAAGAGAAATTGAGAATATTTTATC CAATGGATTAGGGATAAATACAGCTAATTT AAACGAATTATTATCGGAAAATGATAATTTC TCATGCAAAGTTACTGAAACGAATGTATTG TGGAACCAATTGATTCACGACTTACAAGTC CTTTCCATTGAATCTATTATTTTACCTGAAT TTTATATAATAGGAGATATTGGGCAAAAAG AACTACCAATGTTTTATGGTTTTCATGAACC ATTTTTAAAATTAGCGATATTAAGATTTGAA AACTATTGGAAAAATATTCCCGGTATATCA GATAATGTATTTAATAAATTATTAATCTATTT ATATGACCAATTAGCCGAAATTTCATATAG AACTCTTATTTTAGAATTAAATATAGCTAGA GAAGAAAACAAACTAGCAGGAGAAACATCT GAGGAAAGATATAATTATTTTTCGACTCAAT ACTTAAGTGATAATTATTGGTTGATTTTAGA GGAATACCCTGTAATGTTCAGGTTGATGTG TGAGGCAACTCAAAAATGGATAAATAATAC AACACGATTTATAGATAGAATTTTGTCAGAT AAAGATGATTTAGAAAAAACTTTTAAAATTG AGGGAGAATTAAACAGTATTGAATTGAATA CTTCTGATGTGCACAATGGAGGACAAACT GTAATTATACTTCATTTTGAGGAAGGTACTT CTGTTGTCTATAAACCAAGAGATTTAAAAAT GGATGTAACTTTTCAAAGTACAATTAAATG GTTCAATGAAGTGACCAAAAGTCATTTATA TAGTTTAAAAATCATAAATCATACTGAATAC GGTTGGGTAGAGTATATTCCACACGAAGA GTGTAAGGATTATAGTGATTTCAAAAACTA TTATACTGAATTGGGACAATTATTATTTTTG TTTTATCTACTTCGTGGGAATGATATACATT ATGAAAATATTATTGCAAAAGGAAAACATC CAGTATTGATAGACTTGGAGACATTATTTC ATAACAATACTTCAAATACGAGTGGTATAG	MKLIDKFKNGLYSFERDIQHDETNNYKSEN RLQYWKKFLGVNEREIENILSNGLGINTAN LNELLSENDNFSCKVTETNVLWNQLIHDLQ VLSIESIILPEFYIIGDIGQKELPMFYGFHEP FLKLAILRFENYWKNIPGISDNVFNKLLIYLY DQLAEISYRTLILELNIAREENKLAGETSEE RYNYFSTQYLSDNYWLILEEYPVMFRLMC EATQKWINNTTRFIDRILSDKDDLEKTFKIE GELNSIELNTSDVHNGGQTVIILHFEEGTS VVYKPRDLKMDVTFQSTIKWFNEVTKSHL YSLKIINHTEYGWVEYIPHEECKDYSDFKN YYTELGQLLFLFYLLRGNDIHYENIIAKGKH PVLIDLETLFHNNTSNTSGIDTAADRVNELL ENSVRTVGILPNLVWAQNGKNGVDISAIST SENKEIPIEQASITNVNKDNMKVEYKTSTLA SQKNNPYIIGEEISLTSYHKYLKKGFIESYT KIKNINKKEIINQVENYKEIYARQILRPTQYY TTLIQISLHPDFLRSAIDREMLFSKLWIYFD ENNSFRKVSEIEFTSLLKNDIPWLISNVSKN NITTKDGSEIESIFKHSSIALVKEKINILGDK DLTLQVELIETALNYDSEYNKAESQRENDR KIIEINDDKLNKNHLDQQLLEISTNIGDYLIN QSFIGMNGDVSWIDMNVIGEKANDWNMV PTSMDLYSGLSGIMIYFIFLYKETKQNKYLI MVKRCYKSIINYIKNVRKRTNINSEVMFGG FSGETPIIYALTILEEELGGIFDLDELEKIRS WIFKECKKNISVGNEHDIIIGSSGVIAILLRY YDLTSNDAILEVCQQYAEQIIDNYIEMDNN SIAWIGIASRNALGGFAHGVSGIVWALSKL YSYLPDERYIEVIEKALRYEDYLYSEDDKN WVDRRETEEGIEYNNLSSNMPVAWCHGA SGILLSRASLKKHNLPLSEKRKNKIDEDIEIA VRTTLKNGFGHSHCLCHGDLGNMLILKYA SSELNTKNDIDKRYDIYMSHLISQLKDKWE CGIPYKNSPGMMLWLSGIGFGLLSLMNED LPFILLLE

					ATACAGCTGCTGATAGAGTAAATGAACTAT TAGAGAATTCTGTCAGGACAGTAGGAATTT TACCTAATTTAGTTTGGGCACAAAATGGTA AAAACGGTGTAGATATTAGTGCCATATCAA CTAGTGAGAATAAGGAAATTCCTATAGAAC AAGCCAGTATTACTAATGTAAACAAAGATA ATATGAAAGTAGAA TATAAGACATCTACGTTAGCATCTCAAAAA AATAACCCATATATAATAGGAGAAGAAATT AGTCTTACAAGTTATCATAAGTATTTGAAAA AAGGGTTTATTGAATCATACACTAAAATTAA AAACATAAATAAAAAAGAAATAATAAATCAA GTTGAAAATTACAAAGAGATTTATGCAAGA CAAATTTTAAGACCTACACAATATTATACAA CATTAATCCAGATATCATTGCATCCTGATTT TTTGAGAAGTGCTATTGATAGGGAAATGTT GTTTTCCAAGCTATGGATATACTTTGATGA AAATAATTCATTTAGAAAAGTTTCGGAAATA GAGTTTACTTCTTTGTTAAAGAATGATATTC CGTGGTTAATATCAAATGTCAGCAAAAACA ATATAACGACTAAGGATGGTTCAGAAATTG AATCGATATTCAAGCACTCTAGTATTGCTC TCGTTAAAGAAAAGATAAATATTTTAGGGG ACAAAGATTTGACGTTACAGGTGGAATTGA TTGAAACAGCTTTGAATTATGATTCTGAATA TAATAAAGCGGAGTCCCAGCGAGAAAATG ATAGAAAAATTATAGAGATAAATGATGACA AATTAAATAAAAATCATTTAGACCAGCAACT TTTAGAAATATCTACTAATATAGGTGATTAT TTGATTAATCAGTCATTTATTGGAATGAAC GGGGATGTTAGTTGGATTGATATGAATGTA ATTGGAGAAAAAGCTAATGATTGGAACATG GTGCCTACTAGTATGGATTTATACAGTGGA TTATCAGGAATAATGATTTATTTTATATTTC TGTATAAAGAAACTAAACAAAACAAATATTT AATCATGGTAAAAAGATGTTATAAAAGTATT ATAAATTATATAAAAAATGTTAGAAAAAGAA CAAATATTAATTCAGAGGTTATGTTTGGTG GTTTTTCTGGTGAAACACCTATTATATACG CACTTACGATACTAGAAGAAGAATTAGGAG GAATATTTGACTTAGATGAATTAGAAAAAT AAGATCATGGATTTTCAAAGAATGTAAGAA AAATATATCAGTTGGAAATGAGCATGATAT TATAATTGGTTCATCAGGGGTAATAGCTAT TTTATTGAGATATTACGATTTGACATCTAAT	

					GATGCTATTTTAGAAGTGTGTCAACAGTAT GCAGAACAAATCATTGATAACTATATAGAG ATGGATAACAACAGCATTGCTTGGATAGGT ATTGCTTCGAGAAATGCATTGGGAGGTTTT GCGCATGGCGTTTCTGGAATTGTTTGGGC ATTATCGAAACTGTATTCTTACTTACCAGAT GAACGATATATAGAAGTTATTGAAAAAGCA CTAAGATATGAAGATTATTTATATAGTGAAG ATGATAAAAACTGGGTTGATAGACGTGAAA CAGAAGAAGGTATTGAATATAATAACTTGA GTTCCAACATGCCCGTTGCATGGTGTCAT GGAGCAAGTGGAATTCTATTAAGTAGAGC GTCATTAAAAAAACATAATCTACCTTTATCA GAAAAAAGAAAAAATAAAATTGATGAAGAC ATTGAAATTGCTGTTAGAACCACTTTAAAAA ATGGATTTGGACATAGTCACTGTTTGTGTC ATGGAGATTTGGGAAATATGTTAATTCTAA AGTATGCATCATCAGAATTGAATACGAAAA ATGATATAGACAAAAGATACGATATTTATAT GTCACATCTGATTAGCCAATTAAAAGATAA GTGGGAGTGCGGAATTCCTTATAAGAATTC TCCAGGTATGATGTTGTGGCTCTCAGGAAT AGGGTTTGGATTATTAAGTTTAATGAATGA AGACCTACCGTTTATATTATTACTTGAATAA	
21	7000000424	16139 1586	4	C227022 __gene_3 169	ATGAAATTGATAGATAAGTTTAAAAATGGTT TATATTCTTTTGAAAGAGATATTCAGCATGA TGAGACTAATAATTATAAATCAGAAAACAG GTTACAGTATTGGAAAAAATTTCTAGGTGT TAACGAAAGAGAAATTGAGAATATTTTATC CAATGGATTAGGGATAAATACAGCTAATTT AAACGAATTATTATCGGAAAATGATAATTTC TCATGCAAAGTTACTGAAACGAATGTATTG TGGAACCAATTGATTCACGACTTACAAGTC CTTTCCATTGAATCTATTATTTTACCTGAAT TTTATATAATAGGAGATATTGGGCAAAAAG AACTACCAATGTTTTATGGTTTTCATGAACC ATTTTTAAAATTAGCGATATTAAGATTTGAA AACTATTGGAAAAATATTCCCGGTATATCA GATAATGTATTTAATAAATTATTAATCTATTT ATATGACCAATTAGCCGAAATTTCATATAG AACTCTTATTTTAGAATTAAATATAGCTAGA GAAGAAAACAAACTAGCAGGAGAAACATCT GAGGAAAGATATAATTATTTTTCGACTCAAT ACTTAAGTGATAATTATTGGTTGATTTTAGA GGAATACCCTGTAATGTTCAGGTTGATGTG	MKLIDKFKNGLYSFERDIQHDETNNYKSEN RLQYWKKFLGVNEREIENILSNGLGINTAN LNELLSENDNFSCKVTETNVLWNQLIHDLQ VLSIESIILPEFYIIGDIGQKELPMFYGFHEP FLKLAILRFENYWKNIPGISDNVFNKLLIYLY DQLAEISYRTLILELNIAREENKLAGETSEE RYNYFSTQYLSDNYWLILEEYPVMFRLMC EATQKWINNTTRFIDRILSDKDDLEKTFKIE GELNSIELNTSDVHNGGQTVIILHFEEGTS VVYKPRDLKMDVTFQSTIKWFNEVTKSHL YSLKIINHTEYGWVEYIPHEECKDYSDFKN YYTELGQLLFLFYLLRGNDIHYENIIAKGKH PVLIDLETLFHNNTSNTSGIDTAADRVNELL ENSVRTVGILPNLVWAQNGKNGVDISAIST SENKEIPIEQASITNVNKDNMKVEYKTSTLA SQKNNPYIIGEEISLTSYHKYLKKGFIESYT KIKNINKKEIINQVENYKEIYARQILRPTQYY TTLIQISLHPDFLRSAIDREMLFSKLWIYFD ENNSFRKVSEIEFTSLLKNDIPWLISNVSKN NITTKDGSEIESIFKHSSIALVKEKINILGDK DLTLQVELIETALNYDSEYNKAESQRENDR

					TGAGGCAACTCAAAAATGGATAAATAATAC AACACGATTTATAGATAGAATTTTGTCAGAT AAAGATGATTTAGAAAAAACTTTTAAAATTG AGGGAGAATTAAACAGTATTGAATTGAATA CTTCTGATGTGCACAATGGAGGACAAACT GTAATTATACTTCATTTTGAGGAAGGTACTT CTGTTGTCTATAAACCAAGAGATTTAAAAAT GGATGTAACTTTTCAAAGTACAATTAAATG GTTCAATGAAGTGACCAAAAGTCATTTATA TAGTTTAAAAATCATAAATCATACTGAATAC GGTTGGGTAGAGTATATTCCACACGAAGA GTGTAAGGATTATAGTGATTTCAAAAACTA TTATACTGAATTGGGACAATTATTATTTTTG TTTTATCTACTTCGTGGGAATGATATACATT ATGAAAATATTATTGCAAAAGGAAAACATC CAGT ATTGATAGACTTGGAGACATTATTTCATAA CAATACTTCAAATACGAGTGGTATAGATAC AGCTGCTGATAGAGTAAATGAACTATTAGA GAATTCTGTC AGGACAGTAGGAATTTTACCTAATTTAGTT TGGGCACAAAATGGTAAAAACGGTGTAGA TATTAGTGCCATATCAACTAGTGAGAATAA GGAAATTCCTATAGAACAAGCCAGTATTAC TAATGTAAACAAAGATAATATGAAAGTAGA ATATAAGACATCTACGTTAGCATCTCAAAA AAATAACCCATATATAATAGGAGAAGAAAT TAGTCTTACAAGTTATCATAAGTATTTGAAA AAAGGGTTTATTGAATCATACACTAAAATTA AAAACATAAATAAAAAAGAAATAATAAATCA AGTTGAAAATTACAAAGAGATTTATGCAAG ACAAATTTTAAGACCTACACAATATTATACA ACATTAATCCAGATATCATTGCATCCTGATT TTTTGAGAAGTGCTATTGATAGGGAAATGT TGTTTTCCAAGCTATGGATATACTTTGATG AAAATAATTCATTTAGAAAAGTTTCGGAAAT AGAGTTTACTTCTTTGTTAAAGAATGATATT CCGTGGTTAATATCAAATGTCAGCAAAAAC AATATAACGACTAAGGATGGTTCAGAAATT GAATCGATATTCAAGCACTCTAGTATTGCT CTCGTTAAAGAAAAGATAAATATTTTAGGG GACAAAGATTTGACGTTACAGGTGGAATTG ATTGAAACAGCTTTGAATTATGATTCTGAAT ATAATAAAGCGGAGTCCCAGCGAGAAAAT GATAGAAAAATTATAGAGATAAATGATGAC	KIIEINDDKLNKNHLDQQLLEISTNIGDYLIN QSFIGMNGDVSWIDMNVIGEKANDWNMV PTSMDLYSGLSGIMIYFIFLYKETKQNKYLI MVKRCYKSIINYIKNVRKRTNINSEVMFGG FSGETPIIYALTILEEELGGIFDLDELEKIRS WIFKECKKNISVGNEHDIIIGSSGVIAILLRY YDLTSNDAILEVCQQYAEQIIDNYIEMDNN SIAWIGIASRNALGGFAHGVSGIVWALSKL YSYLPDERYIEVIEKALRYEDYLYSEDDKN WVDRRETEEGIEYNNLSSNMPVAWCHGA SGILLSRASLKKHNLPLSEKRKNKIDEDIEIA VRTTLKNGFGHSHCLCHGDLGNMLILKYA SSELNTKNDIDKRYDIYMSHLISQLKDKWE CGIPYKNSPGMMLWLSGIGFGLLSLMNED LPFILLLE

					AAATTAAATAAAAATCATTTAGACCAGCAAC TTTTAGAAATATC TACTAATATAGGTGATTATTTGATTAATCAG TCATTTATTGGAATGAACGGGGATGTTAGT TGGATTGATATGAATGTAATTGGAGAAAAA GCTAATGATTGGAACATGGTGCCTACTAGT ATGGATTTATACAGTGGATTATCAGGAATA ATGATTTATTTTATATTTCTGTATAAAGAAA CTAAACAAAACAAATATTTAATCATGGTAAA AAGATGTTATAAAAGTATTATAAATTATATA AAAAATGTTAGAAAAAGAACAAATATTAATT CAGAGGTTATGTTTGGTGGTTTTTCTGGTG AAACACCTATTATATACGCACTTACGATAC TAGAAGAAGAATTAGGAGGAATATTTGACT TAGATGAATTAGAAAAAATAAGATCATGGA TTTTCAAAGAATGTAAGAAAAATATATCAGT TGGAAATGAGCATGATATTATAATTGGTTC ATCAGGGGTAATAGCTATTTTATTGAGATA TTACGATTTGACATCTAATGATGCTATTTTA GAAGTGTGTCAACAGTATGCAGAACAAATC ATTGATAACTATATAGAGATGGATAACAAC AGCATTGCTTGGATAGGTATTGCTTCGAGA AATGCATTGGGAGGTTTTGCGCATGGCGT TTCTGGAATTGTTTGGGCATTATCGAAACT GTATTCTTACTTACCAGATGAACGATATATA GAAGTTATTGAAAAAGCACTAAGATATGAA GATTATTTATATAGTGAAGATGATAAAAACT GGGTTGATAGACGTGAAACAGAAGAAGGT ATTGAATATAATAACTTGAGTTCCAACATG CCCGTTGCATGGTGTCATGGAGCAAGTGG AATTCTATTAAGTAGAGCGTCATTAAAAAAA CATAATCTACCTTTATCAGAAAAAAGAAAAA ATAAAATTGATGAAGACATTGAAATTGCTG TTAGAACCACTTTAAAAAATGGATTTGGAC ATAGTCACTGTTTGTGTCATGGAGATTTGG GAAATATGTTAATTCTAAAGTATGCATCATC AGAATTGAATACGAAAAATGATATAGACAA AAGATACGATATTTATATGTCACATCTGATT AGCCAATTAAAAGATAAGTGGGAGTGCGG AATTCCTTATAAGAATTCTCCAGGTATGAT GTTGTGGCTCTCAGGAATAGGGTTTGGATT ATTAAGTTTAATGAATGAAGACCTACCGTT TATATTATTACTTGAATAA	

28	7 000000602	16140 4210	11	C255791 __gene_1 5890	TTAAGTACATATTAGCTGTGGCTAAGGATA AAGAATCATTACATAATAAATCATTAGTGG CTCTTTATCATTTCTGCCCCATAACCTTGT GAATTTTTAAAAGAAATGTGTCACAAATCC ACGCGACATTGATAATCGTTATCAACTAAT ATATGATGTACTTGTAATAAGTATTAATTTT TAACAAATAACAAATTTATATTATGTATATC ATCCAGTGAAATAAATAATATAAGCGGCTT AATTTAGATTATTAATTGAAAGGAGTTACAT ATACGATGAACAATCAGACAAATTGGATAA AGATACTCAGCGGATTTGCTTCTGATAGTA AATGGAAAATTATGTTATCAATTTTGTTGTC TATCATCAGTGTCTTTTCTGGATTAGTTCCT TATTGGGCAGTATTTAAAATTATTTTAATGA TGATTAACAATACATATACGATTAATTCGAT TATGGTTTATATCTTTATTGCTTTAATCGCT TATATTTCACAAGTGTGTTGTTTTGGAGCG TCAACGATGCTATCACATATTACGGCATAT GAAATTTTATCTGAAATTCGTAAAAAATTAG CTCAAAAATTAATGCGTCTCCCTTTAGGCG TAGTGGAATCTAAGAAAATAGGTGAATTAA AAAATATATTTGTCGATAAGGTTGAAACAAT AGAATTACCTTTAGCACATATGATTCCTGA AGTTATAGGAAACTTACTTTTGTCAGCTGC TATCTTTTTATACATAATGCTCATTGATTGG CGTATGGCTAGTGCCTTATTAGTAACGATA CCAATTTCAATTTTCGCTTTTAAAAAAGTTA TGTCTGGATTTAATGAGACATATGCTGAAC AAATGAAATCGAATAATTATATGAATAGCG CGATTGTAGAGTTTATCGAAGGCATTGAGG TAATTAAAACATTTAATCAATCTCAAAGTTC ATATAAAAAATATAAAGATGCAGTAGATAAT TATAAGATTCACACTTTGAATTGGTTTAAAA ATACATGGGGGTATATGAATTTAGGAGCTA GTGTATTACCTTCAACATTTTTAGGGATTTT ACCGGTCGGCATGTATTTGATATCTATCAA CCAATTAAACTATGCGGAATTTTTCCTATG CATCGTATTATCTTTAGGTGTAGTAGCACC GATTAAGAATTTTACTAATTATGTAAATCAT TTAAAGTCTATACAATACGCATTAACTGAA GTGAATCAAATATTAAGTCTAGAAGAGTTA GTATTGTCAACAAAATTTAAGAAGCCTCAA CATTATGAAATTGCTTTTAATAATGTTGGAT TTTCATATAACAAAGATAAGGACGATCTGG TCTTTAAGCATTTATCATTTACAGTACCAGA	MRIFITSTNTDVGKTYVTKHLYHALKTRGH RVCVFKPFQTEERQDGTFPDLEVFKNECD LSYDITSLYTFKQPVSPHLAFKMTDQIFLNK QRVLDKVKVLDKEFDFILIEGAGGIAVPIYE GTDDFYMTKDLINDCADCVISVLPSKLGA ISDAIVHQDYVNQNVSASNFLIMNRYTDSYI EKDNQITIGKLTNKTVYTFEEHATYENFSE AFLKQLIGVKNELHTTT

					AAATAATTTCACAGCTATCGTCGGGGCATC TGGTAGCGGTAAGTCAACCATTGCCAAGC TTATATCACGATACTGGGATGTGACTTCAG GTGAAATTACGATTGGCGGTATAAATATAA AAGATATTGAATCGAAACAACTTAACGATT TAGTTGGTTTTGTTGGACAAGATAACTTTTT ATTAAATCTTACTTTTAAAGAAAATATTAAA CTTGGCAATCCAGAAGCTACGGATGAGGC AGTTGAAAAAGCTGCCAAGTTAGCACAATG TCATGAATTTATTGAAAAGTTGCCAGATGG ATATGATACGAATGTTGGTACAGTGGGAGA TAAATTGTCTGGTGGTGAAAAACAACGAGT CACTATTGCAAGAATGATATTAAAAGATGC GCCTATCATTGTATTAGATGAAGCGACTGC TTACGTTGATCCAGACAACGAACAAAAAAT TCAAGAAGCATTAAATGTGTTGACGCAGGA TAAAACATTGATTGTTATTGCACATCGGTTA TCTACGATTCAACATGCAGATCAAATTATT GTTTTAGGTAAACAACAAATTTTGGAGAAA GGCTCACATCAATTGTTGCTAAAATTAGAA GCTCATTACAAAAATATGTGGGATATGCAT ATGCATACAAAAGATTGGGGAATCAATACT GGACATAATTAAACTAATGAGAAAAGGAGG TCAAATATGTTTCAAATTACATTTAAAATTTT AAACTGGATAAGACCATATAAAGCAAGAAT GATTTTAGGATTTAGCATGTCATTTTTAAAT GCTATTTTTATTGCGTTACCTATCTTTTTAG CTGCAAAAATATTTAATAATGTGCTATCTCA TAAACCTATTTATATGAAAGATATATTGAAT GTTGTAATTATCATGGTTTTATTAGTGATTG GACGATTTATTACAGCATATTTCAAAAGCA AAAGCCATGAGAGCATCGCTTATGAAATGA GTGCGAAAGAACGTTTAGATATAGGGGAT AAATTGAAAAATGTAAGGTTAGGTTATTTTA ATTCGCATCATTCAAATGAGTTAACAACAA TAGTAACAACAGATTTAACCTTTTTAGAAAA CTTTGCTATGAAAATGGTGGACGTTGTTGT TAATGGATACATATTAATTACAGTACTCATA TTGTCTCTACTTGTAGTTTCGTGGCAAGTA TCATTATTAGCATGCATTGGCGTATTACTAT CATTTTTTGCGATTCAATTATTAGAAAGAAA GAGTCGACAAAATGCGCCAGCGTATCATA ATGTACAAAACCAATTAGTGGAAAAGGTAT TGGAGGTTATTCGTGGTATTCAAGTAATAA AATCATTCGCGAAAGAAAATACGAGTCTTA	

					AAAGTTTTAACCAAGCAGTCAATGAAAGTA AACGTATAAATACAAAAATAGAAATGCAAT ATATCCCATTCAATTTATTGCATTTACTTAG TTTAAAAGTTGTTTCAATAATGATTGTATTA GTTGCATGCTTATTATATATGAATCATAGTA TTGATTTACCTACCCTTATTATGATTTCAAT TTTTTCATTTGTGATATTTGATAGTGTTGAA AATATTAATAGTGCTGCACACGTACTTGAA ATGATAGATATGACGATAGATGATATTGAA AAGATAAAAAATGCTCCAGAACTGGATGAG AATGGAAAAAATTTGACGATTAAAAATGAA AATATCGCTTTTCAAAACGTGAATTTTTCAT ATGATGATAAACAAGTGATAAAGAATGTGA ATTTTGAGATACCTACACAAACATCAACAG CAATAATTGGACCTTCAGGAAGTGGGAAAT CTACATTATGTCACTTACTCTTGCGCTTTTA TGATATCGATGATGGAAATATTCGCATCGA TGGTGTTGATATTAAAGATATGACATTAAG TACGTTAATGTCGAAAATTAGTGCAGTATT TCAAAAGGTGTATTTATTTAATGATACGATT GAAAATAACATATTGTTTGGCAATCCAGGT GCAACGAAAGAAGAAATTGTTCGTGCAGC GAAGCAAGCATGTTGCCACGACTTCATCAT GCCATTACCTGAGGGATATCAAACAATGCT AAATGAAAAAGGTAGTAATTTATCTGGCGG AGAAAAGCAAAGGATTTCTATTGCTAGGGC GATATTAAAAGATGCACCAATAATTATTTTA GATGAAGCAACTGCAAGTATTGACCCTGAA AATGAACAGCTGATTCAAACGGCAATTAAT GAATTAAGTAAAGGCAAAACAGTAATTACA ATTGCACATAAACTTGAAACTATTAAAAATG CAGATCAGATTATAGTGCTCAATGAAGGTG AAATAATTCAAAAAGGTAGTCATGATGAAT TAATTCGAAAACCAGGAATGTATCAAGACT TTATAAGAATAAAAAGTAAATCAGCAGGCT GGAAATTATAAATAATAGTTTAATATGGAAT CAGAATTTGGAAATAAAAAGGGTGATAGAT AATCAACAGCAAAATAAATCCCCTCACTAG GGCAATAGTGAGGGGATTTGGTGTATTTA GATGTTCAATTTATTATGTAACTAATTAGAG CATTGATGAACAAGGCGCTCAACGTAATAT TACAATTAATGGATTGTTTCACAACATGGT AAAGCTAGGATGTCCTTTTATTAAGAAATA ATGCAATTATTAATTTTAACTGACATATTTT ATCGAATCGCTGTTTATATCTGCCCGAATT	

					CCGTAAGAAGCCTTAGTCATTTCAATAGCA AGGTCAAAATTTTGCCAAAATTTATAGTTAT ACTGAGAAGTTTAAATTCATGTTTTCTTGAA AAACATGCGCCTTAAATGTAAACTTATTAAT TATAAAAGTTTACATTTGGATTGAGGTGCT TATTTTTTATGAGGATTTTTATTACAAGTAC GAATACTGATGTAGGCAAAACCTATGTTAC AAAGCATTTATACCATGCTTTGAAAACACG TGGTCATCGTGTTTGTGTTTTTAAACCATTT CAAACTGAGGAACGCCAAGACGGGACATT TCCAGATTTAGAAGTATTTAAAAATGAATGT GATTTAAGCTATGACATAACGTCACTTTATA CTTTTAAGCAACCTGTATCACCACACCTTG CATTTAAAATGACAGATCAAATTTTTCTAAA TAAGCAGCGTGTATTAGATAAGGTAAAAGT TTTAGATAAGGAATTTGATTTTATCTTAATT GAGGGTGCTGGGGGAATTGCCGTACCAAT ATATGAAGGTACAGATGATTTCTACATGAC TAAAGATTTAATCAATGATTGTGCAGATTGT GTCATCAGTGTGTTGCCATCAAAATTAGGT GCTATTAGCGATGCCATTGTTCACCAAGAT TATGTTAATCAGAATGTATCGGCGAGTAAT TTTTTAATAATGAATCGCTATACAGACAGCT ATATTGAAAAAGACAATCAAATTACGATTG GAAAATTAACAAATAAAACAGTCTATACATT TGAAGAACATGCCACGTATGAAAATTTCTC AGAAGCATTTTTAAAACAATTAATAGGAGTT AAAAATGAATTACACACAACAACTTAAACA AAAAGACTCAGAATATGTTTGGCATCCATT TACACAAATGGGTGTATATAGCAAAGAAGA AGCAATCATCATTGAAAAAGGAAAGGGTAG TTACCTTTACGATACGAATGGCAATAAATA TTTAGATGGTTATGCATCGTTGTGGGTCAA TGTGCATGGTCATAATAACAAATACTTGAA TAAGGTAATTAAAAAGCAACTCAATAAAATT GCCCATTCTACGCTGCTAGGATCATCAAAT ATTCCGTCAATAGAACTTGCGGAAAAATTA ATCGAAATCACGCCAAGTAATCTAAGAAAA GTATTTTATTCTGATACAGGCAGTGCGTCT GTTGAAATCGCAATAAAGATGGCATATCAG TATTGGAAAAATATTGATAGAGAAAAATAT GCCAAGAAAAACAAGTTTATAACGCTAAAT CACGGTTATCATGGGGATACGATTGGTGC GGTAAGTGTTGGTGGTATCAAGACCTTTCA TAAAATATTTAAAGACTTAATATTTGAGAAT	

					ATTCAAGTAGAAAGCCCATCTTTCTATCGC AGTAATTACGATACTGAAAATGAAATGATG ACAGCTATTTTAACGAATATAGAGCAAATT CTAATTGAAAGAAATGATGAAATCGCAGGG TTTATATTGGAACCGTTGATTCAAGGTGCG ACAGGCTTGTTTGTTCATCCTAAAGGCTTT TTGAAAGAAGTCGAGAAATTGTGCAAAAAA TACGATGTCTTATTAATTTGTGATGAGGTA GCAGTTGGTTTTGGGAGAACTGGAAAGAT GTTTGCATGCAATCATGAAGATGTTCAACC GGATATTATGTGTTTAGGTAAGGCGATTAC TGGTGGCTACTTACCACTTGCAGCTACATT GACATCTCAAAAAATATACAATGCATTTTTA AGTGATTCGCATGGTGTGAATACCTTTTTC CATGGTCATACATACACCGGAAATCAAATC GTTTGTACGGTTGCATTAGAAAATATAAGA CTTTATGAAAAACGTAAGTTATTGTCACATA TTGAAACGACATCATCAACACTTGAGAAAC AGTTACATGCGCTGAAGCGTCATCGAAAT GTTGGTGATGTAAGAGGACGAGGCTTAAT GTTTGGAGTTGAATTAGTTACAGATAAAGA TAGTAAAACGCCGTTAGAAATTGAAAAAGT TGAACGTATTGTACGTAATTGTAAAGAAAA CGGGCTAATGATTAGAAATTTAGAAAATGT CATTACGTTTGTGCCAGTGTTAAGTATGTC AAATAAAGAAGTGAAAACGATGGTACGTAT TTTTAAAAAGGCAGTACATAACATTTTAGAT AGGAAGTGTTAATATGAATTTGGCTAAACG CATATTACAAGGGGAACAATTAACAAAAGA GACTGTATTGAAAATTTATGAGGATACTAA TATTGATACCTTAGATTTATTAAATGAGGC GTACATTTTAAGAAAACATTATTTTGGTAAA AAAGTAAAATTAAACATGATTTTAAATGCTA AAAGTGGTATATGTCCTGAAAATTGTGGGT ACTGTGGACAATCACGAGATATTAAACAAA AACAGCGATATGCTTTAATTCCAGAGGAAA AAATTATCGATGGAGCAAAGGTGGCACAT GATAATCATATTGGAACATATTGTATTGTTA TGAGTGGTAGAGGACCGAGCGATAAAGAA GTTGATCATATTAGTAATACTGTAAGAACG ATTAAATCTCAACACCCGCAACTAAAAATC TGTGCATGTTTAGGATTAACGAATGACGAA CAAGCTAAGAAACTTAAGTCAGCTGGTGTA GACAGATATAACCACAATATTAATACAAGT GAAAATTACCATGATAACGTCGTGACAACG	

					CATAGTTATAAAGATAGAACAGATACGATA GAACTAATGAAAGCGAATAATATATCACCA TGTTCTGGCGTGATTTGTGGTATGGGAGA ATCTAATCAAGATATTGTTGATATGGCATTT GCTTTAAAAGAAATGGATGCCGACAGTATT CCGATTAATTTTTTGCATCCAATCAAAGGT ACAAAGTTTGGAAGCATGGATGATTTAACA CCAATGAAATGTTTAAGAATCGTAGCATTA TTCCGATTAATCAATCCTACGAAAGAAATT CGTATTGCTGGAGGAAGAGAGGTCAACTT ACGTTCGTTACAGCCATTAGCATTAAAAGC GGCGAATTCAATATTTGTCGGTGATTATCT AATTACTGGTGGGCAACCGAACCAATTAGA CTACGATATGATTAATGATTTGGGCTTTGA AATTGATTATGACACTTGTGAAAATAAGGA AAACAAGAATGATGTTTCAAGAGCAAATTG ATTTATTACAACAGAAAGGGTTATATCGGT CGCTTAAATCGGTAGGGTATATAGATAGAC AGTATATTGAAGTAGAAAATAAGAGATGTA CGAACTATACATCTAATGATTATTTAGGATT AGGTCAAATAGCGTTTGATAAGGATGATTT CGAAAGATTTATGCGGAAGTATAGTTATCA CTTATCAAGTTCAAGATTAATTAGTGGAAG TTCGACAGCTTATGAAGAAATTGAAACAAT GTTAGCAGGTTGGCTCGGATATAGTGCAT GTACTATCTTAAATAGTGGTTATGATGCTA ATTTGGCGTTATTTAATATTTTCAAAAATAC AAATTATGTCGTGTTTTCAGATCAAGAAAAT CATGCGAGTATTATTGACGGTATTAAGTTA AGTGGTTTAGAAAAAGTGATATATAAGCAT TTAGATATTGCTGATTTAGAAAAAAGGTTA GAGAAATACCCTAATCAAAATATACCAAAA ATAATCATATCTGATAGTGTATTTTCAACGA ATGGCGACGTTGTGGATATTGGTCAATTAG TCTCATTAAAGCATAAATATAACGCAACAC TGATACTCGATGTTTCACATAGTTTTGGAA TAGAGAATTACTCGAATTATCAAGGTGTAG ATATACTCACTTCTAGTTTATCTAAAGCATG TGGTGCATACGGGGGTGTGATTTTAAGTTC AAATGATGTGAAGGATATGTTAATTAATCA CGGTAGACCACTCATCTACTCAAGTAGTTT GCCAATTTATAATTTGTATTTTATAAAAAGA AACATTGAAAAGTTAATAAATGCTGATGAT AGACGCACTAAATTAAATAGTTTGAGTAAA TATTTTAACCAAAAGTTAAAAGCACTCAATG	

					TTAATTATAATAGTTCAAACTCGCCGATTAA GTTTATTGAGTTTGATGACATAGAAGCGGC TGAGAACATTCATCAAACATTATTAAAGAAT CATGTGTTTACAAGTTATTTAAGGTATCCAA CTGTGACTAAGCCAATGTTAAGAATATCAT TGTCGTATTTTCATACTGAACAAGATATAG ATAGGTTGTTTGAAATTTTGCATCAAGAAG ATTGAGGTGATCGTCATGTATAGTATAAAA ATGCGTTCTAGCAATCAAGATGTTCATATT AGTGGTGCTGAAACGATTTGTGAGTTTGAC AAAATAGAACAGACGGTACAAAGGTTTTAT AATAAAGGTTTTTTTCATGAAAATGGGCAA CCGGATTTTCTCAATATAAAAATACAAAAAA TCATGGAACCGATTCAACAAATAAAAGCAT TACAAATTATTGAGGATGATAAAGCAAATC TCCAACATTTAACACAAGAATGTGGTGTAA CTGAACAAGCGCTAAATCAAGGGATGACA TATATTAAAAATGAAACGGTTTATACAGGA GCTATTATTCTATCTGCTATATCTGGTAAAC GTTTAGATTCATTTGGACAAAGGGGGATTA GGGCAACACATTTTTCGTTTGAAGATATAA ATAATAAAGGTGATCTTAATGAAAGAGTGA CTGATGCCTTAGCAATTGCTAGTTGTATCA ATGCGCATCCGTATGTCAAAGGAGAACTTT GCGTGTCCGATGACTTAACGTATACGACA GGTTATTTTGCCTCTGCTAAAATTGGTTAC CATCGATTATTTGATATTAAACCAGTTAATA CGAGATATGGAGGCAGAATAATATTTGTGG ACGATTGTATTGATTTAAATCATTACATATC ATTTTTAGAAAGCACACCGAAGCAAGTTGT TTATGAAACGGTATAGGGGTTTTAGTATGA CATCAAAAGATATTACTCAAATTAGTGTCAT TGCTGCGATTTTAACCATTTTGGCAGTTTT GAAAATACCGTCCATTATACCAGGATTAGA TTTTCAATTATCTGCACCGGCAGCATTATT GATATTAGCTTTCTTTGGAATTAAAAAGTAC TTTTTAGGTGGATTATTATCTAGCCTATTAT TACTAGTATTTGGCGTATTTAATCCAATTAA TGTGATTATCTCTATTATATTTAGAGTTATA GCTATTGCAGTTGTTTATTTATTGAAAATAA ATGTACTATCATTAGTTTTAGCAAGTGTATT AGGCAGTTTGGTATCTAGGCTACTATTATC TATTATTTTAAATTTACCTGTGTGGGTAGTG TTGTTAAACGCGATTCCAGGCGTAATATTC ACTTTAATTGTAGCTATTCCTTTATATCTCA	

					CATTGAGAAAAAGAATGGCAGTCTTACTAA GATAATAAATCAAAACACGGTCGTCACAAT TACTGTTGGCGACCGTGTTTTACTAGCTAT TTATTGTTTTCAGTTTCTTTTGTATCTAACA ATTTCACTTTGTGATTTTCCCAATCAATTTC ATATGTTGATTTAAATGTTCTAGTTTTAAAG TTTTTATAATTTGCGCCTGCCCAGTAGAAG CCATTCCAACGAATTTGGTATAAATCCATTT CACGTTGATAAGTTACTGTAATTTTAGATTT TTTAGCGCCATCTTGTCTGTGTGATAGTAC GCTTAAAAATTCTGGATTGAAGTTACTTCTA GATAATAATGGCATTTGGTGTTGCGCTATG AAGTTTTGGCCAGCGTATGCACTGCTTTGT CTGCCAGCTAAGAAGAGTTCATTACCATAT GTTGGGTGGAAGCTATCTCGTCCATAAGG TCCCCAACCATTATTCATAATTTTATGTGCT TCAACTCCCCAGCCAACATTTTTATAATTTG TGTTGCGACTTAATGTTGTTCTGTAACTTTC TTGTTTATAATTAATTGTTTCAGAAAAAGCT GTATTTCCATTAAGTCCACCAGATAAACCA TTAGAGATACTAATGTCACCACCAAATGTA TAGCCTAAAGTATTTTGAACTTGAAACTCTT CATTTTGATTTTTTGGTGCATAATCAACGAC GTTTACTGAATCATTAGATTGTGAGCTTATA GATACATTGTATTTAGCTCCCCAATATAATT TTGAAAAGTCATAGTCATTAGGATTAGGTT TCACAAAGCCTGAGTTAATATTCCCAGTAG CTTTAAGTACTAAAGTATCTTTATCATAACT TTTATCTTTGATGAAATTAAATGTTAAAATC TGTGAAATTTTAAATTTATCAGAATCTGCTG TGGCTGTTGTTTTGTATAAAGTAACTTTGTC ATCGACTTTTTTTACGCTGACTGGTGTTATT TTACCTTCAGCATTAGCAGTACCAGAAAGT AATAATAATGCCATAGATGTAGTAACGGAT GATTTGACTAATTTATTCATTTTCATATCAA TTCTGTCCTTTCACCTTGATTTCATGAGTCT TCCAATTGACCTCGTATTTCACAGTATAGT TTCTATTTACAAATGCATTATGGACTCTATG TCCGTCTAAATAACTGTTGCCATAATGTGT TGATCTTTTAATGGCATGAGTGACATCCAT ATTTCTTCCATAAGTGATTTCAAATTCGCTC GTGTCTCCTGAACCTTTTTCATGAGATACT GTGGCGATAAATGAAGGGTTAAATCCACTT TGTACAAGAGGTGGTAATTCGCTGTCTGG AACAAAATAATCTCTTGGATCTTTGCTATTT	

					GGTTTGTATCCTACGAATAAATCACTATCA AATGCTGACTTTTGACCTGATGCAGTGACG AATGAATTCGCTTTGACACCCCATAAAACA CTTTTTGAGTTTTGTTGTTCCACTTCACTGA CATAATTTTGTTGTGTATAGCTAATCGATTT AGAGTAGTTAAAGGATCCATTACCACCGAG TGATGGGGCTGATTGGAAATTACCGCCGA TGTTGTATCCTAATGTCTGGCTCACATTTG TAGATTCAATTTTATTTTTCGGTAAATAATT GATTAAAGAAACATTTGAATCATTTGTTTTT AACCCAATATTATATTGGAAGGGCCAACGC ATAGATTTAATATGATTTGTATTTTTATAGTT GTAATATGTTGTTCTAGAGCTAATGAATCC TTGCATCTTTAAGATCAAAGCATCTTTGTTA TATTTTTTATCCTTCACAAAGTCGAATTGGA TATTTTGAGTCACTCCCCATTTATTACTTGT TTTATCTTCTGTCCTTTTGATAATTTCTATAT CGCTTCCTTTACCGATGTCTTCAGTATCGT TGGCAGCTTTAGCATTTTCTAATAACGGAT TGGCAAGAGGGGCAAGTAAGCTTACAGAT AAAGTTGTAGCTAATATTTTATTTTTAAGCA TAAGTTTCACTTTCTTTCTATAATTTTATTTA ACTTAATTAAATAATAATATATTTTGGACGT TATTATAATTAACTGTTCATTAACAATTGAC CATTATCAACTTTTTTTACAATTAAGTTAAA AATGCCATAATTCCGTTCAATTTAACCTTTT TTGTTTGGAATGCTAAAAGCTAATTTTGTG AGAAATATAGTTAATAAAACCTTGGTATTGT TGACAGTTGTAAACAATTTAAGGTAACCGA TTGCACGAATGTCTTGCTTTTATTTTTAATT AATGTTTGTAAAATACAAATTGTTCTTTTTA CGATTTTTGATGAATCCATCTAAAAGTTAAT TATTTATTAACTTGAGAATATTTATTTAAGC TTTCAACTTAAATTGCTTAATTTGTATTATTT GTGTATCTACTATGGTGCGATACTTTGTTC CTATAAAAATCAGTTGTTATTATTAAATTTA ATGTACGATAAATTAATTTAATGTTATGAGG CTGATTCAATATGTTAAAAAATCCTGAGAC AAGTATGAAGTGTCCCAGGATTTTATAAAA TGATTGAACTGTTTACTTAGGTGTGATGCT TTTAATTTTTACTTCATGTGTTTTCCAGTTC ACTTCATATTTAACTGTAACGTTTCGGTTTT TAAAAGCATCATGTTTTCTATCAACGGCTA AACGATGTCTTGTCACGTAAGCATATGTAG CATCCATGTTTCTGCCGTAAGTGATTTCAA	

					ACTCGCTTTTATCACCTTTACCTCTTTCGTG TGACAATGTTGTAATAAATGATGGATTAAA GCCACTTT	
30	7000000387	16138 8644	6	SRS0111 05_Baylor _scaffold _1331__g ene_2336	ATGAGTCAATTCCTCACACAATTGAAAGAT AATGTTTTAGTAGCTGATGGCGCTATTGGA ACGATTTTATACTCTGAAGGGTTAGACACC TGTCCAGAAGCATATAATCTTAGCCATCCA GATAAAGTTGAACGCATCCATCGTTCATAT ATTGAAGCCGGTGCTGATGTCATTCAAACC AATACTTATGGTGCAAATTTTGAAAAGTTAA AACGATTCGGTCTTGAAGATAAAGTTAAAG CTATACATCAAGCCGCCGTTCGCATCGCA AAAAAAGCAGCAAATAAAGATACGTATATA TTAGGCACAGTTGGTGGGTTTAGAGGTAT CAAACAAGAGGATATCAGCTT ACAAACTATTCTTTATCATACTGAAATTCAA ATAGACACCTTAATTGAAGAAGGCGTTGAC GCGCTACTTTTCGAAACGTATTACGACCTA GAAGAGTTAACAAATGTCATTTCACGAACG AGAAAGAAATACGACATTCCAATCATTGCT CAATTAACCGCTTCAAACACAAATTACTTA GTTAATGGTCAGGCAATCAATGAAGGATTA AAACAACTCGTTCAATGTGGTGCAAACATC GTGGGACTCAATTGTCATCATGGCCCGCA CCATATGCAAGAGTCTTTCACACATATTGA ATTACCAGAGCACGCATACTTATCTTGTTA TCCAAATGCCAGCTTATTAGATATTGAAAA TAGTGAATTTAAGTATAGTGACAATGCACA ATATTTCGGCCAAGTTGCTCAAAATTTAATT CGCGAAGGTGTTCGTTTAATTGGTGGTTG CTGTGGTACAACGCCAGAGCACATCAAATT TATTAAAGAATCTATTCAGACACTTAAACCT GTTAATGACAAAAAAGTGATTCCGATACCA AAGAAAGCACTTTTCAATCCATCTCAAAAT AAAGTTAGACAATCATTAACATCTAAGGTT CAAGAACGTCCAACCGTTATTATCGAATTG GATACACCGAAACATTTAGACACGGATAGA TTTTTTGAAAATATCGCTAAACTTGATAAAG CTAATGTAGATGCGGTAACACTCGCAGATA ATTCATTGGCAACTGTCAGAATTAGCAA CATTGCTGCTGCAAGCTTAATTAAACAATA TTACAATATTGAACCACTCGTACATATTACA TGTCGAGACCGAAACTTAATCGGCTTGCA GTCCCATTTACTTGGATTATCGCTCATTGG CGTTAACGAAATATTAGCCATAACTGGTGA	MSQFLTQLKDNVLVADGAIGTILYSEGLDT CPEAYNLSHPDKVERIHRSYIEAGADVIQT NTYGANFEKLKRFGLEDKVKAIHQAAVRIA KKAANKDTYILGTVGGFRGIKQEDISLQTIL YHTEIQIDTLIEEGVDALLFETYYDLEELTN VISRTRKKYDIPIIAQLTASNTNYLVNGQAIN EGLKQLVQCGANIVGLNCHHGPHHMQES FTHIELPEHAYLSCYPNASLLDIENSEFKYS DNAQYFGQVAQNLIREGVRLIGGCCGTTP EHIKFIKESIQTLKPVNDKKVIPIPKKALFNP SQNKVRQSLTSKVQERPTVIIELDTPKHLD TDRFFENIAKLDKANVDAVTLADNSLATVRI SNIAAASLIKQYYNIEPLVHITCRDRNLIGLQ SHLLGLSLIGVNEILAITGDPSKVGHLPGAT NVYDVNSKGLTELALRFNQGINTDGDALK KRTHFNIAGAFNPNVRKLDGAVKRLEKKIE SGMSYFITQPVYSKEKIIEIYHATKHLNKPF FIGIMPIASYKNALFLHNEVPGIKMSDEILQ QFEAVKDDKAKTRELSLKLSKDLIDTVHEY FNGLYIITPFQNVEDSLELAAYSKSITAHKE AIL

					TCCTTCAAAAGTTGGTCACTTACCAGGTGC AACCAATGTCTATGATGTTAATTCTAAAGG ATTAACTGAACTCGCTCTAAGATTTAATCAA GGTATTAACACTGACGGTGATGCGCTGAA GAAACGTACACACTTCAACATCGCTGGCG CCTTTAACCCTAATGTTCGTAAATTAGATG GTGCCGTAAAACGACTAGAGA AAAAGATAGAAAGCGGAATGTCTTATTTTA TAACACAACCCGTGTACAGCAAAGAGAAAA TCATTGAAATTTACCATGCCACTAAGCACT TGAACAAACCATTTTTCATAGGCATTATGC CTATCGCAAGTTACAAAAACGCACTCTTTT TGCATAATGAAGTGCCAGGAATCAAGATGT CAGATGAAATTTTACAACAATTTGAAGCAG TTAAAGATGATAAAGCCAAAACACGCGAAC TAAGTCTTAAGCTTTCAAAGGATTTAATCG ATACTGTTCATGAATATTTTAATGGTTTATA CATTATCACACCGTTTCAAAATGTCGAAGA TTCATTAGAACTTGCAGCATACTCAAAATC TATTACTGCTCACAAGGAGGCAATATTATG A	
31	7000000566	16140 2347	10	C236577 __gene_1 8145	ATGGAACAAAAACAAAGCGTGGCAATTGA GACAGAACAGTTAGAGTTGGTTCAAGCAG TATCTGATGAGAGGTTAGAAGTGTTGGTAG GTGGGAAAGGCCCCGGACTGTTAAAGACA CTTACAAAAGACTGTCCTGGTACAGTTTCC GATGTTTGTGTAAATTTAGCAGTTGTTAGC ACCTGTAAAAAATGTGAGTAA	MEQKQSVAIETEQLELVQAVSDERLEVLV GGKGPGLLKTLTKDCPGTVSD VCVNLAVVSTCKKCE
31	7000000566	16140 2345	9	C236493 __gene_1 7271	ATGCGCTCAATTATCGCGATCATCACCATT CTTCTTCTTGTCGTATCCACCTACGCCTGG ACCAAACTAGACACTCTTAAAGGAAATACT GCTAACGCTTTTCTCAATATCGGCGGAGG GGAAGACGGAGCCACCGATATCCTCCTAG TAGGATCCGACTCCCGCACCGATGCAAAA GGTAACCCTCTTACCCCGGCTGAGCGGAA TATGCTGCACGCCGGTACCGATATTGAGT CCACTAACACGGACACCATTCTGCTCATAC GCATTCCTAACAATGGCCGCTCAGCCACC GCCATT TCTATCCCTCGCGATACCTGGGTCACCACT CCCAAATTGGGGTCGGGGAAGATTAACGG CATCTATGGGCAAACACTTCTCGCCGCTAA ACAAGAGGCGCTTAACCGCGGCGTCCCCA ATCAAAAAGCTGAGGAAGTAGCCACTGAC GCGGGACGCGAGGCACTCCTCAAAGCAGT	MRSIIAIITILLLVVSTYAWTKLDTLKGNTAN AFLNIGGGEDGATDILLVGSDSRTDAKGNP LTPAERNMLHAGTDIESTNTDTILLIRIPNN GRSATAISIPRDTWVTTPKLGSGKINGIYG QTLLAAKQEALNRGVPNQKAEEVATDAGR EALLKAVADLTGVTVDHYAEVGLLGFVLLT NAVGGVEVCLNQAVDDPYSGSHFPAGVQ TLNGSKALAFVRQRHGLMRGDLDRITRQQ AYMASMVRKILSAGVLTNPQALDRLTHAIN RSVVVDDGWDVIQFAQQLQHLSGGEVKF STIPVVNPDAHMQLDNGDMISAIEIDPPAV KQWVQRTIGSTEKKTKKRELPKPGRVNIE RSAVSLTVLNAGSHSGVAGAVAQYLGSLS YQIDNTGNANENIPTSIVIGNPQTMAAAEA VGKDLGGLPVQSSVDVPVGTVRVVLADDY DGPAVKNGRTVFASELKLPSKKATPTPGT NPFINAGGHGPQCVN

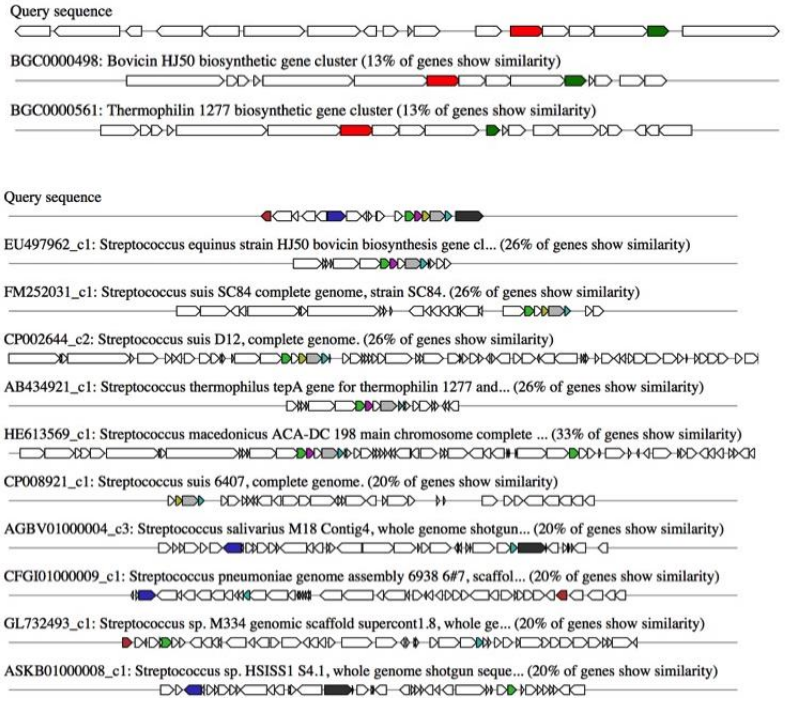
					TGCCGATCTTACCGGCGTAACCGTCGACC ATTACGCTGAAGTTGGGCTCCTCGGTTTC GTCCTCCTGACTAATGCAGTTGGCGGCGT GGAAGTCTGCCTAAACCAAGCAGTTGACG ATCCCTACTCCGGTTCTCATTTCCCCGCGG GGGTTCAGACTCTCAACGGGTCAAAAGCT CTTGCTTTCGTCCGGCAACGCCACGGCTT AATGCGCGGCGATTTAGACCGTATTACCC GCCAGCAGGCCTACATGGCGTCCATGGTG CGGAAGATTCTCTCTGCTGGAGTCCTCAC CAATCCACAAGCTCTTGACCGTCTTACCCA CGCCATTAACCGCTCAGTAGTGGTAGATG ACGGATGGGACGTCATACAGTTCGCGCAG CAACTACAGCACCTTAGCGGTGGCGAAGT GAAATTCTCCACCATTCCCGTTGTTAATCC AGATGCCCACATGCAACTGGACAATGGGG ACATGATAAGCGCCATTGAAATAGATCCCC CCGCGGTAAAACAGTGGGTACAACGCACT ATCGGATCCACTGAGAAAAAGACGAAAAA GCGCGAACTACCGAAACCC GGACGCGTCAATATTGAACGCTCCGCTGT ATCGCTCACTGTTCTCAACGCCGGATCACA CTCTGGAGTAGCAGGTGCCGTAGCCCAAT ACTTGGGAAGCCTGAGCTACCAAATTGATA ACACCGGAAACGCTAACGAGAACATTCCC ACCAGCATCGTTATTGGAAATCCACAAACT ATGGCGGCTGCCGAAGCAGTAGGTAAAGA CCTTGGAGGATTACCTGTCCAAAGCTCTGT AGACGTACCAGTGG GAACTGTTCGCGTTGTTCTTGCAGATGACT ACGATGGTCCGGCAGTAAAG AACGGCCGAACCGTATTCGCCAGCGAGCT CAAACTACCGTCGAAGAAAGC CACACCTACCCCCGGCACCAACCCATTCA TCAATGCCGGCGGGCACGGCC CACAATGCGTTAACTAG	

Appendix 74: Summary of genes in the BGC possibly coding for bacteriocin DpA.

Type of gene	Locus tag	Location	Function predicted by antiSMASH	Amino acid sequence length	Amino acid sequence
Additional biosynthetic gene	BWX42_02000	379197 - 380162	Hypothetical protein	321	MSDYRVLLYYQYVDIEDPERFRKEHKALCDQLELKGRILVSDEGLNGTLSGTVENTQKYMDAMHADERFAEMPFKIDEADEHAFKKMHVRTRPEIVSLNLGEEDVDPNQTTGQHLEPTEFRDALLDEDTIVLDARNDYEYDLGHFRGAIRPDIRNFRELPDWIRENKEQFMEKKMVVYCTGGIRCEKLSGWLLKEGFEDVAQLKGGIHNYGNDEETQGELWDGKMYVFDERISVDINRKEKTIIGRDWFDGEPCERYVNCANPYCNKQILMSEENEHKYLRGCTHECRVHPENRYVKEHNLSTEEVQERLEAIGESLPTFA
Other gene	BWX42_02005	380298 - 382103	oligoendopeptidase	601	MTYEQVWDLESVFSGGSNSPELQEKIEALQTDIDRFGELVKKWTPEEGAVHADKLATIFNLSDDISNGLSQGFTFSMAHLSAKQDDPHAPIHINKLSLLATQFSNYSTELTKKIETISNDDWQALLADEPFQSIAFRLNEERDQAKELLSLDLESAINQLSVDGFNGWSDMYDDLVASVTVKIERDGELKEYSAAQADNLLTSAETPKERAKVLEAWETAWQEKAPLFQTVLNHLSGFRLNNYELHGKDSYMEKPLKINRMKEETLDVMWKTIDKNKDRIVDFLNKKAELMGLDQLGWADVEASLPVEGADLSEYTYDEAADFIVKNFRQVSPQMANLAQRAFDEAWIEAEDRPNKRPGGYCADLPESGQSRIYMTFSGSADNVSTLAHELGHAFHSHVLRDEPRLNQEYAMNVAETASTFAELVVTDATIQAANSDAEKLTLLDQKNSRAAVMFMNIHARYIFERNFYDQRQKGIVTTDQLNDLMKSAQQKAFQDALSSYHPTFWASKGHFYGTDVPFYNFPYTFGFLFSLGIYAQAESDPDSFEERYINLLKDTASMSTEDLAAKHLDVDLTQPEFWQSAIDLVLKDLDEFDQLAAKFA
Other gene	BWX42_02010	382257 - 382724	universal stress protein UspA	155	MLQEYQRILVAVDGSESANRALRKAVAVAKRNDATLFITHVIDTRAFQPYEAFDASVSKSAKSEANSTLNACKLYAENHGLNDVHLLLEHGSPKKLIARDLPNDYDIDLIMLGATGLNTVERFFIGSVSENVVRSALCDVLVVRTDAENNYEIGH
Other gene	BWX42_02020	383126 - 384391	recombinase RarA	421	MKQPLAYRMRPESIDEVVGQQHLVGEGKIIWRMVEAKRLSSMILYGPPGIGKTSIASAIAGSTKYAFRMLNAATESKSDLNVVVEEAKMSGTVILLLDEIHRLNKPKQDFLLPHLESGRIIMIGATTENPYLSIHPAIRSRSQIFELNPLQEADIISALNRALDDDTNGLGHLELTVDEAALRHFARATNGDLRSSLNALELAAESTAPTNGSIHITLSIAEECIGRQALTHDKDGDAHYDVISAFQKSIRGSDVDASLHYLARLMVAGELEVACRRLMVTAYEDIGLANPAGAARAVTAVQAALRLGLPEARIPLANAVVELALSPKSNTAYKALDQAISDIREGKAGEVPKHLKDTHYSGADSLGRGQGYKYPHNYDNAWVKQDYLPDRLAKAQYYMPKTNSKFESAIARQFEKFNTDQ
Other gene	BWX42_02025	384436 - 385419	biotin--[acetyl-CoA-carboxylase] ligase	327	MHSHLIFQLNKQFPNPVPTETLSQQFNCSAAEITEAVEELRALGLVIHSYNSGFQLASPLYSIPGIKTYLNTQTFGTQGLYLFETIDSTHTYVSNNLSTIDHGSVLLAYSQTAGRGRRGRKWSSPIGKSLSLSIVLKEFSAQMKPTLLTQLTAAALQQALSETGLPVAIKWPNDLLINRRKVAGVLTEAMFERQVLQAITIGIGINLNQSQTDFPTEIHT

					KATSLALETESPFFADKLIAQFLKIFEHFYQDYQQSLDPEPFLSICRQESAVIGES VTVVQGDKRRDVEVLGIEATGELLVRDLSSQEQMTLISSEVSLRNPDGSYI
Other gene	BWX42_02030	385538 - 387223	RNase J family beta-CASP ribonuclease	561	MVSNVKIIPLGGVRENSKSLYVVEVDAQIFVLDCGLMFPEGELYGIDAVIPDFTY LKQNKERIAGVFLSHGHEDAIGALPYFIDEFNVPVFGSELTIELARLQVESVIGKG NFSDYHVVDSSLAIDFGEVKVHFFQVTHTIPDAMGISIETPEGHVVYTGNFKFD QSADGPYRTNLNHISDIGKERVLALLSSSLGAESLEENAPDKRVQSVVTEIFQD APTRIIVAAVQSNLLRIQQIINAAHATNRKVFISSSKSQDIINVAIDLGKLTLPKKDIL QSISDLEKFHDDETVILETGSSGEPIKALREMATNQHPNIGIRNEDTVFIATNPSV AMEVEVAECENLIYRSGGEVVKLSNQIKASGHATPNDLQLMMNLLHPEYFIPVQ GEYRVLAAHARLAHETGIPHEHIFTLSNGDVVEYKKGKMHSTGTVPADNTLVD GIGIGDVGNIVLRDRRLLSNDGIFVAVVTIDRRKQKIVAKPHIVTRGFVFVKANYD LINESTTIVEKVIQQNLKHNDFDWGSIKSEIRNELSDYLFKETKRRPIILPVIMEVN QRRWYDK
Other gene	BWX42_02035	387316 - 388770	cardiolipin synthase	484	MSYISTIFWWFIIINSTLAIFTIFRDKTRDIASIWAWLLALIMLPGIGFIAYLFFGRGM SDKDIYDIKQQDQIGLRELKDTLTERYEHQHEAGDEVFSNNKKEMVSLFQGLD NAALTRDNEVDIYSDGKEKFDQLLEDIKQAKHHIHLVYYIFRGDKIGRAIVDALEE KANQGVEVRVLYDPVGTRWMTRSFFKKLKSFGGQACPSFGERMHILNMRLNY RNHRKIVVIDGKIGYTGGFNVGDDYLGEYPEMGYWRDTHVRIVGNGVLPLQTR FLTDWNASADEAQQVPYDNQYFPIQETTGDVDLQIVSSGPDSTEQQIKKGFIK MISLARKSVYIQTPYLVPDDAVLETIDIASKSGVDVHVMIPNKPDHPFIYQATLSY AEQLVEYGAHVHIYDGGFLHAKTIIVDDEMLSIGTANFDIRSFKLNFEVNTFMYS EKLAQAYHAQYMSDLQHAYELTPDIIANYSLWERFKQQFSRLFSPIL
Other gene	BWX42_02040	388860 - 389213	Hypothetical protein	117	MGRHKIFDMAVSAVYPHYVNKAEKKGKTAEQVDELMCWLTGYSQDELHEQLE ANVSFEKFFDQAPKLHPNRTLIKGVICGKRVEDIEDPLMQNIRYLDKLIDELARG KKFENIKRTQ
Other gene	BWX42_02045	389415 - 389825	Hypothetical protein	136	MNQESHYQNLQLQLHDTMTQLDKMISQLKSAKQWSWVDLLGGDFFSSMVKR KKMKNISSDLQKLQPQLNALKHNLQDANIQVHSSFDHSTSRQFFDIFFDNAFTD LKVQDEIKQLLAELTTLRQVLTTYQQRIDKE
core biosynthetic gene	BWX42_02050	390109 - 390207	Hypothetical protein	32	MQEKRQQLVQVVSDTELEVLIGGKGPGLVKTF
Other gene	BWX42_02055	390255 - 390992	Hypothetical protein	245	MIPLIKNEFRTRAKKYGLFLFLGVVGLIQVLLITATLKIFKFDQLPNNPFIATFFRFY NVLFFAYSTMLIPVSAAVIGYYIISVEYISNTWEFLLLGIKDKKKVLMSKYIVSLIIF WIQQLVIYGVFSIIQVIYFKQQLDGNFMVLGFFTVLFFQVVLFTAQIVLHYFINNG VVAIVCTVVFIMLFLLMPEGTSNIFVEKILMLTPTYIGMFDTFNVVNFIGAWWSTY LLLRGCYQWLFISSNCK
Other gene	BWX42_02060	391979 - 392689	Hypothetical protein	236	MKTLLKKEFNTHGQNKLLYIVLLVVGMMQVFLARSLFGAEGGGALAKTIFQDNS IIYSLIATIFTPIIAGVLGYYFISSEFEQRTWEMYFLGIKNKSKILWTKYVVCLFYWI AYNICVWGIFLGLSAFYWQRQVDLFFVWLIPISTTLLGLFLFTAQLSMHFLIPNTQ ISISIVMVFIGIVGMVLESWLIYVLPLSSGFYLNHIDQFNNLEYMRLIIQNIIFSIILISI VVKRFYL
Transport related gene	BWX42_02065	392928 - 393842	ABC transporter	304	MNMNIKFNNVNKTYGRKQALQDINITIPEGKIYGFIGANGAGKTTAMKILTGLTP ASSGTVTFDGKELNQLENKAKAFGAFISTPTYYKNLTAYENLAIIQDVLEEPKNEI DRVLEIVGLADAKSKTIAEYSFGMKQRLGLAFSFLNDPDILILDEPTNGLDPKGIV

					EIRELLYSLSKEQGKTIFISSHNISELESIADMIGIIQNGKLIFEGELDELYASGESS YLLEIDDIDQAKTVLSEEGISFTNKEHKFQIKSSKQRIPEVVKALLDRDIAIFEITPN KNLERIFLDLTDGDDEHVDTH
Other gene	BWX42_02070	393826 - 394551	Hypothetical protein	241	MWTLIKHEFKTHGTQPILYLILGIMGLLQVFLTTIMIKTTDTVSIQGSYIQTVFQTN NAIFLSNSIIFIFSVGAVIGYYIISVEYQNNTWEMLLLGTGSKSKVLWAKYIVSTLY YLSYQVLFYSLFLLVQSTYFNLQIEISFSLLMLVSIMFLSLVLFTAQIACHYLIKNG TTAIACAVGFLIMLVILPSTDLFRYVIRLLTPGYLAGLDEFSVTGFVSVVALNIIVAS SMMSLVVKQFKL
Other gene	BWX42_02075	394564 - 395262	Hypothetical protein	232	MFLKLELLKFKRRISVGALCLFMLLVVGLQTFMALSVQELATDVFYIASLEWFSG IIFPALAPVFITLSIYAATEHKGMQNFVLKGVKPQKLSQSTWKFFLIALPLLYLSYV MIPFVFMFFRGGLTSEGIWSVVIYSGLSVISIVSLVNISFIIYRITMNDFLPVIISVAG VMLAFVPFGQDMWLVNPYLYMYYSAGHSGFNWYHYAIIIGLCAVSFAVYPYVS YYKKTKRA
Regulatory gene	BWX42_02080	395267 - 396751	Hypothetical protein	494	MEKWTPKIDVIYHIVFSVLAVYLFLGYKGVAYLPLMMSLWLLLLLFDYMMMTIIG VRTQIARLIVWTFRLAQVQSISLMTTIQLPLVIRYVLLMMLLNLLCQCIASLLTQYV DNRLLVRRLGIGLVVGIGLWTDYANILFIAYIVYVLFSTILAIKQRINHRVYTKNKH TGYWVLVAAFSLQLIPLLGLVILAVLSSQLFTQVWLFFMAVASTLLMLIAGFNHK AVVGQASKWLSLIFCTLLYIYIFVLILGLNLSLVLFILYIIVSIMYILTLTDVFGKHQV NAIRLIKREEALKQDFSNYLHDDVLQDINVLIKMTTLEPSEKTQAFLREQLSALN DQLRQQMNQYSPQLSKHLTLQENYRLLVRSLEQTYPNQAVSTSFSMNRNLTL PEPYDVLVYRWLRELIHNVYKHAGANQLDIHVAAYDELIIVLVEDDGCFKKGSHL KIGHGLYVIQEQVEAIGGQLYIEANRPRGLRMRVEFSVMGGEIIEDFVN
Regulatory gene	BWX42_02085	396732 - 397331	DNA-binding response regulator	199	MKILLIDDHKLFSQSLALVLDQTTSDVQVDMINSEGELPDDLSELTVYDVLVLDIN LDKGFSEDGFELAERVRAVAVDLPILMLTGFDLPVYEYQAHKLELSGFVNKNID TEDLLSLLKHVKDGGRHFTTENWFIDELTSRERELLEAVATGKKRKVIAEEMYIS ERTLTNHMQSIMDKLEVNSTIEAIQKARELGYLK
Other gene	BWX42_02090	397710 - 400367	DNA polymerase I	196	MSKKKLLLIDGYSIAFRAFYAMHSQLHRMKNKNGLHTNALYGFHNMLEAVMEK EQPTHALVAFDAGKTTFRHELFDDYKGGRESMPAELSEQIPYLKELIAAYGLQT YQLPDYEADDIIGTLSTRASDEFEVVIITGDKDLTQLASDQTRVDITKKGITQLKS YTPESMMEEMGITPGQITDLKGLSGILRIIFQV

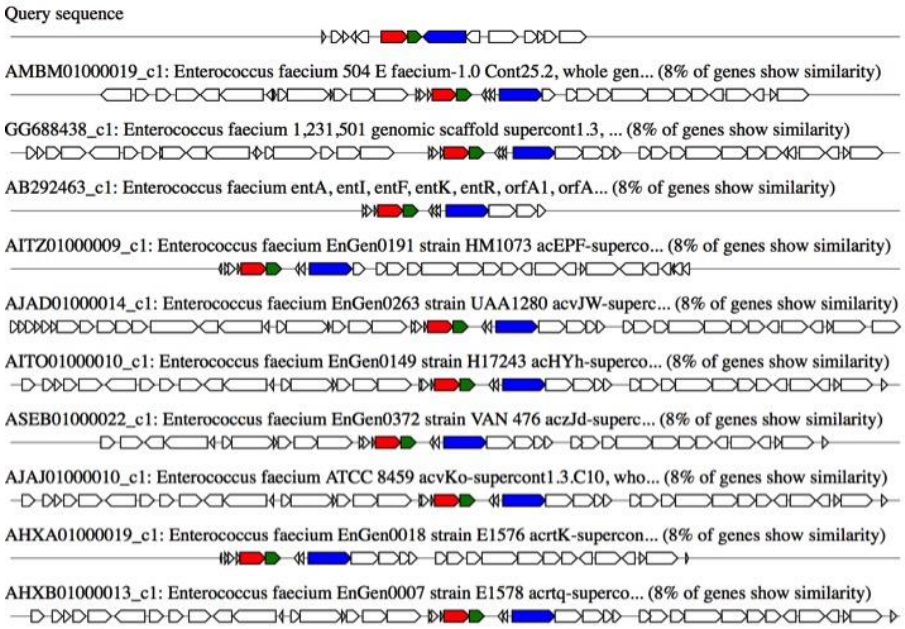


Appendix 75: Bacteriocin DpA biosynthetic gene cluster and its predicted similar gene clusters according to antiSMASH.

Appendix 76: Summary of genes in the BGC possibly coding for bacteriocin DpB

Type of gene	Locus tag	Location	Function predicted by antiSMASH	Amino acid sequence length	Amino acid sequence
Other gene	BWX42_08235	1738676 - 1738894	Hypothetical protein	72	MSNQNKMKLSNFELIPSKNLEKIVGGGVVTAVAVGTAVVFGAGAIDGSLEAHQKKINIMFSKIINVYVLSVR
Other gene	BWX42_08240	1739194 - 1739691	Hypothetical protein	165	MVDCDLKQCFDTLNHDKMMYHVEQFVQDKAILKFIRKSLRCGTIELSGEFTDSETDAPQGGVISPLLCNIYLNELDKELERRGHKFVRYADDFIIFKSSKRVCERILKSITQFIEKDLKLQVNQEKSKTGSPTQLKFLSCLIHNNFGNCRFRPTDETKAKFKKKT
Other gene	BWX42_08245	1739753 - 1740067	Hypothetical protein	104	MTQGWINYFGLDFVKRFVRRTEEWLRRRLRQLILQRWKKCSTKIKMLQKYGLTEDEAKRIAFSRKAYWHLSQTYEVNKAITTKRLHKWGLKSLTTIAESDYARY
Other gene	BWX42_08250	1740193 - 1740435	Hypothetical protein	80	MSDKEKVLNGYMFCAILVLNILWFLLNYSLSAFDFWGTVFSLNSYISIVVIIGYASMLISAKKKGEIDSYWIKDNQEKDK
Other gene	BWX42_08255	1740454 - 1741035	Hypothetical protein	193	MMKKKLGATVATVLLGVSAIPFNVSAHEEVTTEEVQIQHGENNQEHHEEQIGQMTLVNEHTGVLNFNEGITHAFNGDGTVTLTDSQGNSEVLPQETVDKNGENVRLVYEEAQNGQLAVTAVLPQENGRQKRSAGKCLAGIGAGAFTTGTSLGIKGAGAGTVAVPGVGTVGLGVVGAVGGAIGGGLGGAAASCF
Other gene	BWX42_08260	1741714 - 1743033	Hypothetical protein	439	MILINPVVAVILEIGLHYIFMTIAFIYLFNIEKKSIISLFLLGTLFWLLSLFLEGYLRIGNLSMVLYGIAIYRYFLNDRANRKRELIHILLFVLIKGSIEFLTALPINLLALGHLHSHVVVFLLASIVDVGLSVTVYKQVIKYFKNITLSPYETDILRHVLLGYNLALVFYFGYMKYQALYFDLLVFTSVILVGNGIVVGLVLFYVSRYSETRTREEAISKRLDYMVKYTNIIEENQLELKRFKHDYQNLLLSLSILAESGDLDKLKQKTTSLRQYSAQQLLKEYTQYSDLARVHHTLLKSILIAKLTELHHMNIVFRFVCPEPVESLNIADFDLIRMVGILIDNAREHVKDMEGGMIAVSVHKTTRGTHIKIENSYANQVTSIATLMTPGYSSKDNHSGLGLSNVEKIKNKYDHVFLNYHIDDKFEVTLIIMNEEGEY
Regulatory gene	BWX42_08265	1743034 - 1743771	Hypothetical protein	245	MHSIIICEDNQEQLAYLSMLVQNYIQFHENQFRLVYEDTNPQTTLNYIQQEAIKKGLYLIDINLGADMTGVDLAEAIRDSDIQSKIIFITSEEDQATNILNRHIEPLAYISKNEGLETMQSNLHRALDDAYSRFSEITVRTKDVFSFNFDTLTYQFDLDEVISIEVSGDHRLTLKTITGQYDFFNTLAQIEQDYPTLLRIGRSDMINPVNIKHIDYKRRNVTMVNDEQFTIAASRIIQLRQLYKS
Core biosynthetic gene	BWX42_08270	1743828 - 1745981	peptide ABC transporter ATP-binding protein	717	MRDTFKFVKQQDIMDCGVACIQMILKHYDSDMPAHKLRHMTGTDIDGTSALGLKSTLEELQFECLAVQADNSVWTDPDMNYPAIAHVLLEDNRQHYIIIYGFKNNKLLIADPAEGKYKMSPEEFTTIWTNILLIPKPNKQYQPVVEKIGGLTSFLPALFKSKKIMFFTIIASFLATSLSIVSSYYFQGIIDRVIPQEDLSLLNIMSIGLLGVYLLRVLFDYIRSQLLIILGQSMSSHIMLDYFKHVLYLPMQFFNTRKNGDIISRFLDANKIIRALASSALALFLDITMVAIVGTFLFIQNRILFVITLLSLPIY

					FVTVLFFVKRYNKANEEEMSSSAILNSSIIESLDGMETIKSYNSEQNVYET VQNQFLNLMKKSYKTVTLDNMQRSIKQSVQLLTSAGVLWAGSYYILHNS MSIGQLITYNALLVFFTNPLENIINLQAELQTAEIASKRMNEVLAIESEYNLT DKQAAITFDGGIRINHLTFSYNLKESTLKNITCHIPFNQTVALVGMSGSGK STLAKLLLRLHEVAEGTIQYDNILINKIPHNYLRDHVTYLPQESFFFKGTIIE NLLFGLSHQPSELEIITACEQAKVRHVIDSLPLGFNTPLEEGAANLSGGQK QRLAIARALLRDTNIYIFDEATSSVDTVTEQKIIQSISELKNKLIIHITHHLPIA KQCDQILVMHDGQLVEEGTHEQLLANSGTYHTLWQSTFP
Core biosynthetic gene	BWX42_08275	1745997 - 1746710	Hypothetical protein	237	MTSFIKKIGSPVKIIFLGIIFSVIATLIISVGQIPRDIFPAFVGDNTAYISRYIYTI VVFILLIWLLKWWNKSVTPIYMNNTLFQKKNWVYGIGGYVLIFILVLLFILAK GYIFGSNSVEISKTLLDEFAFEQNVYLGSLSILFFIFLGPILEEFIYRLGILGQ LVEANTPTWLAVILSALIFSFSHHNGIISQHLISGLGYSLLMIKTKSIYPVIIAH ILFNTTTLIYRWYMLYYL
Other gene	BWX42_08280	1747157 - 1748628	glutamate--tRNA ligase	419	MSEAIRVRYAPSPTGNLHIGNARTALFNYLFARHNDGTFIIRIEDTDKKRH VDEGEESQLKYLQWLGIDWDEGPVQGGDFGPYRQSERNELYQKYVDEL LARDLAYKCYVSSEELEEMREKQRANGVMPHYDGRHAHLSEEEQAQFE AEGRQPVIRVRVPEQTTYTFDDMVKGDISFESKDIGGDWVIQKVDGTPTY NFACAVDDHHMKISHVLRGDDHISNTPKQLMIYDAFDWEPPRFGHMTLII NAETGKKLSKRDESILQFIEQYNNLGYLPEALFNFITLLGWSPKGEEEIFTK QEFIDMFDYKRLNTSPAAFDQKKLEWITNQYIKELPDEDVVALIQPHLVEA GVFPADPSEADQEWVTKLANLYKEQMSYGQEIVSLADLFFGEGIAFDEG RERYCLERRSQLSWKPSRST
Transport-related gene	BWX42_08285	1749002 - 1749637	Hypothetical protein	211	MDKKAKVWIGVLLVISGLETVFNAATYAMVFDIIEQQKLNLVVVYTVVVILG YILFSGIRHIKDRTINHYVKEEKASIQKKILDNEEQKFKLFEHTGSSDKMSF FQNDLKLFEDNYLRKKFEIISQVIVLLGVLAFSLYSNFLLTVIFSLFAAVPHF VSGFMNKKIQQSTEHWTQATAKANHSLIDLFKNFWTLMVYHATNKEIKSV SEDI
Other gene	BWX42_08290	1749662 - 1750015	Hypothetical protein	117	MQNTISLAQMFTMMVGYTMMMIPIGIGMYLTISGDLPIATFVAVQYSSSMII NSLLTTVRLKNELQSAKPIKDKIDKELSFVPRSTGSKKLNGHIQSVVFKDV SFSFGRRRCSTTSI
Transport-related gene	BWX42_08295	1749994 - 1750575	Hypothetical protein	193	MLDNFNLTLNGVDKVLLQGESGSGKTTIFKLLTKQLLPDSGTIYINGIDTRE LDIEDILQCFGYISQQALVFDDTIEKNITLGHDFSTEEIMKACEMAQIGELIK KTGLDYVIGEDGQNLSGGQLKRLEIARAILFKRKGLLVDEALSSLDKETGR AIVKTLMDLDKLMIDIEHHIAAELTEGYTDVMTLSRS
Other gene	BWX42_08300	1750791 - 1752139	Hypothetical protein	27	MEYHVLEEKLPVKPAGRWCWGWRRCGN



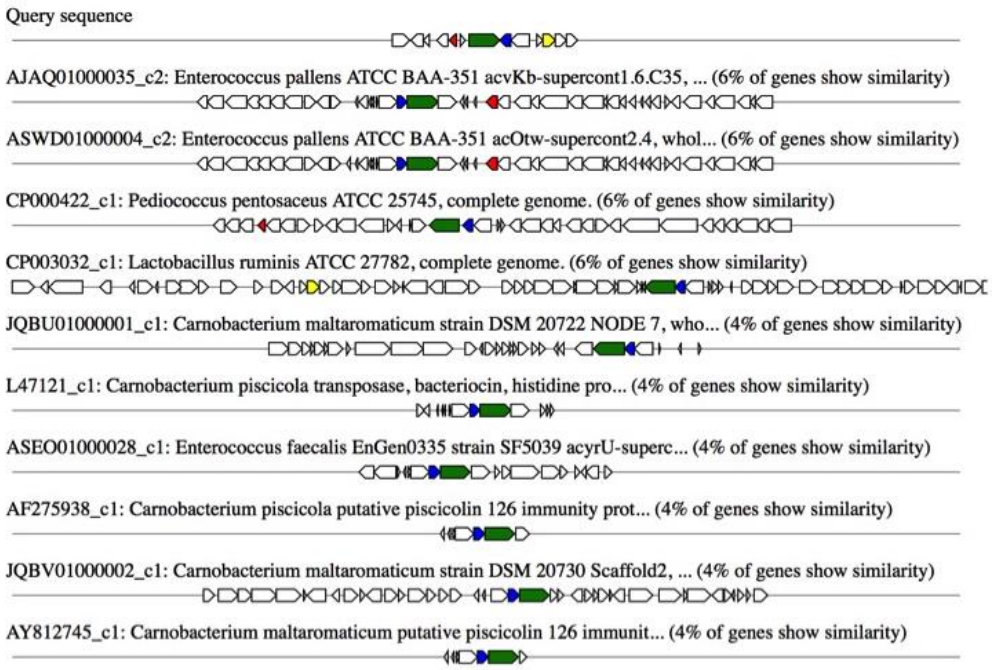
Appendix 77: Bacteriocin DpB biosynthetic gene cluster and its predicted similar gene clusters according to antiSMASH.

Appendix 78: Summary of genes in BGC possibly coding for bacteriocin DpC

Type of gene	Locus tag	Location	Function predicted by antiSMASH	Amino acid sequence length	Amino acid sequence
Core biosynthetic gene	BWX42_08355	1761437 - 1762714	O-acetylhomoserine aminocarboxypropyltransferase	425	MSNTHRFETKALHAAQSVGQEKSRAVPIHQTTSYLFDDTQDGAEKFALSKPGNIYTRLNNPTQTVFESRIAALEGGTAGLAVASGMAAISYTIQTLAQVGDHIASSSSIYGGTYTYLSQQVKNIGLKTSFFDINDVQSIKDSLEPETKLIFIESIGNPTGSIPDIEVIAEIAHEHGIPLVVDNTFPSPYLLNPIKFGADIVIHSATKFIGGHGTSIGGVIVESGQFDWTQNNKFPLLSEPDDSYHGLIFAEAVPDAAFTTKIRAGLLRDTGAAISPFNAFLLLQGVESLHVRVERHVENTEAVAQFLANHEKVEWVDYAGLADSPYYELKQKYLPKGAGSVFTFGVKGGYDKAVEFIEALELFSLLANVGDAKSLVVHPASMTHSQLTEEELKSGGIYPETIRVSIGIENVDDIIDDLTQALAKI
Other gene	BWX42_08360	1762783 - 1763664	homoserine O-succinyltransferase	293	MPIVCDQNRYGSDVSMISNAEIQMNQTHPIKIAILNLMPNKITTEHQLLRLLGKSPLPIEVTLLYTKTHAPTHMIDDYLQKQYRVFDDVKDDTFDAFIITGAPVEHLSFEQVDYWDELTDIFTYVHSNVRSTLFLCWASQAALHFYYDIPRFEVSEKIFGIERFKTRVKNSLTDGFESYFNVPQSRYFYHQLSDIQSHSDLILYAGSEQSGAQIVASKDYRSVFIAGHFEYDADTLHLEYKRDQERGLHTSLPTCYFVNDDPTQGYCSTWQTHAERFFTNWIKHIVDPTVSDA
Other gene	BWX42_08365	1763699 - 1764169	S-ribosylhomocysteine lyase	156	MEKIASFTVDHLNLKPGLYLSRTDQIGERKLYSYDLRFTAPNYEPVMNTAEIHTIEHIVATYLRNDEQFKQEIIYFGPMGCRTGFYLITTDAIAPADLLDLLIRAFQFTANFQGEIPGADARSCGNYLDQNLPMAKYYAKHYITVLENVTDLTFSY
Core biosynthetic gene	BWX42_08370	1764684 - 1765439	3-oxoacyl-ACP reductase	251	MEKKVVIVTGSSRGLGKAIAQKFADQGAYIVINYVKSEDKAQQLVQDIGEDRAIAIQADVRDSEQINHLVTKTIEHFGQLDVIVSNALINFKFDPHAQQSLDTIDWKNYNEQFEGSVKAALNLVQASLPYLKDSPQARIITIGTNLFQNPVVPYHEYTSGKAALLGFTRNIAKDLGPDGITANMISGGLLQATDASSVTTPEVFDLIANSSALQRVTTPEDVAEVVAFIGSEASRGLTGQNITVDSGFTMN
Additional biosynthetic gene	BWX42_08375	1765528 - 1766097	galactoside O-acetyltransferase	189	MNKHEYNKMIRGELYDPRDPYLRDLMTKQHALMDAFNESGRTNHATQNKLLEKLLGQYHGESSHIRRPFFIDYGLNITIGKHFFANYNCTMLDVAPITIGKNVMFGPNVSLNTPEHPLSAAERNMGLEFARPITIEDNIWIGASVTVVGGVTIGQGAVIAAGAVVTKDVPANTVVGGVPARVLKTIENC
Other gene	BWX42_08380	1766318 - 1766671	hypothetical protein	117	MHTRFKEGLKSFIISYVAMILYHLPSAYNPLFDIVHMMVIVGTLLYGGYYFYRNGASSNGMAGLLIQVSLFTIGFNIALPLTREIFSARIMFMTGCSLSLIFLKHIENNLYRIDRKS
Core biosynthetic gene	BWX42_08385	1766971 - 1769124	peptide ABC transporter ATP-binding protein	717	MNTNFRVIYQQDNMDCGTACVQMILHHYNSEFSIHKLKDLTDKGKQGTSALGIKRCLESLNFDCLVIDADNAVWTDSAMTYPAIAHLNVHNGKHYVVVYGYQDNKLFIADPDDGKYNMSIETFSAVWSRTLFVPSPGRHYKPIKEKVQGLMSFVPKIIQYKKFVVFSIIASLLIMGLSIGSSYYFQGIIDSIIPANNISMLNIISFGLMISYVLRVIFGYIRNYLLFILGQKLSRDIILNYFKHVLK

					LPISFFASRQSGEIISRFLDASTIIDALANTSLTLLLDSIMVVVVGIVLLLQ HNVLFGVTLVTLPLYIGTIILFIIKLDKSNEEEMESAAVLNSSIIESLNGIE TIKSYNSEKVIYQRVVGEFEDLMIKTLTTMTLDNMQGAIKTGTQLVGN ALVLWLGAYYVMSGEMSIGALITYNALLVFFTEPLESVINLQVQIQKAQ VASRRMNDILSIEVEGNISNDTVEIEEMGEADLKFDDVYFSYGLGELIL RGITCEIKVGDKIALVGESGSGKSTLAKMLVRFYAPSMGNIYFGKHDIN TIDVHILRHNIVYVPQDSFFLAGTIRDNLLFGLDYEPTQNQIEQACEVA CIKSFIESLPLQYHTIVEEAGSNLSKGQKQRLAIVRALLTNSQVIIFDEAT SGIDVFSEQRIMDNIIKMTERTIIFITHSLPAAKLCEKIMVMDQGDLVEE GIHEVLRYAGGTYQKLWEVIS
Regulatory gene	BWX42_ 08390	1769189 - 1769944	DNA-binding response regulator	251	MYPLIICEDSIGQLNAIKQIINLYTGFYESRFNVSLITSSPHEVLDYLETY QPKKGVFLLDVHLNSTIDGLDIAEVIRKNDVNAKIVFITTDEHAAPLTFK RNLEAMSFIEKPSNVEELRKNIFETLNIAYNRLTTSLDEKKQLFTFSRY GETVNIDFEKVLYLQTSIEESHKLTLTTTLGEYEFSGSLIKCEENFPELF RISKSALINPGNVVSINYKNRDIVLKNNHIIQYAFRRAKSIKEKFSNPNK N
Other gene	BWX42_ 08395	1769962 - 1771272	hypothetical protein	436	MFVNPIILFIYEILLDVSLFFTTFNFLFTQNKPVLKRIFVSISVPVVWIFSM IIHDLFYATDIVALLYGLVLYGIFFSLKRINFEVLNSILLLVYTKGIVEFLVA LPINTLKYYNLHNHILVMITTKILTFIVSIAIYKKIYPYFTSNIDQVKAKKIVT YSLLILNTLFSSYFIFIQWFDLYRKLLYFTLTITIIFVIITVGFLYALHKYLQL KFQHELSNQHLEALTKYNHIVEEETYKTKEFQHDYKNLMLSTTSFIEN NDLAGLKEMMHSLIDYSKDYFSHSTFQYKDIHNIKHPALKSIILAKLVTF TELGITCQFECRKPLDNIYIDTFDFIRVVGILLDNAQEYVTTIEEKHIDILI TSHTKVMELVISNPYIPDHYTLRNFKQEGFSLKGNDRGMGLSTITKIEK KHNNLTVLFEKNTIFSAKIITTKL
Other gene	BWX42_ 08400	1771842 - 1772261	hypothetical protein	139	MYRKILLSLISTLIIFGTGIYLNVSMSLLMFAVISYVLPLMGNIVIDYYVQT ENVLIGSGIISTITTTGYAIFAILMENNINFDGFIRQHTYRSGNMTMGLE PGLADMSQLIFVFALNLSVLYFIQYLSRGDFSHVRRK
Transport-related gene	BWX42_ 08405	1772245 - 1773132	multidrug ABC transporter ATP-binding protein	295	MLDVSNVSKQYSGNDFYSLKDVSFTIEKGEIVGLIGKNGAGKTTLLKLL AKSHFPTTGVIRYRGLDIFSEDNVLRPFGLMIETVYYSHLTVKQNIEFY LDIHDQSEYKKNIQSVLETVELWHKKDFKPDGFSFGMKQRLSLAMSL VTNPEFMILDEPFVGLDPDGVNDLMRILKQWAADRGTSIIVSSHQLNE LEAICDRYLYIEGGALKKQFGKAAVEVTVIELVDALQDKDALFSAYEAI QAISEDGKQIEVDSNSEAFHQILGTITASHNIKKIFTKENALNQYFKGRR DHE
Other gene	BWX42_ 08410	1773125 - 1773889	hypothetical protein	254	MSKLNQVVFHNTLCKKEVWVYLAFTLLPLLVFVTELTDTKFLRLTGDF SEMNFLDFYGTTLGIIDGMILPTIVIAYIASTMFYGEINKGILFLYKDINRK KVMNSKVFSLMKTYGLYLLLVMLSSLMVYYFFVIRMKGGTFSLLGSVA PVTYYSVLGVLGLFGIDIVAILVAANLSVRFSTGVTVLGTIFFLFLMQAIS RVDGLNYLSPMSYRELFNVGQMTFSMSFILLVFLCVVYVVSAYTLLHR QFSKVEY
Other gene	BWX42_ 08415	1773916 - 1774677	hypothetical protein	253	MTTFNQVLLKNNLMKWEVWLYLAFALYPLLVFIGQLFQVNFMQLDNL ANVTLLEAYSKLLVGANQMLAPVIIISYVIATIFYVEINNGLLFLFKDINRK KVFTGKVTALLGVYGVYMLVLAMGTIITYYFYIVPTIGASGALIPTGASA

					WSYLIIELIAAIAIDIIIILVAINLSLYFSPGITIMGAVFVSLLAELAAKLDTLR YIFPNSYKTLFHSGQITFVTAISIVIGLFALYAIILYMNAKSRFVKIEY



Appendix 79: Bacteriocin DpC biosynthetic gene cluster and its predicted similar gene clusters according to antiSMASH

Appendix 80: Summary of genes in BGC possibly coding for polyketide DpA

Type of gene	Locus tag	Location	Function predicted by antiSMASH	Amino acid sequence length	Amino acid sequence
Regulatory gene	BWX42_08600	1808769 - 1811533	transcription antiterminator BglG	617	MERKEEIIHLLSNQPEKPFTTREVAEALGIQRSNASSYLNELTGEGRLTKLGGRPVKYTLNEP SSSSPIPADTAHPRSKTTSDNQRDVFSQVIGHNGSVKNAVMQGKASILYPPNGLNTLITGP TGSGKSFFANTMFQFAKNKGLIAADKELQVFNCADYAKNPELLMSHLFGYTAGAFTGATT EKDGLIQQADGGMLFLDEVHRLPPEGQEMIFYFMDTGQYSRLGEAERKLHASVRIVCATT EDPKSTLLETFIRRIPINIQLPTFKQRPVSEQIDLLKLMVGLEAKRINRKVSMSKDVVRTLISH VDYGNIGQLKSNIQLLCARGFMNQLHNNVIELTTQELPENLTRSFIAMTSDRSASSELIRRL EPTLEIHPEDVDIILDRDSYELPYNIYDIIADKATLLEEEQIDQETINQYISTDINLHLKSFYKNQ GFALQTTSKLAEVVSLEVIDFARDVLVMTDLETAPGQSGSLLYALSLHVSSLIGKLRDGQLR PFNARVRDLALNTPDALRQAEVIRHRIADHFNLAVPESEAYYLTVLLTSFKTDITQQVGIVL AAHGPSTASSMAKVVAELLETPPIPAIDMPLSMSQQKPTPRSKKPSNTLMTEVVFCY
Other gene	BWX42_08605	1811690 - 1812970	RNA polymerase sigma-54 factor	426	MKITQQAMIKQTQAPLPFLIKSVGILQLNRVDLSHYISNYMSENPLMTFANDDALMFGSE SGAEGINLDEINASTQTLFDFITEQVELFYRQTPLRDMTMWWIKQLDTNGYVTKTLEEAV QITGESDYMVLDGLTLLQQLDPPGIGARDLRENLMLQTERLDSAPNLAYIILEENFDRLVN RQWSDLATIYRTDLADIRAIFKFIQTLSPIPANLYQAPQTEMIYPELEVTVTNQQLTISETRF QTPLLTFDQAYYDELAQIDKPSIQDYIKAKKQEFDQLQQALQKRKETILRVGTAILSYQKDY FLTEDASLQPLQINDLVKQLQLHESTISRAIRGTYVQTNRGVIELKSLLSKRSVVKDTSQDAI FTALSDLIEGEDKTHPLSDQQLADTLKQTKHIALSRRTIAKYRSQLGIPSTRERKIR
Additional biosynthetic gene	BWX42_08610	1812972 - 1813976	Hypothetical protein	334	MNTLKWGIIGTGGIAEDFAAGFQDTAITCHGVSSRSLEKAEAFREAFGLNAAYGDYQELL QDEAIDIVYIATPHHTHAQIAADALQAGKHTVVEKPIVTQVDDLAQLKQLAEAQGVYLFE AMTIHYRPVYQTIRELIDQKELGALKMIQVNFGSFKPKDTGYFFQKDLAGGALLDIGVYAL NFVMEFLSTPPTEMKTLGYINEQFGVDESAGIVLKNDQNELATITMTFHAKLPKRGVIAFE GGYFEIDNYPRADHVTFTTPEGKTKHYQAGQSSQALQEEMIAITNLIASGTPQPHLTKTTH VIELMDAIRQEWNLTYPFDNHTLDTPNREKA
Other gene	BWX42_08615	1814123 - 1815076	mannose-6-phosphate isomerase, class I	317	MSTEPIFLTPVLQDKIWGGKKLQTEFGFDLPSDTVGEAWVISTHPHGESTVSSPEQYAGM GLNELYHEYPELFGAQQPDTFPLLTKILDAKEDLSVQVHPDDEYGQEHEGELGKTECWFVI SADEDATIIYGHNARTEEEFRRLVEAGEWDELLREVPVKAGDFFYVPHGTIHAIKGGITILET QQNSDTTYRVYDYNRTDKDGNTRELHLEDSIAVSNIPHIDPAIQQQEDRVGMSAITHYLT NEYFSVYRWQIRDQLVVDLTGNYTLVTVLDGQGMIEIDRESYPIEKAQSFIIPHGVSSVTLS GDLDMIASNPE

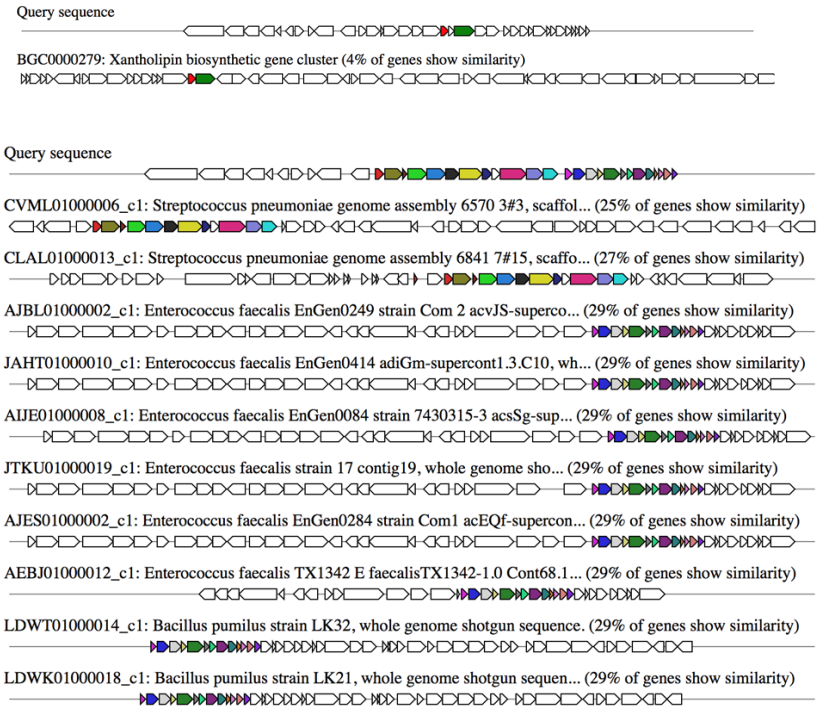
Other gene	BWX42_08620	1815172 - 1815501	Hypothetical protein	109	MIEDRDFLMRQIKQMIQNLGKILDRNTLRELLALSEDDMSNEELDSLYLMGQVDIIAAEKQLTPQHLSVDLGIDIDRLEDLFKGQAFLTEPEATNVKSFLETQDNSGPT
Other gene	BWX42_08625	1815809 - 1816369	heptaprenyl diphosphate synthase	186	MTKTQRMIYISLLSAQAVILGLVENSIPFPFAFAPGAKLGVANLIVIIAIYTLPLKDSFKVLMMRLLMTTLLGGTLSTFMYSFMGALLSYLGMLLIKTLGRSQVSIIGVSATGGMLHNLGQLLVASFIAQTWTVLMYLPVLSFFGILSGMAIGIAANYLMEHVHTVRRFRVEQEQADQAKHSSRLS
Other gene	BWX42_08630	1816515 - 1817116	16S rRNA methyltransferase	48	MSDHYYTKEPTSTSNPEQFKEVVIGQTLSLTSDAGSFRVDRWTLGREH
Other gene	BWX42_08640	1817393 - 1817782	Hypothetical protein	129	MRLWHETLIRDLPRQQLLGQHRECCALRGKGWDRPHATVQYVFDYSPYKLYQYHQLIMEEMKSRTYQPDERWEDPLYRGKACAPYRELEPVTPTKPIYPEHNTTYLAECLENLADKGIELSVRMKQSEK
Other gene	BWX42_08645	1817831 - 1819097	tyrosine--tRNA ligase	54	MNIIDELQWRGAINQQTDEDGLKELTEKQSIALYCGVDPTGIVCILGTSFPSLS
Other gene	BWX42_08650	1819641 - 1820615	L-lactate dehydrogenase	324	MTQSEKKHKKIILIGNGAVGSAYAYALVNQNIGQELGIIDLDPFKVEGDVMDLNSALAFTAPKDIYVADYSDCADADLVVFTAGAGQKPGETRLDLIHKNLKITKTIVDQVMASGFNGIFLVASNPVDILSYAVHKFSGMPAHKVVGSGTSLDSARFRIELAKRLNIDARNVHGYIIGEHGDTEFPVWSHANIAGLQITEWLETNPLEDEGELLEIFEAVRDQAYHIIERKGATYYGIGAALAKITQAIFNNENSVLPLSVLLEGEYGHDDIYIGTPAIINDRGVKRAIEIPLNDYEQGKMNHSINTLRKFIQDAREKAPELGL
Regulatory gene	BWX42_08655	1820933 - 1821394	MarR family transcriptional regulator	153	MDHQTMNEVNEYLVMIFNEVLSIEEEAISRSEFSDLSVKEMHTIEAIGLSGDLNSAQVAKRLDITPGTLTVSIQNLVKKGYVTRVKSETDRRVVKLALTKRGKLIYRLHHKFHMRMVEESLTGFDPEEAQVLIRGLRNLHEFLNDLKNQQRKE
Core biosynt	BWX42_08660	1821400 -	Hypothetical protein	314	MSKVVATGQYLPANRLTNEQLIAQAGIDSSDEWIVQRTGVKARRFAEEESVADLATEAARAIIANIGPDVKEDIRLIIVASMSSGNPTPSIANQVQANLGIAEAWGFDISGACSGFTMAVDIAERMSRTYESGYVLVIGADKMSQILDLSDRTSSIIFGDGAGGLLIACDGAGLPGYRSQLVAL

hetic gene		182234 4			EDTKDAITLDKTVPDRQHLTMLGRDVFNFVVRAVIPGLAKFIDELNESYDYVLSHQANDRL LDVMSQKLGIDRQQIPANIAEVANTSSASIPILLDDLVQAGTITLSGQQQIVMVGFGGGLA YGINYFKL
Core biosynt hetic gene	BWX42_ 08665	182238 4 - 182260 8	acyl carrier protein	74	MVFETVKDIIVEQLGIDESEVTKETDLENGLDADSLDIFQIISDIEDEYDITIDTDLNLQTVGE LVDYVEQLIG
Other gene	BWX42_ 08670	182263 1 - 182360 2	2-nitropropa ne dioxygenas e	323	MKTAITELLGIRYPIIQGAMAWVADADLASAVSNAGGLGIVGTGHDSREVVKEKIDKMKE LTDQPFAVNAMLLNPHIEEVIDYIIEESGVKIVTTGAGNPSQYMKRFQEAGIRVIPVVASVA LAKRMERIGADAIVIEGMEAGGHIGRSTTLTLLPQVVEAVDIPVIAAGGFGNGESLAAALM LGAEAIQVGTRFVVSKESNAHQNFKQKILKANDIATVVTGQITGHPVRVLRNQLTTDYLEL ERLQTSEDEPDFSEMEKLGKGALRRAVVDGDIKQGSMMAGQIAGLVKKEETVAEMIQDY IDGAQQAYRRCQACFTQDDA
Core biosynt hetic gene	BWX42_ 08675	182365 1 - 182459 8	Hypothetic al protein	315	MKVAVIFNGQGAQFEGMGLDFREHFPEARAVFNQASEATGLDMVQLVSEDFSKLRQTK YAQPAIGTVSLAIWASIRETLPQVDYMAGLSLGEYAALMASGIFTVSDGLRLLFERGQVMS DVCEEIAEDEPMQMLAVIGMSREVVEHLVEDLPQTYLANFNSPEQIILAGPKSNLKLFNQ AAKAAGYRKGLPLKVEGPFHTPLMAAACQPLEVLLDSYELQPGCAPVISNTTVEPHDLETL KSTLVRHLIEPVQWEQTIDWLIQAKVTHLIQIGPGQTLQKLLKAHDQAPLCLAISQVEDVS EIEKFLNENKGEKE
Core biosynt hetic gene	BWX42_ 08680	182459 5 - 182531 7	3-oxoacyl-[acyl-carrier-protein] reductase	240	MSQVCVVTGSSRGIGLAIAKQLADDGHQVVLNSRSPLKPEVLEQFADAKLEVGTIVGDVS DFADAERMITAVKEQYGRIDVLVNNAGITRDGLVMRMKEEDFDQVIATNLKGCFNMAR HTTPIMLKQRSGTIINVSSVSGIMGNAGQVNYAASKAGVIGLTKSLARELASRSITVNTIAP GFIETDMTAEMSERVTEAMLGEIPLKRFGQAEEVASAVKFLMENRYVTGQTIEVNGGLHI
Core biosynt hetic gene	BWX42_ 08685	182531 7 - 182657 0	beta-ketoacyl-[acyl-carrier-protein] synthase II	417	MGQADKLNRVVVTGLGTISPLGNNIDEFWKKVRANESGIAPITKFDASAVGVHVAGEVK DFDPTLTMDRKEYKRMDLFCQYGIAASVEAVEMSGYDIAANASRVGTLISSGIGGLIEIEN GIRKMIDKGPKRIPPLFVPLTIGNMAAGNISMKLGAKGISMDIVTACASSTNSIGEAFLKIQ AGFLDACLAGGCEGTINEIGIGGFNALTALSTNEDPTKASRPFDKDRDGFVMGEGAGVLF LESLDSAQERGAHILAEIVGYGATSDAYHMTAPVPDGSGAGEAIKMALASAHITPEQVSYI NAHGTSTPTNDSGETTAIKYALGDAAYNIPVSSSKGHFGHLLGAAGGIEAITCVKALEDGFI PATLGLETSDEACDLDYVPQTGREADLQYVLSNSLGFGGHNAILCFKRWEGK
Addition al biosynt	BWX42_ 08690	182653 3 -	acetyl-CoA carboxylase , biotin	177	MRFCASSVGRGSKVNYDQLLALVDKLDQSSLAYIRYEHDGSKVELSKEVPHSQPVSMPAE TSSVSSPVIRETTPTVPSVVSDDKLHNETSNDVSESVGVEVLSPMVGVAYLQPAPDKDPYV QVGDRVEAGDVVVIIEAMKIMTEIKAECSGIVVDILVANEELVEYDQPLIRIKEDN

hetic gene		1827066	carboxyl carrier protein		
Additional biosynthetic gene	BWX42_08695	1827067 - 1827501	3-hydroxyacyl-[acyl-carrier-protein] dehydratase FabZ	144	MSVLTAQEVMEIIPNRFPIFFIDAVDELVPGKRIVCRKNVTINEHVFQGHFPGEPVLPGVYIVEAMAQAGSIPLLKQEGFEGKTGYLGGLNKVKFRQKVVPGDVLRIEVDIIKQKKNAGIGRGRAFVGDKKVAEADMTFIIGAN
Additional biosynthetic gene	BWX42_08700	1827503 - 1828867	acetyl-CoA carboxylase biotin carboxylase subunit	454	MFKKVLVANRGEVAVRIIRVLREMGISSVAIYSSADKQALHTELADEAICIGPAKTLDSYGNPVAVISAALQMNCDAIHPGYGFLSEKSEFVDLCEEVGLTFIGPSSYVINQMGNKQHARETMRAAGVPCTPGSDGLVQTVEEAEQVAEVIGYPLMVKAADGGGGKGMRRVADASELAHKFSAAMMEAQAVYGNKDVYIEKIIAPAKHIEVQLLADTHGNVIHLGERDCSLQRNNQKVIEMAPATFLDASVREALCDAAVTAAKAIGYTNAGTIEFLVDEQQQFYFMEMNTRLQVEHPVTEMITGIDIVREQINIAMGQPLTITQEDVTFNGMAIECRLNAEDPKQSFRPASGYIERLILPSGGMGLRVESGVYPNYSLPPYYDSMIAKIIVHKPSRQETFKTLQRALVEVVVEGLVTNIELLEELAYSEEVLADQYHTKWLEDEFLPAWMMTD
Additional biosynthetic gene	BWX42_08705	1828890 - 1829762	acetyl-CoA carboxylase subunit beta	290	MGLFRKRKGIKLNKILEEKDQERLAHVPDDLVERCPSCRKMLLHKQIEADCCCPNCGYHMHFAAYDRIEWLVDAGSFSEWEADLQTNNPLDFPHYAEKMAQQREKTGLNEAVITGRALINHQAVAIGVMDSRYIMASMGTVVGEKITRLFERATAQRLPVVLYIASGGARMQEGILSLMQMAKISQAVQQHHQAGLFYLPILTHPTTGGVTASFAMQGDIILAEPAATIGFAGKRVIEQTIKAKLPLDFQTAERVLETGFIDRIVPRERQKSIIQTLIQIHAGDGVNESL
Additional biosynthetic gene	BWX42_08710	1829740 - 1830534	acetyl-CoA carboxylase carboxyl transferase subunit alpha	264	MVLMKAYDIVKTARDINRISTRDVIQALCQEEFIEQFGDRLSGNDGAIIGGVGLANNTPLTIIGIEKGRTVEENIQCNFGSASPSGYRKASRLLKQANKFNRPVLTLINTPGAYAAPESEEAGIGEAIARNLLIMSELTVPTLAIILGEGGSGGAIALALADEVWMMELSVYAILSPEGFASILWKDAKRAPEAAELMKLTSKDLSELAIVDRVIKEVDNNGHSIDKTTILSQLTRQINEKVTQLVEQPVHERLKKRQERFRKF
other gene	BWX42_08715	1830919 - 1831287	30S ribosomal protein S10	122	MLGNFHGARFILKVSEGGKNMAKQKIRIRLKAYEHRVLDQSATKIVETAKRTGAEVAGPVPLPTERKLFTIIRSPHKYKDSREQFEMLTHKRLVDILNPTPKTVDALTKLDLPSGVDIEIKL

other gene	BWX42_08720	1831329 - 1831961	50S ribosomal protein L3	210	MTKGILGKKVGMTQVFTESGELVPVTVIEAKPNVVLQVKTVETDGYNAVQLGFDDKRNVLSNQPEQGHVKKADTSPKRFIREIRDAELGDVEVGSEITVETFKQGDIIDVAGTSKGKGFQGVIKRHNQSRGPETHGSRYHRRPGSMGQAADPARVFKGKKLPGRMGGQTTTIQNLEIVRVDADKNVILVKGNVPGPKKSMVEIRSAKKAD
other gene	BWX42_08725	1831990 - 1832613	50S ribosomal protein L4	207	MPKVTVFNQQGDTNGEVSLNNDIFGIEPNENVLFDAIIMQRASQRQGTHAVKNRSAVRGGGRKPWRQKGTGRARHGSNTSPIWRGGGVIFGPTPRSYSYKLPKKVRRLAILSALSQKALDDEIIVIDELNFDQPKTKDFQHMLDHIGVERKALVVLEKENEFAKLSARNIEGVKVVAPDNVSVLDVVAHDDLILTKTALEAVEEALQ
other gene	BWX42_08730	1832613 - 1832903	50S ribosomal protein L23	96	MQAQDIILRPVITEQSMADMELDKYTFEVDTRATKSQVKRAVKELFDVDVEKVNIMNTKPKPKRMGRYVGYTKKKRKAIVTLKPDSKEIEIFQTEE
other gene	BWX42_08735	1832944 - 1833789	50S ribosomal protein L2	281	MAIKKYKATSNGRRNMTTSSQSDITTNKPEKSLLVSQKRGSGRNNAGKITVRHKGGGHKHKYRLVDFKRRKDGIRGIVKTIEYDPNRSANISLIQYEDGTKAYILAPKGIQVGQEIYSGPDADIKPGNALALKDIPVGTVVHNIELKPGKGGQLVRSAGASAQVLGKEGKYVLVKLPSTENRLILAECRATIGTIGNEQHELVRIGKAGRNRWKGIRPTVRGSVMNPNDHPHGGGEGRAPIGMPSPVSPWGKPTLGKKTRKGKKHSDKLIVRRRRTKKRKK
other gene	BWX42_08740	1833836 - 1834114	30S ribosomal protein S19	92	MARSLKKGPFVDEHLMKKVQAMDSDNKRVIKTWSRRSTIFPNFVGHTIAVYDGRKHVPVYVQEDMVGHKLGEFAPTRTFKGHSKTEKVTKKF
other gene	BWX42_08745	1834136 - 1834492	50S ribosomal protein L22	118	MASSRTEAHATARMVRIAPRKVRLVVDQIRGKDAAEAISILRFTNRGAAEAVEKVLKSAIANAEHNFDMNIENLVVSEAYANEGPTLKRFRPRAKGAASRINKRTSHITVVVSEKKEG
other gene	BWX42_08750	1834499 - 1835155	30S ribosomal protein S3	218	MGQKVNPHGLRVGVIQDWDAKWYADSDFSDKLHEDLAVRELIAKDLEEASVSRVEIERAANRINVSIHTAKPGMVIGKGGSEVDALRNKLSNLTNKRVHVNIIEVKKPDMDATLVAKSIAEQLENRISFRRAQKQAIQRALRAGAEGVRTQVAGRLNGADMARTESFSEGTVPLHTIRADIDYANVEADTTFGKIGVKVWIYKGEVLPEIEEDQKGGK
other gene	BWX42_08755	1835159 - 1835590	50S ribosomal protein L16	143	MLVPKRVKYRREHRGKMRGEAKGGKEIAYGQYGLQSLDSKWITNRQIESARIAMTRYMKRGGKVWIKIFPHKSVTAKGIGVRMGSGKGAPEKWVAPVKRGKIMFEVGGVSEEVAHEALRLASMKLPVRTRVVTREESGDAHEG

other gene	BWX42_08760	1835580 - 1835774	50S ribosomal protein L29	64	MKANELRELSHQELKDKEKEFKDELFNLRFQLATGQLEDTSRIKKVRQNIARVKTVLRQAELAQ
other gene	BWX42_08765	1835811 - 1836074	30S ribosomal protein S17	87	MTERKERKQYVGKVVSDKMDKTITVEIATQKQHKKYKKRMKYSTKLKAHDEKNIAKEGDIVRIMGTRPLSKEKRFRLVEVVEEAVVL
other gene	BWX42_08770	1836115 - 1836483	50S ribosomal protein L14	122	MIQTESRMKVADNSGAREVLVIKVLGGSGAKTANIGDEVVVTVKHATPGGVVKKGDVARAVIVRSKSGLRRKDGSYIKFDENACVIVREDKAPRGTRIFGPVARELRDNDFMRIISLAPEVL
other gene	BWX42_08775	1836517 - 1836819	50S ribosomal protein L24	100	MFIKTGDKVQVITGKEKGKQGTVLKAIPRENRVIVEGLNIAKKHTRPSMESEGGIVETEAPIHVSNVQLVDPKSGEPTRVGFRFEDGKKVRYAKKSGEAI



Appendix 81: Polyketide DpA biosynthetic gene cluster and its predicted similar gene clusters according to antiSMASH.

Appendix 82: Summary of predicted biosynthetic domains in BGC possibly coding for polyketide DpA

Open read frame	Sequence (start-end)	Biosynthetic domain	Start of domain	End of domain	Amino acid sequence
orf_1832	1822383-1822608	Thiolation	4	69	MVFETVKDIIVEQLGIDESEVTKETDLENGLDADSLDIFQIISDIEDEYDITIDT DLNLQTVGELVDYEQLIG
orf_1834	1823650-1824598	Acyltransferase	6	294	MKVAVIFNGQGAQFEGMGLDFREHFPEARAVFNQASEATGLDMVQLVSE DFSKLRQTKYAQPAIGTVSLAIWASIRETLPQVDYMAGLSLGEYAALMASG IFTVSDGLRLLFERGQVMSDVCEEIAEDEPMQMLAVIGMSREVVEHLVED LPQTYLANFNSPEQIILAGPKSNLKLFNQAAKAAGYRKGLPLKVEGPFHTP LMAAACQPLEVLLDSYELQPGCAPVISNTTVEPHDLETLKSTLVRHLIEPV QWEQTIDWLIQAKVTHLIQIGPGQTLQKLLKAHDQAPLCLASQVEDVSEIE KFLNENKGEKE
orf_1836	1825316-1826570	Ketosynthase	10	415	MGQADKLNRVVVTGLGTISPLGNNIDEFWKKVRANESGIAPITKFDASAVG VHVAGEVKDFDPTLTMDRKEYKRMDLFCQYGIAASVEAVEMSGYDIAANA SRVGTLISSGIGGLIEIENGIRKMIDKGPKRIPPLFVPLTIGNMAAGNISMKLG AKGISMDIVTACASSTNSIGEAFLKIQAGFLDACLAGGCEGTINEIGIGGFNA LTALSTNEDPTKASRPFDKDRDGFVMGEGAGVLFLESLDSAQERGAHILA EIVGYGATSDAYHMTAPVPDGSGAGEAIKMALASAHITPEQVSYINAHGTS TPTNDSGETTAIKYALGDAAYNIPVSSSKGHFGHLLGAAGGIEAITCVKALE DGFIPATLGLETSDEACDLDYVPQTGREADLQYVLSNSLGFGGHNAILCFK RWGK

Appendix 83: Summary of genes in BGC possibly coding for bacteriocin DpD as predicted by BAGEL.

Name of gene	Legend function	Function	Amino acid length	Amino acid sequence
orf00001	Blast hit with UniRef90	Organic hydroperoxide resistance protein-like 1	134	MLYEVVAVNETGIEGESYLVDGETYLTSSPMSEAPGTNPEQLMGLALAT CFNATLHAVLKERGLEPESRVQVTVQLHQDEENKEYYFTLDLEAAVAQV DLQEAKEVVKVAHKRCPVAKVVGNYEGMTVKTVPFN
orf00002	Blast hit with UniRef90	Uncharacterized HTH-type transcriptional regulator SA2153	419	MKLKHILKTASVLALLTGPSAITTVESVYAEEAQDKVADSETNLEALSQQV RSSVQTIEDLNDELAQLDTTIADVESEIEQAEADISQQEAVIEDYFDQAKS RLQNLQLSNVNENAVLSFLEADSLQDLLARVNAVVQLSQANSDQMQSA QDQKDKFDQLVATLESKQNDLTDKKEAIEEKKQAVNTELIELEALIEENQG AFETLKNEDEIAEVCLRAEELVKEHQDQETANTETPDQATDQSEELDNQ PEQSEENTATKEATEQSEPEPEAEQAPQVEETPVQTEQPEQPEQAQPES DVQEPVEAQPESSEESVQPQTAGATVAQYSIDDLEFQGIIHALGKKWTFY SERVLPGGGLNIPGRHTADGFVRDGEGYIVLAASSSVGHGTIVDTPFGSQ GKVYDTCASCHADWFDVYTR
ABC	Immunity / Transport	Potential ABC transporter	221	MEKQITFKNVNKTYGKKNALENMDLAIKNNAITGLFGDNGAGKTTLMK ALTTQIKIDSGEMSIENHVIHFNQPTNKAIGVLIEEPSLYPYLSGREHLNLFK KLEEQKNNDLSIEWLIDEFKLEKFIDMPTKKYSMGMRQRLGIAMTLVTDP DIIVLDEPFNGLDPKSVKSLRDTFITLRNKGKTLLISSHLIREIAEIVEFSCYND RCRLKDTTYDVTIII
orf00005	No function determined	-	203	MKALLFKEFQMTFNWKTFVSYIALLFLIGTIFGHIQVALFAFTFALMHSSIL EDDINDADVFINSLPVSRRMVVASKFLATMIEVALMLGTVYGSKIIVPLFS DLNWQTALLTFAGALILIGFYYTCYYVFGLKLLNVILVILFIMATVVFPILLNL NYLTKLIQIIQFLSEGIGLVIATSSSLLLVIVFYVISYKVYQRKEF
ABC	Immunity / Transport	Putative lantibiotic ABC transporter,ATP -binding protein precursor	294	MENVVEFQGVEKTFDHFKLKELNLKIKKGYITGFIGPNGSGKTTTLKLMM QLLRADQGRVTIFGQEMTQQNYQLKDRIGFVYAENVLYKSQTVAQIGKL VSRFYSKWDQAIFDQYIEKFNLPLDSKVDKLSTGMKTKLSLALALSHHAEL IILDEPTSGLDPVVRRHILDVLFELIQDEDKTIFFSSHITQDIEKIADYIIFIKDG EIIFDEEKYNILERYKLVKGPVDLLDGDIRDLFISIEEKQVGFTGLTTESATLIE LFNDLVTIEKASLDDIMVFFDENQSTALNDGRK

orf00008	Blast hit with UniRef90	HTH-type transcriptional repressor YtrA	130	MIHIVLSKQSSDPIYEQIRLQLKQAILRGELASGEQLPSIRALARELEVSVITT KKAYESLEAEQFVESIAGKGTYVAELNPDAIRTKQIIKIKQDIAQIVESAQVI DLTLPELLEMMEETFQQESQKDT
orf00012	Blast hit with UniRef90	Lipase 2	342	MGIIWIIGYYWADLVIGITRVFSKMRQRLKRRRVKPAKQANVYKNVEYKS DYPNNKLDIFIPSHSSHPTRSPEENRIKEQQNTCDNENDMEIRPALYPVIV YFHGGGFAWGDKIDERRYLLEFVKQGYAVISANYALTPTYKYPVPIIQASN VMDFVASEGQRYGLDLDNVVLAGLSAGGHLAGQLALVESSSAYANRLG LRQHATLTFRGLMLNSALIDPKRSNQVGHWLINWLFSVMGFSYMDMN PSRMKENTLAEANLLNHVNDAYPPVYVSDGNRVSFTQQATDLVAQLRA KDIYVQSNIYSDNYPHRMYHGYEAQLWLKEARDNIHKQLDFLDTIIE
orf00014	No function determined	-	77	LKEREIEPKSRVQVTVQLHQDEVSEEYYFTLDLTAAVADIPLDEAETIIQAT HKRCPVAKIVGDYKHLTVETVPFEA
orf00015	No function determined	-	46	MLYEVVAVNETGIDGQSYLVDGESYQTSSPTSDEPGTNPEQLMDYP
orf00016	No function determined	-	67	MSIIARTTVYFMYKIVPYLNKIAHIQNLSFSLITVTNSETYPTFFLPQQNRG PYKKWTEIMTAMTDQ
LanT	Transport & Leader cleavage	MrsT protein	163	LKSYRQLFGTVLQDETLLNDTIQKNIDPTHSHTISKIREAAKLACLDEDVM NRMPSKYNTEIGDNGRNLSGGQRQRVAIARALLTNPKVIILDEGTSQLDV STEREIFNNLKKKNITIIAVTHKLSTTTISDCVYVLKKGKIIAHGKHEKLMQN NRYYSDFFR
LanT	Transport & Leader cleavage	Bacteriocin ABC transporter, ATP-binding/permea se protein, putative	50	VGSMEKNVKELNFTINEGEFISFVGRTGCGKTTIIKLILSLYNNFEGIYY
LanT	Transport & Leader cleavage	MrsT protein	468	MVLNYYGCKLHISDLVEKCSISRDGVHLRDLMNVANEYGLKSKAVELNKK ENIFSEKIVFPCIAVLSMSHYVVIEKAKRGTIYYLDPECGRIIMDGSEFSKKF TGILLLFFPENIKKQKSQNKIIEILTLGEIKKKFIVGVIFFSFIIQLFVLLMPLFTE YMIDNVINGIALVSHVKLIIGILFAILSYGVFSFVRELLTTLLEIKYISTLKKNV VRKLFYLPLSFFDVRSSGDIVSRINNIDSIQQLLSNIITGVFVDVLTIVISISM MLYISRILSIIMIVYGILLCVLLNIFFKKTDVKNKLSLMKREKTQSYLIELSSNI NMMKTSNYGEALYNKWVKNYNEQMHLEFKRQRLLGMYRSLIISYRLLPS

				VTVLFIGSNFVNQGMLTLGQVMSFLALGNTLLSPLAILIQNIFDFQYSKNN IDRLSEIITVKSEKNFDGREINSFKNIKFQDVNFSYSGINGKKC
LanM	Modification	Lantibiotic mersacidin modifying enzyme	818	VHNGGQTVIILHFEEGTSVVYKPRDLKMDVTFQSTIKWFNEVTKSHLYSL KIINHTEYGWVEYIPHEECKDYSDFKNYYTELGQLLFLFYLLRGNDIHYENII AKGKHPVLIDLETLFHNNTSNTSGIDTAADRVNELLENSVRTVGILPNLV WAQNGKNGVDISAISTSENKEIPIEQASITNVNKDNMKVEYKTSTLASQK NNPYIIGEEISLTSYHKYLKKGFIESYTKIKNINKKEIINQVENYKEIYARQILR PTQYYTTLIQISLHPDFLRSAIDREMLFSKLWIYFDENNSFRKVSEIEFTSLL KNDIPWLISNVSKNNITTKDGSEIESIFKHSSIALVKEKINILGDKDLTLQVEL IETALNYDSEYNKAESQRENDRKIIEINDDKLNKNHLDQQLLEISTNIGDYLI NQSFIGMNGDVSWIDMNVIGEKANDWNMVPTSMDLYSGLSGIMIYFI FLYKETKQNKYLIMVKRCYKSIINYIKNVRKRTNINSEVMFGGFSGETPIIYA LTILEEELGGIFDLDELEKIRSWIFKECKKNISVGNEHDIIIGSSGVIAILLRYY DLTSNDAILEVCQQYAEQIIDNYIEMDNNSIAWIGIASRNALGGFAHGVS GIVWALSKLYSYLPDERYIEVIEKALRYEDYLYSEDDKNWVDRRETEEGIEY NNLSSNMPVAWCHGASGILLSRASLKKHNLPLSEKRKNKIDEDIEIAVRTT LKNGFGHSHCLCHGDLGNMLILKYASSELNTKNDIDKRYDIYMSHLISQLK DKWECGIPYKNSPGMMLGLSGIGFGLLSLMNEDLPFILLLE
LanC	Modification	MrsM protein OS=Bacillus sp. (strain HIL-Y85/54728) GN=mrsM PE=4 SV=	247	MKLIDKFKNGLYSFERDIQHDETNNYKSENRLQYWKKFLGVNEREIENILS NGLGINTANLNELLSENDNFSCKVTETNVLWNQLIHDLQVLSIESIILPEFYI IGDIGQKELPMFYGFHEPFLKLAILRFENYWKNIPGISDNVFNKLLIYLYDQ LAEISYRTLILELNIAREENKLAGETSEERYNYFSTQYLSDNYWLILEEYPVM FRLMCEATQKWINNTTRFIDRILSDKDDLEKLLKLREN
42.1;Haloduracin_alpha	Core Peptide	42.1;Haloduracin_alpha	81	MRGGNIMKNKVNNWKTPVYEKEEDYDNPAGDIFRELKSDDIDKVMAS GTANTYCRCYSGRHSCGRACTITAECPVFTVACC
orf00023	No function determined	-	238	MVLEFKKMKRSKLIIVLILIVVLNIGIHYLMGNVKYVGIPYAAEPGWTLQN TLIVGTYYLLLPVFTLIGSASFIIERENSVYINILTIPVRKSHLILNKSKFIWVVS MIFTSSIFVFTLVLEKIHSNILTQEIVLKYLFEYVVHSNGLYVIAMLILSVIIYM DYNMQLAVMVGFVSSFVSVFIEQTAISYLYPVNALFNISGFKESNLYEYSS SVVILCVIAICAVFIYKKIAQQESW

ABC	Immunity / Transport	Putative lantibiotic ABC transporter, permease protein precursor	244	MLDLVSTEITKITRMKQMWVYASVLGIYITMMSVYTAQAVGLFDQFIHL YKFSLSYTGFLLLPFILLSLSSSTISDDYRNEVMKNLTVIPISKQKIIAAKLIAIYI VLSVAMVMIWISVCVIGYVTRFRFSMTWLLIFRYLWLCILTSLVIFLSMLP VTLISVATKGNVVITNLIGSMYIIASFFLTNFMNGIIPLATAPHIIWYGSME GVEVNSNIALMVISIIFYTIIVLFAMNKILKKQEL
ABC	Immunity / Transport	Putative lantibiotic ABC transporter,ATP -binding protein precursor	302	MKNVIEINNLTKSYDNVAVIKSININIVEGTIHALLGKNGAGKTTILKSILGLI KPTSGSIFILGENQLTEKGRNVLKQVGCMIETPGFYPNLTGTENLSIFAKLR ELDKQAVQEALTIVNLPYADSKKFKEYSLGMKQRLAIANAIMHDPDILILD EPTNGLDPSGVIEIRELCKEMKRQGKTLLISSHILSEVEQVADEVSIIDEGEL VKYINLKKMNENSQVTIHIFTSNFDKVKSVLLEAGVSQHNMINTAKGVRL LSKDMDISEVNKLLTQNDVDIEGIMREKLTLEEQFQRVTETR
orf00029	No function determined	-	372	MKLKNILKTASILALLTGPSATMTVDHVYAEEAQDKVADSETNLEDLRQQ VRSSVQTIEDLNNELVQIDETIDGIEAEIAQTEEDISQQEAIIQDYFDQAKS RLQNMQLSNVNENAVLSLLEAESFQDIIARVNAVVQLTQANSDQIQSLQ AQKDELDQLVQSLETKQTDLSNKKETIDAKKQSVHTEIADLERLIEENRED FETLKEESEIVEICRRAEELVEERQAEQEVQEDASSNEEAKQVASQSAEET QQPQAEETVAIEENEEPAEPTDDSAAEQSVEEPQVTETEPAEAEQSQPE PETQKQPDQQVQEESQTQTEPEVQEEPIQPQAAVPTTAQYSIDDLEFQG SLMHLGRNGRSIQNVSYPVAD
orf00032	Blast hit with UniRef90	Iron-dependent repressor IdeR	214	MTPNKENFLKAIYELGGMSKLINNKSLAEYLNVSAAAITDMNTRLVKQSII TYEPYKGVKLTDKGVRIVNQLIRRHRLWEVFLAEKLGYEWDEVHTDADLL EHISSDKLIERLDAFLGHPTVDPHGDTIPTSDGEVIVNQYHALVECKQGES FKVKQVDDDTEFLTYLTDKGIQLNETYQITEIEPYEGPITLTNNDEENILVSY KAAFRIFGQ
orf00033	Blast hit with UniRef90	Probable biotin transporter BioY	186	MRTLTTKDLVYISVLTAMLCVASLVAIPVGIGIPITLQVFFWLLIPALLKAYR GFLSLALYVLIGLIGIPVFAGGTGGFQAVLSPSFGFLLGSLIIALYIGKVAQKR PSLLTMILHMIVAILILYTIGILYQYFIFNVIAESGGSTTLISLIIANVSSFLPLDI LKAILAGIVYDRLIRHTRFKNI
orf00034	Blast hit with UniRef90	Trehalose-6-phosphate hydrolase	551	MGKNFHDKVVYQVYPKSWKDTTGTGMGDLQGVIEKLPYLADLGIDLLW LNPFYHSPQNDNGYDIADYQSIDPMYGDFDTFDKLVQEASKYDIGLMLD MVLNHVSTEHEWFQRALAGEEEYQDYFILRQAQPNGSLPTNWESKFGG PAWNQFGDSNYYYLCLYDKTQADLNWHNPEVREALYDVVNFWIDKGV KGFRFDVLNVIGKSQGLKDAADGVGKKEYTDTPIVHEWVRELNKRTFGP

				YDDIITVGEMSSTNVSNSVQYSHSESNELDMVFSFHHLKVDYKEGSKWT LMDFDFDELKRLLDHWQTGMQAGNGWNALFWNNHDQPRSNSRFAD PVNYPYESATMLATTIHLLRGTPYIYQGEEIGMTNPNYESIDQYDDVETH NAYRTLLAQGMNNSEALAIIQVKSRDNSRTPMQWTSEDQAGFTTGQP WLEVASNYKYINTNVTDPSQHIRDYYKQLIRLRKDYAVISEGTYRSIHLDH SNVYSYIREYNNQQLLVVNNFYGKQTAIDVPEDFLKKDARVLIGNYDQHA LEQQLILQPYESIAFLLE
orf00035	Blast hit with UniRef90	PTS system trehalose-specific EIIBC component	502	MSNFKQDAQSLLKLVGGKDNIQAVTHCATRMRFVLVDPSKADKNQIEE LESVKGSFTQAGQFQVIIGNRVSDFFNEFQAVSGIEGTSKDSVKSAGKGN MNWLQQLMANLAEIFSPLIPAIIIGGLILGFRNVLEGVQIEALGQAIVDGA PAFTNSGEPIYNTIVDVSTFWNGVNHFLWLPGEAIFHFLPVGITWAVTRK MGTTQILGIVLGITLVSPQLLNAYGVEEILAGDVQQWDFGFFQLDMIGY QAQVIPAMLAGFTLAYLERFFRKITPEAISMIVVPFFALVPTIFIAHAVIGPI GWQLGKWLADGVQWGLSGSLNWLFGFIFGGLYAPFVVTGLHHMTNAI DLQLVAEYDQTILWPMIAFSNIAQGSAVLAIWWKNRHDEKESAVSLPAA ISAYLGVTEPALFGINIKYVYPLVAGMIGSAFAGMYSVATSTAAYTIGVGG LPGILSATNSSYLNFGIAMMIAFFIPIILVFVFEKYNILTDQQLEGLPIPKLGS
orf00037	No function determined	-	91	MMADRNYMMISDQMGHIAGKCYVLSPDSERGFDPEQVADHLTESGTP TTVISNVQAIKQYIDTEATDDEHIVIFGSLYLVGDILTLYKKTL
orf00038	No function determined	-	42	MEQLSEQPFVLVDGAHNEEGVMMLSKSLEQIAPIKNGHYSLE

-; not mentioned by BAGEL

Appendix 84: Summary of genes in BGC possibly coding for bacteriocin DpE as predicted by BAGEL.

Name of gene	Function	Amino acid length	Amino acid sequence
orf00002	-	56	LKKKLNEQERKARNKRLIEKGRVFKSIFEESVDFTKDEFYKLIKMLNDEEIIDNRV
orf00003	Uncharacterized HTH-type transcriptional regulator SA2153	150	VEVEMKVDIIIDETIEETRVKIFAKEYSKEVETIKDLLVDRLVDKLVAFRDKEVFILSHEEIIRIFAQDKSVFIKTKNGTFSSRLAIYELDQRLDKKKFIRISRSDIVNLDFVKKLDLSFTGTIAVELKNGEVAYVSRRNLKEFRKALGL
orf00004	-	141	MEKIKNLLIGFLAGVGIGALIEAILSMIIKENIVGVPEFVASHTVGYAKIIQCLVYGGFGLVSVLMGAIFKNKNRATYLNRTIHFCTMLIYFIFAGFYLRWFNYDFSIVTSAIFFVGIYLLIAYISYFYEKNMIKKINKKL
ABC	Uncharacterized ABC transporter ATP-binding protein MJ0412	250	MMLELQSVDFSFQQDQEKRVLSEVSLSVAPNQIVSVVGKSGCGKSTLFKLCTTELTPTRGAIRFQGREIQLGDVAYMPQQDLLLPWQTVLGNVMLPTKLDEANQLTEEHGRRWLEKAGLGEVADQLPHQLSGGMKQRAAFIRTLMTDSDVLLLDEPFGALDYFTQKEMQEWLLTLWMATNKTILVITHNIEEALYLSDVVYVMQPYRPGQSEQVTPIQIDLPRPREEAVRYSREFIAYKRQLEEAIYEKN
orf00007	Formylamino pyrimidine transport permease protein ThiX	192	VRRYYPFGLILMALLGFFEWSVSAGITPHFIIPKPSTVVLTLIEQYELLWRHTLITLLEVAVGLGVSIGLGIPLGILLHYSSWAKRALYPFVLVSQTIPIIALSPIFVMWFGYGLTIKVAVIFLFCFFPLVVSTYDGLKVTAPAYLSLFRNLKASKWQMFKYLQWRMALPSVLSGIKLSVIYALMGQRWVNG
orf00009	Formylamino pyrimidine-binding protein	354	MHNMTTYFRRLMKGITAVLVLLVLAACGQSTQNDADQAADSSADDQTTESVADGELQDVELTLDWYPNANHVPIYTALKHGYFEEAGLNVTLKMPAEADDPIRLVGANQTDIAVSYPGVLMKARAEDIPVKAFGSLVQRRLDAIMYKEESGIQSPKDLEGKKIGYASDSISEEIIYSMVEEDGGDASKVEMIDVGYDLMPALSTDNVDALISAYMNHEYLLLEDEGYNMGHFEFQDYGIPENQELIFIASDQTIDERSDILTKFMTALQKGYETAVENPDEAIETLFENEENEYALDKDIEIQSWKEFLIEYMSADGEFGTIDPAQYEEYAKWIYERGAIDSELTGEDLTAPAL
ABC	ABC transporter ATP-binding protein NatA	245	VYKQGKEPFVAVDDVSMTIHEGEIVALIGPNGAGKTTTVSMIGGYLLPTSGDILLEGKSIVKTSSKHKPKIGVVFGGHSGFYGRATLADNLSFFADLVRIPPKEHEQEVARVLKLVDLYDVREKEAHQLSTGMMQRLHIARAMLGNPSLLLLDEPTTGLDVEIAKEVRDTIKKLAQQGMAILLTSHIMSEIEVLADRIYLIGGGQIKHEGTVADILDLAKVTHIDRPATLEESYLSIAPTLKRGN
orf00011	-	250	MMRFLRLCWFHFKLYTKNSYFVWLPISSTISIFLLQYLGAYASGTLASSNIWLISGTFGMWASTTTAAGSIGFQRYMGTLQYIVNTRIDDRVSVAAAITPASTYGLLAYLVALVMSVILRVGIHGLTIGTVLAIISLWLSALIMSLFIAAFFVFTPNAMTYEELIMIPILLLSGLFSLQLVQLPIFSVFQWLLPLATPIKFLLTEAVEFDILPWLSSLVLWSGLSWWLSSVLLKRANITGQIGGFRCVYN
orf00013	-	63	VLNPVVAVTSVAVAAITISLSSVSQLLTHDSMRGVDKIMVINRPYSPRYWGNKILTAMVVSWD

orf00014	-	137	MGLMAINLGLLAIVGVDWVVISRIVLASPALLVTGVLLGMLGFFLAWRESNPYFYTNLLSAAT PIIYGVIVSVSEYPPVFKVFSQIFPFYNVRDYLQTGQWEHLAIEGIKLVVLIITTGLVYQLKHKKV VDKKGLLF
orf00015	-	249	MKNYMKSEFRRLRKKKVGYVVLLLGIALLVGAAFGLDFMSQRELEFPYTTNMFYYSSIFVAP NFLLMLTGTLAIVLLGRDRDLISVSIGFGVNRSHIFWGKYFVTLINFLIIGAIFFGVAYASGEMIV PNTEVEYLHRFINNSLNLLPILLSALTVTYVTAILFNSEISAFILILLIYRVINYASNAIIGILPQSEP VFDYLPGTLFSELSVNYLTGNVQLEYVHWGINLGIILVFLLLGSLLYRRKSY
ABC	Putative lantibiotic ABC transporter,ATP-binding protein precursor	305	MDFSRKGDHMTFQSIQVQDLTKKFGRKTVIDNANFNIAAGEICAVIGKNGAGKTTLFKLLTEQ LFPNKGHIEFQGSFERPSIGTLIENPVFFPKFSAYHNLAYFSKQITGSIDKERIQEILDLVELEN SRRKFEQFSLGMKQRLGIALALLFKPDLLVLDEPSNGLDPEGVRDIRQILLKVNRERRTTIIVS SHVLTELEEIATDYIILNEGQIVEKVSKDKLIDNMVKTLVIKVDAAKQAATVLNEQFDALEIKIVD DTTLHLSSDDIPSYDINRRLVSKGIQVHSLTIQNETLEDYFFEKVGV
orf00020	-	79	MEFAVDDIETHAEEISDKVDISIDVGHNYSNRSSFDVGVEENYRVVDTSSSNEVNLDVLLEGI NGISAELDGENVFSII
186.2;Propionicin_SM1	-	33	LTLLAVAIVKSPLTSDVSIVYDIFWFSIKGSCT
orf00023	-	77	MIIRRTLISICSLLSVISGAVLAYQSGIALDEMWSFERYFKHNQLYIYMLAISVVVLVITILLEWLA TSRWGNDYHR
orf00025	Protein/nucleic acid deglycase HchA	288	MTMTELSKQPQRDRAEHNAYFPSEYSLDAYTSSKTDFDGYKIDTPYEGSTRKILVVASDERY VQMKNGKFFSTGNHPVETLLPMMHLAHAGFDLEVATLSGNSVKIEMWAMPEDDKPVMDFY YEILPKFEQPHKLADILAAVTAEDSVYEGVFFPGGHGALVNLPESSDVNQVLNWAMTEKKDII TLCHGPAALLAGALDDQEFLFKDFEMVVFPDALDEGANQEIGYMPGKLKWLLAERLEELGA TVLNDEMSGQVHKDRNLLTGDSPLAANALGILAADELLKRLG
orf00026	-	74	MPNEFNVLSPIYTRRMICMHERLFSFIPRFVGWIDKIPTALLLILIFFPLFYRLIVKGHSLSDIFS DAKDRLLK
orf00027	Uncharacterized protein HI_0108	155	MFTYYIMPIFSATIATVGFGILYNIPKRTIPASATTSGLGWIVYFICTQVFHLPLFVGTTLASFTIA LISQLFAKHYRMPVTIFAIPAIIPLVPGGSAYNSMLAFVTGEILSAMRYLIETFIVAGGLALGLTV NSAIFQVLSPRAIIQQGRRYLP
orf00028	Inner membrane protein YjjP n=265 RepID=YJJP_ECOLI	255	MTEATTYQAEEALNICMDIGRLMLSNGAETYRVEDTMHRIATSFRLEYVNVFVVPTAIIMTTK NEVGADVTQLVRVTDRATNLEMVAELNQLSRDLSAQPKSPNEVRAYLYLLQMRIREFPPYLT VLLAAITTGFFPFLFGGSWPDIIPAFLAGGLGEFLFEYVNGSTNITFFAEVVAAFAIGLTAFTLY TLGLGENMNAIIISGVMTLVPGIAITNGIRDLMAGHLLAGVSTLAKALLTAGAIGVGIAVVLTFI
orf00030	Phosphatase YwpJ	291	MIKLIVSDMDGTLLNEEIELSDENLAAIKDAQAHGIHFAIATGRDYQTGYTIVQERGINCSFFGL NGAIGYDEEGNRLYTKNLKPSTVQTLLHVLDRDDVHVNIMTDKGVYSTNYERERYLRHVLS DINKTLAPERLEEKLDLFLEQHNITYLDNFQELIRRDDEEILKVSAQTTAGEEMLTELKDTLLQ AADDIVITASSAKNLEINHKDATKGFAVATYARELGIDHTEVLTIGDNINDLSMLEWAEHGTAM ANAAPEAKETAAYETGSNSEHGVAQIINRVLAGEIY

orf00032	Uracil-DNA glycosylase	231	MNLPVMNDWRPILEEAMQTETYQQLRAFLKQEYREHTVYPAMEHIWNAFEQTPYEKVKAVI LGQDPYHGEGQAHGLSFSVQPGIPIPPSLNNIYKELQSDLGIAPVNHGFLRAWTEEGVLMLN TVLTVRAGEANSHRGKGWEALTDHVIKALNERSTPIVFILWGNQAIAKEAMIDETRHAIIRSSH PSPLAAYRSFFGSQPFSKTNTVLRALEMEPINWELPHHVSSDEV
orf00033	Uncharacterized HTH-type transcriptional regulator YurK	132	VTTKRVPLYLQLTEKIIDQINDGTYEAGDKLPSERELCHIYDMSRITVRSALSELERDGYVKKF QGKGTFIANTTYQQNLLNVYSFTEETKKMGKTPQTNIVSFELVLADKKYARKLGIRVGTRCTE LSDVA
orf00034	HTH-type transcriptional repressor DasR	86	LTEKDLANSPMYDVFNRAYNIQATRAEEEFSITTLRDHEAELLAEAVGDPAMLVKRTAYDKS EEVIEYTISVINGQKYKYKVELQQ
orf00036	Putative D-galactosamine-6-phosphate deaminase AgaS	387	MFNKSTDELKALDALHTTTEIKQQPDLWRETLAIYRENKERIDTFLDQIKEQHKQINIIFTGAGT SAFVGETIQPYLHGKYRNTGISVQSIPTTSIVSNPEDFLSEEVATILVSFARSGNSPESVATVE LAKQIVKDLYQVTITCNADGELAKNAEDDAKNLSLLMPKQANDQGFAMTGAFTAMTLSALLIF DTDADKAAYAEELIPLAENVIARESEISQVADLDFNRIVYLGSGSLEGLSHEASLKLLELTAGK VATFYESSLGFRHGPKSIVDEQTAVIVFQSTDPYTKQYDNDVLREVYHDKITDHVFAVEQGE SNFEGQSILVDKSDMKLPDAYLALPYIVVAQIIALHKAVNIKNGVDNPSPSGTVNRVVQGVIIH DYNRD
orf00038	-	69	MKLEKIINGYMMIALFLLFIMGRLLDYALTMDFWGSVFSSSTFYHLVALSTYIACMINMKRRGII DSYW
orf00039	-	63	MITLSFLFGILNISHMSLTLLYSVYFQINREYGFFEWSYLLHLRGYSPNVIRTEKYFAFLAQS
orf00040	Magnesium transporter MgtE	425	LFLRLHDRDQSDAFRELTADNKRKIAEYVEPLEFRDIFRMMEVEDQEAAIEHLPYQYIKDVMA QMPNDSIAYFINRSNQTKKEWVLKYLDSAKQREVLEILSYASETAGSIMTKECAVVLEHQTVE QVIEYLREIGEYAETIYYVYIVDEHHRLTGVTSLRDILMSSTDTLMKDIMIQQIVSAHVSDDQED VLQTMKDYDLLAIPVVSEDNMMQGIITVDDMIDIMEEESTEDFHEFAGIRKQDMTDETGQKK SIIQTAFSRTPWLAIFMLIAMAIGGLTNLFEDTLQQAVLLSAFIPAIMDTAGNVGTQSLAVTISE KNLTGMTFTELLKTLRVEILAGIYMGVTSAAMMFVVINIFYRDMAIAAVVSLSLIITIVVSTILGVII PLIVDKLGFDISVASGPFITIFADAIGLFLYFSIATILI

-; not mentioned by BAGEL

Appendix 85: Summary of genes in the BGC possibly coding for NRP CsinA.

Type of gene	Locus tag	Location	Function predicted by antiSMASH	Amino acid sequence length	Amino acid sequence
Additional biosynthetic gene	CSING_01090	216954 - 217766	Zn-dependent hydrolase, glyoxylase	270	MEHPAYSQLRPVSQSVGVVLCDNPSYTALEGTNTWIIRAGEDSRAIVVDPGPEDEGHLNVVTAKAGEVALILLTHRHDDHASGAQRLRQLTGAPIRSFDPNYCNGADALVDGEIIDIDGVTPQLEVVHTPGHTADSTCFFVWSAEAKNSTLEGILTGDTIAGRHTVLLSETDGDLGAYLKTLDLLEERGKDVPLFPGHGPDLEDTSAFARKYIDRRHYRLNQIKEIRSRLGEDVDVKTLVDEMYDDVDPVLRHAAEQSTRTALKYLAAQQ
Regulatory gene	CSING_01095	217864 - 218547	cAMP-binding protein	227	MEGVQDILSRAGIFQGVDPVAVQNLIEQMETVRFPRGTTIFDEGEPGDRLYIITSGKIKLARHAPDGRENLLTVMGPSDMFGELSIFDPGPRTSSAVCVTEVTAATMNSDMLKQWVADHPAIAQQLLRVLARRLRRTNANLADLIFTDVPGRVAKTLLQLANRFGVQEGGALRVNHDLTQEEIAQLVGASRETVNKALATFAHRGWIRLEGKSVLIVDTEHLARRAR
Other gene	CSING_01100	218851 - 219489	endonuclease III	212	MSDLDSDHSRIGHIRAALAAEYPDADCELNYSTALELLVATVLSAQCTDERVNQVTPRLFATYPAAADYAAADRAELEAILRPLGFQRAKAGHLIGIGEKLVGDFGGEVPRGIEELTSLPGVGRKTALVVRGNAFGLPGITVDTHVTRLSQRLGLTEAKTPRAIERDVAKLVPEEEHTVFSHRLILHGRRVCTARKPRCGACVLASWCPSRS
Other gene	CSING_01105	219499 - 220068	thiol-disulfide isomerase-like thioredoxin	189	MSQYSRWPGYVKASIAAVLMLAALALVGVVNLLEDDEPEVAVPQSQEQEVTKRPKCPEGPIAGVDLPCLGAASTAAAKDIQIVTVWAWWCEPCRTELPFFQDIARSHSTWNVVGVHADANAANGAALLSDLGVDLPSYQDENGTFAGELGLPNVIPVTVVVRDGKVEKKFVKPFTSADELEQAIDEVLK
Other gene	CSING_01110	220065 - 220742	NTP pyrophosphohydrolase	225	MSRLKPAHAPAWLRPALGVDTRRVQERLVTHVPAAKRRRESAVLVLFKGESFEDGEVLLTHRSPSMRSHSGQIAFPGGRRDDGDASLVDVALREAEEETGLDRSTVTPLEQWGKLDIRATGNTVSPVLAYWHQPGTVWPASPEETDDVFTVPLRELADPANRMMVGFSRWKGPAFRARGYVVWGFTAGVLSGLMDHAGWAVEWDKNKVHDLRESLTRSLNNEKMG
Other gene	CSING_01115	220775 - 221962	Trypsin-like peptidase domain/Colicin V production protein	395	MSLAVDIAIVIAVLLAVWVGWRQGAWTAILAAVGVIAGLVLGTAVAPAAMQLTDQPALRFLLAVGVVILLVGLGQLVGASLGAALRDRMRTRSGQRVDSSIGAVFQAVAAIVVIWLVSLPLASNLGGQPGQALRESRILSELNAAAPDRLAALPNGLAAMLNESGLPPLVSPWERSGADIEVEAPELEIGDPGLLERVRPSVIHVLGDAEACSRRLMGSGFVTEPDYVITNAHVVAGTDKVRLDTVLGLKEADVVYYNSDVDIAVLHSPDLGIDPLPWAENDAVTGDDAIVMGFPQSGPFSAETARIRDRLTIAGPDIYSTGRVERDAYTVRGNIRQGNSGGPMITPEGQVLGVVFGASVEDSDTGYALTADEVRGHVGDVTQLVDAIPTGSCVG
Other gene	CSING_01120	222197 - 222433	integrase family protein	78	MPPMPPSFHTLTRCTSAPPALAANHSDHGSQYVSIIYNERLAEHGITASIGTVGDSYDNALAENVNGSYKNELIHSRK

Other gene	CSING_01125	222758 - 223375	-	205	MRIFYGLSLAGVLTALHADTPSTTSVTTTTGSTAAASPVTGATSRPGLNGHANHNATSAAADSAVDFVGTKPVRVGRTAYTYDDAGRVVRTVTKRLSRKPLVHEFFYATNEQPIGFISSDNKHLGWRYIYDPYGRRVAKEAINTTTNNTSEAIAHGGGHPENNLSDIDTVDQALVDAEFSCHRSDRQPTRTHPPHHREHRRYSPP
Other gene	CSING_01130	223636 - 224220	Transposase IS116/IS110/IS902 family	194	MPEFQKLDAMLAAIHEQTVSIAGAEYAELGVAMSATDALTKLEHRKEIEAQVLKLIQDIPQTEILLSMPGIGPRSAAQILMTVGDMSDFPDAAHLASYAGLSPRTNQSGTSIMSNSPNRAGNKKLKNALWQSSFASIRFHERSRQFYERKRKEGKRHNAAVVALARRRLNVLFAMMRNGELYRDISTVQEAAAA
Other gene	CSING_01135	224278 - 224958	-	226	MPENNIHKEKQKPFTPPNKLTHGIDKLYRNTSLGDGAGAESAAVADAAPNTATVVEYPLPVQEFAVTRYDLPAGTTVSTQLPGPAIALATSGVVTVNDLELAPGEAAWLPVTGGDTTRLSAAAGKDSQLFIAAARPARQVALPRRWGRRRATSSTVALTGAIYDMVGRRNAAAAASSTTVSATIQATKAQTAVPGWPVARISALGVPSMAAPARAAKVSGFHRAMG
Transport-related gene	CSING_01140	224763 - 225584	ABC-type multidrug transport system, permease component	273	MTNHTTATVSHDTTWLSTVATHFGRHLRAARRDSAVITNTVAGPVLMFLVMRWLFGDLMAASQGSATLDALPLTIALILSSELMNGTSAAAQIIKERQRGITTRIATTRYGTSPEIFGRWCFDSLRSLISGCAVLLASVASGLRIHTLAGFGWVLLVIIFGAVVAASLSAMVGAMCATPEAAIGPAPIIMAAMALNGGLVPVEQFAGVVQPIARWNPLTFAARAGAAIDGTPSAEILATGQPGTAVWAFVAWMVALTVVLLAAAAAFRRPTMS
Transport-related gene	CSING_01145	225577 - 226338	ABC-type multidrug transport system, permease component	253	MNALGAVAASWHRNVLRTVRNKAVMVSAVAIPSLFMVIFYATFAKASTELGIDYAAFLLPAGVIQAIVFAAGGSSLAITRDAENGIHDRIRATGTPAWATVVGRLLADLTRAAWSCSIVVLTTILLGARYAAGPGRIILTIALFAALTIILSAFIDGSCLLAPKPTSASLLFQNLVLVMVMFSTAFVPADALPGGIGPIIRHVPLSPILDTARNLLGGAALGARGVEALCWLIAMTVVGVWGFITAFQGRKHD
Transport-related gene	CSING_01150	226335 - 227399	ABC-type multidrug transport system, ATPase component	354	MKENPTTATASGDPSATAVPSTSAHAKGRDLVLDVQDLRCSLGKVKKRTEILRGVTLAIPRGGILAVLGANGAGKTTLVNVLSTLLPATGGTAIIDGHNLATDPSKVRASIALTGQYAAVDEELTGKENLIFFGRLAGLKASDAKARAAELLERFDLVGAADRLVSAYSGGMRRRLDIAASLTIPPALLFLDEPTTGLDPAARSSVWDTVRGLAADGTSILLTTQYLEEADQLADRVAVLAGGKVLDEGTPAELKDRYGTTVCLITMSSPDDAARLTHTLQEHRIDCRQEDTVLRVDAPDGHRTLVRALALWDGDDTSVADVSLVPPSLDDVYFAIAGSGDIAQQAPHTGGGSQ
Additional biosynthetic gene	CSING_01155	227811 - 228440	acetyltransferase (isoleucine patch superfamily)	209	MDSREQIEDKILCGQPFRPIEDPDLYYEMMEAAEKCHDLNQLRPLQTAEKQAVMADLLGELGERLMVFLPFHVQYGSKIKVGDGVTFNRGTTVLDMAPVTFGNEVMVGPWCSFFAGNHALDPVERQYMVCTGGAITIQDHVWLGGNCTVTAGVTIGHGAVIGAGSVVTRDIPPMVLAAGNPCRVIRDIGPEDKLMGESWVPEWPSWITD
Additional biosynthetic gene	CSING_01160	228448 - 229665	glycosyl transferase, UDP-glucuronosyltransferase	405	MKVLAVVLGSEGDMLPFAHLGVALRDRGHQFVLAGFSEFGGSFRANGVDYIELPGDYRALMKRLLGDSKGMMDTVLGIREMISDAHVFDVLEHAMDGVDVVMYTQFSEVARLLGAARGVPSVRVQVFPTEPCLRYSLVDPRKLDGSVGALVTHWMSNTLMAWAMRPVIAAWRRRLGLRQRALSSPPTTIYQFSPALSPPAPEWKNHIHVTGEWLDPHHSELALDPAVEEFVAAGSAPVLAS

					FGSMVSDRVYDLQRWTRDACIEHGLRAIIVDPDHEPGVREGILTVARVP FATVLPWCRAGICHGSLGTTGAILRAGKPCLAVAFGGDQHFHANAVCR NGAGPNYIDAQRGELSAQSLAAGIADLVSGAYDSAARRMPELLATDPGT VAAVEIVLNQSDLAKGEK
Core biosynthetic gene	CSING_01165	229662 - 230447	putative thioesterase involved in non-ribosomal peptide biosynthesis	261	MTEISCEGVFPFVAGHRTGTLLFCFHHAGGSASVYRGWVGVNPSIDVV PVELPGKATRRREKWVSDFDKLADMCATEIVDLAAGAPIALYGHSMGA ALAYQVAACLQQMHRPTIPEAVVVAARQAPGETVPGEYHSSMGFGALR REFEKVGGTPPEILANDDVMKLLLADIRRDYVLHEGFHHATTVLRSPVLA LAGDSDPAVSPEMLSRWENFTSGSFELKVLPGGHFFPLDTGTEFLDLLA GFLAGITGNTAWLARNNA
Transport-related gene	CSING_01170	230475 - 232241	ABC-type multidrug transport system, ATPase and permease component	588	MTEQKPPITFRQISQPAHRQITVSLLLAVLAAVLSLVPVIVLVELVRTVML SFQGEPIDASKLWLLVAALMVATVLHGQATVKSLQVSHQADGLLGEHLR QQQITKLGRLPLSWFTRTPSGTVKTYIEDDVTKIHQLIAHAPHDYSAGVL VPLLSLGYMFVVDWRIGLLGLVPLLLAVATMPFMMREFQEKAEKFKSSQ KQLDAAVVELVRGIPVIKVFVPEGYDESIFLSRSRSFGGFYREWLHATVH PTALMKIFTSTGFGLLVVAAASPWLITSAGVPVADVLAAFLFINNIAAPLL MLSRTNIMFSEAKAAAADLTEFFNIPVLPPAAGGREPANAGIELKDLSFS YVPDTPVLSGINQKLTAGSITAVVGPSGSGKSTLAALIPRLLDPTEGSVRI GGVPSTDLRSDQLYRRVGFVFQDPYLMRMSVRDNIRLAHPEATDADVV RAARAAQIHERITHLPRGYDSVVGEDAQLSGGEQQRLAIARSILLDAPIL VLDEATAFADPDSEAAIQRAITELVAGRTLVVIAHRLHTITGADRILMLEN GEVTEAGTHEELVAAGKGYATMWARYQTAQAGIQGEEK
Transport-related gene	CSING_01175	232241 - 233965	ABC-type multidrug transport system, ATPase and permease component	574	MIRTLYSLGSPADRRRLVLVLTLIGISAVALAIGLILIALFLDTLFSEDPAQA AAWLPWIIISVLVYAAADWPTEVIAQDLGHDYVLRIHRLIADRTAQLPLGY FEEDRAGQIGVVATSGALFAANAPAMMLRPMLHGAASAGLACVFLIAVD WRLGLVTVAVAAVVWFAYNRLMRQYRVAERQKGERNEHGAAEVLEFA QVQPVLRMAGPDSLGERAVRASIREQLSAQQHTQKTGEMIMGRLALIIM VGTVVIDALATVLLLNHWLEAGTYIGVIVLVFILARVAMSGIPYGEGLESA RNTLDEIQKILDARVLPEPAVPAAPRDYGIEFDGVGFGYEPNTPVVRDVS FRVAPGTTTALVGPSGCGKSTLLKLAARFYDVDQGTIRIGGVDIRDLGSR SVLDSLAMVFQYVYLFEDTLYENIRLGCHNATREEVLRAAELAGVTEIAR QLPQGFDTLISEGGQNLSGGERQRVSIARALLKDARIVLLDEATSSLDVQ NEHLVMRGMKTLSAERTIVVIAHRLHTIRDADQIVMMSPSGTVESIGNHE ELMESSPRYRSFWGEKSDATSWKL
Transport-related gene	CSING_01180	233969 - 235447	ATPase component of various ABC-type transport systems with duplicated ATPase domain	492	MIKLHDVSLSYGQNTAEPVLHHVNLTVAKGELVLVCGESGCGKSSLLRLI NGVAHTFCDAQISGEVLLDDEDITHAHPHDIAERVGSVFQNPKSQFFTL EVASELAFGCENLGVAPNEIRQRIGELSADFGMVHLLDRHLFTLSGGQK QKIACASVAAMHPQVLLLDEPSSNLDLAAVDELRRIVAQWKTQGRTVLIA EHRLSYLVDIVDRVLVMQDGQIAHDLSGSEFRTLDKAELYRMGLRSAQQ VPEVTREPRPSSGTMVLKKLQFTYPKATEPSLSISHTELPQGQVIGVVG RNGAGKSTFVRVLTGLESKATGVVDINNRSLSGPRQRLRQSYLVMQDV NHQLFGESVESDVIIGTSGPDGKNEARLTEVLGALDLADKRARHPMSLS GGERQRVAIASAVLSEREVIVFDEPTSGLDLCRMHRVAHLLDALARQGK

					TVLVVTHDMELVARCCDLLLRVEKGRLTACEPCDDLALARTVRYLRGRDS
Other gene	CSING_01185	235444 - 236190	ABC-type cobalt transport system, permease component CbiQ	248	MSPTLPGSGTHTAGLHLDPRTTVLVILVVSSVLISPAGSSGALGSTARW MLIAVPGALFLASGMVAAALRYALTFAVLAGFPAVVVQALPERHIIVDVC ATWFAGLSLILPGITCCWYLLRTVSASEFMAAMQRMRSPDALTIPTSIMF RFFPTILEEYRDIRTAMQMRGISGLRNPIAMLEYRFVPLLASVVSIGNELA MSAVTRGLGSPRRRTTLCEIGFRVGDAIVISVLVVSLILLLVNIVMLP
Core biosynthetic gene	CSING_01195	236908 - 244479	amino acid adenylation enzyme/thioester reductase family protein	2523	MIKEEQIREELLASLHQILGEDAEIGIDDNLLSHGLESLPTVRLLADWMKQ GHRVSFGDFMRAPTVRQWAKMLVESTPNHSTDAIESPKDGFAAPIDDS VPFDLTDVQYAYWIGRNSSQQLGGVGTHGYVEVESRSINIDRLQQSWL TLLRSHPMLRACYTEDGKQYVLPEPPHPTILVHDLTKMDESTREEALLS TRERLSHRLLDIATGHVVSLEVSLLPQDVAVIHFDIDLLVCDVQSFQIILHD LAHHYATGEAPDADPSWSFARYLAGHAREGVADIDRDQAYWRNRLSE LPGAPTLPMSHGTNEEQAHRFVRRSRSFDSATWSRLREVCEHHATTPA MVLLTAYARTIGQWSENKKFLMAVPLFNRGSDTAIKNVVADFTALTLTSI DQSTRRTFSEDLKDIQASFYEDSSHSQYSAVRVLRDLRASRGEQVLAP VVFSCNLGDPLVGQEFIDTFGEISYMISQTPQVWIDLQVFTTVNGFLIVC DAVEQLFPEKMLDDLFATLVMEIDKAITDDLSHSDPVESPGAQARRASR AEVASWRLPDTTLVDEVIAAARCHPQATAIRSASGDVITYQDLEEQATTI ASALVNSGVGRGALIAVMVERGPRQIIGALAAMMAGGAYVPVSLQQPE SRIAALLGASQVTHLITDRPDKVLSETAVQVVDFTSATGTANLPQLHPQD PAYVIYTSGTTGTPKGVEICHGAAWNTISEINRRLGVGPTDRLLSVSSFD FDLSVYDAFGLLSAGGELVTIPDDARRDAKKWVSLVDSLGITIWNSVPTL FEMLLSAADRTPGKLSSIRHVLLSGDWIDTSLPERMRTVTPQAHLLAMG GATEASIWSNGLDLDVVSPEWTSIPYGRPLAKQMYRVVSSNGQDCPDY SVGELWIGGLGVATQYVGDPGLTETKFVISEDSRWYRTGDMGRFWAD GTIEFLGRSDNQVKVRGHRIELGEIESACEALLPIERAVCITHQGASSSPS LVTFAQFTPSHVARTTPEQFATSLRAKVNDVLTEGDIRTSVEHDEHLQT AYAFSVMRRWEEQLTGVGTPNHLREHRNRWQTWLGKADEHPATADLL LDDESFGALERFVTPFEQAFVMAEKQRSIAEFIQSPDSMSVEQFLATRP LGRLVHRVLGAVVRECSTHSTSELKILEIGSRRPEASADYAAIAGTSAYV LADPYRHHLEHAGQRVGNTFTYRQLGVTSTPQPTPGEAVTKADLVLCN QTLHQSEDIEKTLCEAWGLSAPGATMVVVEPTAPSPMSDITAAFIANNTT DARAETGTVLLSARSWKEILQRTGWKPVEHVEITKTTALIIAERASSNES VTLCDSDYAKATNLLATRLPEYMLPKRILELAKFPLTSNGKIDRKALTALV PEYFDNEPAVTELPHTATEKRLIDIWDELLHTSSNVNSDYFRLGGDSLTA TRLRRTIEQCFGVEFPLENIFDVPLLRDMAARIDQIAEVPHQQSDLPKIVH GSEQYAPFPLTEVQQSYLIGSSGAIELGDVSSHCYFEMSTACLDPERVE DAFNALIKRHPMLRTVVCEDGLSQRVLPEVPRYRIALIRSGNADNEDTLD EIREEMSRQKFDPTQWPCFDVRYVAEPDAGRLLLSFDNLFIDGWSMFHI FREWKQAYDHGVDSLDPAIPYSFKDYVEATIELSHSDIHKRDQAYWESA VDTIYPAPQLPVTDTNGANTSQFCRHHALVDAAKWRRIKQRVREEGMT EAVFLAEVYAEVLARYSDEPRLSINLTRFDRTRFAPEVDHIVGDFTSLSIL SVDTQCAPSFRDRAAALHRRMFSNLDHGSVSGVSVQRMLTKQRGARV

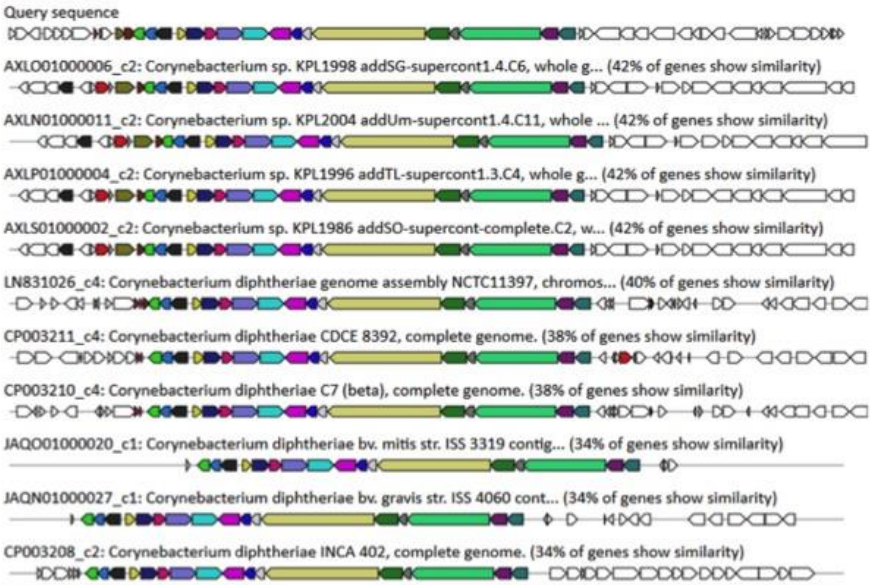
					TMPVVFTCGLGVVEHPESDQSPYLGVIDHGLSQTPQVWMDLQVYEHD GGLMLNMDAVEAIFPDDMVAELFTSLTATLSHLAESPELWNAPTSTIAPT TNAPTADRINDTDRELPGADKSLLGLYQKGLAEHGDNLAVIDATTQWTY EQLNEQSDKWAQLIAATDPAPGDLVGIMMEKSAQQIAAVLGAMKAGCA YLPLSVDQPVGRNTSIINDAGASIVAMDHPDDDFAALAEHCTVITLADVA RHRPGDQALSESSPTPSSLAYVIYTSGTTGTPKGVAITHESAVNTIVDVN ERLGVTPTDRILGISELNFDLSVYDIFGMFARGATLVLPSPADKRDPQCW ADAVTTHSVTLWNSVPALFSMYVEHLRERSLIGSSVRSALLSGDWIPVNI AYQVSTLFRDCTVFAAGGATEASIWSNWYEVGVDDASRTSIPYGTPLA NQRMYILDEALNPRPTHVPGDLYIAGRGLAMGYWKDPEKTAASFITHPR TGERMYRTGDKALYNHLGHIIFLGREDGQVKVNGYRIELGEIESTARKFN ELRDCVAVNDHGIVLYVVTHEGFNMAALNNHLAESLPAYMRPRVISRID GLPRSWNGKIDRKSLEGKTFEQPQTRERSRNHRDSGIITILQDLLGPKEI SIDDDFFTIGADSLTAVRLTNSIRREMSVEISIRDVFNHPTVRELSDLIADI VGSDVEEGEI
Core biosynthetic gene	CSING_01200	244492 - 246123	peptide arylation enzyme	543	MRGRDIDRNSVLKHYEGTGHLARKSLCQELFESSEIYSDHVAVIADDAH LTYAQLEQSLCVFTEELRESGLSPGDHVLLQLPNTAAYVVTLLALMRVG AIPTLLLPAHREAEVAALCESLHPVAYIGGRDHLGFDTVAMVEAMGPGE LGLKELWADNGPTHDKESSYRVLPGLFTAPISTKCSPPTKWPDPRSVAL NLLSGGTTNMPKIIPRVHEAYAYNTRAAAQCCGVGPDTVYLAVLSTSHD FALAQPGILGTLLSGGTVVLCTSAAFDEAFPAIATHGVTLTALVPAVAEV WVEAAEWFPADFSSLERIIIGAAALNDGLGEAIQDRFGVRIHQGYGMGE GITTFTRIDDPPAVILGTQGRPISDADELVIDGPGGEPGEILEKGPYTFFG YEGNRDTPDCFTEDGFFRTGDRGYLTEDGNLVLYGRVVEQINRLGENV SPSEVETLLSGTPGISAAAVFPMPDRALGERTVAAIVAQPGVNRSAILDD FLTRGVARYKVPDQVITVDEIPLINIGKVDKKKLRALAAAQFTDRESEQS
Other gene	CSING_01205	246120 - 246728	Glyoxalase-like domain	202	MANVGHIIYKADDLENSINYFRSRGFDVEPGQQKNAASALIYFCEGPYLE IRERADVPPFFRQLLHLTGNGKFVDSYDRFATMSQGYSRVVLEAQRKE FDHIKEIFDSHQTKSLTIPFSRKDPAGRRLACWCLSPDDWAIPLFVTPFAI DTHRSTPHPNGITHITDIDFAASEKTLSICKHVGVDGGLTCRSGTGTIDLE FA
Core biosynthetic gene	CSING_01210	246721 - 252255	amino acid adenylation enzyme/thioester reductase family protein		MDVTALINDLESRGIALWVNGDRLNYRSPKGSLREEDLAALRSNKEKVL AWLREREAVPHDEQARFAPFPMTDIQRAYATGQNEGYDLGGTGCHSY AEIRTERLDRSRLEQAWHELIQRHDMLSAVVVPPDSLQVVKSRSLPVLQ AVDLAGHNPDVPDAEYLRHRAKLENRSYPLGTWPLHEFQLLQFDECSIL QFSVDMIIADFVSVRVMVEELLTLYAGNVLPELEDTTFRDIITSRNHHSQS AAGFAARTNAKKYWSEIIPSLSGKPLLPTLTSADRTSEMPVRFTRRTWR CSPAAWSKLTDAASTHGVTPSATLLTAYADVLRRWSSTSDFCVNVTSM NRDSAIAGINRIIGDFTEMTLHACHPHTGTFSERVHATQEQLSEELSHAA YSGVDVLRDIARTTGQPAVIPVVFTSALGADTPHNNGPAYNLVSGVSRT PQVWIDCQAFQDGGSCNVNWDVREDVFEPALIDDMWESFTDLLDRLV DDGSAWQETDSVHLPDKTIAIRNRIHKTHVQQTTRCLHDGFWDNVQQH PHQPALVCGGKTYSYQHLAGYVGALQHELSDVGPGDYIAIVLGNGVWQ IAAAVAVVSTGAAYVPIDHEQPAIRQRSMIEACRPANVITNSHFSEENTDI

					SNINVDTLSPIQYSGTIASPVSPTETAYIIFTSGSTGIPKGVVVTHSAAMNT IDSVNNLLGRNKRRTVLGVSKLSFDLSVYDIFGTFASGGTLVLPLDEESR NPSKWIDFLVDNNVDTWNSVPALFQMLVREVEVTRHPNILSLDLVMLSG DRIPGTLPAHAAPHFPNAELISLGGATEGGIWSIFHPMTCHTNETSIPYG TALPNQGMWVLDEACNECPDWVRGQIHISGESLATGYLNDPTSTAEKF FFSEKHGTRMYDTGDIGSYRPDGVIEFHGRRDNQLKINGYRVETGEIEG VLESNDFVERAIVLTQETSDPIKLHAFVTDAQSDKDELKDAGQIRNSELR TMLEQRWTPADTSLDTGIFATWMRLGNEAAMAALLAAFQQAGVFLVAG KYHTLTEITAAIHPSEEYRELITRWLNILTGEGLATKDDEGWTVSQQTLD FFVFGEAWDQFGNMEAEINNSKELFNYQRHAAEALLSQLRGEISPTEVF FPEGDTHNARTIYGENRISKAMNAAAAEAVIGIAEHHADHPVRILEVGAGI GATTEKIVSRLPENVIEYRFTDISTFFLHKAQKMFAHCGAMTYGLFDMNS DCTSQDVEFGGYDIILCANVLHNSVNIEESFTRLKQLRRPGGVIVIVEPIT ELYAALISVSIKMNLVDFTDHRAESHKVFIEDAQWDQVFRDTQMHRIAEY PNTSDPLRECGQRLIIVGADDDDVPTLNSEDILGYLRAHLPGYMVPASV NVLPELPLTSNGKVDRKALAQLCLEPVGSPNNRIDPPRNETEEQIATIWR DVLDTTEVGRNDDFYALGGDSLLMAETVTRLRQEIPGLQQHTWDALMR GVLKVPTIAGISALAQAAGSSCQPEALKAVNSANHTSPELTALSTVASGS PTGSSNLHVYRLPKDATFCRVMIHAGTGRLKDYEFLMPELLQRQPEIAH VGFTAGDADRFLDYTTRTLIRDLAQSYAQELDELDMESYQLVGYCIGGM LALETAKALTELGRDVRQVTCISTHQCPHRVTNELLCELAYGCIFNADLS AMGANFDLKTLAAALEHTLDGINRNISDEELCTLEGPYADIGEFFQKMAV LSPRARRKLIYRSIREFDTDSESTRGMLDILYDVFRHSLLGTIDYVPDVYF GDVVVLQPTEGVTGFYPSLGGDIDWPATVLGNLQIHAVAGSHATCLLQE NVPSLLPFFTEREQRNG
Other gene	CSING_01215	252237 - 253382	thiazolinyl imide reductase	1855	MNTVRVVVCGTTFGRIYINGIKKLADKFSLVAILSQGSEQSRRLAEQLGV PLCTKIEDLPAFDLACVVVRSSVVGGSGTQLALEFLSRGKHVVMEHPIH KKDSVDCYRMAAKNNVQFKLNTFYRWNPTISRYLEITTTLTRQFKIIHIDA ECSIHFLFSTLDIIGRITGGFTPWSFDDDTQVTGIFTTLTGSIRKTPLCLRV VNHNDPEKPDDFAHVGHRITVFTHAGNLVLTETDGTIIWHPNTPIPRDQA GLLSTAADDRLSTLQLHENLLIEPCVTRVALYDHTWPAGIAEFLNEVYDK LACSKNEAQEAEYLLALCAVWARTGQLLGPTCTVQAAMHAEPLSLESL MWPLGESHTSPIERNNPNHRNTGEITWMSQR
Other gene	CSING_01220	253424 - 254533	NADH(P)-binding	369	MVRTGETARVGVLGSHGEVGRQVATLLRASGHMVNCGNRSHHTSDER TAIVDAHDEESVNKFSRDLDVVVNCAGPSCLLKTSVASALPDSVGYIDPF GANSFDNYPTNRPCVVNAGATPGLSGLLLRHLAGLIDDCQSVTLYTGG RDRGGLSGLVDVVLSTHNSYGHPGKMIVDGDLVAYQPGSIHAEDLEPF PSDDGLIASPFVTNELRKVAQDFSIPRLIGVTAIPDRDTQQLLLRALSQTD ISDPTTLMNLCADIAAAKTKLDAGKNHWFAIQATVHGTLRGRPVRVSGS VHAPDSTFITALFVAEATKSVIRDELLTGPKWGYELPAPQTVLSSLASYN VDIKIVGPTATTDAPQWSDDDDAGFI
Other gene	CSIN	254895 - 255158	hypothetical protein	87	MVDFLDTYPTLGDFFPIAYIDGDDYAVLQRTADGVLCGRYDALDNEWW EQPAQSFTEWFYAMAAPRQNNAQPVHPRLPNTTASAPSH

	G_01225				
Other gene	CSING_01230	255179 - 255832	hypothetical protein	217	MKKTLDAWNLRITLTATPAHRIPHVAATRLKELRAKRTAKPGVRITYRAK GPNTLSITDDATTIANMRAVLESTNTTDLLQAANDIFFKEGTGSTPAVRT QVIITLNELDQIINGDGDEIELNLTNGARMTGAEFLTHKFAEIGYATIVHPIE GPINTWRLSRGANLNPSNNTTHKPQQGTRGFRGSLTTVNVPQHAAGC TGTLKLLKNETRQPEDG
Transport-related gene	CSING_01235	255862 - 257571	ATPase component of various ABC-type transport systems with duplicated ATPase domain	569	MTKPLLEMKDVHISFTTSTGVVEAVRGVNMSIYPGQSVAIVGESGSGKS TTAMSILGLLPGTGKVTGGQILFEGEDITHYNNKQFESLRGDKIGLVPQD PMSNLNPVVWRIGTQVEESLKANHVVEGSKRHERVVELLEEAGLPDAER RAKQYPHEFSGGMRQRALIAIGLAARPKLLIADEPTSALDVTVQKTILDHL EHLTEELGNAVLFITHDLGLAAERAEHLIVMHRGRIVESGPSREILRSPQ HPYTRRLVDAAPSLASSRIRAAKKAGVKAKELKSGGAIGAAVKQGTATA DSASAQAPVISVRNLTKEFDIRGQRGGKKLLKAVDDVSFDIRRGTTLALV GESGSGKSTVANMVLGLLEPTSGTIEFEGHDTSTLSKQELFKLRRKMQV VFQNPYGSLDPMYSIYKCIEEPMALHKVGSRKEREARVAELLDMVSMP RSAMRRYPNELSGGQRQRIAIARALALRPEVIVLDEAVSALDVLVQNQII QLLAELQSELSLSYLFITHDLAVVRQTADDVVVMKKGQAVEQGTADDIF ENAQQEYTRNLINSVPGMHLEIGTGH
Transport-related gene	CSING_01240	257571 - 258530	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	319	MPNFEKTTHYPGQEHFVSETDETGLGAVDAVKDESAPSSQWGEAWRY LRRRPLFWIAAVMILVAVLMAVVPGLFTNTDPRLCELSKSLAPAEPGHPF GFNRQGCDIYARVIYGARASVAVGVLTTILVVILGSLIGAIAGFFGGWIDS VLSRITDIFFAIPLVLAAIVVMQMFKEHRTIVTVVLVLGLFGWVSIARITRG AVVSIKNEEFVQSARSIGASNWHILFSHILPNAAAPIISYATVALGTYIVAE ATLSFLGIGLPPTFVSWGGDISDAQASLRVAPAVLFYPAGALGLTVLSFI MMGDVVRDALDPKARKR
Transport-related gene	CSING_01245	258523 - 259449	ABC-type dipeptide/oligopeptide/nickel transport system, permease component	308	MLRYIGRRVLQMIPVFFGATLLIYALVFLMPGDPVEALGGDRGLTEAARA RIEAEYNLDKPFIIQYLLYIKGIFMLDFGTTFSGVPVTQVMANAFPVTVKLT IMAIIFESIFGIFFGVVAGMRRGGFFDSTVLVISLLVIAVPSFVIGFVFQYIV GIKWGVLPVTVGANASVKSLLMPAMVLGALSLAYVIRLTRQSVAENLRA DYVRTARAKGLSNGAVTRRHVLRNSLIPVATFIGADIGALMTGAIVTEGIF GINGVGGTMYQAILRGEPTTIVSFTTVLVIIYIIANLLVDLLYAVLDPRIRYA
Transport-related gene	CSING_01250	259648 - 261249	ABC-type oligopeptide transport system, periplasmic component	533	MTLKKTLAVLSAASLPLTLAACGGDDSGSNSASGSGAGDNYVIVNGSEP QNPLIPANTNETGGGRIVDSLYSGLVYYDGEGEAQNELAESIEPNKDNT EFTVKLKESTFSDGSPVTANNFVDAWNYAVANDQLNASFFSNIKGFKEG VEKLEGLKVVDDLTFTIALNSPEQDFPAQLGYSAFYPLHESAYDDMDAY GQNPITNGPYKLSEWNHNQDATVVPNEEYKGGQTPQNDGIKFVFYASQ DAGYADLLSGNLDVLDAVPDSAFDVYETDLGERAVNQPTAVFQSFTLG ENLEHFSGEEGALRRQAISHAINREEITETIFKGTRTPAKDFTSPVLPGYS EDIKGNEVLKYDPEKAKKLWAEADKIKKWDNPSVEIAYNSDGGHKSWV DATANSIKNTLGIEAVGAPYPDFKSLRDEVTNRTIKTAFRTGWQADYPS QGNFLAPLYKTGGSSNDGDYTNPEFDKLLEEALTAKSDDEATKKYNEA QAILFKDLPSIPLWYSNATGGYSENVDNVVFSWKSQPVYYNITKK

Core biosynthetic gene	CSING_01255	261472 - 262374	putative hydrolase or acyltransferase of alpha/beta superfamily	300	MTSLSPSVVELDGPFEHEFLHTRGIRLHAATAGDPDNPLVVLLHGSFGGWFDFREVIALLADAGFHVAALDMRGFGMSDKPPLEPGQDIRVAVGDVSGAIRALGHDDAFLIGADTGGAVAWALATERPERVRGLVSVSAAHPADLRRAVAARPWDFGWLLLRSTLCRLGWLHAPSLLMRESTYRRELDLDALGSFGGGEREETLRLRVAASQIGNVRRGILWNHRLRTAAVPLSWLDLTVNAPVLFIHADQLLWRPVARRAARRCRGGFTATTIPGTKNLPFLENPRSFVHTLTHWMRST
Other gene	CSING_01260	262465 - 262959	Protein of unknown function (DUF1469)	164	MSNDGLFTDGSNQFAPKVDSIPLSDADTSRRGEASIGQLVSNATEQMSQLVRSEVELAKTELAASAKKGGIGAGMFGVAGTVALYSSFFFFFLAELLAKWLDRWAAFLIVFLIMVVLAAIVAFIGFKQIKQVKAPEKTIESTKELKKLVPGKAEKSIERGLYT
Other gene	CSING_01265	263065 - 263754	hypothetical protein	229	MGMPSSDPLTPLMTLSGVEEGAAAAVDAIARVHRRPAGLRKFEVISSESLLRGARTCAAIDGAPLARDGVPPTVSAYSLLAPEVQASTVRTFARAPLQVLARIDVAAGGPGRPARDSAVAQALAQLITRGAGVDFDRLLPVVLHAEIAARSLFGENSTVVALVAARAAAIHTGFDPRGFAVPETYLNRHRAAYREALMGYEDTPAELFTLLFNAWTAGAEEADGIARAA
Additional biosynthetic gene	CSING_01270	263770 - 264582	HAD-superfamily subfamily IB hydrolase, TIGR01490	270	MIEHPVTPSDSTRIAAFFDLDKTIIATSSAYAFGREFMHNGLITHTEALQLSLAKASYMLAGHSSEHMDATRDQLAAMVAGWSVKEVHDIAVDTMHTVVTPAIYAEARELIAAHREKGHEVVIISASASVLVEPIAQELGIEHVVATELAEEDGRFTGEILFYCKGGAKAEALARMATKLNVDPNASFAYSDSATDIPMLEQVGHPIAVNPDRLLKKHAADNGWETRTFKHPVPLFTAPSAKEVGIGSAVVAGVAALVVGGVLLARQRAD
Other gene	CSING_01280	264905 - 266797	heavy metal-translocating P-type ATPase, Cd/Co/Hg/Pb/Zn-transporting	630	MSSACGCEHKPATEIDELDRPWWKDPELLLPIFSGVALITGLALDWSGLETPATVLFWVGLLLGAYTFAPGAIRNLVTKRKLGIGLLMTISAVGAVILGYVGEAAALAFLYSIAEALEDKAMDRAQSGLRALLKLVPQTATVLRDGTAAEVAAKDLEVGELMLVRPGERIATDGIIRSGRSSLDTSAITGESIPEEVAPGDEVPAGAINSAGSLEVETTAAGTDNSLTTLVDLVEQAQAEKGDRARIADRIARPLVPGVMILAVLVGVIGSLLGDPETWITRALVVLVAASPCALAISVPLTVVAAIGAASQFGVVIKSGAAFERLGGIRHLAVDKTGTLTRNQPEVTSVVPADGFNRAQVLTFAAAVEQQSTHPLAAAIAAAGPEAPAALDISEEAGHGIGGTVEGRRVLVGSPRWIDAGPLKADVERMESEGQTCVLVTVDDALAGAIGVRDELRPEVPEAVKALHDNDVEVSMLTGDNTRTARALAEIAGIDDVRAELRPEDKASIVAELSSKTPTAMIGDGINDAPALAGATVGIAMGATGSDAAIESADVAFTGHDLRLIPQALQHARRGSRIINQNIVLSLAIIIVLMPLAITGVLGLAAVVLVHEVAEVIVILNGLRAARTKR
Regulatory gene	CSING_01285	266794 - 267153	transcriptional regulator, ArsR family	119	MLTIASRLDVMNRLGRAMADPTRSRILMTLLDGPSYPAVLSRDLDLTRSNVSNHLTCLRDCGIIVAEPEGRKTRYEIADPHLTAALNALVNATLAVDENAPCIDPECSVPGCGEKGADA
Other gene	CSING_01290	267340 - 267573	hypothetical protein	77	MTDSLGAQKIRRAVLIVTLLNLAYFFVEFAGSVAIGSASLFADFADFLEDTAINLLVFSLWPGLRPAVARREACSPR

Other gene	CSING_01295	267558 - 267995	Cation efflux family	145	MLAALILIPAIAAIVTLVAKIMNPVAPSPEGLTGIAIGALIVNVICAILLQLRNEGSSLATGAWLAARNDALANILIIAAGLLTFVWETAWFDIIVGAIIAAINLSAAKEVWEASRKEHDFVEDAFADQSAKLLNISHFNSHFEG
Other gene	CSING_01300	268232 - 269251	helicase/secretion neighborhood CpaE-like protein	339	MSTFHLLIAVGDSALRAEAASTAAASTAEVSSVEDPRDFPRYLPKVDAVLADSLTASLVGSHPRVYFLAPDPGPIDYEAALRCHASAAFILPAQSKELLAALAADTHPETPTASAGLTIAVTGSAGGIGTSTLAAALARHAGAALLVDASPYSGGLDLLLGVENKPGARWPDLAAGTGTVDPGDLARALPTTPDGIAVLSAARTTSAGATALSATRRAAIMQAACAHPDIVVVDCPPWDIPDTADHVVVVTAAEVRSAAACAEIVAELRARPQECSVVLRHRQWSGIDEGDIAKLTHADPVTELPTVRGLTRVVETGGLPQRLPRPLSRAAKDVWEVLV
Other gene	CSING_01305	269248 - 270384	helicase/secretion neighborhood ATPase	378	MSLIDTMRTIVATEPQLAHDASALARRIREEAGVISDVDVLDLLQRLRHDSLGLGLLEPVLSLPGLTDVVVNGPDSCFVDCGQGLVRREVGFADDAEVRQLATRLAAVAGVRLDDAQPFADGRLTRLDGTHIRLHALLAPPSASGTCISLRVLRQAQTSLDQLVANGSIDKSVEGLLRGIVDKRASFLVTGGTGSGKTTLLAALLGCVPETERLLIIEDTPELAPAHPHVVTLVSRRANAEGRGEISMSLLLRQALRMRPDRIVVGEIRGAEVVELLAALNTGHDGGAGTVHANSVDEVPARMEALAALGGLDRVALHSQLAAAVHVVLGMERGPEGRRLAHIGVLDGNPVTPQLVWTTDDGPRPGFDELCARMGVKP
Other gene	CSING_01310	270381 - 271133	type II secretion system protein F (GspF)	250	MMWLYLAAACLLSPPGSKARLSAPKPRSVAPAVLFLAVGCFVFFGRPTVVIAVGCILACALWFAHDLAASRRERRGAQALATYFGTLAAELRAGSTTSGALRRGADSLPESTPDNLRSALSTAAGLAAQGGSPGAALTTTELSRFAALLNVSGRHGVALASLIEQAQSQLDTAQRHARETAASLQGPQATALILAFLPVAGILMGGAMGADSLGFLLGGGVGGVLLDVGVALVCIGFTWSRLILRKAARR
Other gene	CSING_01315	271130 - 271678	type II secretion system protein F (GspF)	182	MTSYLAAVLLAAALAVATSSPAGRVGLGTTGTPRPKTPRDGPDYGPLDAASDLELFAACLEAGLSLRTAVAAVGEASSPWKEAAALVGVGVPMKSVLQALASQPHLVELVRLAQLSGESGTAMATGCHRLVAQLRAEASAQATAQAERAGVFIAAPLAACFLPAFLVLGLVPVIISLGQQLL
Other gene	CSING_01320	271719 - 271907	Protein of unknown function (DUF4244)	62	MQKTRTIQAQLANEDGMSTIEYAMGSLAAAALAAVLYAVVNGGQVTSAITSIITDALSNTPV
Other gene	CSING_01325	271907 - 272230	hypothetical protein	107	MRALDDAGSVTIEAALSLASLVIVAAGIVGGIATLSAHLAAIDAAGAAARSAAIGVDYHRDGVDITFTESSGLITAEAAVPAPLGTMRAQAVFPAEMGAGTAEEVRR
Other gene	CSING_01330	272227 - 272571	helicase/secretion neighborhood TadE-like protein	114	MRRRQPGRLKPLGGEDGYATVVTAGIIAAVTSLLLAVAAVGAAVAARHTAQVAADLAAVAGAWDLAKGRDACSKAEEVAALNNARLDSCAVDDRDVEVTALIRGRSAIARAGPI



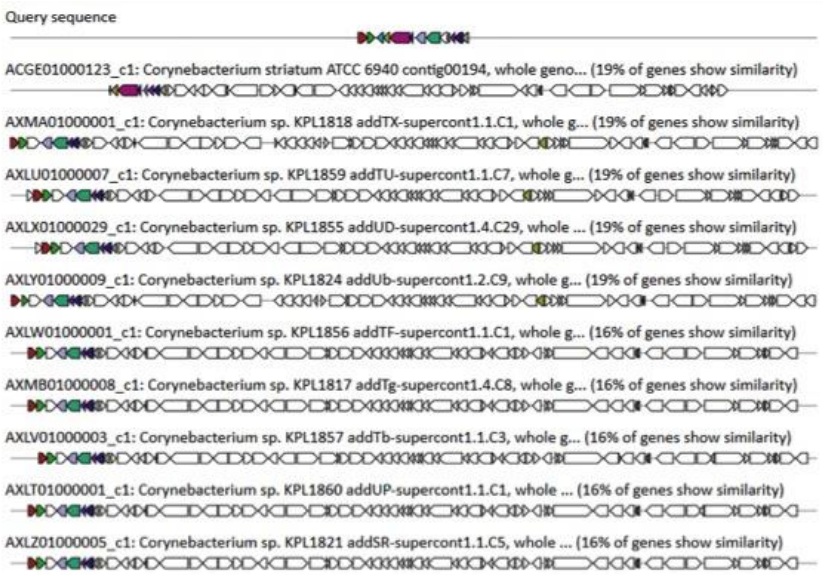
Appendix 86: NRP CsinA biosynthetic gene cluster and its predicted similar gene clusters according to antiSMASH.

Appendix 87: Summary of genes in BGC possibly coding for bacteriocin CsinF.

Type of gene	Locus tag	Location	Function predicted by antiSMASH	Amino acid sequence length	Amino acid sequence
Other gene	CSING_11410	2396072 - 2397025	ABC-type Mn2+/Zn2+ transport system, permease component	317	MGGSSLSKFLEDTQYLLGVDFVQDALIACALLGILSGVMTSLIVLR QMSFSVHATSELALMGAAAALLFGLNIGFGAIAGAIVAAIILAVLGL KGQQDSAIGVVMSFGLGLSVLFLHLYPGNANTAMTLLTGQIVGVS SASVWLLAATTVIIVAAVTILWRPLLFASADPVMAQAAGVPVKALS VVFAVLVGLTAAQSVQIVGSLLVMALLITPGAAAVQITSSPLRAVV WATIFAEVSAVGGFILSLAPGLPVSVFVTTISFVIYLVCRLIGWLRN RRAVRDDIAAARVTKAHFEGRSDSEKHHTSTHDEHCAHPD
Other gene	CSING_11415	2397037 - 2397837	poly(3-hydroxybutyrate) depolymerase	266	MTIVEFQGFKRKYTVVAPEDPGPNQLLFLHGSLQNGSVMRRFTN GTFDELAARTGTVVVYPDGVDRHFNDARGVLPVKARELGIDDVA FLQHVAATVQEEYGTQRTYACGYSNGGQMVIRLLFDAPGFLDGA CIFASTMGSGANHAPSNPEGYKPTPILMMHGTADPLAPYEGGQA GLANSSRGEVTSALWTAQRFATMNGCSKAKVTRPFSDVSLHSW EGDNPVELWTMEGVGHVIPSGNGLDARLGPNTDSFIAAQVVEDF FGF
Regulatory gene	CSING_11420	2397980 - 2398600	response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain	206	MWRLKYSDMPCLLAGIRDDEQPILDELRRYFLWPQVECASTQRE LLNALSRSQEQLLVLCERKDSATWVWDIQELKSSDFIYLVDYSQH SREKLQKAINLGASAHCVVTREARSGHGGLKLSCCFDINGVAVG TMPEPDMQRNLERLSAIQLRILNGIVCGHSNNGIAEETHLSVPMIK KIVSQLLDIFAAENRVQLAIKTLPAMG
Transport-related protein	CSING_11425	2398600 - 2399226	heme ABC exporter, ATP-binding protein CcmA	208	MEIQVEDLHYSVKGKMILSGLTRSFAAGRVTALTGVSGSGKTTLL NLLGGLISPDKGSIMWDGNNIASLSAGRLRKYRRECVGYLFQDY GLVPDLTVVDNVKLAASNVSRRAFDEALDSALAQVGLRGYEKRV VSGLSGGEQQRVALARILVSKPKIVLADEPTGALDEDNAQLVVEH LRAFADQGAVVVVATHSKHVADQADCEVTL
Other gene	CSING_11430	2399230 - 2401215	hypothetical protein	661	MKTSVRLLTVGILLSTVISAFIAFVGASLLSTLDYRLGNEFVQIEDF KPDSHTETLFPELSEFAGDSGFDVSIYYSSLADSDGELRIYSSALP SGVASFESQKFQRGEKLVWYPLSEIPTKEPRQVFNIKGSENDAR DLTTWLEERGYSVIGLQNLTWSFVMTSSLPLLVAAAFLLALVLGG GHCLVRAREIGIHRLLGLKLSQTLRLDVTRLAKQTWLPFFLAIIFIA AALYLYNDWAEATLFWAFFGVSVLLQWVGLALGYVLGQLLVRNI SIPQAIKGKVRARPITYALYAVRLLALLTSLTALSQMVDTGAEVSA REELQPYWSGHANEQEFALSSNITPDEKEQAAAAAPLRQADREG KLLLADPFWLTWPTQLEAPVILSNHRFAVQAGVSPSEEDVTVCAP VALSEESKQTIIDSVAFEAELAKKKVPDFNWKEDCSLGQVFSYDV ELRPVINDPIVVSLPVGLEAIGDRNLLAKASQKVLLAIEPTVVGELK DGTVGKVLATAQPHADSWRSSLEEAKDQFSLWSLNTVVAILLVS LLIVAGLFSFQVAFRRELYVSFICGRSPWKMCRKILLAELFFFVTTI

					GWVLYRLREHKEQLESHYPSTWSMGFENRWSTATVLAVVAFSIV WMCISIGLTWWISARNQMRWTDSEGN
Core biosynthetic gene	CSING_1 1435	2401222 - 2401515	bacteriocin, lactococcin 972 family	97	MKVSMSKRAVSGTLAGLLALGGTGVAGATIVNIDGGTWDYGTGS GQVWSHYFHNGVRHGSTAVGTFKSDSGCVNKNTWSRATAPKK PSGNKSYYRHC
Other gene	CSING_1 1440	2401825 - 2402772	rRNA methylase, putative, group 3	315	MARTHGRGGAGIRKSNKKGATKGSGGQRRKGLVGKGPTPKAE DRVYHAKHKAKLERERRNSGRHQKETAEMVVGRNPVIECLHAK VPATSLYIVQGTRNDARLSEAVAMCNTRNIPILEVQRHEMDRMTG NGMHQGIGLQIPPYKYAEVHDLMDRARDAQEPGMFVILDNITDP RNLGAVIRSVAAFGGHGVIIPERRSASVTAVAWRTSAGTAARLPV ARVTNVTRTVQDFQKAGYQVVGLDAGGEHTLDTYTGGKDPVVIV IGSEGKGISRLVRENCDVIMSIPMTAWVESLNASVAAGAVLSEFA RQRRAAK
Other gene	CSING_1 1445	2402828 - 2404234	cysteinyl-tRNA synthetase	468	MSTLRIFDTATRTQRDFEPVRPGHASIYLCGATPQSQPHIGHLRS GVAFDILRRWLLAQGYDVALVRNVTDIDDKILTKAAENNRPWWE WVSTYEREFTKAYNTLGVLPPSVEPRATGFVTQMVEYMQRLIDA GFAYAVPGEGDAGSVYFDVAAWNAAEGSDYGALSGNRVEEME QGESEAEGKRSPQDFALWKAAKPGEPSWPTPWGRGRPGWHL ECSAMSTYYLGSEFDIHCGGLDLQFPHHENEIAQSHAAGDKFAN YWMHNHWVTMSGEKMSKSLGNVLSIANMLEIVRPVELRYYLGSA HYRSVLEYSESALTEAAAGYRRIEEFVAKFEGVEKGEWTPAFEE AMNDDIAVPKALAEIHTTVRAGNKALAEGRREEAERLAGAVRAM AAVLGFDPCDEQWKDGGETRGADAALDVLVQAELERRTTARAE KDWATADAVRDRLTAAGITITDTADGPTWSLAD
Other gene	CSING_1 1450	2404268 - 2405002	hypothetical protein	244	MTVRMGGLGQHLTAEFFNWVHAVHVKVLPLNEDSDIRVWTDLN PDDGTGLLELHAPLVYGVGSQGGDRFSYSDLFESVVRFHTSLSP DESSLSVKSSKFELRVGTAPGVRFEYERENRTAPTAHIHHSGVA GLLSPALMKNFSGVKNPRKKGRIEDIHFPIGGRRFRVSLEEYFFFL FMECGFRSRPGWKSFLEQSREKWLDDQLRGAVADNQEVAAQV LREMGYTVERKGTAPSAEARGHRPW
Other gene	CSING_1 1455	2405003 - 2405203	hypothetical protein	66	MALTMRIINTEALRARYGELLHALEVAIGANLSSEDLRMLAESGRF STEERDLYDELRRVELMLER
Other gene	CSING_1 1460	2405231 - 2405722	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	163	MTNPIIPRVGIATDAHQIEPGKECWIAGLLFEGADGCEGHSDGDV VAHALVDALLSASHLGDLGSFVGVGRPEYDGVSGAQLLRECREL LGKEGFAIGNASAQLIGQTPKMGPRREEAESVLSEILGAPVSISAT TTDRLGFTGREEGRAAVATAVVWRAPNS
Additional biosynthetic gene	CSING_1 1465	2405715 - 2406464	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	249	MTSPSTTKTKPRVIALVAAAGKGTRLGADVPKAYVELRGRTLVER SVQAMVTSGIVDEVIVLVSPAMEDFAARILERSAAEIPVRLVHGGG ERADSVWAGLQAIPDEDAVVLIHDAARALTPPGMVARVAKRVLD GATAVIPVVPVADTIKEVEGETVLSTPDRSRLRAVQTPQGFNLAA LRRANLDYWEQNPDFIATDDASLMEWHGARVETVQGDTFAFKIT TPIDLVLATAVTDEAEPTIFEVPSD

Other gene	CSING_11470	2406436 - 2407020	transcriptional regulator, CarD family	194	MEFKVGEVVVYPHHGAAKITAIETREMGGETLEYLVLQINQSDLVVRVPSKNVEMVGVRDVVGKEGLEKVFSVLREVDVEEAGNWSRRYKANQERLASGDINKVAEVVRDLWRRDQDRGLSAGEKRMLAKARQILVGELSLATPVDDKKADTMMEEIGATIERHRAAGLVDDKSITTDVDSDIDLDDLSFDDED



Appendix 88: Bacteriocin CsinF biosynthetic gene cluster and its predicted similar gene clusters according to antiSMASH.

Appendix 89: Summary of genes in BGC possibly coding for type 1 PKS CsinA.

Type of gene	Locus tag	Location	Function predicted by antiSMASH	Amino acid sequence length	Amino acid sequence
Other gene	CSING_12240	2604856 - 2605368	universal stress protein UspA-like protein	170	MSNGETLLIAYDGSARADRALEYAAHLLRPTDVKILTAWETSARQAARAVSRSGLHQSAVSSEHVEDDPAYEEALAICRRGITKAESLGLRGHAHLVESVSSIQSAIVDAAQELDVDVIVTGTRALTGFRSLWNSSTAEYIVRNAGLPVFIVPPENDDDSDDSDASGDDA
Other gene	CSING_12245	2605412 - 2605609	Protein of unknown function (DUF2613)	65	MAFDSDSLNKRTLGPALGSAVVGVALGVVTILGIAQFSQADAVPSTSSVKASDAVMGGPEYGSRE
Other gene	CSING_12250	2605612 - 2608713	protein of unknown function (DUF3367)	1033	MRARIGGLVCLALVVVVQPWGLTAADTKHDLVANPSGFLAGALHAYTDTFTLGQLQNQAYGYLFPQGLFFLLTDPLPDWIAQRLWWFLSLAVGFGGFFRLVRAALPAAPAHAALAGGLLYALSPRVLTTLGAISSETWPVMLAPWVVLPFLRAGRLTWRDAAPSVVAVACMGAVNATATLAACLPAGLVLVWCRAGRTGLAWLAGCAAVSAWWIVPLLILGRYAPPFTEFIESAGVTTRWLNLAEILRGTTSWAPFVDTERVAGTALATEPVFVLFTVGVAALGLVGLARLPRMWSVMLLIGVAVLGTHAAWYLDALDGPLAALRNVHKFDPLVRIPLLLGVAAVVSRMGLPHSVHPTRRQAAGLLVGLIMVGATAPAWSGRLLPKGAYEEVPAYWHEAADFMNSLNTRVLIYPPASFARQDWGWTRDEPAQPLLRVPWAVRDAIPLIPPEAIRGLDGAMAQISRDPATAPDVLQRLGIGAVAVRHDLISGSSARIPGEVHTFGEVDVVVFDSDLSMNLAPADLPRVAGGGEALALLQGPHELVPSDASIVTDSPSLRDRNYGTLDGPVSAPLAPTDPSTVRNRLRDYPSAGPLTQVVEHDGRVTASSSAADASAFGGANPARSMTAAVDGENSTAWWPAPGDTGWLELHAATEWTAPTLRLMTTGTTEVTIHNGDSSVTVTMEAFRARDIRVPGARASTIRIEPQRRTGIAEASIVDQPTPRRLVTVPDTSPSAEQFLFQQDNEGVLIRSFTAPRSMTVRVSGEKPVLIDDLPHSPGSTLTLPAGQHLVRTTGAWAALTDTASALPAPLPASSRALSVSDRGYTPLAPDEAIEPADHDRLLITGRAFNEGLRGELRYPGSEPLHLEPRSVDADTQAFLIPAGAAGRVHLSFVADRAYRVSLGLGGALAALTTAFCLLALSRPPRRAVTAALPAAHHSVPSHPTSVIIALMIAATMTLVAGLPGLATAAAAWAVTRWTTLRADYLAGALALAAGAMLARAPWPSDGYAGDSLLVQLLCVAAITCACLPPRGEA
Additional biosynthetic gene	CSING_12255	2608999 - 2610198	putative acyltransferase	399	MRFLPELEGVRALAALAVVVTHVSFQTATGWAWAERLDYGVAVFFVLSGFLLWRRRHAHTVGQYFLSRVARLAPLYVVCVVAVLALLPDASPLSLPQLLTNLTGTQLYIADGLVPGLTQLWSLSVEFAFYAVLPLLAWGLDKLSARGRVAFIAVAGLLSLGWGFLPFVASYESGQVNTQIWPPAYFSWFAVGMLAAEAEAAGVRVRAGRIVRPACWAVALGCVWLASREWFGPQGLVHPAPDEFARRIAVGAVCGACIVVPVALAPRQSWLSSEWMRALGRWSYGVFLWHVAVLAIAFPLLGVSAFSGSAWDFAVVLVFTVVVSCALSAATYVLVEVPGRDLLRSLAASRGRIIVPRSLPRQAADPARSARAGSRSARANTRSAHAKTRSARANSCRNP
Other gene	CSING_12260	2610253 - 2611224		323	MFLRSRLAQVTIAAAVFLILGSVLPPLYNNQARPLATNLNLEVRVEQDGVEFTRHTTTSPAEKDSVTRAQSTTDLVADGETIAHITEDSLLNRESTYPVAGPNTTQHIELPAFDVQ

					VEDRIERDGMSYFFPAGAEQRSYPIFDPLAQAAAPIDFVRDEKLDGIPTYVFYQDVA PVSLLEVFSFDAHQAGPASDFYSPSSLERYNLKPTSHVILEPFYAVSRTVWVEPTTG TIVNEEEHPRIFWATDSHQAQAMVEADDTKERALLDVPLTWPDSTRTARLDAVRRVI ETVKVLSIVGWLGKTSGVVLLAIAAAMFMRRRAQSA
Other gene	CSING_12265	2611238 - 2612722	hypotheticla protein	494	MSASSAVRWALRVWGVVLVLAVCWPFLLPGEFVWRDMVLLDQPALTASAFGSGD LPARNAPQDGVLALVGGAWLARVFVLGAASAAAWSACRWSLGRAHPSPWAAAAA MACAVANPFVIERLLQGQWSLVVAAWLAPVIAWGCLNGHPRTAWLALWASSLTPT GGVFGALSAVLCEREGRRRWLFAGGSVLLWLPWAVPSLLSGSSRAAGDPAAQAA AFAPRAEAYAGTVGSLLGFGGIWNAAAVPASREAGFALVGLVVAVLAVLGASSLDR GELRPLAVLAAVGMGLAILGGIAPGVTAFAVEHIPGAGLLRDSSKLTLLALPLAVAGIG ALHRLPAALAICACLLQAPDAPRELAVLRPHDTGVDHQLVEELDGRLTYFVDRPAMV EVPGGIALDPYSKATNKLDSGALTVDGTVVDHPSPRYLAAAAAWAERDLDRLEELG VGVVVEGGRIVATTDAAPQPAPWMLSAAWCALPLLAVLRNARRSSPRNT
Core biosynthe ti gene	CSING_12270	2612646 - 2613737	glycosyltransf erase	363	MKILLLCWRDSTHPQGGGSERYLERVGEYLAAQGHEVVFRTARHMNAAKRERRN GVLYSRAGAKFSVYPRALLALLFGTRDMRGIDVVVDTHNGIPFFARLATRVPVVILTH HCHREQWPVAGPILARLGWFLESRVVPVLYRGNTWVTVSEASKRDLEKLGIHGAHII ENGVDPLPEHVPTLEREADIHLVTLSRLVPHKQIEHAMDTVARIPGALLDVIGSGWW ESHLREYARTRGVEDRVRFRGQVTEDYKHALLARADVHLMPSRKEGWGLAVMEA AQHGVATVGYAFGLQDSVIDGKTGVLVQREAEFTQAVQKLVEDPSLRHRLGAAAR ELAATYSWEKTGARFEELLEAEVRTRRR
Additiona l biosynthe tic gene	CSING_12275	2613762 - 2614523	methyltransfer ase family protein	253	MKRTRAMATLRRSWRLLRSFRFEQTAPAVFYGGLADDTAALIDALCHDHSLSLHGA RVLDVGGGPGYFASAFAQRGASYVGLEPDAGEMAAAGIEVAHAVRGDGTRLPFAD NTFDITYSSNVAEHIPHPWDMGEEMLRVTKPGGLVILSYTVWLGPFGGHETGLWEH YVGGEFARDRYTRRHGHPPKNVFGTSLFALSAREGLDWARRVSARGAKLVAAFPR YHPWWAWWVAKVPVLREFLTSNLVLVLQAEP
Other gene	CSING_12280	2614580 - 2616400	phosphoenolp yruvate carboxykinase (GTP)	606	MTATIPGLVGELPTKNEKLISWISENVELFQPDEVVFVDGSQEEADRLAAELVEKGT LIKLNEEKRPNSYLARSNPSDVARVESRTFICTETAEDAGPTNNWAPPAAMKEEMT EAFRGSMKGRTMFVVPFCMGPISDPDPKLGVQLTDSPYVVLSMRIMTRMGQQALD KIGENGEFVHALHSVGAPLEPGQEDVAWPCNDKKYITQFPKTKEIWSYGSGYGGN AILAKKCFALRIASVMAKEEGWMAEHMLILKLTNPEGKKYHIAAAFPSACGKTNLAMI TPTIPGWKAEVVGDDIAWMHLREDGLYAVNPENGFFGVAPGTNYESNPIAMKSME PGNTLFTNVALTDDGDVWWEDMGEAPEHLIDWHGNDWTPSSTTKAAHPNSRYCV PITQCPSAAPEFDDWKGVKIDAILFGGRRADTVPLVTQAFDWNHGTMIGSLLSSGQ TAAAEGKVGALRHDPMAMLPFMGYNAGDYLQHWIDMGKEGGDRMPEIFLVNWFR RGEDRRFLWPGFGENSRVLKWIIDRIEGNVEAVETVVGHTARPEDIDLTGLDFDIED VREALAVNPADWAGDMEDNKEYLHFLGSRVPSEVWDQFKALKERVEAAS
Other gene	CSING_12285	2616754 - 2617527	tRNA (guanine-N(7)-)- methyltransfer ase	257	MNSADSSPNPAAEMPHGRGFQTDFNTEFDNDLDYPRLGNVTFRRGTLTDNQNALF EEHWPKLGQMLSDVPLDIPSWFGHDGAKTIVEIGSGTGTSTAAMAPLEKDTNIIAVE LYKPGLAKLLGSIVRNDIENIRMVRGDGIEVLMRMIAPESLDGIRIFFPDPWPKARHH KRRIIQSGTLNLFASRLKKGGVLHVATDHAGYAEWIDELVRVEPSLGYKGWPWPEC PILTDRQVITKFEGKGLDKEHVITEYLWEKK
Other gene	CSING_12290	2617527 - 2618135	NYN domain	202	MAGTYLLVWDAPNMDMGLGAILGGRPTSAHRPRFDAIGRWLAELARDNDATPEAA VFTNVSPGGADAIRPWVEAIRNVGFAVFAKPKLSEDDDVDPDMVEHIKANRENLSG

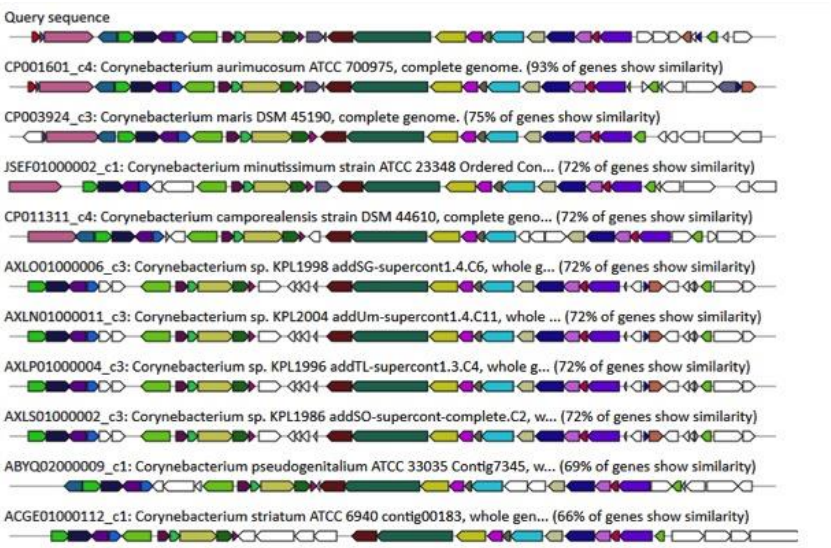
					VIVASADGQNFQHLLEQLAESGLPTCVLGFHEHASWAVTHPEIEFVDLEDIEGVFRE PLPRVNLDNLPPGGAWLQPFRPLSALLQRGPEHS
Additional biosynthetic gene	CSING_12295	2618144 - 2620486	putative RND superfamily drug exporter	780	MFYSWGRFSYAHRKAIPLVLVGLILVLFFGFGTRLSDRMSQEGWEDPGAASTSAAQ IELDTFGRDNSGDVILLFEDPDAHAAEATAFLEDLKSQHPGQIDSVTSYFDSRNPNLI TRDHSRAFAAIGLKGDGEQTLKDFRAIEDSLHASDLPVQVAGATAVADALDEGMAQ DISRAEKAALPLVGLLLLVVFGSVVAAFMPLVVGGLSILGSLGILSVLAGFLQVNVFA QAVVTLLGLGLAIDYGLFMVSRFREELDKGRDTKEAVAITTATAGQTVVFSAAMVAV SLSGLFLFPQAFLKSVAYGSISAVGLAALLSVMVLPSLFGMLGHNIDKWTVRKTSRR GRRIEDTWWYKLPRWAMRRAGLMTAAICALLIALTLPVLGVKFGGINETYLPPDNDV RVAQDQFNEDFPSFRTEPVKLVVKNATNEQLVDVVMQARSIQGLTEPMGPKTATV DGTTVLGAGIQDREDYGRIVHELENIEAPEGVELYVGGTPAMEVESLDALFETLPW MALYVVLATFLLMALVFGSFILPLKAIIMTLLSLGATLGILTLMFVDGLGSGLFNYTAGP LMSPILVLIIAIVFGLSTDYEVFLVSRMVEARRKGASTDEAIAAGTAHTGGIITAAALIMI VVAGAFGFSDIVMMKYIAFGMIFALLIDATIVRMLLVPAVMHLLREDNWWAPRFLTGL THRLGGDSPEPAAAQEPAVPEPTVAEPTGAEPAAAASPAPTPAASQAPAEEGASP SSANDAAVRGDRSADTDEELIPFTELMRRLNSTDKR
Other gene	CSING_12300	2620483 - 2621493	putative integral membrane protein	336	MNKKWITWIISLVILAALAWFFRDQLDFITEGLRRLRHAEPFPVVLVVVFALLSIAGMA EVMKRLIGAGGVNVPLRETYAITLASNAWSTTLPAGPAFSAVLTFQVQRRWGASVA LCSYFLFLSSVISSMFLALIGLGGVFFLNADMALGSLITTIVLMLAALGGIFWLTSHPVT LERWLRKLHPGATRDRVIQEVRNLGEVHLSRGPFAVVAGASLVHRLADMAALWAS VWAVTGEIPWLRAGADDTTIAGVALAFLAAKLAGSAQVTPGGLGTVEAALIAPLVAT GLTVVHATSAAIIYRLISFALITIIGWVIYFAHYARDGFSYAALNQKES
Other gene	CSING_12305	2621494 - 2621841	Protein of unknown function (DUF3054)	115	MRVAPLIDVLALALFAILARLAHGGLSLSTWFDAFWPWTAGAIVGWLIIMATKLGGL WKQGAVVWVSAILGGMALWVLVNGRLPHWSFLIVATVMSALLLFGWRAIATVTSRS RA
Other gene	CSING_12310	2622022 - 2623035	HNH endonuclease	337	MKLGDLLGVLARGMHVVSLCAGHTLDEMIALGATPRVARQLDALHRVYFGQTAFTA KQRRARNTNHALDTLLTIERHVARVKDSRKAWDLRVELCETPVGEIAAVARRRLAEL KPQPSPGVRVRRSATGMHSITITDRPRAIADMVGTLRATNDDLLKAAHDVFAGSGA PRPALHAHVIVRLDELDTIVRGEGDDIILQATDGGTMTGAEFLRTQFADRGFVSLFHP VEGPVNLYYGSRFANEKQRLLLSAEHPTCAWLGCTRPAAECQMHHIQRWQDGGP TNVANLVPLCQYHNAINDDDPTRPTGRGRVDRINGRVHWLPPGGGPPVPSPSPAW R
Additional biosynthetic gene	CSING_12315	2623309 - 2624856	acetyl-CoA carboxylase, carboxyltransferase component (subunits alpha and beta)	515	MTSTADKLADLRERLERAQDPGSERSRARRDEAGRSTPRQRIAALLDEGSFVETGA LGKTPGDPDAIYSDGVVTGHGRINGRPVCIYAHDKTVYGGSVGVTFGKKVTDIMDL AIKIGCPVIGIQDSGGARIQDAVTSLAMYSEIARRQLPLSGRSPQISIMMGKSAGGAV YAPVTTDFIIAVDGEAEMYVTGPNVIKEVTGEEISSHDLGSARQQELNGNVSVVVPS EDDAFDLVRDLLDHLPLTCFDEAPVFDAPSDEEVAQDTELDSFMPDDTNAGYDMM DLLTQLGDDDELIEIQENYAPNVITAFGRIDGKAVGFVANNPMHSAGCIDADAADKG GRFIRICDAYNIPLVFVVDTPGYLPGVDQEKAGLIHRGAKFAFAVVEATVPKVSLIVR KAYGGAYAVMGSKNLTGDINLAWPTAQIAVMGSAAAVVMIAGKQLEAAETPEQRA MTKKMFMDFYDETMTAPYVAAERGYLDGMIKPSESRLALRRALRQLADKQEKDLP KKHTISPL

Core biosynthe ti gene	CSING_1 2320	2624853 - 2629727	mycolic acid condensase	1624	MTVEELRGWLRTWVAQTTGLSAEEITDTKPLENFGLSSRDAVVLSGELENLLGIKLE PTVAYEYPTIAQLADRLVNGATPEAPQATSGQTQTPLIEGGDIAVIGYAGRFPGAKN VAEFWDMLAEGRPGTGPLPVGRWSEYSSDPITSEKMAQQNTDGGYIEDIASFDAE FFGLSPLEAANMDPQQRILLEVAWEALEDAGVPANQLRGTATGVYMGSTNNDYGM LITADPAEMHPYAMTGTSSAIIANRLSYAFDLRGPSINVDTACSASLVAVNQAVKDLR VGAADVALAGGVNILSSPHASIGFSELGVTSPTSAIHAFSDDADGIVRADAAGVLVLK RLADAERDGDTIAAVIKGSAVNSDGHSNGLTAPNPEAQVDVLERAYADAGIRPQDV DYVEAHGTGTILGDPIEATALGRVLGPGRQAATPTLLGSAKTNIGHSESAAGVVGLIK VIEGMRHGVIPPNINYAGPNRYIDFDAEHLEVVEDPREWPEYSGQKIAGVSGFGFG GTNAHVVLTDYRGLTRQDAPQVAVGEGQPVALPVSGLLPSRRAAAAADLADYIEAE DPNLLSLARTLARRNHGRSRAVVVASSSEEAIKRLRQVSEGTVSVGIAAADSPTVHG PVFMYSGFGSQHRKMAKDMIEISPLFKQRLEELDAIVAYESGWSILELVHDDSQTYN TETAQVAITAIQIALTDLLAAFGVRPAGVIGMSMGEIGAAYASGGITAEDAMLIACHRA RLMGEGEASLADDQQGAMAVVELSAEEIDALPGHIEPAVYTGPGMTTVGGPREEVL ALVEKLEGEGKFARALNVKAAGHTSAVDPLLGELYAEIAGIEARPLHTTLFSSVDRG KAYRPGTVVHDADYWIRMTRHSVYLQDATESAFDAGHTQIVEISPNPIALMGLMSTA FAAGKADAQLLYSLKRKVDPTESLLDLLSKLYASGAPVEYTKVFGSGALVDAPSTRF RRQRFWTNARPSSGVSGLPGARVTLPEGAVAFSTNADQAPSALALLEAAAEAVTP GAQIAASEEHLDLPAKGEVTTIVRRTLGGLSVAVHFVDGASTKLIAEGFASTLNLPAP AVTPEPTLAQAVTAPQFADPEVGVDAVRWDPAKETVEERLSLIVSESMGYDVSDLP RELPLIDLGLDSLMGMRIKNRVENDFQIPPLQVQALRDASLADVITMVEDAVAEQTP AAPAADASDATTPADATPAQATAPFALTESDDSAGKAATVSAKGGADELRGDNRG ADGAQNDGERDGEAEGVGVAPRDASERMVFGTWATFTGKAAAGVTSALPEVSDE VAAQIAERLTERAGIEVTAQQVHEADALENLADLVREGLETEVEGNIRVLREADGPA VFMFHPAGGTTVVYQPLTRRLPADVAVYGVERLEGSLEDRAAAYIEDILHYARGRK VVLGGWSFGGALAYEVAYQLQERAARGEDSAEVAFIALLDTTQPAHPAPDTLEETK ARWGRYAAFAKKTYGLDFDPPYEMLETVGEDALMAMLAEFLTSTDASEHGLSAGV LEHQRASFVDNQILGSLDMGRWASVSVPVLLFRSERMHDGAIELEPAYAEINPDGG WGVIVKNLEIVQLPGDHLAVPDEPAIGIVGKHMEDWIEEKIRK
Core biosynthe ti gene	CSING_1 2325	2630030 - 2631889	acyl-CoA synthetase (AMP-forming)/AMP-acid ligase II	619	MDLKALIGQFFDEKGNIALPPHLTLAGLAEHLYGAEQQAGVPDRPVMRQWVYNDN PEGTPRDFTRSQVNTRIKVVAARLQQVGKMGDRVAILAGNSPEYIFGFLGAMYAGM TPIPLYDPNEPGHTDHLRAVFGDANPPVVLTNNVSAAANRAYFSDRPAAERPRIISID SLPDSLAASWTNPLETEEGKAALAALKTPPVDHPAFLQYTSGSTRTPAGVLLTNRNI MTNVLQIFTAVQIKLPPRVVSWLPMHHDMGIILAVFVTILGLNLEIMTPRDFVQQPKR WVDQLKRQPIDAQLQGTYAVVPNFALELAARYGAPAAGEELDLSNVEGIIIGSEPVT ESAVNAFWETFGKEEYGLRRETLRPSYGMAEASLIVTTPQTPDRPIISHFDRERLAH GEAVIVEKSEDSVAYASCGQSVVAQDLTIVDPETRAELADGHVGEIWLHGENRAAG YLGRDEETASTFHNTLGKRLAEGSRVANAPDDNNWLATGDLATIVDNQVYITGRLK DLIVVAGRNHYPQDIEGTVQEASAHVRADSIAAFSVEGGSTEELVLLIERADGANPA EDEAASEVIRSAVSSHHGVTPAAIQWFNANEIKRTSSGKIARRVAKKEYLAS
Other gene	CSING_1 2330	2631979 - 2632968	Cutinase	329	MSGIPQGIPHLPKLDFPTMRKTFTVLAVVVLLAVIGGGAVHFLNTRNEGEPGVLVAP SDDPAQPGEEEPAQPDWCPRFEFVSAPGTWESAADDDPINPGANPNSFMLSITNP LKESYVPNDVKVWTLPYTAQFKNINAQQEMSYDESRDEGTSRLEGELTYMHETCP DTQFILSGFSQGAVIVGDVADRIGGGNGPIPASSVAGVAVIADGRREEGVGINPGAH

					VGGVGAEIALQPVSTLIQGIVPGATMRGARANGFGELADRTFQICAPNDSICDAPLD VSNGLDRAMDLIEANGVHAMYASNPHVIEGTTASQWVTQWAKDIIDAH
Other gene	CSING_12335	2632994 - 2633506	hypothetical protein	170	MRRLSHPTVAAALLAAGALLAGCGSATVDESEATETSVAPLERSASASSAASGASS ETSTSASENSPAGSSSRGTGDEPEDRGAREVSEIPAPAPDEGPNQHFLAAISQGGV NTDGVEDQLIGAAQASCQGDAVTVPAVAGQLIEQGRSELSHEQLTQLITEQGRALC AG
Additional biosynthetic gene	CSING_12340	2633516 - 2635516	putative esterase	666	MPNTAPSQKSAQPSTRAGLRSPSAARKGLTIAALPTAVAVGLALLPMANAQSSGSS SGEGLTDILGSSNFSDSFAPSNPPQRTPIETEYPDIEGLPEGVSVSRVEYLTNRHLK VYIKSAAMPDQEQVVQIQLARDWYSSPDKTFPEVWALDGLRARDDESGWTIETNIL SQYADRNVNLIMPVGGESSFYSDWQKADNGKHYKWETFLMQELIPVLEKAYRSNG QRAVTGISMGGTAAMNLAERNPWAFKFVGSFSGYLDTTTTGMPEAVTAAQADAGG YTSTNMWGELYSQDWVDHDPKLGIENLKDMKVYVSAGNGKDDYGQIGSVAKGQA NMAGVGLEAISRMSTQTFVDYAKRAGVEPVVKFRPTGVHSWEYWQFEMTEAWPFI ADSLGLDKSDRGADCTPVGAIADATKSGIIGTCLNNEYDVADRGKAQDFQAGTAYW SPETGAHAVIGRIGARYSELGGPTSWLGFPVTGEEKTPDGRGRYVHFEHGSIYWTP ENGAWPIPKDIFDEWGKNGYEGGDLKYPTSDVRKIGDGVVQEFEDGVLTHNPDGS FHIVHGAIGAKYKELKAAESDLGYPVGEEKQVNGGFFQEFEHGNIYWSMDSGAHYI LKGDIFDEWGKHGYEQGEFGWPTEDYKEIPAGGLTQTFQHGVIRKVMGTVQADKK
Additional biosynthetic gene	CSING_12345	2635777 - 2636808	putative esterase	343	MSALSSALRTRVVAVVAAIAVALGLVVTAGTQEASAANRDWLRGDATGACEWDRV GYWVQRCDVWSAAMGRTVPVQVQPAKNGGNAGLYLLDGLRATDRTNAWINDVNA ARSYVDHNITLVMPVGGAASFYADWRAPATYDLVKPVNYKWETFLTQELPAYLEGN FGVARNNNSIAGLSMGGTAAITLAGKHPEQFRQTLSYSGYLTTTLPGAQTFMRLALL DAGGFNINAMYGSVINPKRFSNDPFNLIPDLAAHGTDVFVSAASGIPGPLDQQRYLP QHIASGVVLEGFANATTRLWELKARASGVRLQTNYPAQGLHNWEQFGYELEHSKP QVLNVMNAW
Other gene	CSING_12350	2636935 - 2638695	hypotheticla protein	586	MRKPDSRTVAALSALASVLVIGVFSFWGGFARRWISDDGLIVLRTVRNLEAGNGPV FNAGERVEANTSTLWQYLILLAHWLSNAPLESIAVNVGLLLSVAAMMLGAWATARG LYRSHASAARGAAPLVLPAGALVYLALPPARDFFTSGLEWGLSIFYLAALWALLQNW AQSHSTPSAYWLALWAGLSWLVRPELALYGGLAGLVLLAAHRNWKVWLGILGAAL PIPGAYQIFRMGYYGLLTPHTAVAKSASESAWASGLNYLQDFAGPYALYLPVLVLAV AGAWASREHLRPTALRSPTAVTYVFLGAAAVHTLYVLRVGGDFMHGRMLLLPLFAA LLPVFVLPLKGLATWFAVAFCAVWALVVILRGHETDWDAYEQPLDIVDERDFWTYA TKREAGYPPRSARDYLTMKFMNGYPSAMENLEDGDALAVIYVEDKDADRLNWLTV PREEGRKDQPTLFLVNLGMTSMNAPLDVRVQDTIGLATPLAARMPREDNGRIGHDK SLALEWQAAESGADIDYLPPWYDKEKTRQAREALETPELKELVASYSEPMSWERF RANLGFSLTTGRTLQLSEDPSDYLPAG
Core biosynthetic gene	CSING_12355	2638732 - 2639733	decaprenyl-phosphate phosphoribosyltransferase	333	MSTRRVDDGEQKVFHSEPHTSGIDTGRKRKPPKNLADGMVKALRPKQWVKNVLVL AAPAAAGADALFHTRVLLDVLLAFVVFCMGASSIYLINDARDVEADREHPTKRFRPIA AGVLPVSLAYVMAAVLIVGSIGLSFLATAGPQLAIVMAVYIGLQLGYCFGWKHLPVIDI ALVSSGFMLRTMAGGVAAGIELSQWFLLVAAFGSLFMASGKRYSEILLAERTGAKIR KSLEGYTPTYLRFVWTLAATAVVMSYSLWGFQLSNLAEGSAGVWYQVSMVPFTIAI LRYAADVDRGNGGAPDEIALEDRVLQILALAWLACIVMAVYVMPSIGG
Other gene	CSING_12360	2639730 - 2640257	5'-phosphoribos	175	MPSKLSVAESHVLEKIQDVAFDAPGILPAARGLSHFGEHALGWMALSALGAAVNYS KPQKRRQWAYVGVGAFTAHAASVVIKRIVRRKRPDYDYVKVGVKTPSKLSFPSSHA

			yl-monophospho-decaprenol phosphatase		TSTSAFLVGAAHVLGTPAPLVGVPVMMASRMVLGVHYPSDTALGALLGAATAEACHRYERKSR
Other gene	CSING_12365	2640247 - 2642220	putative glycosyltransferase	657	MWRNEDNVTANNSAAYEPVQRILLPKRGEPVDVRMLYLIESEQNRERLSWNDRTSVTIPAGEEASFETYFNAFPAAYWRRWSQLKSIILSMDVEGEANISLYRSKQDGQRIAVANHVVTTGHHEFELPLKNFEDGGLLWFDASAVEDTTLVDAAWCAPHAPNPQLLPDGSEYPAQEKRVAVGIPTFNRPTDAVAALQALAEDPVVDGIIDYVLMPDQGNQHPADEPGYDDAVAHFGERFREFRQGNLGGSGGYSRIMFEALENTDSPYILYMDDDIAIEPDSILRAVQAARYAAKPIIVGGEMLNLQERSQLRTTGERVNRADFMWGAAEHAVYDHDFAKYPLRAIGTKESRLDPKKYDSRALHRRIDVEYNGWWMCLFPRIVAETNGQPLPLFIKWDDTEYSLRAAANGFPTVTWPGAAIWHMAWADKDDAIDWQAYFHLRNRLIVAALYHEGDPRGITRSIFKSTLKHTMCMEYSTMAIQLEAMKDFLAGPDRLFDILESSLPRIAEIRKGYSDAVIIESADQLPAPTGAPGVPTRNIGGRLGKLKKIPWLLKSAKHLVSKEDPAHHEAPQLNLTPEEARWFTLSRVDSATVSTAGGTGVAFRKRDRDLAKELVQSTRELLKEIEDNFDALRAEYRAALPELTSRESWRKVFDAQ
transport-related gene	CSING_12370	2642623 - 2643621	ABC-type Fe3+-hydroxamate transport system, periplasmic component	332	MHRISISLSAVLLAPLIATSCSSPEPQTSTHNANTHIIDNCGYKVSVNVPVQRATTLEQGATDTLLLLGAKDQIAGYGHQKDAPPEAFDLEGIRELSPGVPNSEQLRDADTDFIYSPFALSWTTDSAGAREEWEKLGVSTYQSNVECPNMGDNVGKSKFDLIERDITELGMLFGREDAAKKLIEKQNRALETATQAPEGTTFMLLYSSIGGAPYVAGGPSIITEIGEASGMTNVFGDLSEEWPQVSWEAVAEANPDVILLADLPKRGEPGDKWQEKVDALEATPGTKEMEAVKSGKYIVVPGVATSAAGRSYEILEIVSESISDDLFSNKAR
transport-related gene	CSING_12375	2643624 - 2644607	ABC-type Fe3+-siderophore transport system, permease component	327	MRKTWITAGVSIVVLLVTMTLSLTMGSSSMTLAEFRGYIDPAYPRPSELRTSILTELRIPRVLMAAFVGAILAVCGVVMQAITHNDLAEPYLLGVSSGASTGAVVAILFSTWQYGIIVGAGIGALVSFLLLMLLLRHSAADATKVVLTGVLIGFLFQALTSLVVTASGNAESTRGIMFWLLGVLGAARWHTLAAVIVVGTIGMVLMWILSRYLDALSLGSATVSTMGVPVVALRYTVLIAVSLMTAVTVASVGAIGFVGLIVPHAVRMLNTPRHSSLIPLSALVGAITLVVADAIGRVLFAPQELPVGVITALIGVPMFFFILQRKH
transport-related gene	CSING_12380	2644607 - 2645383	ABC-type cobalamin/Fe3+-siderophore transport system, ATPase component	258	MLEIKHLSSGFSGKTVLDDISFKVAAPSLVGLVGPNGSGKSTLLKAIAGVQPFKGTVVFSGKPLHSYPRTQRVRVLSYVAQHSDKDIPFTVREVVQLSRDAARAPFSPVTTEDKHIVDRALFHSSLEDIANERLTELSGGQIQRVMVARAMAQQASIMLLDEPTNHLDLHHQYTLMELLRHLSREHNTVIVLAIHDLALAARYCDRLLLIHDSHLVGDGHPSKVLDPATLATVFGVQGKLSTATDGVPLLHIKHALNS
Other gene	CSING_12395	2646317 - 2646616	transposase	99	MPRYSEQFKRDAVALYENNEDLSLHAASAELGVNRSSLFSWLKRYGTGKRARTKAVHDHAQVTTESERIRQLEKEVSKLREERDILRKAAKYFAEETRW
Other gene	CSING_12400	2646971 - 2647579	hypotheticla protein	202	MTFPNARRFDTLRQDLLADYGHQCSAEEINSFLDAAIERHEANATLDEFVPVMVEREVREHFGNHRIHVRFAAGNDNALAQAAVALTKKYAGSALLVDAAVTHPENEADSHMAHVLQERGIAEHPHRYLEDLRLLTIPDYIVYLGRDIPRDEAGKDIKIWDISRAKTVEETRELADDLGARVHYMLNRLGIEPISEDATVNV

Other gene	CSING_12405	2647979 - 2648248	hypotheticla protein	89	MKLWQKVSLAIFGVCFLMFTTTALLGDDPLMWFYPIVALVGIGITLYDGPGRKEKRCQEEEARARADREELARRKQMERRERRKRGKGK
transport-related gene	CSING_12410	2648635 - 2649786	DMT(drug/metabolite transporter) superfamily permease	383	MANNTMTAGVSTPPHPSLASNPVKPFSHGKAVAILLLGTMFLSGTPLWVKGSNMDPATQAFLRVSIGFLLLLPFGIVEMRKKMVLPKKGIMFSCIAGLFLGLDFTNWNFSIFLIGSGVAAILLNLQVVIVPILTTIFDKYKLPKSFIVILPLMIVGVLFTGGVFEPAAELAGPAEISGIKTSVLGTACGLTSGICYSFYLYFSRKAGVTAPRNDIYVQPMMFTMLAQCVVPTIIMIIGNGFNITTGVLVEDANGQMVLPELPEGVGLFEAVSQGDPIDGGNWVHLISLIVLGQAMAWTFVQMGTVWLEPVLSAGILLLSPVTSVIIAWPLFGEIPSWLQWLGVVIILGCVMFQNGMFDPLLGKKKKEEPVVAETVNEAAGPAH



Appendix 90: Type 1 PKS CsinA biosynthetic gene cluster and its predicted similar gene clusters according to antiSMASH.

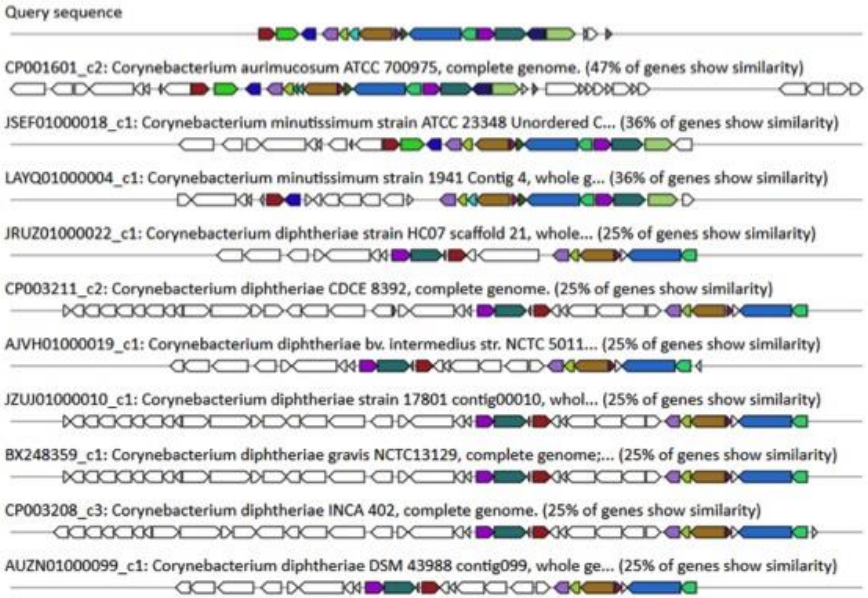
Appendix 91: Summary of genes in BGC possibly coding for terpene CsinA.

Type of gene	Locus tag	Location	Function predicted by antiSMASH	Amino acid sequence length	Amino acid sequence
Other gene	CSING_10935	2300058 - 2300876	NH(3)-dependent NAD(+) synthetase	272	MVVMSHESTDLQRSIIKALGTKPLIDPAEEVERRVTFLADYLRDTGAKGYVLG ISGGQDSTLAGRLAQLAVDRVEGTHFWAIRLPHGVQADEDDAQVALDFIQPD HRLTINIADATSALDGAVASALESSTLSDFNRGNLKARLRMAAQYAVAGEVG ALVIGSDHAAENITAFFTKWGDGAADLIPLEGLNKRQGAQLLEYLGAPESTW TKIPTADLEDDRPLLSDEEALGVTYTHIDDYLEGKAVPADARERIETLWRRGA HKRIMPPGPQ
Other gene	CSING_10940	2300925 - 2302043	HNH endonuclease	372	MGVLETYFHYRNSGIALVENASSSPDELRALGADASDAAELAHLHRTYFGPTR FSGKQRKARTAAQTQKHSLATLTLIESYTAKVKKDLDAWNLRIKLAGTPAHKIR DIAVKRLKELRNKRTPKPGVRFTYRATGPNSMTITDDASVIADIRGTLESVNKH NLLEAARTVIVTGSVGTKPAIHAQVVVTLDEFDRIINGDGDEIELQLTNGARM TGAEFLSYQFASIGYATIVHPVEGPINSYRAERSASWKQRLTLAAETQTCSRPG CNKPADYCHVHHIIPWQAGGYTNINNLTFLCAYHNSINDDDPERPTGRGYVF RIPGGIGYIPPWGLPITATPEYQQAVARLTAEKARAEQTAEEPPPDPPPSTG
Other gene	CSING_10945	2302102 - 2302836	fructosamine-3-kinase	244	MLGCMDVYTKKVSRPDQAGAEAAGLRWLAAARADVVVEVAGVTETSISTRR VMESRPSLKAARSFGAALREVHEAGAPAFGAAPEGWEGANFIGRVEQPCQP CESWGEFYVEQRVLPFVETLPKDLQDVVRVASDAIRSMDWEVAPARIHGDL WAGNVLFTAESAVMIDPAAHGGHPWTDLAMLELFGAPLYEEILRGYGVPKE YERWVPMHQLHPLAVHATTHGPAYYRPLGKAAEATIDALR
Other gene	CSING_10960	2303136 - 2303888	protein-disulfide isomerase	250	MSTVKDPNSKGNSGFLWGLAVLLVIVAVVIGYIVYQGRGAQTDALGDYAAED VSMDMSLADNAVTLKSPDAAKDAVEVELYEDYSCPHCGDLAKETDGQMKD AIDEGKLIVHVRTLNFIDGSQNGLESIKSNKGHSSKAAAAMEQVAKSGDATLY WNMRKYLLENQSKVYNKWELNDFADAAKALGADQDIVKSIQEAKVGEGNEL ATANYKKLDADTGQISSPRIIKDGKDIPEDKNTSIMEWVDLVVE
Other gene	CSING_10965	2303943 - 2304410	DoxX protein	155	MAQKKISVALVLDIISAFARFYMAYVWIKAGLSKFSDQLAVTQSIEAYEIFTPE WSHYLSYLIGPLEVCGGLLLLLGLFLRPAAWVGQVVLTLFMIGIAQAWLRGLGI DCGCFAPNPNDDAQVMNYMMTLLRDLFYSVLMVWTIKRPFTKFALHP
Other gene	CSING_10970	2304455 - 2304916	DoxX protein	153	MASLQLARVADIISALARFGMAAVWIIAGVQKLNAQMDMAQAIQAYEIFTP EWSGYLAVAIGPLEVMGGVLLLLGLFLREASAVGAVVLVLFMVGMAQAWAR GLVIDCGCFGYTANSGSQGMNYMLTLLRDIVFLALTMWTIRRPYRRFAIYS
Other gene	CSING_10975	2304940 - 2306556	phosphoglucomutase, alpha-D-glucose	538	MAHERAGQLAQPSDLIDIAELVTAYYTRTPDADNPDQQVAFGTSGHRGSAL DTAFNEAHILAITQAIVDYRTDQGTTGAIFIGRDTHALSEPAMVSALEVLLANE

			phosphate-specific		VEVRVDDRGRYTPTPAVSHAILTNPGTDGIVITPSHNPPRDGGFKYNPPTGGP ADTDATDWIAAKANEYLRAGLEGVKRVPVSGVLDERCVKHNYLDTYVADLEN VVDMKAIKEAGVRIGADPMGGASVDYWAAIAEHYGLAMTVVNPEVDGTFR FMTLDTDGKIRMDCSSPNAMASLVDNRAKYDLATGNDADADRHGIVTPDA GLMNPNHYLAVAIEYLFSHRPQWGNAGVGKTLVSSSMIDRVVANLGRELVE VPVGFKWFVPGLVDGSLGFGGEESAGASFLRFDGSVWSTDKDGIILDLLAAEI LAVTGTTPSQRYAELAEEYGAPAYARTDAPANREQKAILKKLSPDQVTATELA GEEILAKLTEAPGNGAAIGGLKVTTENAWFAARPSGTEDKYKIYAESFRGEEHL KQVQEEAQALVSQVLGS
Other gene	CSING_10980	2306636 - 2306947	Integral membrane protein possibly involved in chromosome condensation	103	MPQTILVGCGAALGALARFALSTILGGGALPLLVINLVGSAVMGYTKPSAFWG TGVLGGFTSFATFAFLTSGFDPAQAGAYVLATVVGCTGAYLLGDVSRRPA
Other gene	CSING_10985	2306944 - 2307306	Integral membrane protein possibly involved in chromosome condensation	120	MIIISALAVLVGGFLGGCGRFALTRILPSPTCTFAANITGAAVAGVAYGYAAHT QAPEYLLPLLAAGLAGGLSTWSTLAKELGEMIKAKRWWRLARYLFWTVGLGI VVAWRGAWVGSLLA
transport-related gene	CSING_10990	2307323 - 2309869	putative ABC-type transport system involved in lysophospholipase L1 biosynthesis, permease component	848	MATKNSMRTVSVRNILAHKLRLALTLLAVVLGTAFISGSFMFTNALSNTFDSA VDTAFTGVDAAVSQKEGGPVLDSKMREDIAHDPEVRAVNVQSSQTVVVAN KDAEAFQTGGGTSSVQPYYSADQSVGEPAELVDGSEPHGTGEVVINDSAAEK FGITIGDTILVVHPDQRDEVKVVGVVKPAVDQGASLTLHMDHEGFTERYGDS SQLKVAAAEGVSAEELVDHLNETFDVDAEAGEELAKEISDQISSALKFVNYFLIA FGLIALLVGTFIIANTFSMIVAQRTKEFALLRALGASRGQITRSVVTEAIIVGIVGS AVGVVAGVGLVAAIKAVFTAQGMPMGGGLGLSLSAVIVPLILGTVVTVISAW APARRAGAVEPVEAMRTTESAAGSSLLVRTVLGALLLVAGIAFAVLGVLIDAKT GLRAALVGLGSLNLIVGFFLAGPAISLPIVPSIGRVVGAPFGAVGKLAATNSAR NPRRTAATAFALMLGVALVTAIGMLGATMKASVSDVIDEDVRADFLLTGPTT GNFPTPNETVDRARDTEGAGQVVALSSAPIMVDGQASMNYGPQVAFSQVI GSNADDILNLTVAEGSLKLGDESGFIADTDLAAEQGWQVGQSYELTGMDPN RTAQVTLIGTYEPFQMLPRLAVSESAATEVINPEGLDTMMVAVNAAEGFDKE KLRSNLEESVKDLVVVQVRTGEEYAGEAAGMIDQMLTILYALLALAVIIAILGIV NTLTLGVIERRQEIGMLRAVGTQRRQIRTMITVESVQIALFGAIMGMLLGLGL

					GWAFISVLADDGLDSATVPWPMLIIMLVGSAFVGVLAAIWPAQRAAKTPPL DAIAD
transport-related gene	CSING_10995	2309873 - 2310613	ABC-type antimicrobial peptide transport system, ATPase component	246	MTTPTMTAAARAVDLFKQYGQGDTAVTALDHVNVEFARNAFTAIMGPSGS GKSTLMHTMAGLDSATSGSAFIGDTDMSQLSDKDITALRRDRLGFIFQSFNLV PTLTAAENITLPTDIAGKKVDKEWFDEVTTRLGLAKRLEHRPAELSGGQQQRV ACARALVSRPEIIFGDEPTGNLDSNSSAEVLDILRTAVDKDGQTVVIVTHDAKA ASYADRVVFLADGRLVNELHNPTMESIHAVMAEIEG
core biosynthetic gene	CSING_11000	2310686 - 2311555	phytoene/squalene synthetase	289	MWSKIWAMRASLSALERYDRSAVAAAAQVMRHYSTSFSLATRLLSPQVRVD IRNLYAMVRTADEIVDAGLDPHDAAALLDAYEATVRHSPSQRFHTDPIVHAY ALSARRCGFKDDHVAQFFSSMRSDLTRSTHTQESFEDYICGSAEVIGLLCVDAF FSGGPRPEGVDNGARRLGAAFQKINFLRDLHEDSAELGRSYFPSLPDPTRLDE EAKAAIIADIRTDLSEARTATALLPRGSRIGVAAATALFTELTDLLEATPAADIME TRVSVPAARKAILVSRAAAGALR
additional biosynthetic gene	CSING_11005	2311552 - 2313060	four-step phytoene desaturase	502	MRAVVIGAGIAGLASAALLLREGWDVVVVDKQSSVGGRAGKLEVDGFRFDT GPSWYLMPEAFDEFFRACGTSTERELDLVDLSPAYRVYNEDGEHLDVESGVD EVARLFESLEPGAGERVRTYLAQASEVYAIALRNFLYTTFSRPTELLTRDVLSRL GTLALLLTRSLDGLVAARFQDYRLRQILTYPAVFLSTEPKAAPALYSLMSHTDLV EGVRYPQGGFAAVVAAIARQASDATFRLGTAVTAINCAHGRATGVTLDDGHI IPADAVISAADLHHTENALLPTSARTYPECYFARRNPGLGTVLAFLGVKGALPE LAHHTLLFSKDWTPDFRAVYHGPEASRPLGASQSIYVSKTSATDSSVAPDGYE NLFILVPVPADESYGHGDAYGPEESARVSAIANAAVAQLGSWLGIEDLEGRVV VKRTLGPSDFAEQYNSWSGGAVGPAHTLRQSAFFRGRNVSRKLSNLYYAGGT TVPGVGVPMCLISAQNVLTRMREQR
Other gene	CSING_11010	2313057 - 2313950	hypothetical protein	297	MLSLYRDFPDDEILQATLTNEAAEFLFIAPTAPQRSESESPYSEDRTHIQLLSMG GLALRFLPNGELITGDIDVRYPQLFQTSGGKTIWDPRELADLEPFTVYRARGAR HFHERSQQHYVDIWEILPTIADPAMDALVAKLAAPVILNSDHGPLVLDRASNT FDGRAEDLGVDISIVYEGEELLLADTKKNFKAQLKIINKVVNAGFLDKARRFAA KKALAAANRWRTDVAQAEGIDAPIPELSVEDVYRAVTPVAVEVPQRGHIIVEF DDGYLFGGHPVTVTTRRNGTPVGCELDA
Other gene	CSING_11015	2314026 - 2315390	Mg2+ transporter MgtE	454	MADTTEQASEVVEAWLKQDDAIDPKKAPKLQKLLAEVPRQELIAIVERQNAV RAALALRLLPRAQSIAVFDALDAKHQADIIDELGNTDVYEFFDELDPEDRVALL DELPAEIADRLLRSLTKSKRDVTGVVLGYTKGSVGRRMSPEVPEILPSMTVDD ALSTLRDTADELETIYTIPVTRSDRRLVGVVSLREIFVAQSGVLIESIMQDPIYAH ASDDAEETVRWFLPLDILALPIVDDSHRLVGLLTWDDATDIVEEADSEDSARA

					GGTEALQQPYLSTPLLKLVRSRIVWLLVLAVSALLTVQVLDSFEGTLAKAVVLSL FIPLLTGTGGNTGNQAATTVTRALALGDVRTRDLIAVMWRELRVGMLLGAVL GSAGLVLATLVYGLDIGIVIGSTLFLICSISATVGGLMPIVAKTVGADPAVFSNPF ISTFCDATGLIIYFLIAKSVLGI
Other gene	CSING_11020	2315843 - 2315986	hypothetical protein	47	MRHFDLEVVVTRWTPKDVKDAAIERVAAGEAVSLVARDVGVGPDILS
Other gene	CSING_11025	2316014 - 2316472	hypothetical protein	152	MRRMMARIGFTSIVSLSLSSCSFNDPLASHLASLGRSGHAIPVTESITLEELYGP EWTEFSLVCPYTPYQTVKETLMVDDPPVPEHGFDDHENFLFLRGKGDSEQW VRFSRQALDLCSPSSPPSVIMESTTVALSFSLDNPGSAWTLTGLNK
Other gene	CSING_11030	2316920 - 2317168	hypothetical protein	82	MKKPLALIGATALLVTAAPTANAENSTAHNTNQGSTTDLNELVVDSPIYGIIAI PVTLSSIALSFLGIPQCGLHDTRGCKRG



Appendix 92: Terpene CsinA biosynthetic gene cluster and its predicted similar gene clusters according to antiSMASH.

Appendix 93: Summary of genes in BGC possibly coding for bacteriocin CsinF as predicted by BAGEL.

Name of gene	Function	Amino acid sequence length	Location	Amino acid sequence
orf00001	Trehalose-6-phosphate synthase	164	2410931-2411425	VQLATPSRERIDHYRATRSKVEEAVGRINGRFSQLGQPVVHYQHRSVPKDVLRRYYRLADVMLVTPFKDGMNLVAKEYVASHGDGSGALVLSEFAGAAAELTQANLCNPFDIESIKRALLSAVKALDESPELMREHMQAMHRQVHEHDIVLWSESFLGALEEEA
orf00003		166	2410431-2410931	MRKLAVMAAVLCLATLPGCSSDDAPAPRGDDLIVGKQWQVTNLYTTPEARSAIAPDTAQVPHMSFGEHTVVGSTGCVPFTAEVSFTEGEKDSTIWDADGMLVEKVDYESVAAGEECTGSAQWADKLLRSLIAQGHEFDFTLNQNNQLVLTLRTDKVDSPIIRMASL
orf00004	Trehalose-6-phosphate phosphatase	247	2409671-2410414	MINALASTQHLAVVSDFDGTLAGFARSPYAVTPEQRSLDALAALAQLPNTTVAVLSGRHLDGLKRVCPLREPVLMGGSHGAESSWEDSSLTPEMEAHLARKEAEIKEIIARHPGAELEVKPFQRVLHLRALELQDPAAAAAAYEEGAALSADGYPKTTGKCVVEFSATQATKGTWIAGLRERVGATAVVFLGDDTTDEDGFAVLNQPPDLGVKVGEGATQAAHRVPDIAAVSDFLEQLATARGRALH
orf00006	Catabolite control protein A	388	2408508-2409674	LSNDNELKKGKVTGMSTRSTTLASLAAELGVSRTTVSNAYNRPDQLAPATRERILAAAAARGYTGPDPTARSLRTRRQGSVGVLLTEHLSYAFEDMASVDFLAGMAEASAGASTTLTLIPAGPDTVGAPDDAHAAQLVGSAAVDGFVVYSVAAGDAYLEAVRRRGLPVVVCDQPTDSGLPFVGIDDRAAIALAARALVRASHTRIGILAIRLHRERLDGVVSPEQLAEADMHVQRSRVVGALDVFADAGLPPESIPVVTRHINDARTAYAAAEELLNAHPELTAVLCTTDSMALGVLAYARDHGIAVPEQLSVTGFDGIAPALAVGLSTVVQPNKAKGAAAGHMLADLIDRAASEPAPRAASVKREESPRKVLRTSFHEGRTVAPPNV
orf00007	Metal ABC transporter substrate-binding lipoprotein	381	2407340-2408485	MGNMTTPLSRRAWRRPTALLFAATLGCASLTACSSSSDNANGSDGADAENDTIHVVASTSIWADVAQAVVDTAKTDVNVDVEAIVTGNGVDPHHFEPTAADIAKANEADVIIAGGGGYDAWLYQAVKNQDNIIHALPLTEHNHDHDHEGHDHEHAHEEGHEHEHAHEEGHEHEHEEGHEHEHAHSDEHVSMDEAKKIAAEHPEKITNIEGNEHVWYDAAAIEKVATEVAAMVNKTNPDAKATANPLLKRVEKIKARVEKLPTKNYAQTEPIADYIMKYTDSTDVTPEGYRKATVAEGEPSAADLASFLEAITNGEVDLLIFNPQTETDMTQRIRQAAEDKDIPVVEIGETPPENTNFLDYYEEAVSALEAA
ABC	Zinc import ATP-binding protein ZnuC	233	2406638 2407339	VLVSFHNAAVEPLWSELNLAVDRGEFIAVLGPNGVGKSTLLGTILGTRKLTTGTVDVDCRVGFIPQQRMFPADLPLRARDLVSLSLAHGAFRRRRPPHHRVDALLEEVGATGIADRRVGQLSGGQQQLIRQAQALANDPELILADEPLLSLDPGRQQATVEKLNAWRHERGTSILFVTHGINPVLGVVDKVLYLAPNGHMFGTVDEVMRSEVLSELYGSSVKVLNVDGRLIVV
orf00010	Manganese transport system membrane	312	2405707-2406645	LSKFLEDTQYLLGVDFVQDALIACALLGILSGVMTSLIVLRQMSFSVHATSELALMGAAAALLFGLNIGFGAIAGAIVAAIILAVLGLKGQQDSAIGVVMSFGLGLSVLFLHLYPGNANTAMTLLTGQIVGVSSASVWLLAATTVIIVAAVTILWRPLLFASADPVMAQAAGVPVKALSVVFAVLVGLTAAQSVQIVGSLLVMALLITPGAAAVQITSSPLRAVVW

	protein MntC			ATIFAEVSAVGGFILSLAPGLPVSVFVTTISFVIYLVCRLIGWLRNRRAVRDDIAAAR VTKAHFEGRSDSEKHHTSTHDEHCAHPD
orf00011		266	2404895-2405695	MTIVEFQGFKRKYTVVAPEDPGPNQLLFLHGSLQNGSVMRRFTNGTFDELAART GTVVVYPDGVDRHFNDARGVLPVKARELGIDDVAFLQHVAATVQEEYGTQRTYA CGYSNGGQMVIRLLFDAPGFLDGACIFASTMGSGANHAPSNPEGYKPTPILMMH GTADPLAPYEGGQAGLANSSRGEVTSALWTAQRFATMNGCSKAKVTRPFSDVS LHSWEGDNPVELWTMEGVGHVIPSGNGLDARLGPNTDSFIAAQVVEDFFGF
orf00014		198	2404156-2404752	MPCLLAGIRDDEQPILDELRRYFLWPQVECASTQRELLNALSRSQEQLLVLCERK DSATWVWDIQELKSSDFIYLVDYSQHSREKLQKAINLGASAHCVVTREARSGHG GLKLSCCFDINGVAVGTMPEPDMQRNLERLSAIQLRILNGIVCGHSNNGIAEETHL SVPMIKKIVSQLLDIFAAENRVQLAIKTLPAMG
ABC	Macrolide export ATP-binding/permease protein MacB	208	2403506-2404132	MEIQVEDLHYSVKGKMILSGLTRSFAAGRVTALTGVSGSGKTTLLNLLGGLISPDK GSIMWDGNNIASLSAGRLRKYRRECVGYLFQDYGLVPDLTVVDNVKLAASNVSR RAFDEALDSALAQVGLRGYEKRVVSGLSGGEQQRVALARILVSKPKIVLADEPTG ALDEDNAQLVVEHLRAFADQGAVVVVATHSKHVADQADCEVTL
orf00016		661	2401517-2403502	MKTSVRLLTVGILLSTVISAFIAFVGASLLSTLDYRLGNEFVQIEDFKPDSHTETLFP ELSEFAGDSGFDVSIYYSSLADSDGELRIYSSALPSGVASFESQKFQRGEKLVWY PLSEIPTKEPRQVFNIKGSENDARDLTTWLEERGYSVIGLQNLTWSFVMTSSLPL LVAAAFLLALVLGGGHCLVRAREIGIHRLLGLKLSQTLRLDVTRLAKQTWLPFFLAI IFIAAALYLYNDWAEATLFWAFFGVSVLLQWVGLALGYVLGQLLVRNISIPQAIKG KVRARPITYALYAVRLLALLTSLTALSQMVDTGAEVSAREELQPYWSGHANEQEF ALSSNITPDEKEQAAAAAPLRQADREGKLLLADPFWLTWPTQLEAPVILSNHRFA VQAGVSPSEEDVTVCAPVALSEESKQTIIDSVAFEAELAKKKVPDFNWKEDCSLG QVFSYDVELRPVINDPIVVSLPVGLEAIGDRNLLAKASQKVLLAIEPTVVGELKDGT VGKVLATAQPHADSWRSSLEEAKDQFSLWSLNTVVAILLVSLLIVAGLFSFQVAF RRELYVSFICGRSPWKMCRKILLAELFFFVTTIGWVLYRLREHKEQLESHYPSTW SMGFENRWSTATVLAVVAFSIVWMCISIGLTWWISARNQMRWTDSEGN
193.2;putative_bacteriocin	Lactococcin_972; 193.2;putative_bacteriocin	47	2401217-2401510	VPHPIVKVVRPNLATTRTVVPRAAIDIYNRGPSNTCTTQRQQTCESS
orf00024	Uncharacterized tRNA/rRNA methyltransferase MAB_0572	315	2399960-2400907	MARTHGRGGAGIRKSNKKGATKGSGGQRRKGLVGKGPTPKAEDRVYHAKHKA KLERERRNSGRHQKETAEMVVGRNPVIECLHAKVPATSLYIVQGTRNDARLSEA VAMCNTRNIPILEVQRHEMDRMTGNGMHQGIGLQIPPYKYAEVHDLMDRARDA QEPGMFVILDNITDPRNLGAVIRSVAAFGGHGVIIPERRSASVTAVAWRTSAGTA ARLPVARVTNVTRTVQDFQKAGYQVVGLDAGGEHTLDTYTGGKDPVVIVIGSEG KGISRLVRENCDVIMSIPMTAWVESLNASVAAGAVLSEFARQRRAAK

orf00025	Cysteine--tRNA ligase	468	2398498-2399904	VSTLRIFDTATRTQRDFEPVRPGHASIYLCGATPQSQPHIGHLRSGVAFDILRRW LLAQGYDVALVRNVTDIDDKILTKAAENNRPWWEWVSTYEREFTKAYNTLGVLP PSVEPRATGFVTQMVEYMQRLIDAGFAYAVPGEGDAGSVYFDVAAWNAAEGSD YGALSGNRVEEMEQGESEAEGKRSPQDFALWKAAKPGEPSWPTPWGRGRPG WHLECSAMSTYYLGSEFDIHCGGLDLQFPHHENEIAQSHAAGDKFANYWMHNH WVTMSGEKMSKSLGNVLSIANMLEIVRPVELRYYLGSAHYRSVLEYSESALTEAA AGYRRIEEFVAKFEGVEKGEWTPAFEEAMNDDIAVPKALAEIHTTVRAGNKALAE GRREEAERLAGAVRAMAAVLGFDPCDEQWKDGGETRGADAALDVLVQAELER RTTARAEKDWATADAVRDRLTAAGITITDTADGPTWSLAD
orf00027		244	2397730-2398464	MTVRMGGLGQHLTAEFFNWVHAVHVKVLPLNEDSDIRVWTDLNPDDGTGLLEL HAPLVYGVGSQGGDRFSYSDLFESVVRFHTSLSPDESSLSVKSSKFELRVGTAP GVRFEYERENRTAPTAHIHHSGVAGLLSPALMKNFSGVKNPRKKGRIEDIHFPIG GRRFRVSLEEYFFFLFMECGFRSRPGWKSFLEQSREKWLDDQLRGAVADNQEV AAQVLREMGYTVERKGTAPSAEARGHRPW
orf00028		62	2397541-2397729	MRIINTEALRARYGELLHALEVAIGANLSSEDLRMLAESGRFSTEERDLYDELRRV ELMLER
orf00029	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	163	2397010-2397501	VTNPIIPRVGIATDAHQIEPGKECWIAGLLFEGADGCEGHSDGDVVAHALVDALLS ASHLGDLGSFVGVGRPEYDGVSGAQLLRECRELLGKEGFAIGNASAQLIGQTPK MGPRREEAESVLSEILGAPVSISATTTDRLGFTGREEGRAAVATAVVWRAPNS
orf00031	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	220	2396355-2397017	VPKAYVELRGRTLVERSVQAMVTSGIVDEVIVLVSPAMEDFAARILERSAAEIPVR LVHGGGERADSVWAGLQAIPDEDAVVLIHDAARALTPPGMVARVAKRVLDGATA VIPVVPVADTIKEVEGETVLSTPDRSRLRAVQTPQGFNLAALRRANLDYWEQNPD FIATDDASLMEWHGARVETVQGDTFAFKITTPIDLVLATAVTDEAEPTIFEVPSD
orf00033	RNA polymerase-binding transcription factor CarD	194	2395712-2396296	MEFKVGEVVVYPHHGAAKITAIETREMGGETLEYLVLQINQSDLVVRVPSKNVEM VGVRDVVGKEGLEKVFSVLREVDVEEAGNWSRRYKANQERLASGDINKVAEVV RDLWRRDQDRGLSAGEKRMLAKARQILVGELSLATPVDDKKADTMMEEIGATIE RHRAAGLVDDKSITTDVDSDIDLDDLSFDDED
orf00035	Uncharacterized protein Cgl2664/cg2949	191	2394854-2395429	VQSLKSAARCGAIMSVTALSALALASCSAGHVAQTAEKVAAVDGAAASTEDGKV TVRDVTIQVEPDSGETSLKFIAVNQGYKTDEVTLDSVTVDGQDVVLEGATPIARD QALYGDSATNLENVPKDEKDSNVTYVTTSLENDGFAFGGSRPVTFTFNTGTIELD ATVSASPLQSGEFDRDSQSEEGYTESK
orf00038		75	2394462-2394689	MKKIRRSLIALAAASAIAVAGTPAASAATYPEPLNQVMDKMIEYGGPQAPNLLLVP ATLSTIGWLQSIIASSSVR

orf00039		69	2394210-2394419	MKIRRTLVAAGAAVAIALAGTPAASAAQAPQPLADVLTQITNVAGYDAQNVILVPAALSSLGFISNVLF
orf00041	DNA repair protein RadA	482	2392623-2394071	MGVAKLFADRTTKSSVSEWLTRVWGMAKKVRPVHTCSECGFVSPKWLGRCPECGSWGTLQETAVAQESSAAQAAVTGRMPKGLTPTSPALPITKVGAAQTKALNTGIGELDRVLGRGIVPGSVVLMAGEPGVGKSTLLLEVASRWAQLGRTALYATAEESAGQVRARAERTGALHETLYLAAESNLDVVFGHGEQLKPSLIIVDSVQTMHAAGVEGVAGGVAQSRAVTAALTTLAKTTGIPILLVGHVTKDGNVAGPRVLEHLVDVVLNFEGDRQSSLRLLRGLKNRFGATDEMGCFEQTAAGIREVADPSGLFLSHHGSTPDGSAVTVAMDGVRPILAEVQALTVDPVAKNPRRVVTGLDANRVPMVLAVLQARAGERTNDKDAYVATVGGMKIQEPATDLAVALATWSSLHERPLPPKTVVIGEVGLAGEVRRVPNLDRRLAEAARLGYRHAIVPPGDVEVSGIRVRHAATLSEAIAALN
orf00042		247	2391865-2392608	VTEPTHTPKPLPREIYVRRRLAALVAILVLVVILIWAATALFGGSSEEKEEGSQPAAASSSLAQATTSQAEETATETEEATSESADEKPAKDGADNADAADKPEESQPTELAVDPSKTECSLEDLKVMAKTNEDSVPGGKMPTFYMEVFNPTTKDCSINLDEQQLRFEVYKMGSNERVWSDTDCYASVETGKQTFKSGETRFFQADWSRKRSEPGKCTDRPEADPGAYFLHTVIGENYSAALPFNLL
orf00043	Carbonic anhydrase 2	156	2391304-2391774	VLACSDSRAPIEHVFNIGFGDAFVIRTAGHILDSAVMASLDYALENLNANLLVVMGHQSCGAVGAASEFLAGGDLPTGLQRPIIEQVATASLVAQRDGREERADFERENTAQTVSQIISDIPAARELLDAGTLGVVGLRYLLEDSSVETVVLHGVE

Appendix 94: Domains involved in NRP CsinA BGC as predicted by PRISM.

ORF	Start	End	Biosynthetic assembly	Start of domain	End of domain	Amino acid sequence
orf_223	229661	230447	Thioesterase	65	119	MTEISCEGVFPFVAGHRTGTLLFCFHHAGGSASVYRGWVGVNPSIDVVPVE LPGKATRRREKWVSDFDKLADMCATEIVDLAAGAPIALYGHSMGAALAYQV AACLQQMHRPTIPEAVVAARQAPGETVPGEYHSSMGFGALRREFEKVGGT PPEILANDDVMKLLLADIRRDYVLHEGFHHATTVLRSPVLALAGDSDPAVSPE MLSRWENFTSGSFELKVLPGGHFFPLDTGTEFLDLLAGFLAGITGNTAWLAR NNA
orf_229	236907	244479	Thiolation	9	73	MIKEEQIREELLASLHQILGEDAEIGIDDNLLSHGLESLPTVRLLADWMKQGH RVSFGDFMRAPTVRQWAKMLVESTPNHSTDAIESPKDGFAAPIDDSVPFDL TDVQYAYWIGRNSSQQLGGVGTHGYVEVESRSINIDRLQQSWLTLLRSHPM LRACYTEDGKQYVLPEPPHPTILVHDLTKMDESTREEALLSTRERLSHRLLDI ATGHVVSLEVSLLPQDVAVIHFDIDLLVCDVQSFQIILHDLAHHYATGEAPDA DPSWSFARYLAGHAREGVADIDRDQAYWRNRLSELPGAPTLPMSHGTNEE QAHRFVRRSRSFDSATWSRLREVCEHHATTPAMVLLTAYARTIGQWSENKK FLMAVPLFNRGSDTAIKNVVADFTALTLTSIDQSTRRTFSEDLKDIQASFYED SSHSQYSAVRVLRDLRASRGEQVLAPVVFSCNLGDPLVGQEFIDTFGEISYM ISQTPQVWIDLQVFTTVNGFLIVCDAVEQLFPEKMLDDLFATLVMEIDKAITDD LSHSDPVESPGAQARRASRAEVASWRLPDTTLVDEVIAAARCHPQATAIRSA SGDVITYQDLEEQATTIASALVNSGVGRGALIAVMVERGPRQIIGALAAMMA GGAYVPVSLQQPESRIAALLGASQVTHLITDRPDKVLSETAVQVVDFTSATG TANLPQLHPQDPAYVIYTSGTTGTPKGVEICHGAAWNTISEINRRLGVGPTD RLLSVSSFDFDLSVYDAFGLLSAGGELVTIPDDARRDAKKWVSLVDSLGITIW NSVPTLFEMLLSAADRTPGKLSSIRHVLLSGDWIDTSLPERMRTVTPQAHLL AMGGATEASIWSNGLDLDVVSPEWTSIPYGRPLAKQMYRVVSSNGQDCPD YSVGELWIGGLGVATQYVGDPGLTETKFVISEDSRWYRTGDMGRFWADGTI EFLGRSDNQVKVRGHRIELGEIESACEALLPIERAVCITHQGASSSPSLVTFA QFTPSHVARTTPEQFATSLRAKVNDVLTEGDIRTSVEHDEHLQTAYAFSVMR RWEEQLTGVGTPNHLREHRNRWQTWLGKADEHPATADLLLDDESFGALER FVTPFEQAFVMAEKQRSIAEFIQSPDSMSVEQFLATRPLGRLVHRVLGAVVR ECSTHSTSELKILEIGSRRPEASADYAAIAGTSAYVLADPYRHHLEHAGQRV GNTFTYRQLGVTSTPQPTPGEAVTKADLVLCNQTLHQSEDIEKTLCEAWGL SAPGATMVVVEPTAPSPMSDITAAFIANNTTDARAETGTVLLSARSWKEILQ RTGWKPVEHVEITKTTALIIAERASSNESVTLCDSDYAKATNLLATRLPEYML PKRILELAKFPLTSNGKIDRKALTALVPEYFDNEPAVTELPHTATEKRLIDIWD ELLHTSSNVNSDYFRLGGDSLTATRLRRTIEQCFGVEFPLENIFDVPLLRDMA ARIDQIAEVPHQQSDLPKIVHGSEQYAPFPLTEVQQSYLIGSSGAIELGDVSS HCYFEMSTACLDPERVEDAFNALIKRHPMLRTVVCEDGLSQRVLPEVPRYRI ALIRSGNADNEDTLDEIREEMSRQKFDPTQWPCFDVRYVAEPDAGRLLLSF DNLFIDGWSMFHIFREWKQAYDHGVDSLDPAIPYSFKDYVEATIELSHSDIHK RDQAYWESAVDTIYPAPQLPVTDTNGANTSQFCRHHALVDAAKWRRIKQRV REEGMTEAVFLAEVYAEVLARYSDEPRLSINLTRFDRTRFAPEVDHIVGDFT

						SLSILSVDTQCAPSFRDRAAALHRRMFSNLDHGSVSGVSVQRMLTKQRGAR VTMPVVFTCGLGVVEHPESDQSPYLGVIDHGLSQTPQVWMDLQVYEHDGG LMLNMDAVEAIFPDDMVAELFTSLTATLSHLAESPELWNAPTSTIAPTTNAPT ADRINDTDRELPGADKSLLGLYQKGLAEHGDNLAVIDATTQWTYEQLNEQS DKWAQLIAATDPAPGDLVGIMMEKSAQQIAAVLGAMKAGCAYLPLSVDQPV GRNTSIINDAGASIVAMDHPDDDFAALAEHCTVITLADVARHRPGDQALSES SPTPSSLAYVIYTSGTTGTPKGVAITHESAVNTIVDVNERLGVTPTDRILGISE LNFDLSVYDIFGMFARGATLVLPSPADKRDPQCWADAVTTHSVTLWNSVPA LFSMYVEHLRERSLIGSSVRSALLSGDWIPVNIAYQVSTLFRDCTVFAAGGAT EASIWSNWYEVGVDDASRTSIPYGTPLANQRMYILDEALNPRPTHVPGDLYI AGRGLAMGYWKDPEKTAASFITHPRTGERMYRTGDKALYNHLGHIIFLGRE DGQVKVNGYRIELGEIESTARKFNELRDCVAVNDHGIVLYVVTHEGFNMAAL NNHLAESLPAYMRPRVISRIDGLPRSWNGKIDRKSLEGKTFEQPQTRERSR NHRDSGIITILQDLLGPKEISIDDDFFTIGADSLTAVRLTNSIRREMSVEISIRDV FNHPTVRELSDLADIVGSDVEEGEI
orf_230	244491	246123	Acyl adenylating enzyme+	32	439	MRGRDIDRNSVLKHYEGTGHLARKSLCQELFESSEIYSDHVAVIADDAHLTY AQLEQSLCVFTEELRESGLSPGDHVLLQLPNTAAYVVTLLALMRVGAIPTLLL PAHREAEVAALCESLHPVAYIGGRDHLGFDTVAMVEAMGPGELGLKELWAD NGPTHDKESSYRVLPGLFTAPISTKCSPPTKWPDPRSVALNLLSGGTTNMP KIIPRVHEAYAYNTRAAAQCCGVGPDTVYLAVLSTSHDFALAQPGILGTLLSG GTVVLCTSAAFDEAFPAIATHGVTLTALVPAVAEVWVEAAEWFPADFSSLERI IIGAAALNDGLGEAIQDRFGVRIHQGYGMGEGITTFTRIDDPPAVILGTQGRPI SDADELVIDGPGGEPGEILEKGPYTFFGYEGNRDTPDCFTEDGFFRTGDRG YLTEDGNLVLYGRVVEQINRGENVSPSEVETLLSGTPGISAAAVFPMPDRAL GERTVAAIVAQPGVNRSAILDDFLTRGVARYKVPDQVITVDEIPLINIGKVDKK KLRALAAAQFTDRESEQS
orf_232	246720	252255	Condensation +	67	368	MDVTALINDLESRGIALWVNGDRLNYRSPKGSLREEDLAALRSNKEKVLAWL REREAVPHDEQARFAPFPMTDIQRAYATGQNEGYDLGGTGCHSYAEIRTER LDRSRLEQAWHELIQRHDMLSAVVVPPDSLQVVKSRSLPVLQAVDLAGHNP DVPDAEYLRHRAKLENRSYPLGTWPLHEFQLLQFDECSILQFSVDMIIADFVS VRVMVEELLTLYAGNVLPELEDTTFRDIITSRNHHSQSAAGFAARTNAKKYW SEIIPSLSGKPLLPTLTSADRTSEMPVRFTRRTWRCSPAAWSKLTDAASTHG VTPSATLLTAYADVLRRWSSTSDFCVNVTSMNRDSAIAGINRIIGDFTEMTLH ACHPHTGTFSERVHATQEQLSEELSHAAYSGVDVLRDIARTTGQPAVIPVVF TSALGADTPHNNGPAYNLVSGVSRTPQVWIDCQAFQDGGSCNVNWDVRE DVFEPALIDDMWESFTDLLDRLVDDGSAWQETDSVHLPDKTIAIRNRIHKTH VQQTTRCLHDGFWDNVQQHPHQPALVCGGKTYSYQHLAGYVGALQHELS DVGPGDYIAIVLGNGVWQIAAAVAVVSTGAAYVPIDHEQPAIRQRSMIEACRP ANVITNSHFSEENTDISNINVDTLSPIQYSGTIASPVSPTETAYIIFTSGSTGIPK GVVVTHSAAMNTIDSVNNLLGRNKRRTVLGVSKLSFDLSVYDIFGTFASGGT LVLPLDEESRNPSKWIDFLVDNNVDTWNSVPALFQMLVREVEVTRHPNILSL DLVMLSGDRIPGTLPAHAAPHFPNAELISLGGATEGGIWSIFHPMTCHTNETS IPYGTALPNQGMWVLDEACNECPDWVRGQIHISGESLATGYLNDPTSTAEK FFFSEKHGTRMYDTGDIGSYRPDGVIEFHGRRDNQLKINGYRVETGEIEGVL ESNDFVERAIVLTQETSDPIKLHAFVTDAQSDKDELKDAGQIRNSELRTMLEQ
			Adenylation+	529	922	
			C-methyltransferase	1138	1283	
			Thiolation	1424	1472	

						RWTPADTSLDTGIFATWMRLGNEAAMAALLAAFQQAGVFLVAGKYHTLTEIT AAIHPSEEYRELITRWLNILTGEGLATKDDEGWTVSQQTLDFFVFGEAWDQF GNMEAEINNSKELFNYQRHAAEALLSQLRGEISPTEVFFPEGDTHNARTIYG ENRISKAMNAAAAEAVIGIAEHHADHPVRILEVGAGIGATTEKIVSRLPENVIE YRFTDISTFFLHKAQKMFAHCGAMTYGLFDMNSDCTSQDVEFGGYDIILCAN VLHNSVNIEESFTRLKQLRRPGGVIVIVEPITELYAALISVSIKMNLVDFTDHRA ESHKVFIEDAQWDQVFRDTQMHRIAEYPNTSDPLRECGQRLIIVGADDDDVP TLNSEDILGYLRAHLPGYMVPASVNVLPELPLTSNGKVDRKALAQLCLEPVG SPNNRIDPPRNETEEQIATIWRDVLDTTEVGRNDDFYALGGDSLLMAETVTR LRQEIPGLQQTWDALMRGVLKVPTIAGISALAQAAGSSCQPEALKAVNSANH TSPELTALSTVASGSPTGSSNLHVYRLPKDATFCRVMIHAGTGRLKDYEFLM PELLQRQPEIAHVGFTAGDADRFLDYTTRTLIRDLAQSYAQELDELDMESYQ LVGYCIGGMLALETAKALTELGRDVRQVTCISTHQCPHRVTNELLCELAYGCI FNADLSAMGANFDLKTLAAALEHTLDGINRNISDEELCTLEGPYADIGEFFQK MAVLSPRARRKLIYRSIREFDTDSESTRGMLDILYDVFRHSLLGTIDYVPDVY FGDVVVLQPTEGVTGFYPSLGGDIDWPATVLGNLQIHAVAGSHATCLLQEN VPSLLPFFTEREQRNG

Appendix 95: Domains involved in NRP CsinB BGC as predicted by PRISM.

ORF	Start	End	Biosynthetic assembly	Start of domain	End of domain	Amino acid sequence
orf_311	330454	331090	Ketosynthase+	14	129	MVAGGIDERFISRANDPNESELHSLLWPAIGRDEDQPMFVISQKTLTGHSKA GAALFQTGGIIDVFRTHRIPANVSLDCVDPLIAPKARNLVWLRSPLDLGSAG HSVKAAALTSLGFGHVSALIVYAHGVFEQAVAQQRGADAATSWRERAEQR LRAGRAHFEAGMLGRAPLFEVIEGRRLPADNAKAAEIAMLLDDTARLGTDG TYSES
orf_317	336724	345727	Ketosynthase+	2472	2905	MPITPLHSMSHPAVLFAGQGSEWQSIIASATQTPATAQRLQEVLSQARTLIA PVARTVSSSCPGAAERLKQLIGGEDDRELGDVLPAVSIPGIVLGQIAAIEQLR DLGIDIDERAGHSQGSLGALAVDKPAEALALAILMGTASTAVNGTDPRSQML SVRGLERSFVTEHLHGTAAIAVVNGRRHFALSGSPEDLAATRTAIEEAVKSF NDELDKRTIGGDELSPRFDELPVALPFHNPVLAPAAERTVALAQKCGLDVD EARAQAEAILVAEHDWPATLDALSSEHLIVLDRALTNLTRRVVEGTGVTVISA ATPRELNALATPGTELPEALNYADFAPRTIELPNGRTYTQTRFSDLTGLSPIM LGGMTPTSADGEIVAAAANAGHWTEMAGGGMYSEEVFRAHLATMEQHLR PGRTAQFNTMFFDRFLWNLHFGQARIVPRSRAAGAPFNGVCISAGIPEVEE AGELLDRLHADGFPFISFKPGTAKQIRDVLAIATAYPEDRIIMQVEDGHAGGH HSWVNLDDMLLETYGEIRQHPTVYLAVGGGIASPDRATSYLTGEWAKKHA MPAMPVDAVFLGTVAMATKEAKATDSVKDLLVNTEGISPSKNGGWVGRGT GANGVASSQSHLLADIHDLDNSFAAASRLITALSIDEYATHREEIIAALDKTCK PCFGDVESMTYAEWLERFVELAHPFVDPSWDDRFLDLLHRVEARLNPADH GQIETLFTSGEDVHDGPAAVAKLLEAYPAARATKVSARDAAWWISLHYKHV KPMPWVPAIDGDLKSWFGKDTLWQAQDERYTADQVRIIPGPVAVAGITKKN EPVADLLTRFEQATTDALLKQGSTPQKAFSRLNSAADAAAFLRSAPSLLWH GHLMANPAYEMDENAFDLRQDADGNWDIVITADSYWDDLPEEQRPFYVRE VTVPVDLPEDVATGGSPVVSEERLPDSVFALLEGLAGVGSTAEQGDEITDM PSVESGYSFHFPASLLSAHSAVTCGNSAANKAGTPDVLVGPCWPAIYAALG SGRLMDGYPVIEGLLNAVHLDHVIDVRVPLEQLADGRQIDVTSSCTAVEESA SGRIVTVELELTDHASGDIVATQMHRFAIRGRATGTAAPKPAPEWGGGKSA TKVVATPRSFVDRATVTAPQDMTPFALVSGDYNPIHTSYNAAQLVNLDAPL VHGMWLSAAAQQVAGRHGRVVGWTYSMYGMVQLGDDIEITVERIGRKGIH QALEVTCRIDGEVVSRGQALLAAPTTAYVYPGQGIQTEGMGTGDRQASPA AREAWRRADAHTRENLGFSIQKIIDENPTELTVRGTTFRHPKGVLHLTQFTQ VALAVVAYAQTERLRAENTLAPTSYFAGHSLGEYTALASLANIFDLEGVIDIV YSRGSAMGSLVPRDEKGNSEYAMAALRPNMAGIDADNVDAWVAEVAETT GEFLEIVNYNIRGQQYSVAGTKKGLKALVDKANAIAPRAAVMVPGIDVPFHS RVLREGVPAFAEKLDELLPQELDLDALVGRYIPNLVARPFELTQDFVDAVAP LAPSGKLDGLRVEDLSEHALARLLLIELLSWQFASPVRWIETQELLFGKVEQI IEVGLASSPTLTNLAERSLAVAGIPEGTIRVLNVERDQEQVMLADVSEAPAA EPAPEADAEEAPQANEAPAEVPAEAAPTAPTPAPTASAASDAPELRFGAAE AIMVLFAFQNKIRVDQIMDSDTVEELTNGVSSRRNQLLMDMSAEIGVPAIDG AADADVASLREKVNAAAPGYAPFGSVLGEAVTARLRQLLGAAGLKPAAVAE

						YVTGTWGLPESWVAHVEAEILLGSRDEDSVRGGSLNTVPSSATSKSAAHEL IDAAVQGVANAHGASVSKDTAAGGASGGVVDSAALEAYAETVTGENGVLA TVARQVLTQLGHVAEPAEATAPDTEVIEAVEAELGSGWLKSVTPSFDAAHA VLFDDAWAIARERLARVALGEIDADDERAQPAAFQGAGATVAEQARWWAR SGNTAVDSAYFEQIAAAAESTEHGAYAGDVALVTGAAPGSIAAALVERLLAG GATVIMTASRVTQSRKEFARRLYAEHGRQGSALWLVPANLTSFRDVDSLVE WIGSEQRESVGNEVKILKPALTPTLAFPFAAPSVSGSLADAGSSTESQARLL LWSVERLIGGLSDLAVKAPSPTRCHVVLPGSPNRGTFGGDGAYGEVKAAL DAILAKWNVEAGWPAGVTLAQAKIGWVAGTHLMGGNDVLVPAAQKAGIHV WSPEEISSELLDLASAESRARATERPIEADLTGGLEGFSLTSLEVEKPAAQS AAEGAAKDTTARIKALPSPARPVQPQLAEELGDITTDLDDMVVIAGIGEVSS WGSGRTRFEAEYGLQRDGSAELTAAGVIELAWMTGLIAWREEPAPGWFVG ETDEQIAEEDIYERFRDEVVARAGVRELTDKYHLVDRGSIDLTTVFLDRDITF TVDSEATARDIAEADPEFTSIREADGEWEVTRRKGATAKVPRKATLTRTVAA QMPDNFDAARWGIPDHMLDSLDRMAVWNLVTAVDAFIQAGFSPAELLQSIH PGQVSTTQGTGIGGMESLHKVFVSRFLGEDRPSDILQEALPNVIAAHTMQS LVGGYGSMIHPIGACATAAVSIEEGVDKIALGKADVVVAGGIDDVQVESLQG FGDMNATAETAAMTAKGIDKRFISRANDRRRGGFLEGEGGGTVLLVRGSLA QELGLPVLAVVAHAASYGDGAHTSIPAPGLGALGAGRGRENSRLARSLRGL SLTPNDVSVLSKHDTSTNANDPNESELHSILWPAIGRDADQPMFVISQKTLT GHSKAGAALFQTGGIIDVFRTGRIPANVSLDCVDPLIAPKAKNLVWLRSPLDL AAAGRSVKAAALTSLGFGHVGALIVYAHGVFEEAIKQQRGTDAAAAWRERA EQRLAAGHARFEAGMLGRAPLFEVIDGRRMPHADTKVMIDGYGLVDADKA AEISMLLDADARLGATGEFPSA
orf_319	347022	348762	Acyl adenylating enzyme+	182	476	MTESSAHAPSAYETKAWLQYYPEWTKPHLDYGDKTLLDSYQETVEAYGDR PATWFFGHQMTYRELDTHVRAAAAGLKAFGIRPGDRVAVALPNCPQHVAV FYAILKLGATVVEHNPLYTAPELEPLFKDHAARVAVVWDKSSPTFEKLRDST PLETIVTVNMIDAMPRRKQVLLRLPIPFIKDKREELSVPAPNTVPWSALTGRA IGGHGHKLVNADVDLNDTALILYTSGTTGTPKGAELTHSNLYCNMKMAESW VPSLGDKPERMLAALPLFHVYGLTLIAALGVQIGGELVLTPAPKIPLLLEIMKK RRPTWMPGVPTIYAKVMEAAKKEGIDLHGIENSLSGAAALPPEIVEEWEALT GGLLVEGYGLTETSPIVTVNPLNKNRRPGYIGIPFPDTEVRIGNPENLDETQP DGTAGELLVRGPQVFKGYFNMPEATENAFHDGWYRTGDMAIMESDGFIKIV SRIKEMIITGFNIYPAEVEDVIRQHPDVNAVAVVGLPRADGSEDVVGCIVLND GAVLDPEGLKDYCRERLTRYKVPRRFYHFEHLASDQLGKVRRREVQKDLM ALIEEHAKA

Appendix 96: Summary of genes in BGC possibly coding for siderophore CpA as predicted by antiSMASH.

Type of gene	Locus tag	Location	Function predicted by antiSMASH	smCOG	Amino acid sequence length	Amino acid sequence
Other gene	ctg1_99	106659 - 107027	-	-	122	MSTNEQVADFAAETYPSYSQELHHSYIEGYNPVSLAAPHSSLLRN STWIGMGLLLGFFPFAGTLIWGISTGIWDAGTAANYSTILTIIGAIGV VVTLVGGFGSIHFGRRYYRKYVKETGRTS
Other gene	ctg1_100	107159 - 108541	-	-	460	MNAPQHRARPGVSPTRQSPSTTHQGASTSRPGANTMPGFNSGP TDEQRAADLRKHKIFVTGLLVLAAVIFLACSWWQSQPGATPAWVG YVRAAAEAGMVGGLADWFAVTALFRHPMGLPIPHTALIRRKKDQL GDSLSQFVGDNFLNAELITSKVREAKLPEWLGKWLSEGENADKVS REAGKLTTNIVRAIDAQEAEQLIQTHLIDKAAEPIWGPPLGRTLDSLI DEGKLEPAVDEVVTWAKNKIDESEDSIVTLINDRMPTWAPRFARQ LVGEKVFREMASFIREVEADPQHPARKSIRKGLRQLAYDLQFDGT MISRVEDWKADVRGSRAVQALPGLLWEKGSSGLIAAAEDPDSTL RQKIVELCVRWGENIQSDDELRTSLDRRINSAVEFLANNYSSEVT GIISETIARWDADEASEKIELMVGKDLQFIRLNGTVVGALAGLVIYT VNQLLFGA
Other gene	ctg1_101	108627 - 109094	-	-	155	MTDKRQRFSTSEIVESLKETGSRVRSVASEFSENAAENYRARGTA ETVKNAARQTVRDFRGIESFEDAKDTGTKLWKRGKHLGREFSSSL SDAVGKTRNSEAARTGTKGTTGEHAGTTNDGRPRIIEGEVIDVDG QKVNGQSSSDKSFPRKTQNP
Other gene	ctg1_102	109091 - 109480	-	-	129	MIVDVLSQASLAAEQTQHLAQLPPSGPMLEPWQQNLFDGVFWLQ QGLYSLVAIAGLVGAISAAMTRDDAFDADSRQSKMVWVAMLAGS AFIVPMPIPILPWVGMVIIGIYWFDVRPSLRALINGEYRWD
Other gene	ctg1_103	109484 - 110458	-	-	324	MENNWLHRFLAWDFRPAHFCLHDVRAGVPAAIAAGIGAIVVSPNH VATAAKALEKYIRQNHVGPEANARVAANDGVGDSDSAGDTDESA VLTIIAAVGFPTGRHHILVKASEARLAISQGAHRVWACVDTSTADA NAALSDLISLREAVPYPAHLGIVVPDYAGLQPSGAESYSDARANV QRRNELVSTCEFAKVDQLVVVPPEHSSAPAATETPAATKTPAATK LPVATGLLAAAQQLPDRIPIVLDAGAGKAATTEQNGSRHAASLAAA SLAEWLVAESADLAANKKALEKTLEKGLEKAETAKPILAVVPVDPE EIRAVSS
Other gene	ctg1_104	110455 - 111375	-	-	306	MPSSSQVPANSQAQSSAATAWKVIVAVLVALLVVVLLAEFGARYFI GQQLRSSFREDIEAQGIEMTEDPEVSFGSSPVLFGLIGGKLSEVNI DTPSTLQQDPEGNYKGMPASHVHMESMALKSGEAEFLRTRSELP DDYLTHSIQEGLREQLGGVGFLGDIVVSDITTDHEAETITVQFLSGA ANLTLRPLVDANGNVAFDASEAKVLGFSLPEGAADGIANSLNQNL REATGGQLDITAIEFNDHGLVITLEGHNVNPSHMSKELNADPGSG EAGAGSGAGSGTGAGSEAGGTSGSNSDAPVRNT

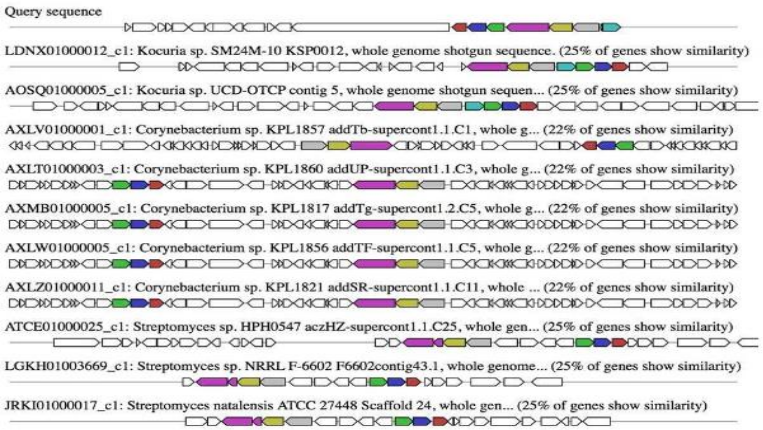
Other gene	ctg1_105	111405 - 111896	-	-	163	MSTRSENTVTINHSAEKVHQALTNEKYWAYIAENLSPEPGELNEF SNNEATLFELLPVTILPEAVRAMVSQSLKVKRTVKVGELSGENAEL SYTADVKGTPVDFKGTVNLAGQGETTTLSYTNDVSVNIPMMGAAI EPKVGEALEELFDNEGQLTEKWINENL
Other gene	ctg1_106	111923 - 113110	-	-	395	MNDFMTQANTALDTPELRTLADAGVIERVDGVEFSQLTTFHLGGA PRAAVRAATGEAAAAVVRALDAAGVQLLIVGGGSNLLVTDEPLDV VAVIMGDDHLEIDASTGVLRAGAGAVWDEVVAASIDAGLGGLECL SGIPGSAGATPVQNVGAYGAEIADVLTRVQLLDRATGTVEWVPAS ALDLAYRYSNLKFTSRGVVLGVEFQLHTDGLSAPLRFGQLAGKPG QRREPAAVREEVLGLRRSKGMVHDPDDHDTWSAGSFFTNPVVP EALADEITQVVAAEYGADEAAGMPRFPVAANAGATSDAGADAEN TADTETMVKLSAAWLIDRAGFAKGYPGEDAPARLSTKHTLALTNR GAARAADVVTLARDVRAGVEKKFGVQLEPEPLWIGLEI
Other gene	ctg1_107	113107 - 113619	-	-	170	MKLSRTFTALAATSVLTLAACSNAQETADEAAGSATSQASSAVES ATDAAGSAADEAKDDDQNSASNADQKHPDILEAELSGSDGEYRIS VTVSSPYDSPERYADGWRVLDEDGNVLAEHELGHDHANEQPFTR SRGPFEIPDNVQEMTIEGHDQEHGYGGATVTIEVPR
Additional biosynthetic gene	ctg1_108	113668 - 114906	-	putative esterase	412	MPSSLSRRAAALALTAAFSLPAAAAVSSTVATPQAHAQLGSSDTS SRIAFDGSSEPMAEWSKDPAVQGHVLDVIQFFQNDTPWHQLPM WNLWGEQTPYTIDKEQDFAQPRLINREPDERFDVERLYVESPAM RRVVQVQVQYPKDRETAAPMLYLLDGVSAPRQSGWLRKGDVQG TMANEHVTTIMPVEAGGTNYTDWNDTDPYLGRAKWETFLTQELP GVLEQQAGIEFNGERYIGGLSMGGSAAIRLANLHPELYSGTFSVS GCYSSTSTAGRAYFNLASRVMGGNPDLMWGPGTTPERLRNDVV ADPKGIASMPLYIYSANGHAAERDIDLARSEGITTFFGNITLEKLTN QCSAELKQSMQEKGMMHDRVEFDFAPTGIHDWPYYKQQLPVAW ASITQGKYTYPTPGE
Other gene	ctg1_109	115087 - 115542	-	-	151	MANQQSRLHSQVFGYNAARWVWTAVTVAITIWFLLMMSWVAFPL VIVTAIVVLFHIDTQLDQEQQTLKIKVLWVPVKTCNLADIEEAAIPRD PLPLERNTFGVHFIGGALSMHAGAANLALEMKDGKTYRVTVYRPQ QFVKRLHNAAQRDA
Additional biosynthetic gene	ctg1_110	115545 - 116510	-	(polysaccharide deacetylase	321	VKKLAAIFGSFTIIASLFSSPIEAHANPAGSSNSARSGAVVNLASSA PALPGSSIGPRWDQFQLGSQILEFFHNSTDEAYRAFFASLPKEVQ QLSSTPQPRSKPIPPNLQPGYQPPAPVTPAGDCSNCVALTFDDGP QPPTNRILDTLDAKQAKATFFVQTPMVHALPDTARRIVNSGHTIGN HGTTHVNFALSSPTRVHTEISQNTKAVQAVTGATPRWLRPPYGAH NDNVVQAARAHGMGVALWDVIVFDWRDRNANTVCERSVNQANG GDVVLLHDIHNTTADATTCIIDGLRAKGLAPVTLDRLHGAPEPGRV YGTR
Biosynthetic gene	ctg1_111	116569 - 125847	-	malonyl CoA-acyl carrier protein	3092	MSLTPLDSMRAPAIIFAGQASEWEKQLRNAAASHHAVGRLDNYLT QARSLVSPYARTIASTIPGAMERLEALVHADSQDTGASDHTASTS NTANDVFPAVSIPGIVLTQLAAIDQLRDLGLDLAGCDLEQGSTQNN PTTAFGHSQGSLGITAIKHPVHALALAILMGTAANVSHGSTDSRSH MLLIRDLPREFVDECLENPNTGSAAISVVNGTRAFVVSGTPDELAA

				transacylase		VENNVAAHASAYNDALEAREFGGSELTPNFTYLPVALPFHHDSLT EAAELTIEWATACNLDLGETDIATVVRDILVRPFDWTTQTARMVDQ QISHALVLDDGLAKLTTPLVAGTGLAVVPAWTPSQRDNLATPGAQ LPTARDWSAFAPRLVTLPDGKTYTQTKFSALTGLSPIMLGGMTPT TVDPAIVAAAANAGFWTEMAGGGMYSEELFTELKNSLIRQLQPGR TVAFNTMFFDRFMWNLQFGQTRIVPKARANGAPITGVTISAGVPE ADEADELLAQLRADGFPYISFKPGTTAQIRDILAIADANPDHQLILQI EDGHAGGHHSWLDLDEMLIDTYAAARERDNVTITVGGGIYSAERA AEYLTGTWSQKYSLPRMPVDGVFVGTVAMATKEATTSPQVKQLL VDTPGISPETNGGWVGRGKANGGVASGQSHLLADLHEIDNSFAA AARLITSLDPADYQDHRDEIIAALDKTSKPYFGDVETMTYQQWME RFVDLAFPFMDPTWQDRWFDLLHRIEARLHPADHGEIDTLFPTLE SVADAPANLRTLLEAFPAAATTTVAARDAAWFVTLNHKHNKPMP WTAAIDGDLKKWFSKDTLWQAQEARYDADAVRIIPGPVSVAGITK ADEPVAELLARFETGTTQALQQAGVEPSAQYSRLADAKTAEDFLR HAPSLQWHGHTIANPAVELPAEAFDIINDGTEAAPQWSIRINTDSY WDDLPAEQRPFYVKSVSIPVDLSDAVATGASPVVDRERLPKAVFD LLAGLAGVGSTSESGDEITALPAIVDGSVSEQYPFGYAEDSFTLPT TLLQAHTAVTGAGLGDKLTTAPGTPDVLVGPAWPAIYTALGSGRI GEEHGEPAGTDYPVIEGLLNAVHLNHVIDLQVPLHELARGTKGNG RRIDVKSWCSAIDESNAGRIVTVELELRDHESGELVATQMQRFAIR GRATGASVPAPAPEWGGSALATTVEPTPRSFVTRATVTAPQDMT PFALVSGDYNPIHTSTNAARLVNLDDALVHGMWLSATAQHLANTH GTVTSWTYSMFGMVQLGDKVEITTERVGRAGIHSALEVTCRIDGE VVSVGQALLAQPATAYVYPGQGVQAEGMGRGDRNASPAAREAW RRADHHTRTQHGFSIRQIIDDNPTELTVRGTTFRHPQGVLHLTQFT QVALAVVAYGQTERLRENNALAADAMYAGHSLGEYTALASLANIF DLEAVIDIVYSRGSAMGTLVERDEHGNSNYGMGALRPNMIGVSAD DVQDYVAKVAAETGEFLEIVNYNISGQQYSIAGTKAGLAELGKRAT AIAPRAFVTVPGIDVPFHSRVLRDGVPAFAEKLDELLPETLDLDALV GRYVPNLVARPFALTQDFVDATAALAPSGRLDGIDAATMEPQKLA RLLLIELLSWQFASPVRWIETQELLFGRVDQIIEVGLASSPTLTNLA QRSMAVAGVSLPVYNVERDQDVVMLADVTEAPTELDDDGDSSSA AADQASQSPEVGADAARPTENPAPMSEAGSDSTAESAESAAAAP ASASTSASANQNANASAGGGAPAGELSFTAAEAIMTLFAFSNKIRL DQINDSDTVEELTNGVSSRRNQLLMDMSTELGVPAIDGAADADVA TLRGRVTTAAPGYSPFGSVLSEAIGSRLRQLTGAAGAKPAAVGER VTSAWALPQSWIAHVEAEILLGSREEDSVRGGSLSTIPTSASSKAE VNELVDAAVQQVAARHGVAVSQATGGASAGGGSVVDSAALNEF AEQVTGENGVLATAARTVLAQLGLSEPAPETPETDNTLFETVEAE LGAGWVKTVTPVFDARKAVLFDDRWASAREDLARYALGQADIAL ERFRGTGTTVAKQARWWAANGGQGDLEKIATLAESELQGAYSED IALVTGAAPGSIATALIARLLEGGATVVMTASRVSQARKEFARKLYA DHAVPGAALWLVPANLSSYRDVDAVINWIGTEQVESVGNEVKLLK PALIPTLAFPFAAPPVTGSLADAGPQAENQTRLLLWSVERTIAGLA

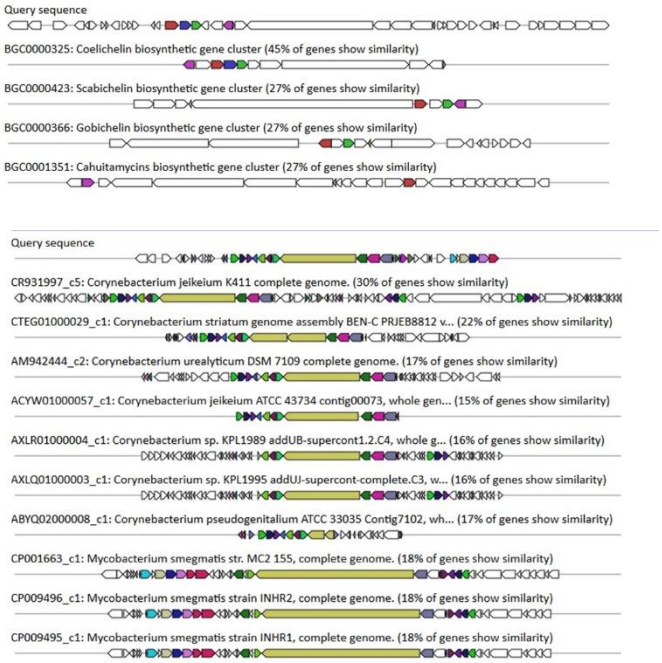
						GLAQAGVDKRCHVVLPGSPNRGVFGGDGAYGEVKAALDAIMNK WKVEAGWSDGVTLAQVLIGWVAGTHLMAGNDGLVPAVEAAGIRV WNPEEISHELMGLASAESRAQAAQKPLHLDLTGGLGDTLNMAELT AQAAANAEAAAAAAAEAAANAPAAPATITALPNIAVPAQPERVDTG EITTDLDDMIVIAGVGEVSSWGSGRTRFEAEYGIQRDGSVELTAA GVLELAWMTGLVNWAEDPTPGWYDANGAEVAEEDIYARFRDEV VARAGVRTLSDKYHLQDQGSIDLATVFLDRDITFAVPTEAQARDIE AADPAFTRVAEADGEWQVTRLQGATAKVPRKATLTRTVAGQMPD NFDPTRWGIPDHMVDALDRMAVWNLVTAVDAFIGAGFSPAELLQ SIHPGDIASTQGTGIGGMESLHKVFVSRFLDEERPSDILQEALPNVI AAHTMQSLVGGYGSMIHPIGACATSAVSIEEGCDKIKLGKADVVVA GGIDDVQVESLAGFGDMNATAETKKMTDQGIDERFISRANDRRR GGFLEAEGGGTVLLVRGSLARELGLPVYAVVAHAASYADGAHTSI PAPGLGALGAGRGRENSRLAKSLRGLGLHPNDVTVVSKHDTSTN ANDPNESELHSMLWPAIGRDADAPLFVISQKSLTGHSKAGAALFQ TGGLIDVFRTGRVPHNAALDCVDPLIEPKAKNLVWLRDGVSLADA GRPVKAAVLTSLGFGHVGAVVVYGHPGVFEAAVAAQLGEDAAGA WREQANARLAAGAAHREAGMIGRRALFSLIEGRRLPERGAKEAEI ALLLDEDARLGADGTYPLPGADGAASLL
Transport-related gene	ctg1_112	126028 - 126867	-	ABC transporter ATP-binding protein	279	MNNPISTPPNRGVSVNAQNICVGYGDNIIIDDLSVDFPSGEITTIIGP NGCGKSTLLRAVSRLIPLRQGTVTVDGKDATTMKRKNLAKTVGVL PQTPSAPEGLLVSDLVARGRHPHQSWINQWSAADEQEVLHALEL TNTAHLAERSLESLSGGQRQRVWISMVLAQQTDVLFLDEPTTYLD LSHSIDVLNLVTRLKHDLGRTIVMVLHDLNLAIRYSDHLVVMKDGAI HASGSPAQVITPELLSEVFNLDAHVNPDPVMSEARGTTVPLIVPAA PPGM
Transport-related gene	ctg1_113	126924 - 128033	-	transport system permease protein	369	MTNSPHLLSRPVVPGRPALRKGRISLVWRPWVLFSTIAMICAAVIIF AASIAIGDYPLTVPEVLRIIFLGDGSRIERIAVFDWRMPRALTALTVG FALGLAGALTQSVTGNPLASPDILGITSGASAAAVTVLTFGGGVGA FAAVAGKVATFGLPLAALAGGLITGAVVWILSSRGSFDAYRLVLFGI IITALMNAYIGFLMTRAELRDAATAQQWLAGSLNSTNWDRAVPVA LVILICVPLLAWMAYTLTAAVLGQDLAAALGLRVSRTQLTFLVIAVTL ASIAVAAAGPIGFVAFVAPQLAVRITKLSTPPLVASAFMGAILLLGA DLVVRTNIPVDLPVGIVTSAFGGVFLLYLLISSNLKRSRSS
Transport-related gene	ctg1_114	128030 - 129091	-	transport system permease protein	353	MRKNSREYTRSSRSLRTRRITGLLALALLLAAALLASLLFGARPIAF ADAVQALINLPSLLDEPQSGSADARVIADLRWPRTLIGLFVGAALG VAGALIQGHTRNPLADPGLLGVSAGAAFAVVLGFFAFGLTSALSTS VVAFCGAVLATALVFGLASLGGGQINPLTLILGGAALTAVLHSMTT ALTLIDDNSLDRMRFWSVGALSGRDLSIFWGTAPFIAVGLIVALATA PTLNLLNLGDDAASALGVNTYRARLAGMVVIALLAGAATAAAGPIG FIGLVVPHIARSLCGPDYRWILPYSALTGASLLLLSDILGRLIARPGE LQVGIVLAFVGAPFFMYLIYRRKVVAL
Biosyntheti	ctg1_115	129202 - 131754	-	putative siderop	850	MSKSQQSCHPISLTPGAMHTSLGTVNIRPVQPTKDAKTLHTWVTH PRSHFWGALDASEQEILAEYRRIETSNAEQAWLLYLDGNPIALSET

c gene				hore biosynthesis protein		YDPKHVVLNTCSEFTLHEGDIGMHILVSPPIGKSIHGLTDNIFSALV RWIFSHGEVNRIVVEPDVSNLGIHRKNARTGFTQPGLTTTLTMSG KPKTALIQYCTRPDFLGSPNAPLSYPIAPTPWELVKPTDSSSPVTL GIGHLAKTADAAHRELFAKALREFTHERLISPRKIKESADETESND GNHSGWRSYQVSWGSRKLIFRAREHRLLHLSVDPLSIHDPADLT WTPDIVTGIADAAEQLGIAKSSRETYLEEISATFAARARTLSIPRPTS TELADATRDPAELFQAIESAMVIGHPGFLANSGRGGMGETQLRAF SPELSSTTDLVWCAVDKNYAHLATIEYERAQESGSSHSSNEAILTL VPDFKDRCRLAGLNPDDYIPLPVHPWQWEQRFTTTFSADLAAGRI VPLGREGEAYRPQQSLRTFFNTSTPAAPYVKTAVAVRNMGFTRG LSAHYMESTPRVNLWLLNLIGDDPEFFQSGIGFLPEIASVGYTGSV YHRVTNDGSHTKMTAALWRQSPFSPETTPALDDGDTLSTLAGILH SDDEGLPLISAWINQSGLSAVDWINQLLTLYVRPLIHALVRYGIVFM PHTENVILRLNQNKPVGIYHKDLGEEIAVVSNETPVPSGLERLRAR CGSDSPEDLSQQALSIHTDVIDGVLRHLAALLDDHAILEEDVFWEQ LRHVASVYAKDHPELTTGDSRQARLWQALLAPRFRHSTLNRLQL RNPHTMVTLGDQDSSLIYAGELENPLHGV
Additional biosynthetic gene	ctg1_116	131751 - 133169	-	ysine/ornithine N-monooxygenase	472	MTKKIHEQPSTNITQRVAADEVFDIVGIGIGPFNLGMAALAHPLVQS NELKAIFFDENQGFCWHPGVMFPNSTIQVPFMADLVSLADPTSPW SFLNYLKRQGRIYQFYIRESFYPYRTEYSNYCAWVAGQLSTLQWS SPVTDVHRTKDGLWRVQVGGTNPRAVLAKNLVNGTGTQPYVPTS LQPNLVESRNQTKNIIHSSDFLFHAERLSNSEDIQSLTIVGSGQSAA EIYRHLMEPFAATGRRLNWLTRSPRFFPMEYTALTLELTSTDYVR HFHRLPEARRDQINRQQRNLYKGISGDLVNDIYDTLYQLSITTPLR GELRPDVTVELFTESDLKSSYERHINEQAAESQILLKLTNNETGNA DIHATDAVILATGYGRPAIPSHLITAAEAGEINTDKQGRLDVNLDFT VNQRRDLYVLNAEEHTHSLNAPDLGMGAWRNSIILNNICGRTVYE VEDSWTFQQFGGENV
Additional biosynthetic gene	ctg1_117	133162 - 134700	-	Decarboxylase, pyridoxal-dependent	512	MDSYSRKFAPQAQPGLQFLGDGYSHQHALSSAGQTLNQALAPTF SPSPETFCPNKDLGVDELSAAFNAVEIDTPLGSLDEALKELNDLWV SHAIRYDSPRYLAHLNCPITAAGFAGATISAGLNTAVESWDQGRG AAIIEDQIIKWFAASAGMSTNNATGTFTSGGTQSNLQALYTSREKA LNSGSSLDSLRVLCSEEAHYSIARSAHILGLPKEAVITIDCDSDGAL KISSLRRTINSLIETDGKHTIACLALTAGTTDLGAIDPIAQATQIAQQH GIYVHVDCAYGGALLLSDKNQKLLSGLHQANSFSLDFHKTFFQPV ACSALIYRDLTELEHVTWHSDYLNPETDTRLNLANRSLQTTRRFD ALKLWLSLRADGPKALGQAFDRCCEIATDAAELIHEFPELELLTTPT LSTVLFRFLPSLPGSPSSHTCMRETEVINKINDAIRRQLFNANELVI ASTRRNGKTYLKFTILNPRLTKDDIRYALEKICAAGTDECAKLQSKE EIDD
Additional biosyn	ctg1_118	134823 - 134939	-	-	38	MCRAPYILKMFIRLSDDLTENRANNVSYGGARNQRKKE

thetic gene						
Trans port-relate d gene	ctg1_11 9	134942 - 136015	-	(iron compou nd ABC transpor ter, periplas mic	357	MSSLMPHSVLHTSNSSGLIARLAAVVSSMTLMLGLAACAENGESN DAADNNSGETRTIEHALGVAEIEGTPERVVTLGQGSTETTLALGVV PIAVESYEWGADESGYLPWVNEELEEMGVKDDEKPELITGAEELS AEEIAALEPDLILAPWSGLTQEQYDQISEIAPTVAYEEEPWVITWEE QINTVAEALGKGDQAQDLIDGINQQFDDAKEESYGEHTFSFIYNSG VGNYGIFLPTEQRVQFVSKLGLQVDPAVEEFRDSVVAGTDSATVS AENLDKLNDSDLIFTFYSDEASRKELHQDPAYSSIAAIAQGAEVAP TDQSFVTGSSMINPLTVPWAIERYKEQIDDALAKAQQ
smCOG; bacterial specialised metabolite Clusters of Orthologous Genes -; not predicted by antiSMASH						



Appendix 97: Siderophore CpA biosynthetic gene cluster and its predicted similar gene clusters according to antiSMASH.



Appendix 98: NRP CsimC biosynthetic gene cluster and its predicted similar gene clusters according to antiSMASH.

Appendix 99: Summary of genes in BGC possibly coding for NRP CsimC as predicted by antiSMASH.

Type of gene	Locus tag	Function predicted by antiSMASH	Start of gene	End of gene	Amino acid sequence length	Amino acid sequence
Other gene	WM42_1357	hypothetical protein	1379979	1381710	576	MISPDYMDSCAAVIEKSGAHRMIEFYFHQDRGVGGRRSTGPKYSMLGVLTVGLALIGIRRVPSMAEIWRTLWTLDPAQQERLDLDLSCGEGSYRAFAMWLTRRLEPLDSLPDAPARRVKNGDHRRMIAARSPEQEQASEMASERLHQVVNSLVAGSIHEQAPKGYRGDIVADETIIDLAGQSMGLGSRDTKHRGAAYSGAYYIRDREDHSLHAELGTRRSTKGGVAVGITAVCRVGPPKAVYSVAPVITGISIDKPSSGSVQGLARAIRFHQENGFDQRITRGRLPLLTVDMGYNVKRHFNDELLAAGYAPVVRYPKSWRTVFASDSAAGDEPASGPLQVAGEFYCPAARAIAGDGKIVRRTMELLGEDDGFERHDARLESLFPLIMGTNSRPYRARQGRGRPRKQEENSEKPVKIDLVCPAVQGRVRCPLKPASMTQSPEEAPEISPSWSADHYKCCAKSSITHTFTPEQWRRAQWGLVPGSWEHATYYEAARAITEQRFSIMKSPHVTGMESFKRGVRREPMLYITLALWVASTNLAIQEQFERKTAGQDSMTRRLRMVREDTGRDLAKVPPRT
Other gene	WM42_1358	Putative secreted protein	1381739	1382816	358	MGRFTALAAAPVVVEESKRSSIPAVSVIGQASPSRVRRLTRRTLPDTDRYLRLREIGIQRPAGPLRSDAIDRAALVQAHTRAIDEVADVKAARVVARYLYDEAIVFGSVDEERALTRPAVDKWLFRDDVFSRRSQRTYRTILYAAGRTLYPHEFPEPHRQRGPRRKAVPAAKTSLEDELYAIAPSLRPMLRAQLLTILDLTTGAGLSSQEIRHLRGSDISPLELAERTVVQIAVRKRGVISRVVPVVCPVRGHRLLTRAREVGSGFVFPIVGDSMPRNAICHVAEELVEQGFPGFNAAALRNRWVVKLASDPGVPAAMLMKMAGIADLRVLHDLEADLPQFSPEDGARALLHCKGEDL
Other gene	WM42_1359	istB-like ATP binding family protein	1383909	1385040	376	MVRLPNGVLLAGDQEIKRLSPKVGDILVYSSGWEIASDHSWEVKAQIGVIEQKAGPDEVLIKTSDGYAQKLVHRNTRWSMGDYALFNESSSILKVIPQEAARKSTTEDEDSASKFRIDLDPERFHWDYFVGSHEILSDAKRIVELYTTPEGRDFISQMKVDPVHGILFAGPPGTGKTFLAQIMAAQSKSAFYLVTTASLGGQLVGQSEERLEAIYEDAAKQEMSIIFVDEIDVLTKDRSNAYEHGSRLVNVFLTNMDGVGAPENVLTIGATNRISDIDRALRRPGRFDREVYFRLPNQMDRLEILKSRTPTLANDIDFDKISSQTEGWTAAELRSILQYSGELAVIEGRSRIYNDHFLLGFEKATTARNARKGEEK
Other gene	WM42_1360	integrase core domain protein	1385743	1386586	280	MCHVLKLNRSSFYKWVQTREKRRLKMYSDALIGARITTIFDDEHGLYGAKRIAASLKEDTTYTPINHKKVARIMKSMGLKGFSKRRRCITTRRKPGHRVMPDLVGRKFTAGEPNRVYVGDITYLPCRGGKNMYLATVIDTYSRKLAGYALAD

						HMRVSLVIDALAHAHGVRGSLDGTIFHSDHGSVYTSQAFRNYCSSLGVRQSMGAVGTSADNALAESFNATLKREVLRDRKVFNNPISCRQEVFRWCMRYNTRRRHSWCNLVAPDVFEAETSAILTTAA
Other gene	WM42_1361	transposase family protein	1386637	1386937	99	MSRYSEQFKRDAVALYENNEDLSLNSASAELGINRASLHSWVKKYGTGKRARTKAVHDQAQAANDSARIRQLEKEVSKLREERDILRKAAKYFTEETHW
Other gene	WM42_1362	hypothetical protein	1386994	1387657	220	MGVRSDWDWDKKSSTLIRRADGTANSPIRLVGLTQLDNYWGDVRSFLFDQSECIALVRISQDTPSTSWSPLKLFDAARGLPSFDGFDTKIAELNSILSSKQKAETPVSNAAERTLSHYLQIRTDNEELKVQLEPQVNEIAQMVPDGSQRQLNDAFDQVDFLLSSQGSTETTPDEIGSLRSKAIRLEQGTGAPLKSEENVLRAPNDDSDLYLVGDIIALYW
Other gene	WM42_1363	DDE domain protein	1388063	1388386	100	MPVDHTTIYRWVQKYAPELDKQTRWYRQVPDWQASSWRVDETYIRVGGRWCCLYRAITAGGQTLDFYLSSKQNVAAAKCFLAKAVRSNASAGYPRVINTD
Other gene	WM42_1364	DDE domain protein	1388441	1388657	71	MKYLNNILEGDHGRLNRILGPKGAFKNRTSAYRTLKGMEAMHSLRKGQGAMFAYGQPNPDAVIVNQVFKRA
Other gene	WM42_1365	hypothetical protein	1388919	1389579	219	MSGNEFSQNLPDDTDALKALFYLCADQFGQDPNKRKQAKQILKAISEKIIHQTDLEKYRGSSEMLNDQNLRQIGLVISQAAATEHVAGQIIAVSKSGPFDSPIDKAWTRSGTQLQESLKPHVPESLLDRLKTAIDLRNEVAHGFHSEITDTEAYELLGSPVDAIQSGDRITVKREPKKDAAPFKTLTWRDDALKELHKELIAIEGELEKTLWEKINSLL
Regulatory protein	WM42_1366	transcriptional regulator, TetR family	1389890	1390343	150	MYHHFSGKQDLAVAALEASATAMRQDAEALLHGDGTATERLVAYLGRQRDSLMGCRMGRMTYDADVLATPKLLTPVSETLAWLVQTIKLVIDEGIVNGEFPRHTDAHRLASMVVATVQGGYVLARAQQDPAEFDAAVQGATVLLCAWSQR
Regulatory protein	WM42_1367	cupin domain protein	1390339	1390723	127	MSTVPVHFTGTVNATTLATPTPQSPLAVYAVTFTAGAHTHWHTHPKGQGLYVTDGVALVHIEGQTAKHLTKGESIWIDAGRRHWHGAAPGQPMTHVAYQQAADDLSTIDWHKPVTPTTYAQATKENH
Other gene	WM42_1368	hypothetical protein	1390722	1391280	185	MYAMQYQITLPTDYSMQIIRDRVTQTGHFMDGYPGLQFKAYLTQEKTKGAPRNAYAPFYIWRDIDGMRQFCWGEPGYSAIVRDFGRHSIQDWTVHQLVDGPADYSAARSLTVKTAHLPTNAAPSQCIDDITAEFLATTTDSTVARVTAVDVTTWNAILVELSTHEADQSSTGVTAYEVLHVSTTA
Other gene	WM42_1371	ribosomal L6 family protein	1392322	1392655	110	MSRIGKAPVTVPSGVTVTINGQNVEVKGPKGTLAEEIPAPITVAQEGEELVVSRPDDRRKNRSLHGLSRSLINNMVVGVTTGYTIKMEIFGVLHKTYAHSRNGTSACHIK
Other gene	WM42_1372	30S ribosomal protein S8	1392669	1393068	132	MTMTDPIADMLSRVRNANNAFHDTVSMPSSKIKANIAEILKQEGYIADYSVNDETVGKTLELNLKYGPSRERSIAGVRRVSKPGLRVYAKSTNLPKVLGGLGVAIISTSQGLLTDREANNKGVGGEVLAYVW
Transport-related genes	WM42_1373	fecCD transport family protein	1393592	1394711	372	MAISTPHAQSPSQSHLGSRADATADASASAPSGKIPRRLVGLGVLFLLLLASIVASIVFGSRQIPFGEVSAVFRDLGTAFGHAEGLNVDQRVIVELRIPRTLLGLVAGAALGASGALIQGHTRNPLADTGILGINYGASLAVVASFSLLGVTSVWATSMWAFGGAIAATALVFSLASIGGGQANPMTLVLGGAALSAVLSAIISGFILTDDANLDRMRFWTVGSIAGRDLTVFYGVLPFILVGLLLAFITAPQLNLLNLGDDIASGLGINTQRARLIGMALIALLAGAATAAAGPITFIGLVVPHLVRAITG

						PDYRWILPYSALTGAVMMLFADVVGRLIARPGELQVGIILAFVGAPFFIALIY RRRVVAI
Transport-related genes	WM42_1374	fecCD transport family protein	1394821	1395805	327	MLLVAIVLLAAVSIGLGDYPVSPARVLEVLFTGQGTRIERLVVLDWRMPRAL TAILVGCALGLSGALTQSVTRNALASPDILGFTTGASAAAVTVITLGGGAGG FLGWLSSIGIPLAAVLGAAVTATVMWALAWRRSTDSFRLVLFGIIISALLTSY INFLMIRTELRDAAAAQFWLTGSLSTADWSKMWPIAIVVLVFTPLLAWIGHQ LLATLLGSDTARALGQNVQGVQVLLLAAAVALAAVAVSAAGPIGFVAFVAP QVALRLCNCSAPPLLASALTGAALLLLADISTQTLLPVELPVGILTSAIGGAF LIYLLVQRNRSTTA
Transport-related genes	WM42_1375	ABC transporter family protein	1395873	1396638	254	MSARGLAVGYGDRTVIEGLDVDFPRGQITTIIGPNGCGKSTLLRAMSRLLP ANEGEVLLDGADISSIRRKDLARTISVLPQTPTAPEGLNVADLVSRGRHPH QSWIRQWSSTDEAEVHKALEMTGSMGLAERTLDSLSGGQRQRVWISMVL AQNTDILFLDEPTTYLDLATSVEILELVQRLRRELDRTVVMVLHDLNLAVRY SDNLVVMKDGQVLATGRPSEVITPELLLEAFALNALVIEDPVTGGPLIVPK
other gene	WM42_1376	hypothetical protein	1396648	1397287	212	MHLSDLMDATTLLQNFGGWALGGIALIIFIESGVLFPFLPGDSMLVTAAILR DQLGLNVPILVGVAIVAAFLGDQVGFWLGHRFGRKLFKPDAKILITEHLEQA EAFFLKYGPLALVLGRFIPIVRTYIPVAAGTAEMPYKKFVGWNVTGAVLWIV SMVGIGVLLGDIPGIADRIDMIAIVIVLVSVTPVVISAMINWRKSKKTPAQELM ED
Transport-related genes	WM42_1377	periplasmic binding family protein	1397460	1398477	338	MTLKKILVSTAAVLTLGAALTACSTDDQSGNSGDSQAVNAEQGAFPTTIEH RYGSTEIKEAPQRVVSLGYTDQDALLALGVTPVSVKYWDGMTPDGQAAG NWSNDKIEGDTPRIDKDTEVNAEAIAKDNPDLIVAVYSDIDEDTYKKLSEIAP VVVQKGEYEELQQPWDVTTEEIGQAVGKPEEAKRQVEQVKEKFAELKGR HPEWAEKELGVATVSDESLAVFAEGDPRSRFFTELGFKINPAYAGITKDKF YGEVSKENADQINSDVLVWDQLSYSPKQSKASVTEDPIVGKLPAVKDGHS VYLEGDLEKAFGWQTVLSLNYLLDKIEQPLADATK
Additional biosynthetic gene	WM42_1378	formyl transferase, C-terminal domain protein	1398542	1399364	273	MWADSVEELASEHELQVCVTERVGQDVIEALRDAAPDIIVANNWRTWLPP EVFSLAKHGALNVHDGLLPEYAGFSPILWALLNRETHVGVTVHEMDEVLD GGPIVAQRAIPVGPQDTTTDLVAKTIDLIEPLVERALSDVAQGTATAQPQDP TRATYFHKRGEQESRIDFTQPAEDIALLVRAQSDPYPNAYFEFRGQRVRV LSAHVSQGRFGGTPGRVTIPHEGGIAVVCGSPRANVPAPAIVLDRVRLDD NSELTATEFFGHRAGYIQPKV
Other gene	WM42_1379	putative formyltransferase domain protein	1399364	1399520	51	MLIRVIAGRVRNYQNYFMTSIDDRRQRPVSPRLIAENSNSHAGDVSLGLPM
Additional biosynthetic gene	WM42_1380	alpha/beta hydrolase fold family protein	1399546	1400551	334	MPDSPDVTTLHRISVPADAEALVFPDPYRSNVELTESMSTVDVRIPAHSAV TYSFQSGELKYPDPHNAHGAGPQASLLSGDSVDQTLWPPRSPEVIADLP GARLSIDRKVFGRRCTARLDDRGADTTVVFLDGDDWIHLHDLTGALDRAV TAGLIPQINRVFLPAAKDRSEEYTSQTFASALATELPPIINSSHIVLVGQSFG GLSALRAALTASTETTPAIKGAIAQSPAVWWSADRSAELADTLSDGPAGG DIAAQLTGPSLENPAAHSGSGLARIVLTCGAEEPPMQRHVDAVADALTRR GFPTTNHRTPGGHDPAMWRHGIIPALAELLG

Core biosynthetic gene	WM42_1381	D-alanine--poly(phosphoribitol) ligase, subunit 1	1400553	1411410	3618	MPTRSSAHTPFSSGDTLASRDGGAVIGHGRSEQDRIAQQRIWNDTDQDR DRPDLISLLLQHAQETPDALAVVDDRHRLTYAQLVAHATAVARNLREHGID AGQSVGISLPRSAEMVVGIVATLLAGGSFVPLDPSWPQARRESVTHDASL SFVLTPDNCALTEDALFDLDATRELFTPPSTDSVAYVIFTSGSTGRPKGAMI RHGAIVERLLWQRDQILFFGRDDASLFKAPLAFDISINEIFLPLVCGGRVVV AAPGVEQDPQRLARLIHREGVTFAYLVSSVLDVMLKQAEGTNLLDSLRHV WCGGEMLTQALFRRFRQQLAIPLYHGYGPAEATIGVSHVIYRDDEDRLNT SIGVANPNCRLYVLDEHLRVVPDQEIGELYVAGFLLAKGYINAPGLTASRFV ADVFASDGTRMYRTGDLVRRHNDGSLEFVGRADNQVKIRGMRLELEDVE SALVGHPDVEAASVIAREGRLLGYVTVTAGLVGAAIRSWCAEVLPEYMVP AIITVMDELPRTANGKVDRKALPEPDWSSLTDAPADAERDGETVALLAEA MAEALGVSAVGAETDFFDIGGDSLRAITLVSALGRRGVEVSVGDIFSARTP QQLARCAEERGTFVQDPDDEPTGEVQSLPILRWFDSITDHVDGFIQSVEF SVPEDVDAQLAGRMVADVLQAHPALRARVQRNPLRLELPEESAEEAVAIH PTTDIAALSELLDPAVGVVVAAGLVPGRLRLVVHHLVVDGVSWNIIGEDLA AAYRGEQLAPERTSLRRWTQLLQQAVDTGEFAEDANSSLPPLPSADEPLR DPQVPALADSPEKAPTVREERSVVHEASVQITDELLGAVPHAFRTGANSV LLTALSVALARWRRNKQTWTLVEMEGHGRETRFVPGPQGREADLSRTVG WFTCLYPMLIDPTRAAVQEASTEVEGSIAPLALALNAVKDQLAAIPGNGVP YQAHTWLHQSAKTPPQAQVLFNYLGRVSAGAQDFAPAGSTGQLGEQRD PDQPLVRELEFNAIAEDTGEGYVLRTTISWARGRISPERIDELVAHWDVAL REVAALADHGVLSVGDVAPAPVNSADLARITAQSSAELQDVLPLTPLQHG MYFHSLFEESASSYVEQQVLRVECSEPFDRERFARAARNLIRRHPALSTR PWETDGGDVVAVIDPGIAEHLRVDFRDVTVPAELAGPGLDGWLVQRTEEI AADDLSRGISLQPPGDAAPEPLMRWTVVLPTSVDGAVCGQDIAVIQTVHH LIADGWSVPIMLRDLLEIYRDDDARIPRYDPDAGMAGSVRWVARRDAEAD MSVWREEMREVRPTVLCPNPSSSLERRELLVDDPRTVGLSERARVAGVG LPDVVHAAWGLVLRTLVGCEPGADVVFATSVSGRDIPVLGVADAVGMQL NTIPVVAPGQSDPTLPVTSMLHAMVHHNNQVRDVQHVSLADIARDLGTNA SELLDTLMVVEVPLSPQDVGCPGSPLQVADVRNNGAPHFPLSVVVNPSAE HPLRLIYDPQRITEVRAERIAQMLAVSVDSLLSESGAVATVGEVAEALGAV SGVDTLPSLWRRSFEHSRDRPALTSIGEDGAAEHWTYEELDDAAQRIRAV LDRKVAIHTPRVALLMERDAWQVAAILATTMSAGTYVPVDPLSPQARVELI LEDCQPDAVLVSPSAEKMVSELVDCPVLVVSEQTMSGEAKPPAGRSASV ARANDIAYVIYTSGSTGRPKGVAVTHANVTAMLGNARSHVEFSQEDVWSI SHSFAFDFSVWEMWAALSSGGRAVVMPYALMRSPEDAAEVLRAEAITVL SQTPTAFAALEPHLGQDSAVRTVIFGGEALEARAEAAYCSAHPNVRFINMY GITETTVHVTAHECSENAGEARSPIGRPMDGLRTYVLDAQLQPVQPGETG MMYVAGPQVTAGYWGLASTTASRFVADPFVGGGARMYCSNDMAKVLNN GHLDYVGRADRQVQLRGYRVELGEIESALEKVSGVREATVVVVDLPEGQ VPGALLITDSRADAKAITSRAAAAARDALPAYMVPQLFAVSTQVPQTINGK RDERAILDLLGEIPAAQQASGTSDLVEAISQAIADALRLDRAEVEPDSDFFR LGGDSILAIRVTHALARADLNVTPRDFFLGRTPRKIAERVTPQATQQKIQET RQSQEEALPKDGHQEISGAFPIPAMLRRQMERGMSDRFVQARRLDLGPV AVEDLEQALQQVAQAHPMLRTRVDTSSAFAHFVGAEGDGPMVVRALDVD

						ALIDHIDIAAGRSVVLGCVDGYVKAVAHHAVIDSASWMILEDDLRDALAGR RILSGQASYKDLCLQELHDAHTAAAQDALHWSELATLPRPVEDSNQWSTD HVSTLEVNIDGQTAQVLQSVAPDVMGVDVQDLVAGLVTVAVARTMGTDA RSWAVDVEGHGRPADHSYGRTLGWFTTIAPVEIPLADPADAARAAAEMR ALHEDGTAPDRRTYQALRYTHPQGQRTLAHGAQILVNYLGRGETGAVLHA PGDAACQWTDYLVEADVWSSERAVSMELSVVNAVISAERLRSAIGEVARE LQEHVSREQSSGRRAPLSVLQRGIWFQSQVAAPGAYVAQTALTFDRRLD AEAVVEAFRDTVTVHPAMGAEFHTDASGQPVQNLPWSGQGIDLPVETTE GDLEAIMHADRSAGIDLASVPLAKATIVTGRSDGQPGDTLLLTYHLVLVDG WSRAVMLQTFLERLTLRSGQARTQGAGELVPRGCSIADALLESVDTRKEQ ADSDYWVHRLETLSQPTLVAPQAADISSEHAGSELPRQVFAEVSEQLTGA LQEKIRAQAVTLTSVVNAAVAVALGAVTGDSDVVFGQSVSGRDALSDPAM SDVVGVLLNTVPVRVTARPGQSIEELVQDVYRQRIEDMDHDSADLGAIQR QLGVGTLFDSMVVVQNFLDPQAAAELRERHGVVEERAEDSTHFPLTWVF TPGPKLGIKLEFRHDVVDESLAHSVLKAATEVLTAFVDTPEVPLAQLAQLA PARAEDSSPQSTTEQMASWQKAEGIDRTIADELKDTAQRFPDRIALADDA QQWTFGELIARCSDIAEKIKNCGVTSGDTVAIAVERSAHSVVALLGALWAG VRYAPLDLTHPDGRLRVLVEDSQPAAALVDSSSRERMERIGALPCVDVTT ADSHATTHTPAAVPGDDAYLMYTSGSTGKPKGVVIKHRGLHNMLDNHRR KIFAPAAADGRTLRIAHAISFAFDMSWEELFWLVEGHEVRIFSEDLRRDAA AMVEAIRAHQVDVINVTPTVAEQLLAEGMLESGAHRPRLVLLGGEAVSHG VWETLRKADDVRGYNLYGPTEYTINALGAGTDESATPVIGMPVDRTAAFV LDPWLRPVPTGAPGELYLAGSGLAQEYHGLAARTASSMVACPWGAPGE RMYRTGDIVRVRADGMFEYLGRSDDQVKIRGHRVDPGDVSAAVSRGVDP RILHCVTVPVRISDATLLACHLVAPQLRDADQGERQSFLTGVRNALREELP SYMIPDRWSIVDELPVTSNGKTDLAALGEGERITEKGREPANETEEITAELF AESLDIEPEDVPVDADYFDMGGHSMAVIKLCALLRGELGVEVGVREFYGL RTVERVAEFVEARS
Additional biosynthetic gene	WM42_1382	pyridine nucleotide-disulfide oxidoreductase family protein	1411510	1412929	472	MTIIDRQAQTSAERSTHENSVRDIVGIGIGPGNLGLAVAIEEQAPELDALFV EARPEFNWHPGMLLEGSNMQISFLKDLVTVRNPQSRFSFINYLHHSDRLID FINRQTFTPERVEFADYLRWIADHITVDTQYNTTVTSIETLPELAADGARFV VHVRRKLGSGESEGQRAQQSESIRCRNVVVARGLEAKMPAWAEDSSLDT SRIFHNIDLLPRTKKLLNSGWDIRRALVIGAGQSAAEAIRYFHDCPHIDTVT GSLNSYGFIPADDSPFANRVFDPEAVDDFFHAPDAIRNELLIRHRYTNYAC VESELLDELHDRQYRESVTGRQRLDIRRTTEVLGARNCSDGSVDVDIRHR VTGDVVTENFDVVVCATGFRSRGLAGIHADSHGSEEFTVTRDYGAVLNGE RVPGLFVQGATEATHGLGSTLLSNIATRSGELVEAITGQQRTHRAPADEDL RSEQHRDSSHLIAG
Transport-related genes	WM42_1383	ABC transporter family protein	1413008	1414757	582	MSRGGRRVSDNKVVTLRGLVRKNSRAMALSGLLLSIYHVAEAMIAVLLGW LAHSLIASENVWHLVGGIAALGATLATVSVSWQTGFRILQATSARNVCELR AGITSQVVSHGGHSDDADSLPRQELPTVVGEDVVQAVDIIEVVPVGISALV GAIFCTIVLALIDLPLGVVVLVLSAVVLIVLHRLSQIIERRAERQQTLLARVTA RMTDILQGLPVISGVAGAHPAYRGYARKSEEACADARKLAWVSGGYEAV AMGSNVLLLSAVGLYAGYRTISGDVTLGELVTVVALSQFIAEPMRQCSRM PRFIGLARASVRRLQRVAEAQRLREGQGVPSAQIGTPAISMRGGDGEAEG

						DGEVEGSEQASVAFAEGRLTVVHCSAAWADALVDALVAGESMELGSLTR PQTQQLWIRGRDICEISVDDVRSAVLAEPRKPALFGDTVGQAVLRSSGEG RAEDAVDILNALGLDELAPGAAHSAGVLDHELTEGARNLSGGQRQRLGLA RALLAEPEMLVMVDPVSSVDSMTGMKVARAVRDIRRGRTTVVLCVGRAF QSVAEDVVDVKPLAGSLRTRGEQFSLGGC
Transport-related genes	WM42_1384	ABC transporter family protein	1414728	1416546	605	MSAVLPTALPGESAAEIGRLLRRTATPAISAVLFTFAGALLSLVPIYLLASVID AVAAGDGKSGVLKVIVWAAVACVGTAVVAGLAEALTGVTIAQVVAKLRER AVAAVLNLPSTTVESLGRGEVLGRVGADVAALVSSARKSVPSTLSALVMV VVASAGIAGLDWRLALAGLCGIPFYALGLRWYLPRSAPLYRRQRELEAGVI GSLQGSMEGIRSVRSHRLVDSRQGLTRRYAQASRDESIAAFRVFSGLVAR ENFAEFMGLSALSVVGWLLFREDAVTVGEISAALILFHRQFVPIGTILFTFDE LQRSGAALGRIVGLIRSAGADTPQPIDDYSSHRQSPAVEVKGLNYRYEDG PEILHDLAFHIPAGCTVCVVGGSGAGKSTVAEIVSGTLEMAEPGVITVGGC DVVGMSAQERSSIFCVASQENYVFAMSLRDNLLLAAEGASDAEIWDALRR TGAEDWCTSLSHGDKQGLDTMLGEGGLHVDAVAAQRLALARVALSRAGV VILDESTAEDDGDLEETQQSHEAFSMSLEDAARAAIRGRTAMVIAHRLSQA TSADSVVVMERGRVVETGTHEELAARNGTYADMWTAWNEQGRQARV
Additional biosynthetic gene	WM42_1385	protein mbtH	1416546	1416777	76	MLSSNPFDDEQGSFFALINDEGQYSLWPTFAAVPDGWTVALGDPSRGVD GGVSRDEAMEFIDREWTTLQPAGKSHA
Other gene	WM42_1386	putative transposase for IS3503g	1417045	1417579	177	MTTNRPSCPLCGNNTKKNGTTSKSTTRWRCTHCGHSFTRNTQTHNKNTA TMALFIQWATGTQSLTTFAAHHGVTRQTMHHRFRWCWWIIPTPTIDSFRIH DQIFLDATYLKSGCLLIAASKTHVINWTWARHETTAAYTELRLPSFSGHPIC TDLVSVREDVHCESTAQEVHAGVPA
Other gene	WM42_1387	transposase family protein	1417535	1417844	102	MSQQRKKYTPEYRREAANLVIESERPIAHVAKEIGVSAGLLGRWVKLERE RRGSSDGMSEADLRAENARLRRELAEAKMDNEFLSKATAFFAAKQREQK SSN
Other gene	WM42_1388	integrase core domain protein	1417843	1418743	299	MQQEKANYSIKRMARLLKVSRSGYYKWAHAQQKRLSGEDDRAAFYDDV DRKIHQIWKDSDEVYGAPRITAELAERYQITLNRKTVAKRMRMMGIEGISP RAFVPVTTIQAKRKSTLPDLVKRMFDTGQLNRVWMSDITYLRTGEGWLYL CAVRDGHSRRVLGWAMDSVQDTHLVERALRMAHTLRGDVPDGLVFHAD RGTQFTSEKLWEVCRNLGIAQSVGRTGVCFDNAMAESFWSTLKTEFYDR QRWATRDAARKAVAYWIEVVYNRRRRHSALGMVSPVDFENHIGLTTSRK EIAA
Other gene	WM42_1389	transposase, Mutator family protein	1418774	1419557	260	MRATPELLRPIAAPLIAVTDGGQGAQSAIHHCWPTTRIQRCLVHAQRTVRR HTTSNPRTDAGKTLYRLALKLTRITDLDQASTWVAHLHEFDHTYREWMNE KTTIKDPATGAYTKVYTHQRVRAAYQSLLSLHRRDLLFTYLQPPPTTIDPDN LAATTNSLEGGINAPIKELARRHRGLSLPHQRTVMDWWLYLHTEVPDDPV KIARDQRWGQDALSTATDLITHNTTATTNDIGAPAEYDTAIDTSYQHNLGIQ KGWIK
Other gene	WM42_1390	abi-like family protein	1419719	1420418	232	MNSNENLTINLDKWFSPARMSTYAHHPDPESLYLWNTRVTKAFLEDLQHV EVLLRNCVDAAVAPRYGPRWYTHPAIPFEKPAERAIKKAERRACTRREQA

						PPPGRVIAELSFDFWAYLFTKTYASTLWPLVRKDLVGTQAPATGGSKPDV LVPSLDEFRSEVGVVYNLRNRCAHHEPIIKKDLQDENDNLDRAQQAIEKLS TWIDPSAAEWIVANSRLPEVRDDRPSPKVWG
Other gene	WM42_1391	putative transposase family protein	1420650	1421655	334	MGKPQRKTHRSRKSVTIHPDSRMGPKGCPPPRGQPQITGYKHLYLANLF SHASKPDKKDRILRVRVHGSTRRLVKALQNHPDAKISNFTISQKGGYWYV AVMVKSAQRQPSLTRRQKQADTLGIDIGIHNYLSLSDGSTVIQLPPLLQRNI KRSKNLRKKLARSQKGSNRRRKLIATIKRVDHELKLQRDGFVHQITAELAR NYALIGIEDLNVKGMTRSARGTVDAPGTNVRAKSSLNRRMLEGIPGEFRR QLEYKTKRTGAALEVIDRFYPSSKTCSRCGWKNENLSLSNRQFTCLECGL SLDRDHNAALNIAAEAERKHLKDTKSPKPSHD
Other gene	WM42_1392	hypothetical protein	1421822	1421936	37	MTIVSRGLVATAVLLPPGDAGFRELKHLSQMKNGAVF
Other gene	WM42_1393	helix-turn-helix domain protein	1421894	1422206	103	MTNSPQAKKSSPNTNHDHEGFVFRLIPTPQQEALFEEMTRAAHWGFNAF TEAWQKYDENYRARKDQLLAAGVNASSVNKLIKKEAETNPALKNRTVFHL RQMF
Other gene	WM42_1394	transposase IS116/IS110/IS902 family protein	1422910	1424119	402	MAYDFVIGMDVGKYFHHACVLDPQGRQVLSKRINQHEGSLRKLFGKFLAN DAEVLVVVDQPNNIGRLTVAVAQAMGADVRYLPGLAMRQLSRIHVGNSKT DVRDAYVIAHAGLNLPDALRSVDRVEEVFLQLKVLNGIDEDLARAYTRLINQ MRSALVGTYPAFEHVLRGQMIHRKWILHLLAKYGGPTKIRRVGKARLTAFA RGHRARNPEPVIDAMLAAIHGQTVSIAGAEYAELGVAMSAKDALAKLEHRK EIEAQVLELIQDIPQTEILLSMPGIGPRSAAQILMTVGDMSDFPNAAHLASYA GLTPQTNQSGTSIMSNSPNRAGNKKLKNALWQSSFASIRFHERSRQFYER KRKEGKRHNAAVVALARRRLNVLFAMMRSGELYRDIPTAQEAAAA
Other gene	WM42_1396	FHA domain protein	1424904	1425738	150	MDSVVLLALRIGLLALLWLFILVALNAMRRDANKAAGVYQPSAQAKATPRR REAPKQINIVDGPLRGSHMQLGTLEECTMGRAQDCDFVTGDDFSSGHHA RLFRRGSEWFVEDLESRNGTFVGGVRIDQPERVGVGTDIKLGRTTVRLMA
Other gene	WM42_1397	FHA domain protein	1425766	1426219	277	MAIFEKLAKLDSAMQRGLDNGMAFVFGGKVVPAEIEELLKQEAQDNLARG DDENLYSPNVMTVGVSSKDLENLSRDPDLPADFADQLTRFVRNHGWSFA GPVIVRIAEESGLRTGQLRVSSFIDHEPAEETGFDAIFHEDDGQEDHMSNP FNADEAATTTFMAPDQNPEQPHSQGPAVNLMLQDGSSRVYHVREGSNII GRSNDADLRLPDTGVSRQHAEITWNGQDAVLVDLQSTNGTTVNETPIEN WLLADGDVITMGHSHIEVRITGTENNHNY
Other gene	WM42_1398	phosphatase 2C family protein	1426215	1427559	447	MTLSLKFTAITDRGLVRQNNEDSAHAGPHLLLLADGMGGHAAGEVASQLM VEHLEHLDQDPEDNDMLALLGAAAEDANASISAHVEQHPETEGMGTTLTA MMFNGTEFGVCHVGDSRGYRLRNGKLTQVTRDDTYVQSLVDEGRLDPE DVSSHPQKSLILKAYTGRPVDPTLFMLDAKVGDRLLLCSDGLSDPVTAATI ELALSQGTLDDAAVTLRDLALRSGGPDNVTIVLAEVVEGKGDSSPVKVGAI AGVVPEPTHPDSSASRAAALTRPREASPAPQEEEDDTEEPHSPKKKVGW KVIGALVVVLALVIGGGFWTKNEISNNFYVATDEEGALSIEQGVDFSVFGR SLHHKYQNACVDKANTLRLGSTPCEGDFAPFKVTDLPDSERAAVDNLPSG SYAEVQTQLSSLSSKALKPCISTPSKKTDEKLEKKTAKPGITCREVS
Other gene	WM42_1399	cell cycle family protein	1427559	1428912	450	MKKIFSRGTELGLLILAAVVFAITTVSLELSQGNILTMDVLYLIGGFIGVFTVA HLVLCFLAPHSDQLMLPIVSLLNGTGLVMLARLDLVNDGVLARRQVMWTIV GLILFVLVLVVLRDHRSLTRYSYILGAAGLILLALPLVWPQPPGVEARIWLW

						LGPFSIQPGEFSKIMLILFFAMLLTQKRSLFTVAGYRFLGLSLPRLRDLAPIL VIWGIAIVIMGISNDFGPALLLFSTVLGMLFMATGRVSWLLIGLLLVGVGGF GIYQISSKIQTRFSNFLDPLGNYDNGGYQLSQALFGMSSGGISGTGLGQG HPEIVPVAHSDFILAGIGEEFGLIGLAAVLVMFGMLASRGFNTALKSRDSYG KLVASGLSLTLAVQVFVVTGGISAMLPMTGLTTPFMSAGGSSLMANYMLL AILLRISNAARRPARETSSNAPTDTSMFPAVQEATR
Other gene	WM42_1400	penicillin-binding protein A	1428908	1430336	475	MNKSIRLVSLFAILLTAVLLVNLTVVQAFSEDKYAHNPKNMRGFYAMQTTP RGQIFAGDTVLAKSTEDSEGKYSRSYPTDSPAFSNITGYLSTQFGASQLEA SQNEILNGTDDSLLTQNWLDTISGKEKVGANVEVTIDPALQQAAYDQLVGP GYNGAAVAIQPSSGKILAMASSPSYNTNDLLGENADETWSALQEQKGTPL LNHATQETLPPGSIFKIITTAAGLNNGFNPGSSLTGANSIVLPDSVTELTNY GNQMCGGSQSVTLQTAFALSCNTAFVEMSESIGADELRKYAEAFGVGEK YDLGVNSAAGNLGELADGAQVAQSAIGQRDVTMSALQAAIMAGTVANKG KRMEPYLINRITDAQMNEIRATKPRQAEQAVTEDVANTIKELMYGSEHSTF GYDGNSFASKTGTAEHGDGLAPHVWYVAFDPDRDIAVAVVVKDGGNLGE SATGGQVSAPIGRAILRAYGGQ
Core biosynthetic gene	WM42_1401	putative serine/threonine-kinase pknA domain protein	1430338	1431772	477	MSNSDNKEHLQALIGDDYQLQWIIGHGGMSTVWLADDVRGDREVAIKVLR PEFSDNTEFLSRFRNEAQSAESITSENVVATYDYRELEDNGRTFCFMALE YVRGESLADLLAREGALPETLALDVMEQAAHGLAVIHRMGLVHRDIKPGNL MITQNGRVKITDFGIAKAAAAVPLTRTGMVVGTAQYVSPEQAQGLDVTPA SDVYSLGVVGYEMLAGKRPFSGDSSVSVALAHISQAPEPLSTSISAPAREL IGISLRKDPTTRFADGNEFTNAVAAVRNGQRPPQPKSGALAPMAAEPSPS ASTEMLASVANPKTATPERAPVKDTPHPVPKEKKKSSGFGVGLLIAAALAA LAALGFFAANFFKGAGESTEPTTPPESIVVTEYRDPVTTEVVTPEETTEDT PEQVTVTTTHSVEEETPTQEKTTQTQPVHTPRQQEIPTTVPQETEVNETPT SPTAVDSLPDNPQENLLPQDGA

Appendix 100: Summary of genes in BGC possibly coding for NRP CsimD as predicted by antiSMASH.

Type of gene	Locus tag	Function predicted by antiSMASH	Start of gene	End of gene	Amino acid sequence length	Amino acid sequence
Transport-related gene	WM42_1794	periplasmic solute binding family protein	1863380	1864334	317	MSFSLKSLAAPAALLCAASLTLAGCSSTDSADEDTIKVVASTSIWADVAQEVIDSAQRQDVKIEVEPIVKGNGVDPHHFEPTAADIAKANEADVLIVGGGGYDSWLYQAVKNQDEIIHALPLTDHGKLDEEPDVVTAAEAKVIAKKEPSKVTNIEGNEHVWYDAAAIEKVATEVADLVNEANPEAKASADPLLERVEDIKKRVEKLPKLNYAQTEPIADYIMKYTPALDVTPEGFRKATVSEGEPTAADLARFLEAIKEDKVDLLIYNPQTETDISSRIHKAAEERHIPIVEIGETPPEGTNFLDYYEQAVDHIEAA
Regulatory gene	WM42_1795	bacterial regulatory s, lacI family protein	1864452	1865562	369	MVKRSQNRKTLASLAAELGVSRTTVSNAYNRPDQLAPATRERILSAAAAHGYTGPDPTARSLRTRRAGSTGVLLTEHLSYAFEDAASVDFLAGMAEASTGAATTLTLIPAGPDAPSDDAAALLTRAAVDGFVIYSVSSESVEAARTRGLPLVVCDQPKDSGLPFVGIDDREAIKPAARALLAAGHRRIGILCIRLHHERIDGFVSPEQVERADMHVQRDRVLGALDVFAQAGIGEIPIVTRHINDASTARAAAEELLAARPELTAVLCTTDSMALGVLGVLRDAGLAAPKDLSVTGFDGITTALLAGLTTVVQPNKAKGAAAGHMLAEHIDAALSETAPAHDHKLLPTRFNEGTTVGSPREQHLLDWRL
Additional biosynthetic gene	WM42_1796	trehalose-phosphatase	1865677	1866424	248	MIEKLAATKHLAVVSDFDGTLAGFANDIYEVHAEPRSLAALERLAQLPDTTVAALSGRHLEGLKRVFPLREPVLLGGSHGAESSWQDSSLSPEAQAHLDRKEAEIRELMERFPGAEIEIKPFQRVFHLRRLELTDPELAAEAYAAGRALDPAGFPMTAGKSVIEFSATQATKGSWISELRERVGATATVFLGDDVTDEDGFAVLNQPPDLGVKVGEGETLAAHRVPDIAAVSDFLEELAAARAVHLGA
Other gene	WM42_1797	hypothetical protein	1866444	1866819	124	MAIYTKPDEPSAIPQTAKAVPQMSFGESSIVGSTGCAPFQAKVSYSEAEDAANIRDADTMHIDKVRMDTRPDDCTGSALWADNLLRNLLAEDHDFEVRLNPNNQLVLTLDTAEVDSPAIRMVSL
Other gene	WM42_1798	trehalose-phosphate synthase	1866938	1868369	476	MSGHDNSFVVVANRLPVDLEIQPDGTRTWKASPGGLVTALSPVLEAEQGCWVGWPGITNDAPEPSHTENGVLLHPVKLTDVDYEGFYEGFSNATLWPLYHDLIVTPQYKREWWYSYREVNLRFAEEVAKVAAPNATVWVQDYQLQLLPGILRQLRPDLTIGFFLHIPFPSADLFLQLPWREEVVRGLMGADVIGFHLESNARNFLELATRLGLEVTGEVSTRDITAGIKVDGRTVGVGAFPISIKSSDMTRFSKAEDRDVAKLRAEFGGHHRVILGVDRLDYTKGILQRLKAFEELLESQALDPSEVTLVQIATPSRERIDHYRVTRSQVEEAVGRINGRFGQIGRPVVHYVHRAVDKDMLRKYYRLADVMLVTPFKDGMNLVAKEYVACHDDGRGALVLSEFAGAAAELKQANLCNPFDIESIKRALLSALKALDDSPDTMRQHMLDLHAQVVQHDVDVWSRAFLSALEKAAKNK
Other gene	WM42_1799	putative histone H1	1868374	1868722	115	MPPKVTDTQPAADQNNAVEEETARAARRIVATYAEDFLDGVTLMSMLGVEPDGLVHKKVLAEQEDAAPKKTSKKSSKKATKKSSKKTTKKATKKSTKKSTKKTAKKASKKTTKKA

Other gene	WM42_1800	integrase core domain protein	1869399	1870296	298	MIRFQFVDDHRTEYSVKRMCDVLKLNRSSFYKWVSTRKKRRLKMYSDAVIGARIKTIFDDEHGLYGAKRIAASLKEDTTYTPINHKKVARIMKSMGLKGFSKRRRCITTRRKPGHRVMPDLVGRKFTAGEPNRVYVGDITYLPCRGGKNMYLATVIDTYSRKLAGYALADHMRVSLVIDALAHAHGVRGSLDGTIFHSDHGSVYTSQAFRNYCSSLGVRQSMGAVGTSADNALAESFNATLKREVLRDRKVFNNPISCRQEVFRWCMRYNTRRRHSWCNLVAPDVFEAETSAILTTAA
Other gene	WM42_1801	transposase family protein	1870292	1870592	99	MSRYSEQFKRDAVALYENNEDLSLNSASAELGINRASLHSWVKKYGTGKRARTKAVHDQAQAANDSARIRQLEKEVSKLREERDILRKAAKYFAEETHW
Other gene	WM42_1802	transposase IS116/IS110/IS902 family protein	1870590	1871232	213	MVVDYLSSQLDWAGYQVSTTWGSGPIDAMLAAIHEQTVSIAGAEYAELGVAMSATDALTKLEHRKEIEAQVLKLIQDIPQTEILLSMPGIGPRSAAQILMTVGDMSDFPDAAHLASYAGLSPRTNQSGTSIMSNSPNRAGNKKLKNALWQSSFASIRFHERSRQFYERKRKEGKRHNAAVVALARRRLNVLFAMMRNGELYRDIPTVQEAAAA
Transport-related gene	WM42_1803	ABC-2 type transporter family protein	1871775	1872597	273	MTNHTTATVSHDTTWLSTVATHFGRHLRAARRDSAVITNTVAGPVLMFLVMRWLFGDLMAASQGSATLDALPLTIALILSSELMNGTSAAAQIIKERQRGITTRIATTRYGTSPEIFGRWCFDSLRSLISGCAVLLASVASGLRIHTLAGFGWVLLVIIFGAVVAASLSAMVGAMCATPEAAIGPAPIIMAAMALNGGLVPVEQFAGVVQPIARWNPLTFAARAGAAIDGTPSAEILATGQPGTAVWAFVAWMVALTVVLLAAAAAFRRPTMS
Transport-related gene	WM42_1804	ABC-2 type transporter family protein	1872589	1873351	253	MNALGAVAASWHRNVLRTVRNKAVMVSAVAIPSLFMVIFYATFAKASTELGIDYAAFLLPAGVIQAIVFAAGGSSLAITRDAENGIHDRIRATGTPAWATVVGRLLADLTRAAWSCSIVVLTTILLGARYAAGPGRIILTIALFAALTIILSAFIDGSCLLAPKPTSASLLFQNLVLVMVMFSTAFVPADALPGGIGPIIRHVPLSPILDTARNLLGGAALGARGVEALCWLIAMTVVGVWGFITAFQGRKHD
Transport-related gene	WM42_1805	daunorubicin/doxorubicin resistance ATP-binding protein DrrA	1873347	1874412	354	MKENPTTATASGDPSATAVPSTSAHAKGRDLVLDVQDLRCSLGKVKKRTEILRGVTLAIPRGGILAVLGANGAGKTTLVNVLSTLLPATGGTAIIDGHNLATDPSKVRASIALTGQYAAVDEELTGKENLIFFGRLAGLKASDAKARAAELLERFDLVGAADRLVSAYSGGMRRRLDIAASLTVPPALLFLDEPTTGLDPAARSSVWDTVRGLAADGTSILLTTQYLEEADQLADRVAVLAGGKVLDEGTPAELKDRYGTTVCLITMSSPDDAARLTHTLQEHRIDCRQEDTVLRVDAPDGHRTLVRALALWDGDDTSVADVSLVPPSLDDVYFAIAGSGDIAQQAPHTGGGSQ
Additional biosynthetic gene	WM42_1806	hypothetical protein	1875499	1876678	392	MLPFAHLGVALRDRGHQFVLAGFSEFGGSFRANGVDYIELPGDYRALMKRLLGDSKGMMDTVLGIREMISDAHVFDVLEHAMDGVDVVMYTQFSEVARLLGAARGVPSVRVQVFPTEPCLRYSLVDPRKLDGSVGALVTHWMSNTLMAWAMRPVIAAWRRRLGLRQRALSSPPTTIYQFSPALSPPAPEWKNHIHVTGEWLDPHHSELALDPAVEEFVAAGSAPVLASFGSMVSDRVYDLQRWTRDACIEHGLRAIIVDPDHEPGVREGILTVARVPFATVLPWCRAGICHGSLGTTGAILRAGKPCLAVAFGGDQHFHANAVCRNGAGPNYIDAQRGELSAQSLAAGIADLVSGAYDSAARRMPELLATDPGTVAAVEIVLNQSDLAKGEK
Core biosynthetic gene	WM42_1807	alpha/beta hydrolase family protein	1876788	1877460	223	MGVNPSIDVVPVELPGKATRRREKWVSDFDKLADMCATEIVDLAAGAPIALYGHSMGAALAYQVAACLQQMHRPTIPEAVVVAARQAPGETVPGEYHSSMGFGALRREFEKVGGTPPEILANDDVMKLLLADIRRDYVLHEGFHHATTVLRSPV

						LALAGDSDPAVSPEMLSRWENFTSGSFELKVLPGGHFFPLDTGTEFLDLLA GFLAGITGNTAWLARNNA
Transport-related gene	WM42_1808	ABC transporter family protein	1877487	1879254	588	MTEQKPPITFRQISQPAHRQITVSLLLAVLAAVLSLVPVIVLVELVRTVMLSFQ GEPIDASKLWLLVAALMVATVLHGQATVKSLQVSHQADGLLGEHLRQQQITK LGRLPLSWFTRTPSGTVKTYIEDDVTKIHQLIAHAPHDYSAGVLVPLLSLGYM FVVDWRIGLLGLVPLLLAVATMPFMMREFQEKAEKFKSSQKQLDAAVVELV RGIPVIKVFVPEGYDESIFLSRSRSFGGFYREWLHATVHPTALMKIFTSTGFG LLVVAAASPWLITSAGVPVADVLAAFLFINNIAAPLLMLSRTNIMFSEAKAAAA DLTEFFNIPVLPPAAGGREPANAGIELKDLSFSYVPDTPVLSGINQKLTAGSIT AVVGPSGSGKSTLAALIPRLLDPTEGSVRIGGVPSTDLRSDQLYRRVGFVFQ DPYLMRMSVRDNIRLAHPEATDADVVRAARAAQIHERITHLPRGYDSVVGE DAQLSGGEQQRLAIARSILLDAPILVLDEATAFADPDSEAAIQRAITELVAGRT LVVIAHRLHTITGADRILMLENGEVTEAGTHEELVAAGKGYATMWARYQTAQ AGIQGEEK
Transport-related gene	WM42_1809	ABC transporter family protein	1879253	1880978	574	MIRTLYSLGSPADRRRLVLVLTLIGISAVALAIGLILIALFLDTLFSEDPAQAAA WLPWIIISVLVYAAADWPTEVIAQDLGHDYVLRIHRLIADRTAQLPLGYFEEDR AGQIGVVATSGALFAANAPAMMLRPMLHGAASAGLACVFLIAVDWRLGLVT VAVAAVVWFAYNRLMRQYRVAERQKGERNEHGAAEVLEFAQVQPVLRTAG PDSLGERAVRASIREQLSAQQHTQKTGEMIMGRLALIIMVGTVVIDALATVLL LNHWLEAGTYIGVIVLVFILARVAMSGIPYGEGLESARNTLDEIQKILDARVLP EPAVPAAPRDYGIEFDGVGFGYEPNTPVVRDVSFRVAPGTTTALVGPSGCG KSTLLKLAARFYDVDQGTIRIGGVDIRDLGSRAVLDSLAMVFQYVYLFEDTLY ENIRLGCHNATREEVLRAAELAGVTEIARQLPQGFDTLISEGGQNLSGGERQ RVSIARALLKDARIVLLDEATSSLDVQNEHLVMRGMKTLSAERTIVVIAHRLHT IRDADQIVMMSPSGTVESIGNHEELMESSPRYRSFWGEKSDATSWKL
Transport-related gene	WM42_1810	heme ABC exporter, ATP-binding protein CcmA	1880981	1882406	474	MLHHVNLTVAKGELVLVCGESGCGKSSLLRLINGVAHTFCDAQISGEVLLDD EDITHAHPHDIAERVGSVFQNPKSQFFTLEVASELAFGCENLGVAPNEIRQRI GELSADFGMVHLLDRHLFTLSGGQKQKIACASVAAMHPQVLLLDEPSSNLD LAAVDELRRIVAQWKTQGRTVLIAEHRLSYLVDIVDRVLVMQDGQIAHDLSG SEFRTLDKAELYRMGLRSAQQVPEVTREPRPSSGTMVLKKLQFTYPKATEP SLSISHTELPQGQVIGVVGRNGAGKSTFVRVLTGLESKATGVVDINNRSLSG PRQRLRQSYLVMQDVNHQLFGESVESDVIIGTSGPDGKNEARLTEVLGALD LADKRARHPMSLSGGERQRVAIASAVLSEREVIVFDEPTSGLDLCRMHRVA HLLDALARQGKTVLVVTHDMELVARCCDLLLRVEKGRLTACEPCDDLALART VRYLRGRDS
Other gene	WM42_1811	cobalt transport family protein	1882456	1883137	226	MLVILVVSSVLISPAGSSGALGSTARWMLIAVPGALFLASGMVAAALRYALTF AVLAGFPAVVVQALPERHIIVDVCATWFAGLSLILPGITCCWYLLRTVSASEF MAAMQRMRSPDALTIPTSIMFRFFPTILEEYRDIRTAMQMRGISGLRNPIAML EYRFVPLLASVVSIGNELAMSAVTRGLGSPRRRTTLCEIGFRVGDAIVISVLV VSLILLLVNIVMLP
Other gene	WM42_1812	conserved hypothetical integral membrane	1883202	1883829	208	MPSTPHTTPKDAKAGRLNITDLMNIGIFFVINLVAGVVIAFIGITPITYVMITSVQ ALILGIPMMLFLAKVHKPGMVFIFLLLSGLASLLLGLGIWPLLLSLVMAVFAELI LRAGKYTSAIHSVIAYAFIAVSPTASYIPLFFTTRRYIESSGMQTKYGASFADG LTSIGQMTWLFLIIVIVTFVCGILGGILGRRIFTKHFQRAGIA

		TIGR02185 family protein				
Core biosynthetic gene	WM42_1813	amino acid adenylation domain protein	1883920	1891492	2523	MIKEEQIREELLASLHQILGEDAEIGIDDNLLSHGLESLPTVRLLADWMKQGH RVSFGDFMRAPTVRQWAKMLVESTPNHSTDAIESPKDGFAAPIDDSVPFDL TDVQYAYWIGRNSSQQLGGVGTHGYVEVESRSINIDRLQQSWLTLLRSHPM LRACYTEDGKQYVLPEPPHPTILVHDLTKMDESTREEALLSTRERLSHRLLDI ATGHVVSLEVSLLPQDVAVIHFDIDLLVCDVQSFQIILHDLAHHYATGEAPDA DPSWSFARYLAGHAREGVADIDRDQAYWRNRLSELPGAPTLPMSHGTNEE QAHRFVRRSRSFDSATWSRLREVCEHHATTPAMVLLTAYARTIGQWSENK KFLMAVPLFNRGSDTAIKNVVADFTALTLTSIDQSTRRTFSEDLKDIQASFYE DSSHSQYSAVRVLRDLRASRGEQVLAPVVFSCNLGDPLVGQEFIDTFGEISY MISQTPQVWIDLQVFTTVNGFLIVCDAVEQLFPEKMLDDLFATLVMEIDKAIT DDLSHSDPVESPGAQARRASRAEVASWRLPDTTLVDEVIAAARCHPQATAI RSASGDVITYQDLEEQATTIASALVNSGVGRGALIAVMVERGPRQIIGALAAM MAGGAYVPVSLQQPESRIAALLGASQVTHLITDRPDKVLSETAVQVVDFTSA TGTANLPQLHPQDPAYVIYTSGTTGTPKGVEICHGAAWNTISEINRRLGVGP TDRLLSVSSFDFDLSVYDAFGLLSAGGELVTIPDDARRDAKKWVSLVDSLGI TIWNSVPTLFEMLLSAADRTPGKLSSIRHVLLSGDWIDTSLPERMRTVTPQA HLLAMGGATEASIWSNGLDLDVVSPEWTSIPYGRPLAKQMYRVVSSNGQD CPDYSVGELWIGGLGVATQYVGDPGLTETKFVISEDSRWYRTGDMGRFWA DGTIEFLGRSDNQVKVRGHRIELGEIESACEALLPIERAVCITHQGASSSPSL VTFAQFTPSHVARTTPEQFATSLRAKVNDVLTEGDIRTSVEHDEHLQTAYAF SVMRRWEEQLTGVGTPNHLREHRNRWQTWLGKADEHPATADLLLDDESF GALERFVTPFEQAFVMAEKQRSIAEFIQSPDSMSVEQFLATRPLGRLVHRVL GAVVRECSTHSTSELKILEIGSRRPEASADYAAIAGTSAYVLADPYRHHLEHA GQRVGNTFTYRQLGVTSTPQPIPGEAVTKADLVLCNQTLHQSEDIEKTLCEA WGLSAPGATMVVVEPTAPSPMSDITAAFIANNTTDARAETGTVLLSARSWK EILQRTGWKPVEHVEITKTTALIIAERASSNESVTLCDSDYAKATNLLATRLPE YMLPKRILELAKFPLTSNGKIDRKALTALVPEYFDNEPAVTELPHTATEKRLIDI WDELLHTSSNVNSDYFRLGGDSLTATRLRRTIEQCFGVEFPLENIFDVPLLR DMAARIDQIAEVPHQQSDLPKIVHGSEQYAPFPLTEVQQSYLIGSSGAIELGD VSSHCYFEMSTACLDPERVEDAFNALIKRHPMLRTVVCEDGLSQRVLPEVP RYRIALIRSGNADNEDTLDEIREEMSRQKFDPTQWPCFDVRYVAEPDAGRLL LSFDNLFIDGWSMFHIFREWKQAYDHGVDSLDPAIPYSFKDYVEATIELSHS DIHKRDQAYWESAVDTIYPAPQLPVTDTNGANTSQFCRHHALVDAAKWRRI KQRVREEGMTEAVFLAEVYAEVLARYSDEPRLSINLTRFDRTRFAPEVDHIV GDFTSLSILSVDTQCAPSFRDRAAALHRRMFSNLDHGSVSGVSVQRMLTKQ RGARVTMPVVFTCGLGVVEHPESDQSPYLGVIDHGLSQTPQVWMDLQVYE HDGGLMLNMDAVEAIFPDDMVAELFTSLTATLSHLAESPELWNAPTSTIAPT TNAPTADRINDTDRELPGADKSLLGLYQKGLAEHGDNLAVIDATTQWTYEQL NEQSDKWAQLIAATDPAPGDLVGIMMEKSAQQIAAVLGAMKAGCAYLPLSV DQPVGRNTSIINDAGASIVAMDHPDDDFAALAEHCTVITLADVARHRPGDQA LSESSPTPSSLAYVIYTSGTTGTPKGVAITHESAVNTIVDVNERLGVTPTDRIL GISELNFDLSVYDIFGMFARGATLVLPSPADKRDPQCWADAVTTHSVTLWN

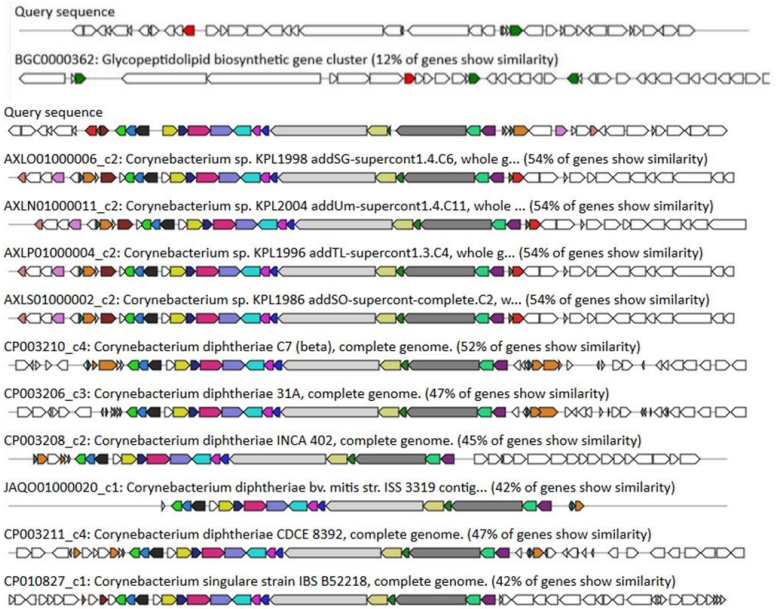
						SVPALFSMYVEHLRERSLIGSSVRSALLSGDWIPVNIAYQVSTLFRDCTVFAA GGATEASIWSNWYEVGVDDASRTSIPYGTPLANQRMYILDEALNPRPTHVP GDLYIAGRGLAMGYWKDPEKTAASFITHPRTGERMYRTGDKALYNHLGHIIF LGREDGQVKVNGYRIELGEIESTARKFNELRDCVAVNDHGIVLYVVTHEGFN MAALNNHLAESLPAYMRPRVISRIDGLPRSWNGKIDRKSLEGKTFEQPQTRE RSRNHRDSGIVTILQELLGPKEISIDDDFFTIGADSLTAVRLTNSIRREMSVEIS IRDVFNHPTVRELSDLIANIVGSDVEEGEI
Core biosynthetic gene	WM42_1814	AMP-binding enzyme family protein	1891504	1893136	543	MRGRDIDRNSVLKHYEGTGHLARKSLCQELFESSEIYSDHVAVIADDAHLTY AQLEQSLCVFTEELRESGLSPGDHVLLQLPNTAAYVVTLLALMRVGAIPTLLL PAHREAEVAALCESLHPVAYIGGRDHLGFDTVAMVEAMGPGELGLKELWAD NGPTHDKESSYRVLPGLFTAPISTKCSPPTKWPDPRSVALNLLSGGTTNMP KIIPRVHEAYAYNTRAAAQCCGVGPDTVYLAVLSTSHDFALAQPGILGTLLSG GTVVLCTSAAFDEAFPAIATHGVTLTALVPAVAQVWVEAAEWFPADFSSLER IIIGAAALNDGLGEAIQDRFGVRIHQGYGMGEGITTFTRIDDPPAVILGTQGRP ISDADELVIDGPGGEPGEILEKGPYTFFGYEGNRDTPDCFTEDGFFRTGDRG YLTEDGNLVLCGRVVEQINRLGENVSPSEVETLLSGTPGISAAAVFPMPDRA LGERTVAAIVAQPGVNRSAILDDFLTRGVARYKVPDQVITVDEIPLINIGKVDK KKLRALAAAQFTDRESEQS
Other gene	WM42_1815	hypothetical protein	1893132	1893312	59	MTPFAIDTHRSTPHPNGITHITDIDFAASEKTLSICKHVGVDGGLTCRSGTGTI DLEFA
Core biosynthetic gene	WM42_1816	amino acid adenylation domain protein	1893733	1899268	1844	MDVTALINDLESRGIALWVNGDRLNYRSPKGSLREEDLAALRSNKEKVLAWL RREREAVPHDEQARFAPFPMTDIQRAYATGQNEGYDLGGTGCHSYAEIRTER LDRSRLEQAWHELIQRHDMLSAVVVPPDSLQVVKSRSLPVLQAVDLAGHNP DVPDAEYLRHRAKLENRSYPLGTWPLHEFQLLQFDECSILQFSVDMIIADFV SVRVMVEELLTLYAGNVLPELEDTTFRDIITSRNHHSQSAAGFAARTNAKKY WSEIIPSLSGKPLLPTLTSADRTSEMPVRFTRRTWRCSPAAWSKLTDAASTH GVTPSATLLTAYADVLRRWSSTSDFCVNVTSMNRDSAIAGINRIIGDFTEMTL HACHPHTGTFSERVHATQEQLSEELSHAAYSGVDVLRDIARTTGQPAVIPVV FTSALGADTPHNNGPAYNLVSGVSRTPQVWIDCQAFQDGGSCNVNWDVR EDVFEPALIDDMWESFTDLLDRLVDDGSAWQETDSVHLPDKTIAIRNRIHKT HVQQTTRCLHDGFWDNVQQHPHQPALVCGGKTYSYQHLAGYVGALQHEL SDVGPGDYIAIVLGNGVWQIAAAVAVVSTGAAYVPIDHEQPAIRQRSMIEAC RPANVITNSHFSEENTDISNINVDTLSPIQYSGTIASPVSPTETAYIIFTSGSTGI PKGVVVTHSAAMNTIDSVNNLLGRNKRRTVLGVSKLSFDLSVYDIFGTFASG GTLVLPLDEESRNPSKWIDFLVDNNVDTWNSVPALFQMLVREVEVTRHPNIL SLDLVMLSGDRIPGTLPAHAAPHFPNAELISLGGATEGGIWSIFHPMTCHTNE TSIPYGTALPNQGMWVLDEACNECPDWVRGQIHISGESLATGYLNDPTSTA EKFFFSEKHGTRMYATGDIGSYRPDGVIEFHGRRDNQLKINGYRVETGEIEG VLESNDFVERAIVLTQETSDPIKLHAFVTDAQSDKDELKDAGQIRNSELRTML EQRWTPADTSLDTGIFATWMRLGNEAAMAALLAAFQQAGVFLVAGKYHTLT EITAAIHPSEEYRELITRWLNILTGEGLATKDDEGWTVSQQTLDFFVFGEAW DQFGNMEAEINNSKELFNYQRHAAEALLSQLRGEISPTEVFFPEGDTHNART IYGENRISKAMNAAAAEAVIGIAEHHADHPVRILEVGAGIGATTEKIVSRLPEN VIEYRFTDISTFFLHKAQKMFAHCDAMTYGLFDMNSDCTSQDVEFGGYDIIL

						CANVLHNSVNIEESFTRLKQLRRPGGVIVIVEPITELYAALISVSIKMNLVDFTD HRAESHKVFIEDAQWDQVFRDTQMHRIAEYPNTSDPLRECGQRLIIVGADD DDVPTLNSEDILGYLRAHLPGYMVPASVNVLPELPLTSNGKVDRKALAQLCL EPVGSPNNRIDPPRNETEEQIATIWRDVLDTTEVGRNDDFYALGGDSLLMAE TVTRLRQEIPGLQQHTWDALMRGVLKVPTIAGISALAQAAGSSCQPEALKAV NSANHTSPELTALSTVASGSPTGSSNLHVYRLPKDAMFCRVMIHAGTGRLK DYEFLMPELLQRQPEIAHVGFTAGDADRFLDYTTRTLIRDLAQSYAQELDEL DMESYQLVGYCIGGMLALETAKALTELGRDVRQVTCISTHQCPHRVTNELLC ELAYGCIFNADLSAMGANFDLKTLAAALEHTLDGINRNISDEELCTLEGPYADI GEFFQKMAVLSPRARRKLIYRSIREFDTDSESTRGMLDILYDVFRHSLLGTID YVPDVYFGDVVVLQPTEGVTGFYPSLGGDIDWPATVLGNLQIHAVAGSHAT CLLQENVPSLLPFFTEREQRNG
Other gene	WM42_1817	thiazolinyl imide reductase family protein	1899249	1900395	381	MNTVRVVVCGTTFGRIYINGIKKLADKFSLVAILSQGSEQSRRLAEQLGVPLC TKIEDLPAFDLACVVVRSSVVGGSGTQLALEFLSRGKHVVMEHPIHKKDSVD CYRMAAKNNVQFKLNTFYRWNPTISRYLEIATTLTRQFKIIHIDAECSIHFLFS TLDIIGRITGGFTPWSFDDDTQVTGIFTTLTGSIRKTPLCLRVVNHNDPEKPD DFAHVGHRITVFTHAGNLVLTETDGTIIWHPNTPIPRDQAGLLSTAADDRLST LQLHENLLIEPCVTRVALYDHTWPAGIAEFLNEVYDKLACSKNEAQEAEYLLA LCAVWARTGQLLGPTCTVQAAMHAEPLSLESLMWPLGESHTSPIERNNPNH RNTGEITWMSQR
Other gene	WM42_1818	NAD dependent epimerase/dehydratase family protein	1900437	1901520	360	MGVLGSHGEVGRQVAALLRASGHMVNCGNRSHHTSDERTAIVDAHDEESV NKFSRDLDVVVNCAGPSCLLKTSVASALPDSVGYIDPFGANSFDNYPTNRP CVVNAGATPGLSGLLLRHLAGLIDDCQSVTLCTGGRDRGGLSGLVDVVLST HNSYGHPGKMIVDGDLVAYQPGSIHAEDLEPFPSDDGLIASPFVTNELRKMA QDFSIPRLIGVTAIPDRDTQQLLLRALSQTDISDPTTLMNLCADIAAAKTKLDA GKNHWFAIQATVHGTLRGRPVRVSGSVHAPDSTFITALFVAEATKSVIRDEL LTGPKWGYELPAPQTVLSSLASYNVDIKIVGPTATTDAPQWSDDDDAGFI
Other gene	WM42_1819	putative transposase A	1902155	1902290	44	MHDQAQAANDSARIRQLEKEVSKLREERDILRKAAKYFAEETHW
Other gene	WM42_1820	HTH-like domain protein	1902340	1902628	95	MCDVLKLNRSSFYKWVSTRKKRRLKMYSDAVIGARIKTIFDDEHGLYGAKRI AASLKEDTTYTPINHKKVARIMKSMGLKGFSKRRRCITTRRKP
Other gene	WM42_1821	transposase family protein	1902711	1902993	93	MVRLVEDRILAENISMQAACKIVAPKLGVSWHTARQWTQAARREGRVVESM PEDLAAENARLRRENQELRDTNELLKAASAFFASELDPKRRK
Other gene	WM42_1822	integrase core domain protein	1902989	1904153	387	MIRFIDEHRNRFSVEFICQTLNTHREGGFLTSRGYRQSKARGLSARRLRDAV LIEHISAVHRDNYGVYGVRKMWHALRRDGIDIGREQTARLMRLAGVSGKGK GRSPITTRKPNVPDLRPDLVNREFRAPGPHRLWVADITYVRTRKGFVYTAFV TDVFSRRIVGWALSDSMRTEALPLQALNQAIVCAKETTGLIHHSDHGSQYVS IVYNERLAEHGIAASTGTVGDSYDKDRAVPGHWEGDLIIGQGGTTALVTCVE RSTRYTMIRKLDVHDSATTVDKLIEMFTGKIRNITKTLAWDQGVELAQVDKLS IARGLAVYFCDPHSPWQRPTNENTNGLVRDFLPEGSDFSALTDKDVQHIQD LLNGRPRKVLNWYKPEEKIQELFK

Other gene	WM42_1823	hypothetical protein	1904188	1905922	577	MSFISALRERFSASSGYVATIDAAKAAPPPSPLAPVDLTDAAQVTGVMEIAAR VGEILISAGTANSDAKAQIHLVASSYGLHYCHVDIMMNTITIHTAMGTGSNRR NLHVFRVAPSISVDFSKLSAVDRLIYSIRSGATPPAMAEHVLDEINNMKPPYR TSTVLLGWGVMGGILSVMLGGDALVGVMAFFVSLLVMGVNAWAARFRVPP FYQNIMGGFLAVVPAAILYNVAAGIGITFSPSQIIGSGIIVLVAGLTLVQCLVDGI TRAPVTSAARFFEAMLATGAIIAGVGVGIQFSDWMGFTLPPLATLAPPVYHQ VPLLILCGSIGSGAFALACGASWIEVTVSGLTAGAGMLFYYFVVIPFGVSDVV ACGISATAVGLAGGLLSRRYMMPPLITMIVGYTPMLPGLTLYRGMYASLNEQ MISGFTNLSRALAIAGALAAGVVLGERVARRLRRPQYFRPYATIKRVGHFSF RQASRLAAHKPRIPRVPLSPFAPKVNRPVLPPKLGPKPPMPAKQVRMLQDQ AIGEDWKEEWESITEMWPVTTQWEAQPQNGESAQPAADGAEDAAETAPD AKS
Transport-related gene	WM42_1825	ABC transporter family protein	1906302	1907202	299	MANTESTLRVESLTKSFGEVCAVKDLSFSIERGEVVALLGPNGAGKTSTIDLI LGLSKPDSGNIEVCGMTPRNAVVLGRVSAMMQSGGLLPNLKVRETVELMAS FHRDSLSAEEALERAGAGEFAERYVGKCSGGQQQKLRFAMALVSNPELIVL DEPTAGVDVEARRDFWANIREDAREGRTVIFATHYLEEADDFADRIVMMNH GVLVADGSPQELRAQVSGKTIHAKLQASQSEVEQALRDFPQISLDMRREEF VLVAEDTDKIARLLLTRGLARDISITARSLDDAFIDLAHQER
Other gene	WM42_1826	putative aBC-type multidrug transport system, permease component	1907528	1907927	132	MLTVYICGPLVHAHMPALAWAQSFVIAWLGSIIFMSFGLFMGYLLPAETTMQ VVGPLLALLSFLGGVFMPLTPGSTFDKIGSLTPLYGLHKLALSPMDASNFSW AAVANVVVWLFVFLVGAARKMRYDTARV
Regulatory gene	WM42_1827	histidine kinase family protein	1907943	1909020	358	MTRYITWPNSIWAIFWLALTVPILLTVRSHPYPVSSAVLVLCCVFCLIGSWAIAI DDLRLGVMRVNLRAVSLLVVGTFLALIVLNLVGTDAIQLAYFIASAVAFVFPW QVSLPFTAVAAIAIPQSLWPFGHLSLITTGIACLASRMAIISQAQRKIQEARSQ ELEVNEERNRMARDMHDILGHSLTTVTLKAELAKKLIELDPNTAKSQIAEIEEI SRSALAEVRTAINGYRELSLSGELARALSLLQSAGIKAKMPNSVDEVQVDLR EVFAWVVREGTTNVVRHSGAKRCSISLTSKSVLIEDDGVGIAEIREGNGLRG LRERCQNNGVELCLGETTMGGLKLTAQAVQTGPWAKEH
Regulatory gene	WM42_1828	bacterial regulatory s, luxR family protein	1909019	1909631	203	MIKVMLADDQAMVRGALSALLSLESDIEVVADVGDGSEVMEVASVACPDVIL MDVDMPNVDGLTATKRLREALPQVKILIVTTFGRPGFLRRAVSAGAHGFIVK DAPAAELADAVRRVHSGLRVVDPKLAADSFLFGESPLTVRESEVLQAAADG AVVAEIAKRVNLSEGTVRNHLSRAMAKTGADTRAAAVNLAIERGWIIS
Regulatory gene	WM42_1829	bacterial regulatory s, tetR family protein	1909637	1910195	185	MPRISKKTEVLKAALTIIENEGVHAVTYDALATATGMSKSGLIYHFPSRHDLLI ECHRFCAARWEAELEELAGGKQAAELDKKERLRALVQSLGKNDPLIELLMSI HSQQHPDFMEQWRDVDKRWLPDPAADDQTALIATILGNGLWVHDHLTERKI PTATRKKTVELVLEFLDSGKLQTAPLKD
Transport-related gene	WM42_1830	sugar (and other) transporter	1910178	1911678	499	MSATDLVHSNSTATTTESTAVQRWTFFAVISLGLLMVGLDNSILYTALPALTD QLHTTSTQQLWIINAYALVLAGLLLGTGTLGDRIGHRRMFVIGLVLFGSASLA AALAPGAWFLVAARAFLGLGAAVMMPATLALIRLTFDDEIERNTAIGIWASVA VVGAAAGPTVGGFLLEHFWWGSVFLINVPIVLVALLLTFLLAPPNQPNPDKH

		family protein				WDFASSLYALITLSSLVLAIKSVAGSHYGLAGGALVACGIGSFAFARRQKQLS EPMLTFDIFRSPIFTGGVIAAGGAMFGMSGLEMLTTQKLQMVNGLSPLHAGI TISAVAIAALPMSTLGGANLHRWGFLPIIAGGFVFMAAGIAIAMWAGHHGIFW CFVAGLLCMGIGAGLTMSVASTAIIGAAPPYRSGMAAGVEEVSYELGTLLSIA ITGSIVPLLYSRNLTADIEGMQALYDAATHDLAAAAYDNAYLTTLGGLLAMML LFAAVTGYLFKNNPKSGGYDAAHQ
Additional biosynthetic gene	WM42_1831	thiamine pyrophosphate enzyme, C-terminal TPP binding domain protein	1911851	1913588	578	MAKNYAEQIVETLEAQGVERIYGLVGDSLNPIVDAVRRSSIEWVHVRNEEAA AFAAEADSLTTGKLAVCAASCGPGNTHLVQGLYDAHRNGAKVLAIASHIPSR QIGSKFFQETHPEMLFTECSAYCEMVNSADQGGVILHHAIQSTMAGKGVSV LVIPGDVSTQEAEDDTFTTSTISAGRPVVFPDPAEAAALTQAINEAKTVALFV GAGVKDAREQVLALAEKIKAPIGHALGGKMYIQYDNPFDVGMSGLLGYGAA HDATHEADLLILLGTDFPYNDFLPKGNVAQVDIDGSHIGRRTKIKYPVTGDVA ATIENILPHIDEKKDRRFLDKMLKEHYNKLNHVVEAYTKKAEKMKPIHPEFVA NIIDEEAADNAVFTVDTGMCNVWGARYITPNGKREQIGSFRHGTMANALPM AIGAQAANEGRQVISFSGDGGLSMLMGELLTVKLHNLPVKTFVFNNSSLGM VKLEMLVQGLPEHETDHEHVDYAAIAEAAGIKHIHIEDPKKARKQIREAMDFD GPVLVDMITDPNALSIPPTLTFEQLLGFSKAATRTVFGGGVGQMLQLAQSNL RNIPRP
Other gene	WM42_1832	hypothetical protein	1913836	1914076	79	MNSDSGENGSFGQHVAFKDLVGELVDESELFFGQREFLAHLESFFRLVAVI AIGEVDLVFTLRIPAWLLAPLARMGVHR
Regulatory gene	WM42_1833	response regulator	1914166	1914871	234	MDDNKEVKVLVVDDEPNIVELLTVSLKFQGFEVFSANSGTEALRVAREVNPD AYIMDVMMPGMDGFELLGKLRQEGLDGPVLYLTAKDSVDQRIHGLTIGADD YVTKPFSLEEVITRLRVILRRGANVDDANSDATISYADLTLNDDTHEVTKGGEI IELSPTEFNLLRYLMQNKEVVLSKSKILDNVWHYDFGGDGNVVESYISYLRR KIDTGDTQLIHTVRGVGYVLRMPRN
Regulatory gene	WM42_1834	HAMP domain protein	1915013	1916366	450	MPLRTWLLVLMVLVSGLGLAGASLAVSSIMSDVMYGNVDDQLKDATNGWA RNMSNDYYLGDYSRRPPTEYVALSYLPNGTVLYTGPKTSMPDAKNLYIGDE PATVGSVGNDETKWRAVAIKVDGTVTVVAKDLTHERLILRGLAIVQIIITAVVM GIIAVVGMWFIRRALRPLRVVEKTASEIAAGDLDKRVPEWPKHTEVGQLSAA LNVMLGQLQRSVVQAQDKEEQMRRFVGDASHELRTPLTSLRGYTELYRSG ATKDVDLVFSKIDDESKRMSLLVEDLLSLTRAEGSRLDMRTVDMLELVLSVG SSARAAFAGRQIDVVNETHDIPVVKGDPDRLHQVLLNLVSNGIRHGGEEAKV KLTIREEGEGTDKAVLIDVADDGKGMSPEDTSHIFERFYRADSSRTRDTGGS GLGLAIVKSLVEQHGGSIAVSSQLGEGSTFTVRLPAS
Other gene	WM42_1835	HIT domain protein	1916409	1916832	140	MSSVFTKIINGELPARFVYRDETCVAFLSIEPLNYGHTLVVPIEEIDKWTDLDP QTWAHLNEVALEIGAAIKTAFNSPRTGYIIAGFDVPHTHIHLFPTEKMEEYDFA KAFAADATDPAAMDEAATRIRQHLGTDEEGRR
Other gene	WM42_1836	phosphoribosylamine--glycine ligase	1916866	1918165	432	MRILVIGSGGREHALLKGLKADPLTTELHVAPGSPNFDSLATVHPEYSKVDD PAQMLELAQKIDAELVVVGPEVPLVAGVADELRANGIAVFGPSKEAAQIEGS KAFAKDVMNAAGVRTARAEQLVPGASDADIESALDRFGPHFVVKDDGLAGG KGVVVTDDRAAAKEHVNEVHAAGNPVLLESFLDGPEVSLFCLVDGETVVPL LPAQDHKRAYDNDEGPNTGGMGAYTPLPWLPADGVQRIVDEVCKPVAAEM VRRGTPYSGLLYAGLAWGKEGIAVVEFNCRFGDPETQAVLSMLESPLAEAL NATAAGKLAELDELKWKDGYAVTVVLAAEGYPASPRKGDVITGPGLDDPNK

						VLHAGTSQAEGSRGVEEEVDVISNGGRVLNVLGQGATLAEARAAAYEVLEG LKLDGSFYRRDIGKRAEDGEISI
Additional biosynthetic gene	WM42_1837	adenylosuccinate lyase	1918236	1919676	479	MFPVAEKKNISNVLSSRYASAELSNIWSPEYKIILERQLWIAVMRAQKDLGVD IPAEAIAAYESVVEQVDLESIAARERVTRHDVKARIEEFNALAGFEHIHKGMT SRDLTENVEQLQIHSSLRLIRDKAIAVVARIGKHAAAYQSQVMAGRSHNVAA QATTLGKRFASAADEMLLGIERVESLLAAYPLRGIKGPMGTSQDMLDLMGG SEDKLASLETAIADHLGFHRIFNSVGQVYPRSLDFDAVSALVELGAGPSSLAT TIRLMAGNETVTEGFKEGQVGSSAMPHKMNARSCERVCGFQVILRGYLTMV ADLSGQQWNEGDVFCSVVRRVALPDAFFALDGQFETFLTVLDEFGAFPAMI DRELERYLPFLATTRILMAAVRAGVGRETAHEVIKENAVAVALNMRENGGD QDLVERLAADERLPMSAEDLEAALADKHAFIGAAESQVNQVLTRINNLVSEH PKAAAYTPGEIL



Appendix 101: NRP CsimD biosynthetic gene cluster and its predicted similar gene clusters according to antiSMASH.

Appendix 102: Summary of genes in BGC possibly coding for type 1 PKS CsimB as predicted by antiSMASH.

Type of gene	Locus tag	Function predicted by antiSMASH	Start of gene	End of gene	Amino acid sequence length	Amino acid sequence
Other gene	WM42_1629	N-acetylmuramoyl-L-alanine amidase family protein	1660759	1662907	715	MQLRRRLVPTKSKWSTPIIAAVTSVSMVAAAAFGGHQVLRTQENGSGPIEVA SASTSFGDGETVVVDDAAISAQGEGDGPRAVKQFHRDEPFSMFAVTWKGA RDVAAFVRAKQADGSWSQWYDMDNSGYTNDDPNATNGTELIYTGTTNDVQ VSINNVDLVSGTNLDKSFEETPVEEDGTQAPESAAPETSAPAASASQSPASA PAEDSNQSVIEQAAEAAANPQPAPLPYNVGDIAPVADVQEMDTKDSGAQDS DAQDSGASANNTSTEGMEAVFIDGNAQAGEAIEQTAVTDGMPKVVSRAGW GADESKRCMGPDYDDGVKALTLHHTAGSNNYTKADAAAQVRGIYQYHAQN LGWCDIGYNALVDKYGTIYEGRYGGLDKAVQGAHVGGFNTNTWGISMIGNY ETAAPTEQMLNSVAEIAGWKAAISNINPKGSTDLVSGGFGGSRFPAGAVAHV PTFHGHNDLHYTACPGQYTVARWGDIRNATYRKYQSIKNGTAGSGAQLPSN DGTNTNTGTNTGTTTPGTNNPGGNTGGTTGTNGATSSLGGVQVPVAVVQA VAGIAATLVGIFLARSGQKVDGNKTVVGGLTTSEIPNIVSKVVQISGNPGLQQ SWTAILNAFGPVLGLAVGGPDFKAGVVSQLFENGVVLSSEETGTHALTGQIA KAWSEGDNSTKLGLPTSDEVSTGNGREIRVDFQGGYIQYDPATEKINVFTD
Other gene	WM42_1630	UDP-galactopyranose mutase	1663111	1664308	398	MTSYDLIVVGSGFFGLTVAERAASQLGKKVLIVERRNHLGGNAYSEAEPETGI EVHKYGAHLFHTSNERVWNYVNQFTDFTDYQHRVFAMHDGTAYQFPMGLG LINQFFGKYYSPDEARQLIKDQAGEFDVDEAQNLEEKAIALIGRPLYEAFIRDY TAKQWQTDPKELPAGNITRLPVRYTFNNRYFNDTYEGLPVDGYAAWLENMA KHELIDVQLDTDWFEVREQVREENPDAPVVYTGPLDRYFDFSEGELGWRTL DFDLEVLPTGDFQGTPVMNYNDADVDYTRIHEFRHFHPERQDKYPKDKTVI MKEYSRFAESGDEPYYPINTPEDRTKLEAYRRLAAAEAKENQVLFGGRLGTY QYLDMHMAIASALSMFDNKLAPFWNEGKALEQERGH
regulatory gene	WM42_1631	divergent AAA domain protein	1664582	1666079	498	MIVVWDDDQAEIAEIIGSLRSQDTDTRAVEVKAAAGGFPKKLVRSISAFANGT GGIIILGLDEENGFVTVPGFDATAMADALAGACADLVTPSVRAQIGIVEWEGA SIVVGTIPECAPFEKPCWVVSQSKYHGSYIRVHDGDRVLKPYEIDRLEENKT QPEWDLEPVPDADIDDLDPDIVAEILARERSIHERIFGRLSDEEAMLALHLITR NDNGKLVPTLGGLMAAGTYPQQFFPRLNVTFAAYPGTDKSAGMGKQRFLD NESLVGPIPVLVADAVRAVRRNMRVGGVIEGVFRKDLPDYPPEAVREAVANA LMHRDYSPQARGTQVQVNLYVDRLEILNPGGLYGTVTVDRLGTAGMSSARN QHLSALLEVTPAGDGDGYVAENRGTGYIEILDQLERQLLPPPVPRDSLTTFEL TFARRNPTTPERTAALGGGTRGRVLDYLREHRTASSRELAGAAGLSLNGVR RTINELVEEGVIVRTEPLKSPKQHYRLQG
Other gene	WM42_1632	hypothetical protein	1666075	1666264	62	MDDEVLAEIARQLVSMLRRDAKTDWTVRDDVRAKLRRSGRRLLRDYKYSAA AREQADVERRG
Other gene	WM42_1633	hypothetical protein	1666579	1667122	180	MTHLNDHRLATLRQDLHDLYGNAHAAAEIDAEVDAAIARHTAGATVEEFIPVL VEREVKDYFGEHRLHVRFSAGANQALAQEVVELAKKYAGDALLVDAAAPHE

						EADPEGRMVEMPDFIVYLGREIPRDEPGKDIKIWDIAEANTEEEKRELHDDLG ARVLYMLGKLGIEPVSDKAPVEA
Other gene	WM42_1634	hypothetical protein	1667130	1668441	436	MKYLFQLLAVAFLLVTLPATAHAATPTVEIFDEGNLLSPSDESTLRQESEELDL PPEVSRILYFTYEHNHKDVSTDALEYLRDNARGLLNTDGRSLREGTLLIGVGL DPMSSTARCSNDVCEAISIEEPGRLNGILDQMEAPVRGGNYTVGLLLAAKAA GDPSIYRKSTTRLPGWSIFIAMSVLAVAMLVASFLYSRRVTARRLLEQLDFNR SALPAATKFIIEGDKIVAELNTPLINEAFKAQWEQIKADYRRGKPNIVALDSLDP SASASLLFQHSGSITQGYRQLHRLSTAKAQLDDLQRMAKGEKLTRSTELGW LKADMQRASAVAESGDLPATSAISALLERIEILEKDLSRADFSAQFARLLADY RPIIGALPERLYAHGELNHQRAVPPQLGEQDWHPGMGTHYMPYNYAEFWV SNNAISANLPPAYRF
Other gene	WM42_1635	UDP-galactofuranosyl transferase GlfT2	1668651	1670631	659	MTANNPSASAAGSETRVFEPVQRILLPKRGEPFDVRMLYLIESEQNRERLSW TNRTAVTIPAGEEASFETYFNAFPASYWRRWSQLKSIVLSLELEGQANVSIYR SKQDGQRISVANHLVGSGHHDFELPLKNFEDGGWLWFDITAEKETVLSNAA WCAPHEPGPQIMPDGTEIPASEKRVAVGIPTFNRPTDAVAALQALAEDPVVD GIIDYVLMPDQGNQHPADEPGYEQAVEHFGPRFREFRQGNLGGSGGYSRI MYEALENTDSPFILYMDDDIAIEPDSILRAVQAARYAAKPIIVGGQMLNLQERS QLRTTGEQVNKHDFMWGGAPHAVYDHDFAKYPLRAIGDAQSHLDPRKYDS RALHRRVDVEYNGWWMCLFPRVVAETNGQPLPLFIKWDDTEYSLRAAANG FPTVTWPGAAIWHMAWADKDDAIDWQAYFHLRNRLIVAAMYHDGDVKGITK SIFKSTLKHTMCMEYSTMAIQLEAMKDFLAGPDHLFDILESSLPRIAEIRKNYS DAVIIESADQLPAPTGAPGVPTRNIGGRLGKVKKLPWLLKSAKHLVTKEDRAH HEAPQLNLTPDEARWFTLSRVDSATVSTAGGTGVAFRKRDRDLATDLVKQT RGLLKEIEENFDDLKAQYRAAMPELTSRESWKKVFDAQ
Other gene	WM42_1636	PAP2 superfamily protein	1670620	1671148	175	MPSKLSVAESHVLESIQNAAFDVPGVVPAARGLSHLGEHALGWMGTSALMY VVNRADKAKARQWAYVGASAFTAHAASVVLKRIVRRKRPDYPYVRVGVATP SKLSFPSSHSTSTTAFLVAVAHLLSTPAPLVGVPVMMASRMVLGVHYPTDTA VGAAIGAVTAEALTRYERKTA
Core biosynthetic gene	WM42_1637	ubiA prenyltransferase family protein	1671144	1672140	331	MSDSHHDDGTQRVFHSEPHTAGIDTGRKRKPPKNLADGMVKALRPKQWVK NVLVLAAPAAAGADALFHSRVLLDVLLAFIVFCMGASSIYLINDARDVEADRA HPTKRFRPIAAGVLPVGLAYAMAAVLIVGSIALSFLATAGPQLAIVMAVYIALQL GYCFGWKHMPVIDIALVSSGFMLRTMAGGVAAGIVLSQWFLLVAAFGSLFM ASGKRYSEILLAERTGAKIRKSLEGYTPTYLRFVWTLAATALVMSYSLWGFQL SNAADGPAGVWYQVSMVPFTIAVLRYAADVDRGNGGAPDEIALEDRALQVL ALAWLACIAMAVYIMPML
Other gene	WM42_1638	putative membrane protein	1672250	1674032	593	MLNFYPVNNTNNKNLVRLSAAASVIVIGVFSFWGGFVRRWISDDGLIVLRTVR NLEAGNGPVFNMGERVEANTSTLWQYLILLLRWITGADLAGIAIYLGLFLAVC AMALGAFATAGLVSRGKSATLVAPAGALVYLALPPARDFFTSGLEWGLAIFY LAVLWWMLLKWARGVDKHAANASDSMPYWLAVWAGLSWLVRPELALYGG LVGVLLLAAHRNWKAWLGILAAALPLPLGYQIFRMGYYGLLTPHTAVAKSAS GAVWSHGFNYLGDFALPYALYVPLIVLAAWALWSLRRELRPSGFRTPATAAY LLLAAGLLHLIYVLRVGGDFMHGRMLLLPLFALLLPVFVLPLRSILGGLTAAICA VWACVIVLRGHSVDWDDFSGPLNIVDEREFWTHATKRQTGHPPRSLEDFGH MRFIEHYKEGMDELEAGDALAVLYAKKGEEDAYGWMGIERDPSIDELPTVYF

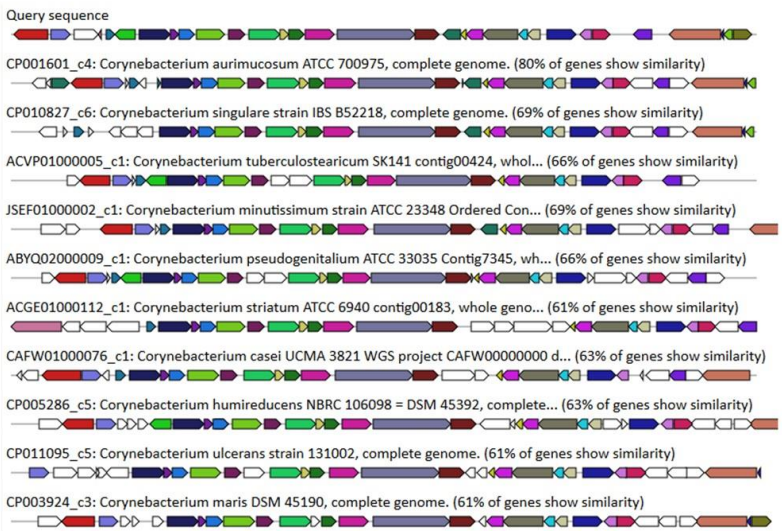
						INMGMTSMYAPLEVRVLDDIGLATPLAARQPRIADGRIGHDKSLPTYWQAAQ TAVDIDELPSWYDKEETRKARKALQTEDFQKLFATYKDPLDAKRFFENIKFAL TDGRTLTFSEDPDDYLGKR
Additional biosynthetic gene	WM42_1639	esterase family protein	1674254	1675277	340	MTSLSSGLKARVLTVVMAVAVALGLAVAAGSQEASAANRDFLRADATGTCD WDAVGWWVQRCDVWSPAMGRNIPVQIQPAKNGGNAGLYLLDGLRATNRT NAWVNDVNAARTYEPHNITLVMPVGGEASFYADWEGPATYNAVSPINYKWE TFLTSELPGYLERNFGVARNNNSIAGLSMGGTAAITLAGKHPGQFRQVLSYS GYLTTTLPGAQTFMRLALLDAGGFNINAMYGSLVSPRRFENDPFRVMGGLR NTDVYISAASGIPGAGDAGYLPQHQLSGAVLEMFAGATTRIWEAKARATGLR VTSNYPMQGLHNWAQFGYQLEHSKPQVLNVMNAW
Additional biosynthetic gene	WM42_1640	esterase family protein	1675540	1677508	655	MRKNASASSLRSPGSPTKARKGLAIAALPTAVAVGLSLLPNATAQSSLGSLT QNLGSSNLSDAFAPGTPPERTPIQTEYPEVEGLPEGVDISRVEYLTNRNLRV YIKSAAMPDKEQVVQIQLARDWYSSPDKRFPEVWALDGLRARDDESGWTIE TDIETQFADRNVNLIMPVGGESSFYSDWQKPDNGRNYKWETFLTKELVPILD KAYRSNQKRAVTGISMGGTAAMNLAERNPHLFKFVGSFSGYLDTTTQGMPE AIAAAQMDAGGFTSTNMWGPHYSQDWIDHDPKLGIEALKDMKVYVSAGSG KDDYGNLKSVAKGPANAAGVGLEVISRMSTQTFVDYAKRAGVPVVSRFRPS GVHSWEYWQFEMREAWPVMADALGIAKEDRGADCTPVGAIAEATKSGILGS CLNNEYDVAGGKAEDFQAGTAYWSPETGAHAIFGRIGARYAESGGPTSWL GFPTTGETKTPDGKGRFVHFQHGSIYWTAETGAWAIPGDMVEAWGKNGFE GGDLKYPTGPVTKVGEGFAQDFQNGVLTRNPDESHSIVHGAIGAKYKELGG PSSPLGFPKGGEKPIKGGFFQEFEHGNIYWSAESGAHYILYGAIMDEWGKR GFEQGEFGWPTSDYNQIAAGGISQTFQHGEIREIMGSVQAEKK
Other gene	WM42_1641	hypothetical protein	1677510	1678044	177	MRTSFRVPVMCAAFMAAALGLSACGSATVESENSPEQTKVAPLERSSQPAS ASESASEASSQSADSSDEKSSSESSAASASDEPEDRGAREISAIPSPSQTQG PEQDFLAAVQKAGVETEGAEDQIIGAGQAACNEGDAVTIPAVAGQLIEQGRA PLSHEELTAVLTAQARGTLCAK
Other gene	WM42_1642	hypothetical protein	1678055	1678991	311	MRKTLTVVAVLVVLAVIGAGAVHFLNANKAGESNNLSEPAPTEQGGGVEAPA QPDWCPRFEVISAPGTWESKADDDPINPSANPKSFMLSISNPLKEAYVPEDV KVWTLPYTAQFKNINAQQEMSYDESRDEGTSRLEGELTTMHENCPATKFIIA GFSQGAVIAGDVADRIGGGNGPIPAEAVSGVALVADGRRQAGVGQNPGVPV AGVGAEIALQPVGMLIQGIVPGASMRGPRANGFGSLADRTFQICAPDDSICD APLDGPNGLERARGLIEANGVHAMYATNPNVIPGTTANQWVVGWAREIIDRT
Core biosynthetic gene	WM42_1643	long-chain-fatty-acid--AMP ligase FadD32	1679093	1680932	612	MDLQALIGQFFDGNGRIVLPPHLTLPGLGEHVFEMEQQSGVPDRPVLRQWL FDDDPAGTPRDFTRSEVNKRIKVVAARLQQVGAVGDRVAILAGNSPEYLFGF MGAMYAGMTPLPLYDPNEPGHSEHLRAVFGDSNPSIVLTNNVSAAAVRRYF ADLPSAQRPRVISIDSLPDSLAASWKNPVEQGAGEGAAAPVDAPAFLQYTSG STRTPAGVLLTNRSILTNVLQIFAAVHIKMPPRIVSWLPLHHDMGIILASFATIL GLNFEMMTPRDFVQQPKRWVRQISRQPGDESFAGIYAVVPNFALELAARYG APVAGEDIDLSHVDGIVIGSEPVTETAVDNFWEVFSEYGLKREAMRPSYGLA EASLMVATPQEGERPLITHFNREQLTAGKAVVEDKSEQTVAFASNGQGMPA QYLTIVDPETKAELADAQIGELWLHGDNMAAGYLGREDETASTFRNTLATRL PEGSRVVGAPEDDKWMATGDLATIVDGHLYITGRLKDLIVVAGRNHYPQDIE

						ATVQEASAHVRPDSIAAFSVEGGNTEQLVLLIERADDADSAGDEAATEVIRTA VSKAHGITPEVIRWFSAHEINRTSSGKIARRVAKKHFLAS
Core biosynthetic gene	WM42_1644	alpha/beta hydrolase family protein	1681113	1685970	1618	MTVEELRGWLRNWVAQTTGLSAEEITDSKPLENFGLSSRDAVVLSGELENLL GMKLEPTVAYEYPTIAQLADRLVNGVASSSNSAVDSSSGAASTRNAGIIGGDI AIIGEAGRFPGAQNVKEFWDMLVEGRAGTGPLPMGRWSEYAAEPTVSEKIA NQNTDGGYLEDIASFDAEFFGLSPLEAANMDPQQRILLELAWEALEDAGLPA NQLRGTATGVYMGSTNNDYGMLIVADPAEAHPYAMTGTSSAVVANRISYAL DLRGPSINVDTACSSSLVAVSQAVNDLRTGASDVALAGGVNILAAPHASTGF SELGVISPTSAIHAFSDDADGIVRSDAAGVLVLKRLADAERDGDNILAVIKGTA VNSDGHSNGLTAPNPDAQVDVLERAYADAGIDPQLVDYVEAHGTGTILGDPI EATALGRVLGSGREMAAPMLLGSAKSNIGHSESAAGVVGLIKVIESMRHGVI PPSINFSAPNRYIDFDNEHLEVVEDPREWPEYSGQKVAGVSGFGFGGTNAH VVLTDYRGLTAETAAQVNIDEGAPVSLPISGLLPSRRAAAAGQIADYIEQEHP DLLSLARTLARRNHARSRAVVTGVSEEDVVKRLRQVADGKVSMGIASADSP SVPGPVFIYSGFGSQHRKMIKDVLEISPAFKERLEELDRIVEFESGWSILDIVT DDAQTYNTETAQVAITAIQIALTDLLASFGVRPAGVIGMSMGEIAAAYASGGIT AEDAMLIAAHRSRLMGEGENSLGEEEQGAMAVVELTAAEIEALDGNIEPAVY TGPGMTTVGGPRPEVLALVEKLEAEGKFARALNVKGAGHTSAVDPILGELYA DIAGIEPRPLHTTLYSSVDRGEIYRPGALVHDEDYWIRMTRQPVMLQDATEA AFASGHTQIVEISPNPIALMGLMSTAFAVGKSDAQLLFSLKRKVDPTESLLDLL AKLYVAGAPVDYTAVFGSGALVDAPHTQFKRQRFWTNARPSSGISGLPGAR VNLPEGKVAFSTNADQAPSALAILEAAAEAVTPGSQIVASEEHGDLPPQGEV TTVVSKSLGGLSIAVYFVNGPATQLVAEGFASALNLTEASIPGVAAIAEPAAPS APASFTDVSDVEAVRWDPSKETVEERLSLIVSESMGYDVSDLPRELPLIDLGL DSLMGMRIKNRVENDFQIPPMQVQALRDASLADVITMVEDAVAGKSVAGAA ETGAVDVGAGEVGAVDTGNAGANSAESSAAEVGDVEADAEKDGVDKHDAE NSAEGVGVAPRDASERMVFGTWATFTGKAAAGVTSQLPQIDESVAADIAQR LSERAGIEVTTQQVLDAQTLEPLADLVREGLETEVEGNIRVLREAEGPAVFMF HPAGGTTVVYQPLARRLPSDVAVYGVERLEGSLEERATAYIDDIIKYARGRKV VLGGWSFGGALAYEVAYQLADRTKRGEESAEVAFIALLDTTQPSDPAPDTLE ETKARWGRYAAFAKKTYGLDFEPPYEMLETMGEDALMTMLAEFLSNTDASE HGLAAGVLEHQRASFVDNQILGKLDMQRWADVDVPVILFRSERMHDGAIEL EPRYAEIDPDGGWGAIVEDLEIVQLQGDHLAVPDEPAIGIVGKHINDWIEEKIR
Additional biosynthetic gene	WM42_1645	carboxyl transferase domain protein	1685969	1687514	514	MTTTAHKLEDLRQRLERAQDPGSERSRKRRDDAGRSTPRQRINALLDEGSF VETGALAKTPGDPDAIYSDGVVTGYGRVNGRPVCIYAHDKTVYGGSVGVIFG KKVTEIMDMAIKIGCPVIGIQDSGGARIQDAVTSLAMYSEIARRQLPLSGRSPQ ISIMMGKSAGGAVYAPVTTDFVIAVDGEAEMYVTGPNVIKEVTGEEITSHDLG SARQQELNGNVSAVVASEDDAFDLVRDLLDHLPLTCFDEAPEFAAPEDSELE DPELNEFMPDDTNAGYDMLELLAQLGDDEEIIELQENYAPNMITAFGRIDGKA VGFVANNPMHLAGCIDADAADKGARFIRICDAYNIPLVFVVDTPGYLPGVEQE KAGLIHRGAKFAFAVVEATVPKVSLIVRKAYGGAYAVMGSKNLTGDINLAWP TAQIAVMGSAAAVVMIAGKQLEAAETPEQRALTKKMFMDFYDETMTAPYVA AERGYLDAMVEPSQSRLALRQALRQLATKTESDLPKKHTIAPM

Other gene	WM42_1646	HNH endonuclease family protein	1687754	1688819	354	MKLGDLLGVLARGMHVVSLCAGHNRSELVAMGATESFAQSLLSLHEIYFGKTAFSAMQRAARNTSHSLESLREIERYVGRVKDKRKAWRLREELCATRERDIAKVAVARLKELRKPPQVARGVKVLRRDNGPHSLVITDSPRAIADLFGTIKATAKDPTAAATAAEASAAGPDLLGAVRRVFSGAGSSAPQLHTHVIVRLEQLDKIARGDGDDIQLLATDGGTLSGAEFIQKRFADIGFVTLVHPEVGPVNLYRTRRFASAKQRTMLAAEHPTCAWLGCRRPANECQFHHLDRWEDGGMTNINNLVPLCHYHNAINDDSPDKPTGRGRMGRLRGRICWLPPGGGRPVVIPSPAM
Other gene	WM42_1647	hypothetical protein	1688861	1689206	114	MRTAPIIDIIALAIFAVLARIAHGGLSFSSWVDAFWPWTVGALIGWVIILATKMEGRWKEGLVVWLSAVIGGMALWMLVNGRLPHYSFLIVATTMSALFFFGWRGIAALASRRH
Other gene	WM42_1648	hypothetical protein	1689212	1690229	338	MKKNWLTWIVSLLILIALGWFFRDHLSFIAEGLSRLRHAEPLPVVLVVLFAFGSIAGMAEVMRLLIKAGKIEVPLRETYAITLASNSWSTTLPAGPAFAAILTFQVQRGWGASVALCSYFLFLSSIISSMFLALIGVAGVFFLNADMALGSLLTTIVLMLAAMGAIFWITSHPATIQRWLNHQRVLKGAKLARVRQEVNNLDEVHLSRGPFAVICISSLLHRLCDMLALWASVWAITGEIPWLRAAEDHTTMAGIALAFLAAKLAGSAQVTPGGLGTVEAALIAPLVATGLTAAHATSAAIIYRLISFALVTIIGWVIYFVHYASKGLTYQALNRKDT
Additional biosynthetic gene	WM42_1649	MMPL family protein	1690225	1692457	743	MFYAWGRFSYAHRKVVPLVIVGVILLLFIGFGTRLGERMSQEGWEDPGAASTSAARIEQETFGRDNSGDVILLFSDPDQNFAAAKDHLATLKEQHPQEIDGITSYFDTKNPNLVNRDHSTAFAAISLKGDGEQTLKDFRAIESDLTATDIPVKVAGATAVADALDEGMANDISRAEKAALPLVGLLLLIVFGSVVAACMPLVVGGLSILGSLGILSVLAGFQQVNVFAQAVVTLLGLGLAIDYGLFMVSRFREELDKGRSTPDAVATTTATAGQTVVFSAGMVTVALSGLFLFPQAFLKSVAYGSISAVGLAALLSVTVLPALFGLLGKRIDKFAVRKTSRKARRLEDTVWYKLPRWAMRHAKLMTVAICGLLIALTLPVLGVKFGGINETYLPPQHETRMAQDEFNEQFPSFRTEPVKLVVTNATNQQLVDVVMQVRQLEGLTKPMGPVTPTKDGTTVLAAGIKDRDDYARIVHELEQVEAPEGVELYVGGTPAMEVESLDALFHKLPWMALYVVLATFVLMSLVFGSFILPIKAILMTLLSLGATLGILTLMFVDGLGASALSFTAGPLMSPILVLIIAIVFGLSTDYEVFLVSRMVEARRRTDGADTDEAIAIGTAHTGGIITAAALIMIVVAGAFGFSDIVMMKYIAFGMIFALLIDATIVRMLLVPAVMHLLREDNWWAPRFIHRAYDKLGHGREPAPAQDKYDGPERSGRSLAEDSSLIPFSELMQRLQADKRTGER
Other gene	WM42_1650	NYN domain protein	1692464	1693061	198	MQGSYLLIWDAPNMDMGLGAILGGRPTAAHRPRFDAIGRWLVDIALQNDAQPEAAVFTNVTPGGADVIRPWVEAIRNVGFAVFAKPKLHEDDDVDPDMVDYIQANRDKLEGVIVASADGQNFQPLLEEIAAEGKPVCVLGFHEHASWAVTHDDIEFVDLEDIEGVFREPLPRINLDNLPEGGAWLQPFRPLTALLKQR
Other gene	WM42_1651	tRNA (guanine-N(7)-)-methyltransferase	1693060	1693837	258	MNTADNPQNSPAGEMPHGRPLQTDFNAKFGNDLDYPRLGNVTFRRGTLTDNQNALFEEHWPRLGAMLADERLDIDSWFGRTGAKTIVEIGSGTGTSTAAMAPLEQDTNIIAVELYKPGLAKLLGSVVRNDIDNIRMIRGDGIEVMVRMFAPESLDGIRVFFPDPWPKARHHKRRIIQSGTLNLFASRLKKGGVLHVATDHAGYAEWINELVEVEPTLRYKGWPWPECPQLTDRQLITKFEGKGLDKDHVITEYLWEKI
Other gene	WM42_1652	phosphoenolpyruvate	1694243	1696061	605	MTATIKGLVGELPTSNEKLISWISESVELFQPERVVFVDGSQEEADRLAGELVEKGTLIKLNEEKRPNSYLARSNPSDVARVESRTFICTESEEDAGPTNNWAPPAAMKEEMTEAFRGSMKGRTMYVVPFCMGPISDPEPKLGVQLTDSEYVVLS

		carboxykinase				MRIMTRMGAQALKKIGADGDFVHALHSVGAPLEEGQEDVAWPCNETKYITQ FPETKEIWSYGSGYGGNAILAKKCYALRIASVMGKEEGWMAEHMLILKLTNP EGKNYHVAAAFPSACGKTNLAMITPTIPGWSAEVVGDDIAWMHLREDGLYA VNPENGFFGVAPGTNYDSNPIAMKTMEPGNTIFTNVALTDDGDVWWEDMG EAPAHLIDWLGNDWTPESTTNAAHPNSRYCVPITQCPTAAPEFDDWKGVKI DAILFGGRRADTVPLVTQAFSWNHGTMIGSLLASGQTAAAEGKVGALRHDP MAMLPFIGYNAGDYLQHWIDMGEKGGDRMPSIFLVNWFRRGEDGRFLWPG FGENSRVLKWIVDRIEGKAEAVETVVGHTARAEDIDIEGLDFDIQDVREALSV NASDWAGDMEDNTEWLHFLGSRVPSEVWDEFNALKSRVENA
Additional biosynthetic gene	WM42_1653	methyltransferase domain protein	1696308	1697076	255	MATLARSWRLLRSFRFEQTAPEIFYGGLAEDTALLIDALCHDVGTQLAGARV LDVGGGPGYFAEAFARRGARYFGVEPDAGEMSAAGIQLTNAVRGDGTQLP FAADTFDITYSSNVAEHIPHPWDMGEEMLRVTKPGGLVVLSYTVWLGPFGG HETGLWEHYVGGAFARDRYTRRHGHLPKNVFGTSLFAVSAREGLKWAEAIT SSGAAPGGASVDSSTLVATFPRYHPSWAWWVTRVPVLREFLTSNLVLVLRR
Core biosynthetic gene	WM42_1654	glycosyl transferases group 1 family protein	1697118	1698240	373	MKILLLCWRDSTHPQGGGSERYLERVGEYLAAQGHEVVFRTARHMNAAKR EHREGFLYSRGGAKFSVYPRAWLAILAGRLGLGDAKNIDVIVDTQNGIPFFAK LVSKAPTVLLTHHCHREQWPVAGPLLARLGWFLESKVAPSVYRNSPYVTVS QASKQDLEELGIKGAHIIENGVDPLPEHVPTLQRETAIHLVTLSRLVPHKQIEH AMDTVAKMPGAVLDVIGSGWWEAQLREYAERIGVASRVRFRGQVTEDYKH ALLALADVHLMPSRKEGWGLAVMEAAQHGVPTVGYSFGLRDSVIPGQTGVL VDTEEEFVRAVRELVADANLRRQLGENARELAAHYSWEKTGRAFEKLLTQT ARATAGSQKARTD
Other gene	WM42_1655	hypothetical protein	1699681	1700833	383	MKRLINWSSRMTKVIVAAIVFLLLGSVLPPLYNNQARPLPDDLNFTIDTAPTQ GPAINVSRAVQHKKPADTHGNPDCDAKDAPLYCYAEERDVSLKRTTRTSPT ADDAVASADSLLQVTSNGQTIAEIKENSLLNRESAYPVAGPNSSQQVDIAPLN ISAAGHDFSRDGISYFFPAKAEQRSYPYFDPMTQTTAPIDFTGTEKRETIPTYI FHQEIEPVALAYSLGVLQQQTPDEKATRTSPEKLQRSGPAQAMFDEESLRR FGLDPTEHVSLEPFYTVTRDVWVEPTTGTIVDAHEDIKIFLATDQDQAKKMVA EDDTADRSLFQANLNWSDATKKERLDTVRTTINNVKILSIVGWLGKVIGVILLA YAAFMYMRRHRVGA
Other gene	WM42_1656	hypothetical protein	1701970	1705153	1060	MNAQSRWRIGGLLALAVFVFTQPFGLVAADTKHDLTANPLGFLRGATSAYTE VFTLGQLQNQAYGYLFPQGPFFLLTSPLPDWVAQRLWWLLVLGVGFWGFH RLIYASLPPQFRGLRREGIRREISPVWPFLAALLYALSPRALTTIGAISSETWP VMLAPWVILPFLRRELTWRCAAAAVLPVAAMGAVNATATLAACVPAAVVLLY RRAWRPGLVWLAGCAAVSLWWIVPLLVLGRYAPPFTEFIESSYVTTRWLNL PEILRGTTSWAPFVDTERVAGYALATQPFFGLVTMTVAAIGLYGLARLPRVW PVMACVGIAILGTHAAWYLHALDGPLAALRNLHKFDPLVRIPLLLGFAAATSAL PLPRVRDEWLRPGRRQAVAVLVAAIACASFSPAWTGRLLPKGAYEKVPEYW HEAADFLNTHAQGTRTLLFPEASFARQNWGWTRDEPAQPLLDVPWAVRDAI PLVPPEAIRGLDGVVAALHEDPATGSAALRRLGIGAVLLRHDLASTEGANGR DDSPLAAWRDASSYGGQVHSFGPDDEVEVIILPPADDAVGTLTSANPVRVA GGGESLAFLDAHGDGSAPAYELVDRDADIVTDTPTLTDRNYGTLDGPISAPL AEGDPSTVNNRLRDYPSAGPLTKVVSHGGTVRASSSAADADAFGGANPAR SRTAAVDGENSTAWWPAPGDTGWLELNKDPNTPAWHAPVVKLMATAHTEV

						TVRSGSAKVKVKLKPYRSREVRVPGGDTRSIRIDLSERVGIANAQVVGSPIER VVTVPATSPDVHQFFFQQLAQDTRILIRDFTVPRTMRLELDARKPVLIDGVRY NPGAHLTLRSGTHRVRSTGGWISLREVGWSPAAASKRTSGTIEATGQDRLLI TGNAFNPGLQGHLDSEAGSIDLAPREINASLQAFVIPAGRSGSFRMSFAAAR TYKAALGLGGALALLTVAGCLLSLRRRPTAPWQPDDGRGAHILAAVCGLGTL SLLGLPAVLAGVAAWVVLRWTSLSAAPLAAALTAAAGTMLARAPWASGSYA GDSVLVTCLCAASLACVLWPSRRAN
Other gene	WM42_1657	hypothetical protein	1705152	1705350	65	MALETDSLNKRTLGPAVGSTVVGIALGVITIIGIAQFSNADAVPSDGAVSASDA VMGGPEYGSRN
Other gene	WM42_1658	universal stress family protein	1705383	1705881	165	MSNGETMLIAYDGSTRAARAMEHAARLLRPRHVEILTAWEPMARQAARAVS RTGMHQSTVAPEAVEDDPAYEEALRICRQGVEVAENLGLSGRAHLVESVTTI ASAIVDAAQELDVDIIVTGTRALSGFRGWWTNSTAEHIVRNAGLPVFIVPPEL DDDAEEEDS
Other gene	WM42_1659	glycosyl hydrolase family 3 N terminal domain protein	1705940	1707110	389	MELSPSPMFSARTAATAVLTLGLGFGLASCDSEQPESTQEPAASSSAPGAA SASAAPSTTPSEEVPVPVDRRQLAASVLMPPAVNYDDALAKLQAGAGGIFIP SWADPALLSEPGRDINALRAAVGRDFQVSIDFEGGRVQRFSEILGEHPAPAQ MAAEHSPEEVEQMGAEIGRVLKEHGINVDFAPVLDVDGGELEVVGDRSFST DPKQAGEYGAAFARGLESAGVKAVFKHFPGHGRASGDTHLGEAVTPPLEEL YGHEFVPFQEAIPAAPAAGLMMGHLVVPGLGDGQTPASINPAAYQLAREQL HYGGPIYTDDIGGMASIADTMNVPQAVVAALASGADMPLWSTDAEFPAAVD AVAAALDDGTLTDVQLQASAQRLTGGKLPE



Appendix 103: Type 1 PKS CsimB biosynthetic gene cluster and its predicted similar gene clusters according to antiSMASH.

Appendix 104: Summary of genes in BGC possibly coding for NRP CsimC as predicted by PRISM.

Open read frame	Start	End	Sequence
orf_1332	1398542	1399487	MRVAVFGYQSWGHRTLSAVINAGHEVVLVVTHPASDHPYEQMWADSVEELASEHELQVCVTERVGQDVI EALRDAAPDIIVANNWRTWLPPEVFSLAKHGALNVHDGLLPEYAGFSPILWALLNRETHVGVTVHEMDEVL DGGPIVAQRAIPVGPQDTTTDLVAKTIDLIEPLVERALSDVAQGTATAQPQDPTRATYFHKRGEQESRIDFT QPAEDIALLVRAQSDPYPNAYEFRGQRVRVLSAHVSQGRFGGTPGRVTIPHEGGIAVVCGSPRANVPAPA IVLDRVRLDDNSELTATEFFGHRAGYIQPKV
orf_1334	1400553	1411410	MPTRSSAHTPFSSGDTLASRDGGAVIGHGRSEQDRIAQQRIWNDTDQDRDRPDLISLLLQHAQETPDALA VVDDRHRLTYAQLVAHATAVARNLREHGIDAGQSVGISLPRSAEMVVGIVATLLAGGSFVPLDPSWPQAR RESVTHDASLSFVLTPDNCALTEDALFDLDATRELFTPPSTDSVAYVIFTSGSTGRPKGAMIRHGAIVERLL WQRDQILFFGRDDASLFKAPLAFDISINEIFLPLVCGGRVVVAAPGVEQDPQRLARLIHREGVTFAYLVSSV LDVMLKQAEGTNLLDSLRHVWCGGEMLTQALFRRFRQQLAIPLYHGYGPAEATIGVSHVIYRDDEDRLNT SIGVANPNCRLYVLDEHLRVVPDQEIGELYVAGFLLAKGYINAPGLTASRFVADVFASDGTRMYRTGDLVR RHNDGSLEFVGRADNQVKIRGMRLELEDVESALVGHPDVEAASVIAREGRLLGYVTVTAGLVGAAIRSWC AEVLPEYMVPAIITVMDELPRTANGKVDRKALPEPDWSSLTDAPADAERDGETVALLAEAMAEALGVSAV GAETDFFDIGGDSLRAITLVSALGRRGVEVSVGDIFSARTPQQLARCAEERGTFVQDPDDEPTGEVQSLPI LRWFDSITDHVDGFIQSVEFSVPEDVDAQLAGRMVADVLQAHPALRARVQRNPLRLELPEESAEEAVAIHP TTDIAALSELLDPAVGVVVAAGLVPGRLRLVVHHLVVDGVSWNIIGEDLAAAYRGEQLAPERTSLRRWTQL LQQAVDTGEFAEDANSSLPPLPSADEPLRDPQVPALADSPEKAPTVREERSVVHEASVQITDELLGAVPHA FRTGANSVLLTALSVALARWRRNKQTWTLVEMEGHGRETRFVPGPQGREADLSRTVGWFTCLYPMLIDP TRAAVQEASTEVEGSIAPLALALNAVKDQLAAIPGNGVPYQAHTWLHQSAKTPPQAQVLFNYLGRVSAGA QDFAPAGSTGQLGEQRDPDQPLVRELEFNAIAEDTGEGYVLRTTISWARGRISPERIDELVAHWDVALREV AALADHGVLSVGDVAPAPVNSADLARITAQSSAELQDVLPLTPLQHGMYFHSLFEESASSYVEQQVLRVE CSEPFDRERFARAARNLIRRHPALSTRPWETDGGDVVAVIDPGIAEHLRVDFRDVTVPAELAGPGLDGWL VQRTEEIAADDLSRGISLQPPGDAAPEPLMRWTVVLPTSVDGAVCGQDIAVIQTVHHLIADGWSVPIMLRD LLEIYRDDDARIPRYDPDAGMAGSVRWVARRDAEADMSVWREEMREVRPTVLCPNPSSSLERRELLVDD PRTVGLSERARVAGVGLPDVVHAAWGLVLRTLVGCEPGADVVFATSVSGRDIPVLGVADAVGMQLNTIPV VAPGQSDPTLPVTSMLHAMVHHNNQVRDVQHVSLADIARDLGTNASELLDTLMVVEVPLSPQDVGCPGS PLQVADVRNNGAPHFPLSVVVNPSAEHPLRLIYDPQRITEVRAERIAQMLAVSVDSLLSESGAVATVGEVA EALGAVSGVDTLPSLWRRSFEHSRDRPALTSIGEDGAAEHWTYEELDDAAQRIRAVLDRKVAIHTPRVALL MERDAWQVAAILATTMSAGTYVPVDPLSPQARVELILEDCQPDAVLVSPSAEKMVSELVDCPVLVVSEQT MSGEAKPPAGRSASVARANDIAYVIYTSGSTGRPKGVAVTHANVTAMLGNARSHVEFSQEDVWSISHSFA FDFSVWEMWAALSSGGRAVVMPYALMRSPEDAAEVLRAEAITVLSQTPTAFAALEPHLGQDSAVRTVIFG GEALEARAEAAYCSAHPNVRFINMYGITETTVHVTAHECSENAGEARSPIGRPMDGLRTYVLDAQLQPVQ PGETGMMYVAGPQVTAGYWGLASTTASRFVADPFVGGGARMYCSNDMAKVLNNGHLDYVGRADRQVQ LRGYRVELGEIESALEKVSGVREATVVVVDLPEGQVPGALLITDSRADAKAITSRAAAAARDALPAYMVPQL FAVSTQVPQTINGKRDERAILDLLGEIPAAQQASGTSDLVEAISQAIADALRLDRAEVEPDSDFFRLGGDSIL AIRVTHALARADLNVTPRDFFLGRTPRKIAERVTPQATQQKIQETRQSQEEALPKDGHQEISGAFPIPAMLR RQMERGMSDRFVQARRLDLGPVAVEDLEQALQQVAQAHPMLRTRVDTSSAFAHFVGAEGDGPMVVRAL DVDALIDHIDIAAGRSVVLGCVDGYVKAVAHHAVIDSASWMILEDDLRDALAGRRILSGQASYKDLCLQELH DAHTAAAQDALHWSELATLPRPVEDSNQWSTDHVSTLEVNIDGQTAQVLQSVAPDVMGVDVQDLVAGLV

			TVAVARTMGTDARSWAVDVEGHGRPADHSYGRTLGWFTTIAPVEIPLADPADAARAAAEMRALHEDGTA PDRRTYQALRYTHPQGQRTLAHGAQILVNYLGRGETGAVLHAPGDAACQWTDYLVEADVWSSERAVSM ELSVVNAVISAERLRSAIGEVARELQEHVSREQSSGRRAPLSVLQRGIWFQSQVAAPGAYVAQTALTFDR RLDAEAVVEAFRDTVTVHPAMGAEFHTDASGQPVQNLPWSGQGIDLPVETTEGDLEAIMHADRSAGIDLA SVPLAKATIVTGRSDGQPGDTLLLTYHLVLVDGWSRAVMLQTFLERLTLRSGQARTQGAGELVPRGCSIA DALLESVDTRKEQADSDYWVHRLETLSQPTLVAPQAADISSEHAGSELPRQVFAEVSEQLTGALQEKIRAQ AVTLTSVVNAAVAVALGAVTGDSDVVFGQSVSGRDALSDPAMSDVVGVLLNTVPVRVTARPGQSIEELVQ DVYRQRIEDMDHDSADLGAIQRQLGVGTLFDSMVVVQNFLDPQAAAELRERHGVVEERAEDSTHFPLTW VFTPGPKLGIKLEFRHDVVDESLAHSVLKAATEVLTAFVDTPEVPLAQLAQLAPARAEDSSPQSTTEQMAS WQKAEGIDRTIADELKDTAQRFPDRIALADDAQQWTFGELIARCSDIAEKIKNCGVTSGDTVAIAVERSAHS VVALLGALWAGVRYAPLDLTHPDGRLRVLVEDSQPAAALVDSSSRERMERIGALPCVDVTTADSHATTHT PAAVPGDDAYLMYTSGSTGKPKGVVIKHRGLHNMLDNHRRKIFAPAAADGRTLRIAHAISFAFDMSWEELF WLVEGHEVRIFSEDLRRDAAAMVEAIRAHQVDVINVTPTVAEQLLAEGMLESGAHRPRLVLLGGEAVSHG VWETLRKADDVRGYNLYGPTEYTINALGAGTDESATPVIGMPVDRTAAFVLDPWLRPVPTGAPGELYLAG SGLAQEYHGLAARTASSMVACPWGAPGERMYRTGDIVRVRADGMFEYLGRSDDQVKIRGHRVDPGDVS AAVSRGVDPRILHCVTVPVRISDATLLACHLVAPQLRDADQGERQSFLTGVRNALREELPSYMIPDRWSIV DELPVTSNGKTDLAALGEGERITEKGREPANETEEITAELFAESLDIEPEDVPVDADYFDMGGHSMAVIKLC ALLRGELGVEVGVREFYGLRTVERVAEFEARS

Appendix 105: Summary of genes in BGC possibly coding for NRP CsimD as predicted by PRISM.

Open read frame	Start	End	Amino acid Sequence
orf_1759	1876674	1877460	MTEISCEGVFPFVAGHRTGTLLFCFHHAGGSASVYRGWVGVNPSIDVVPVELPGKATRRREKWVSDFDK LADMCATEIVDLAAGAPIALYGHSMGAALAYQVAACLQQMHRPTIPEAVVAARQAPGETVPGEYHSSMG FGALRREFEKVGGTPPEILANDDVMKLLLADIRRDYVLHEGFHHATTVLRSPVLALAGDSDPAVSPEMLS RWENFTSGSFELKVLPGGHFFPLDTGTEFLDLLAGFLAGITGNTAWLARNNA
orf_1765	1883920	1891492	MIKEEQIREELLASLHQILGEDAEIGIDDNLLSHGLESLPTVRLLADWMKQGHRVSFGDFMRAPTVRQWA KMLVESTPNHSTDAIESPKDGFAAPIDDSVPFDLTDVQYAYWIGRNSSQQLGGVGTHGYVEVESRSINID RLQQSWLTLLRSHPMLRACYTEDGKQYVLPEPPHPTILVHDLTKMDESTREEALLSTRERLSHRLLDIATG HVVSLEVSLLPQDVAVIHFDIDLLVCDVQSFQIILHDLAHHYATGEAPDADPSWSFARYLAGHAREGVADI DRDQAYWRNRLSELPGAPTLPMSHGTNEEQAHRFVRRSRSFDSATWSRLREVCEHHATTPAMVLLTAY ARTIGQWSENKKFLMAVPLFNRGSDTAIKNVVADFTALTLTSIDQSTRRTFSEDLKDIQASFYEDSSHSQY SAVRVLRDLRASRGEQVLAPVVFSCNLGDPLVGQEFIDTFGEISYMISQTPQVWIDLQVFTTVNGFLIVCD AVEQLFPEKMLDDLFATLVMEIDKAITDDLSHSDPVESPGAQARRASRAEVASWRLPDTTLVDEVIAAAR CHPQATAIRSASGDVITYQDLEEQATTIASALVNSGVGRGALIAVMVERGPRQIIGALAAMMAGGAYVPVS LQQPESRIAALLGASQVTHLITDRPDKVLSETAVQVVDFTSATGTANLPQLHPQDPAYVIYTSGTTGTPKG VEICHGAAWNTISEINRRLGVGPTDRLLSVSSFDFDLSVYDAFGLLSAGGELVTIPDDARRDAKKWVSLVD SLGITIWNSVPTLFEMLLSAADRTPGKLSSIRHVLLSGDWIDTSLPERMRTVTPQAHLLAMGGATEASIWS NGLDLDVVSPEWTSIPYGRPLAKQMYRVVSSNGQDCPDYSVGELWIGGLGVATQYVGDPGLTETKFVIS EDSRWYRTGDMGRFWADGTIEFLGRSDNQVKVRGHRIELGEIESACEALLPIERAVCITHQGASSSPSLV TFAQFTPSHVARTTPEQFATSLRAKVNDVLTEGDIRTSVEHDEHLQTAYAFSVMRRWEEQLTGVGTPNH LREHRNRWQTWLGKADEHPATADLLLDDESFGALERFVTPFEQAFVMAEKQRSIAEFIQSPDSMSVEQF LATRPLGRLVHRVLGAVVRECSTHSTSELKILEIGSRRPEASADYAAIAGTSAYVLADPYRHHLEHAGQRV GNTFTYRQLGVTSTPQPIPGEAVTKADLVLCNQTLHQSEDIEKTLCEAWGLSAPGATMVVVEPTAPSPMS DITAAFIANNTTDARAETGTVLLSARSWKEILQRTGWKPVEHVEITKTTALIIAERASSNESVTLCDSDYAKA TNLLATRLPEYMLPKRILELAKFPLTSNGKIDRKALTALVPEYFDNEPAVTELPHTATEKRLIDIWDELLHTS SNVNSDYFRLGGDSLTATRLRRTIEQCFGVEFPLENIFDVPLLRDMAARIDQIAEVPHQQSDLPKIVHGSE QYAPFPLTEVQQSYLIGSSGAIELGDVSSHCYFEMSTACLDPERVEDAFNALIKRHPMLRTVVCEDGLSQ RVLPEVPRYRIALIRSGNADNEDTLDEIREEMSRQKFDPTQWPCFDVRYVAEPDAGRLLLSFDNLFIDGW SMFHIFREWKQAYDHGVDSLDPAIPYSFKDYVEATIELSHSDIHKRDQAYWESAVDTIYPAPQLPVTDTNG ANTSQFCRHHALVDAAKWRRIKQRVREEGMTEAVFLAEVYAEVLARYSDEPRLSINLTRFDRTRFAPEVD HIVGDFTSLSILSVDTQCAPSFRDRAAALHRRMFSNLDHGSVSGVSVQRMLTKQRGARVTMPVVFTCGL GVVEHPESDQSPYLGVIDHGLSQTPQVWMDLQVYEHDGGLMLNMDAVEAIFPDDMVAELFTSLTATLSH LAESPELWNAPTSTIAPTTNAPTADRINDTDRELPGADKSLLGLYQKGLAEHGDNLAVIDATTQWTYEQLN EQSDKWAQLIAATDPAPGDLVGIMMEKSAQQIAAVLGAMKAGCAYLPLSVDQPVGRNTSIINDAGASIVA MDHPDDDFAALAEHCTVITLADVARHRPGDQALSESSPTPSSLAYVIYTSGTTGTPKGVAITHESAVNTIV DVNERLGVTPTDRILGISELNFDLSVYDIFGMFARGATLVLPSPADKRDPQCWADAVTTHSVTLWNSVPA LFSMYVEHLRERSLIGSSVRSALLSGDWIPVNIAYQVSTLFRDCTVFAAGGATEASIWSNWYEVGVDDAS

			RTSIPYGTPLANQRMYILDEALNPRPTHVPGDLYIAGRGLAMGYWKDPEKTAASFITHPRTGERMYRTGD KALYNHLGHIIFLGREDGQVKVNGYRIELGEIESTARKFNELRDCVAVNDHGIVLYVVTHEGFNMAALNNH LAESLPAYMRPRVISRIDGLPRSWNGKIDRKSLEGKTFEQPQTRERSRNHRDSGIVTILQELLGPKEISIDD DFFTIGADSLTAVRLTNSIRREMSVEISIRDVFNHPTVRELSDLANIVGSDVEEGEI
orf_1766	1891504	1893136	MRGRDIDRNSVLKHYEGTGHLARKSLCQELFESSEIYSDHVAVIADDAHLTYAQLEQSLCVFTEELRESGL SPGDHVLLQLPNTAAYVVTLLALMRVGAIPTLLLPAHREAEVAALCESLHPVAYIGGRDHLGFDTVAMVEA MGPGELGLKELWADNGPTHDKESSYRVLPGLFTAPISTKCSPPTKWPDPRSVALNLLSGGTTNMPKIIPR VHEAYAYNTRAAAQCCGVGPDTVYLAVLSTSHDFALAQPGILGTLLSGGTVVLCTSAAFDEAFPAIATHG VTLTALVPAVAQVWVEAAEWFPADFSSLERIIIGAAALNDGLGEAIQDRFGVRIHQGYGMGEGITTFTRIDD PPAVILGTQGRPISDADELVIDGPGGEPGEILEKGPYTFFGYEGNRDTPDCFTEDGFFRTGDRGYLTEDG NLVLCGRVVEQINRGENVSPSEVETLLSGTPGISAAAVFPMPDRALGERTVAAIVAQPGVNRSAILDDFLT RGVARYKVPDQVITVDEIPLINIGKVDKKKLRALAAAQFTDRESEQS
orf_1768	1893733	1899268	MDVTALINDLESRGIALWVNGDRLNYRSPKGSLREEDLAALRSNKEKVLAWLREREAVPHDEQARFAPF PMTDIQRAYATGQNEGYDLGGTGCHSYAEIRTERLDRSRLEQAWHELIQRHDMLSAVVVPPDSLQVVKS RSLPVLQAVDLAGHNPDVPDAEYLRHRAKLENRSYPLGTWPLHEFQLLQFDECSILQFSVDMIIADFVSV RVMVEELLTLYAGNVLPELEDTTFRDIITSRNHHSQSAAGFAARTNAKKYWSEIIPSLSGKPLLPTLTSADR TSEMPVRFTRRTWRCSPAAWSKLTDAASTHGVTPSATLLTAYADVLRRWSSTSDFCVNVTSMNRDSAIA GINRIIGDFTEMTLHACHPHTGTFSERVHATQEQLSEELSHAAYSGVDVLRDIARTTGQPAVIPVVFTSAL GADTPHNNGPAYNLVSGVSRTPQVWIDCQAFQDGGSCNVNWDVREDVFEPALIDDMWESFTDLLDRLV DDGSAWQETDSVHLPDKTIAIRNRIHKTHVQQTTRCLHDGFWDNVQQHPHQPALVCGGKTYSYQHLAG YVGALQHELSDVGPGDYIAIVLGNGVWQIAAAVAVVSTGAAYVPIDHEQPAIRQRSMIEACRPANVITNSH FSEENTDISNINVDTLSPIQYSGTIASPVSPTETAYIIFTSGSTGIPKGVVVTHSAAMNTIDSVNNLLGRNKR RTVLGVSKLSFDLSVYDIFGTFASGGTLVLPLDEESRNPSKWIDFLVDNNVDTWNSVPALFQMLVREVEV TRHPNILSLDLVMLSGDRIPGTLPAHAAPHFPNAELISLGGATEGGIWSIFHPMTCHTNETSIPYGTALPNQ GMWVLDEACNECPDWVRGQIHISGESLATGYLNDPTSTAEKFFFSEKHGTRMYATGDIGSYRPDGVIEF HGRRDNQLKINGYRVETGEIEGVLESNDFVERAIVLTQETSDPIKLHAFVTDAQSDKDELKDAGQIRNSEL RTMLEQRWTPADTSLDTGIFATWMRLGNEAAMAALLAAFQQAGVFLVAGKYHTLTEITAAIHPSEEYRELI TRWLNILTGEGLATKDDEGWTVSQQTLDFFVFGEAWDQFGNMEAEINNSKELFNYQRHAAEALLSQLRG EISPTEVFFPEGDTHNARTIYGENRISKAMNAAAAEAVIGIAEHHADHPVRILEVGAGIGATTEKIVSRLPEN VIEYRFTDISTFFLHKAQKMFAHCDAMTYGLFDMNSDCTSQDVEFGGYDIILCANVLHNSVNIEESFTRLK QLRRPGGVIVIVEPITELYAALISVSIKMNLVDFTDHRAESHKVFIEDAQWDQVFRDTQMHRIAEYPNTSDP LRECGQRLIIVGADDDDVPTLNSEDILGYLRAHLPGYMVPASVNVLPELPLTSNGKVDRKALAQLCLEPVG SPNNRIDPPRNETEEQIATIWRDVLDTTEVGRNDDFYALGGDSLLMAETVTRLRQEIPGLQQTWDALMRG VLKVPTIAGISALAQAAGSSCQPEALKAVNSANHTSPELTALSTVASGSPTGSSNLHVYRLPKDAMFCRV MIHAGTGRLKDYEFLMPELLQRQPEIAHVGFTAGDADRFLDYTTRTLIRDLAQSYAQELDELDMESYQLV GYCIGGMLALETAKALTELGRDVRQVTCISTHQCPHRVTNELLCELAYGCIFNADLSAMGANFDLKTLAAA LEHTLDGINRNISDEELCTLEGPYADIGEFFQKMAVLSPRARRKLIYRSIREFDTDSESTRGMLDILYDVFR HSLLGTIDYVPDVYFGDVVVLQPTEGVTGFYPSLGGDIDWPATVLGNLQIHAVAGSHATCLLQENVPSLLP FFTEREQRNG

Appendix 106: Summary of genes in BGC possibly coding for NRP CsimE as predicted by PRISM.

Open read frame	Start	End	Amino acid Sequence
orf_1032	1081195	1082914	MSAIDNKEWLQFYPEWTPHHLDYGDTTLLDIYDNNLAINPDKPATYFFGKTQTFAELDRQVRRAAAGLRAFGV RPGDKVAIVAPNCPQYVAAFYAILKLGAIAVLHNPLYTAHELEGLFKDHGARIAIAWDKTASTLEKLRGTTNLET VISINMIDAMPTTQKLALKYLPFLKSKREQLSTHAKNTVTWDSLIGNAIGGDGRDVKTPEDITKDTVAVIMYTSG TTGQPKGAQLSHGNFFAVILQGKHWVPGLGDKPERMLCALPFFHAYGLVMNVILQPLIGGELVIIPKPDISLIMD LMKKHTPTWIPGVPTLYERIVKAAEEDNIDIKGVRAAFSGASTLPVDTVEKWEEYTGGLLVEGYGLTECSPIIVG NPMSTDRRPGYVGIPFPDTEIRIANPDNLDETQPDGVEGEVLARGHQIFKGYLNNPEATEAAFHDDWFRTGD MGIMEEDGFIRLVSRIKEIIITGFNVYPGEVEEVLRTHPAIDDVAVVGRPRSDGSEDVVACIDLANGVALDPEGL KEFCRERLTRYKVPRTFYHFEELAKDQMGKIRRREVQADLLERLEKEKAGQ
orf_1034	1083796	1085527	MSDTQPSAYESKAWLASYAEWTNHHLDYGEQTLLDIYNENLAKHGTWPATWFFGRQQSYHELDVQVRAAAA GLKAFGIRPGDRVAMALPNCPQHIAAFYAVLKLGAIVVEHNPLYTAHELEGLFQDHAARVAIVWDKMTATFEQL RETTPLETIISVNMTEAMPRSKQLLLRIPIPPITEQREQLTAPAPNTVPWSTLTGTAIGGDGTKLADAEIHPEDTAI ILYTSGTTGTPKGAEITHANVCANVKQGVAWVPEFGSKPERMLAALPMFHIYGLTLIAALGVYVGGELILTPAPK IPLIMDIMKKHRPTWLPGVPTLYSKIIEAAEEQGADITGIHNSLSGAAALPPETVEEWESLTDGLLVEGYGLTETS PVLTGNPMNNSRRPGYIGIPFPDTEIRIANPDNLDETMADGEPGELLARGPQVFKGYFKNPEATEKAFHDGWF CTGDMAVMESDGFIKIVSRIKEMIITGFNVYPGEVEKVIGMHPDVVDVAVVGRPREDGSEDVVACVILRDGAAL DPEGMKDFARQNLTRYKVPRTFYHFEHLAADQLGKVRRREVQADLLQRLKSGNHS
orf_1035	1085571	1087110	MQFSSSQPWHRFYGEWTPTSLDYGNATLLSQLAETIAAHPERVALRYMGEELSYAQLDAEINRAASALSARG VQRGDFVALALANRPAHVIAFYAALRLGAAVVEHNPQFTTHELTPMVAHHGARVAIAEGPAVGIFRELPQLEH VIDADADWEDFLAEGKKQTPRAEVNKDDLALILYTSGTTGTPKGAMLSHGNLAANALQMAHWRLPEEHPQAM AVLPLFHSYGLTMSLTVGVVTGASISLVPAPRLELILAAIAAAPPTWLPAVPTVYQRIVDAAPELPKITQAISGAAT LPSATIGQWETQAHGFLVEGYGLTEAAPVVTCNPLNEHRRAGCIGIPMPDTQVRIDPDTGELLVRGPQVFSGY LHNPEATRTAFQDGWLRTGDLATQDADGFITLRARLKETIITGFNVYPGEVEEVLLSHPALADAAVIGRPRPDG SEDVVACIVLASAQPGAPADAAQPTQEEFREWCRGQLARYKVPREFIVVPDLGRDGLGKLRRRAVAERFL
orf_1036	1087297	1096399	MHLSSLTPLHRLTGATILFAGQGSAWQKALAQAAAQPQVRTRLVELLESTRTTTAPVARQIASTCPGVFDRLQ ELIDGEDNVQDADVFPAYSIPGIVLGQIAAVEHLAQLGLRPERALGHSQGSLGVLALHDPQEALSLALVMGTAA SASHGSVDARSHMLALRGLTREFVEERLQGDAAIAVVNGRRAFAVSGAPDCLAATRAALEGAVDEYNATLAE RTVGGDEIELHATELPVALPFHHPSLHAAAELTEEWAATCGIADHGLAREILVEEHDWLSRIQANIAEKQAQGS EFLLSLDASLTRLSEPLVTGSGSVIVEIATPEQRDALATPGTELPEPLDYRDFAPRLVELPDGRFYTQTRFSDLT GLSPIMLGGMTPTTADGEIVAAAANAGHWTEMAGGGMYSDEVFRAHLKVMEEHLNPGRTAQFNTMFFDRYL WNLQFGQTRIVPKARAAGASFNGVCISAGIPEVEEAGSLLNQLHSDGFPYISFKPGTAQQIRDVLKIAAAYPDD QIIMQVEDGHAGGHHSWVNLDDMLLETYAAIREHNNVVLAVGGGISTPQRAADYLTGKWALRYGQQRLPVDA VFIGTVAMATKEAKATDSVKELLVGTTGISPEDNGGWVGRGTGTNGVASSQSHLLADIHDLDNSFAAASRLITS LSTEDYPAHRAEIIAALNKTSKPYFGDVEKMTYAQWLERFVELADPFVDPTWDDRFFDLLHRVEARLNPVDHG EIDTLFPSIDDVHDAPAALKLLLETYPEAASTSVSPRDAAWWISLHYKHVKPMPWVPAIDGDLKTWFGKDTLW QAQDPRYTADQVRIIPGPVSVAGITRKDEPVAELLARFEAATTQQLRDTNAEPQRLFSRLHAVADAAEFIKASP TLVWHGHLMDNPAFHMNDSAFELRQDETGQWEIVIHADSYWDNLPDKQRPFYVREVTVPVDLPADVATGGS PVVSHERLPDAVFALLRGLAGVGSVAEQGDVIDEMPRIAEGSIDADAPFGRAHYSFHLPASLLEAHTAVTGAG LPVAEKALPVGTPDVLVGPCWPAIYTALGSGLLPDGYPVIEGLLNAVHLDHLVDLRVPLAQLADGRRIDVTSQC TSIEESASGRIVTVELELRNGPDVPEKAGEIVAMQVQRFAIRGRASGTAAPTPAPEWGGGKSSTKVTSTPRSFI

			DRAVVTAPQDMTPFALVSGDYNPIHTSYNAAQLVALEAPLVHGMWLSATAQHLAARHGQVVGWTYSMYGMV QLNDVIDIVVERVGRKGIHAALEVTCRIDGEVVSRGQALLAQPRTAYIYPGQGIQAEGMGSTDRAASPAAREIW RRADQHTRAALGFSIRHIIDENPTSLRVGEQTFVHPQGVLHLTQFTQVALAVVAYAQTERLRENNAIGSNSMY AGHSLGEYTALASLANIFDLEAVIDIVYSRGSAMGSLVPRDAEGNSEYAMAALRPNMIGVSGDEVDTYVASIAE ETGEFLEIVNYNIRGQQYSVAGTKAGLRELVRRANAVKERAAVMVPGIDVPFHSRVLRDGVAAFAEKLDELLP AELDLDALVGRYVPNLVAVPFELTQEFVDTVAPMAPSGKLDGLDASSLEPQALARLLLIELLSWQFASPVRWIE TQELLFDEVDQLIEVGLATSSTLTNLAKRSLAVAGRDLPVFNVERDQDQVMLQDVQAAPEPETAPETEEPAVE AASVPAAPATPAAESASAPAEFIPAVPAPAATSAGAAPELTFRAAEAIMVLFAFQNKIRIDQINDSDTIEELTNGV SSRRNQLLMDMSAEIGVPAIDGAADADVASLREKVNAAAPGYAPFGSVLGEAVTARLRQLLGAAGLKPAAVA EYVTGTWGLPESWVAHVEAEILLGSRDEDSVRGGSLNTVASSATSKSAAHELIDAAVQGVANAHGASVSKGA AAGGASGGVVDSAALEAYAETVTGENGVLATVARQVLTQLGHVAEPAEATAPDTEVIEAVEAELGSGWLKSV TPSFDAAHAVLFDDAWATARERLARVALGEIAADDERAQPAAFQGAGATVAEQARWWARSGNTAVDSAHFE QIAAAAESTKRGAYADDVALVTGAAPGSIATALVERLLAGGATVIMTASRVTQSRKEFARRLYAEHGRQGSAL WLVPANLTSFRDVDSLVEWIGSEQRESVGNEVKILKPALTPTLAFPFAAPSVSGSLADAGSSTESQARLLLWS VERLIGGLSDLAVKAPSPTRCHVVLPGSPNRGTFGGDGAYGEVKAALDAILAKWNVEAGWPAGVTLAQAKIG WVAGTHLMGGNDVLVPAAQKAGIHVWSPEEISSKLLDLASAESRARATERPIEADLTGGLEGFSLTSLEVEKP AAQSAAESAVKNTTARIKALPSPARPVQPQLAEELGDITTDLDDMVVIAGIGEVSSWGSGRTRFEAEYGLQRD GSCELTAAGVIELAWMTGLIAWHEEPTPGWFVGETDEQIAEEDIYERFRDEVVARAGIRELTDKYHLVDRGSID LTTVFLDRDITFTVDSEATARDIAEADPEFTSIREADGEWEVTRHKGATAKVPRKATLTRTVAAQMPDNFDAAR WGIPDHMLDSLDRMAVWNLVTAVDAFTQAGFSPAELLQAVHPGQVATTQGTGIGGMESLHKVFVSRFLGED RPSDILQEALPNVIAAHTMQSLVGGYGSMIHPIGACATAAVSIEEGVDKITLGKADVVVAGGIDDVQVESLQGF GDMNATAQTSAMTAQGIDERFISRANDRRRGGFLEGEGGGTVLLVRGSLAAELGLPVLAVVAHAASYGDGAH TSIPAPGLGALGAGRGRENSRLARSLRGLGLSPNDVTVLSKHDTSTNANDPNESELHSILWPAIGRDADQPMF VISQKTLTGHSKAGAALFQTGGLIDVFRTGRIPANVSLDCVDPLIAPKAKNLVWLRSPLDLAAAGRSVKAAALTS LGFGHVGALIVYAHGVFEKAVAKQRGTDAAAAWRERAEQRLAAGHARFEAGMLGRAPLFEVIDGRRLPHADA KVMIDGYGLVDADKAAEISMLLDADARLGSNGEFPSV

Appendix 107: Summary of genes in BGC possibly coding for Type 1 PKS CsimB as predicted by PRISM.

Open read frame	Start	End	Amino acid Sequence
orf_1592	1679093	1680932	MDLQALIGQFFDGNGRIVLPPHLTLPGLGEHVFEMEQQSGVPDRPVLRQWLFDDDPAGTPRDFTRSEVNK RIKVVAARLQQVGAVGDRVAILAGNSPEYLFGFMGAMYAGMTPLPLYDPNEPGHSEHLRAVFGDSNPSIVL TNNVSAAAVRRYFADLPSAQRPRVISIDSLPDSLAASWKNPVEQGAGEGAAAPVDAPAFLQYTSGSTRTPA GVLLTNRSILTNVLQIFAAVHIKMPPRIVSWLPLHHDMGIILASFATILGLNFEMMTPRDFVQQPKRWVRQISR QPGDESFAGIYAVVPNFALELAARYGAPVAGEDIDLSHVDGIVIGSEPVTETAVDNFWEVFSEYGLKREAMR PSYGLAEASLMVATPQEGERPLITHFREQLTAGKAVVEDKSEQTVAFASNGQGMPAQYLTIVDPETKAELA DAQIGELWLHGDNMAAGYLGREDETASTFRNTLATRLPEGSRVVGAPEDDKWMATGDLATIVDGHLYITG RLKDLIVVAGRNHYPQDIEATVQEASAHVRPDSIAAFSVEGGNTEQLVLLIERADDADSAGDEAATEVIRTAV SKAHGITPEVIRWFSAHEINRTSSGKIARRVAKKHFLAS
orf_1593	1681113	1685970	MTVEELRGWLRNWVAQTTGLSAEEITDSKPLENFGLSSRDAVVLSGELENLLGMKLEPTVAYEYPTIAQLAD RLVNGVASSSNSAVDSSSGAASTRNAGIIGGDIAIIGEAGRFPGAQNVKEFWDMLVEGRAGTGPLPMGRW SEYAAEPTVSEKIANQNTDGGYLEDIASFDAEFFGLSPLEAANMDPQQRILLELAWEALEDAGLPANQLRGT ATGVYMGSTNNDYGMLIVADPAEAHPYAMTGTSSAVVANRISYALDLRGPSINVDTACSSSLVAVSQAVND LRTGASDVALAGGVNILAAPHASTGFSELGVISPTSAIHAFSDDADGIVRSDAAGVLVLKRLADAERDGDNIL AVIKGTAVNSDGHSNGLTAPNPDAQVDVLERAYADAGIDPQLVDYVEAHGTGTILGDPIEATALGRVLGSGR EMAAPMLLGSAKSNIGHSESAAGVVGLIKVIESMRHGVIPPSINFSAPNRYIDFDNEHLEVVEDPREWPEYS GQKVAGVSGFGFGGTNAHVVLTDYRGLTAETAAQVNIDEGAPVSLPISGLLPSRRAAAAGQIADYIEQEHPD LLSLARTLARRNHARSRAVVTGVSEEDVVKRLRQVADGKVSMGIASADSPSVPGPVFIYSGFGSQHRKMIK DVLEISPAFKERLEELDRIVEFESGWSILDIVTDDAQTYNTETAQVAITAIQIALTDLLASFGVRPAGVIGMSMG EIAAAYASGGITAEDAMLIAAHRSRLMGEGENSLGEEEQGAMAVVELTAAEIEALDGNIEPAVYTGPGMTTV GGPRPEVLALVEKLEAEGKFARALNVKGAGHTSAVDPILGELYADIAGIEPRPLHTTLYSSVDRGEIYRPGAL VHDEDYWIRMTRQPVMLQDATEAAFASGHTQIVEISPNPIALMGLMSTAFAVGKSDAQLLFSLKRKVDPTES LLDLLAKLYVAGAPVDYTAVFGSGALVDAPHTQFKRQRFWTNARPSSGISGLPGARVNLPEGKVAFSTNAD QAPSALAILEAAAEAVTPGSQIVASEEHGDLPPQGEVTTVVSKSLGGLSIAVYFVNGPATQLVAEGFASALNL TEASIPGVAAIAEPAAPSAPASFTDVSDVEAVRWDPSKETVEERLSLIVSESMGYDVSDLPRELPLIDLGLDS LMGMRIKNRVENDFQIPPMQVQALRDASLADVITMEDAVAGKSVAGAAETGAVDVGAGEVGAVDTGNAGA NSAESSAAEVGDVEADAEKDGVDKHDAENSAEGVGVAPRDASERMVFGTWATFTGKAAAGVTSQLPQID ESVAADIAQRLSERAGIEVTTQQVLDAQTLEPLADLVREGLETEVEGNIRVLREAEGPAVFMFHPAGGTTVV YQPLARRLPSDVAVYGVERLEGSLEERATAYIDDIIKYARGRKVVLGGWSFGGALAYEVAYQLADRTKRGE ESAEVAFIALLDTTQPSDPAPDTLEETKARWGRYAAFAKKTYGLDFEPPYEMLETMGEDALMTMLAEFLSN TDASEHGLAAGVLEHQRASFVDNQILGKLDMQRWADVDVPVILFRSERMHDGAIELEPRYAEIDPDGGWG AIVEDLEIVQLQGDHLAVPDEPAIGIVGKHINDWIEEKIR

Appendix 108: Summary of genes in BGC possibly coding for NPR/type 1 PKS LcA as predicted by antiSMASH.

Type of gene	Locus tag	Start of gene	End of gene	Amino acid length	Amino acid sequence
Other gene	ctg1_758	839455	840675	406	VKKFMIPGAAAVATAALSLLLGGGVANAQEVAEVPAPVQVPAIHAVDALGSAIKDAPQSPAV PASDVLSFLSDSINLLGRDILSAKGLSAKGLSALVTTENPPVDPQDPNAQSEDEGVMVVIGQ VAPYSLALLGIPALAIPGGVVGAVAGAVVGAGVAIFAGIAAAVIVASPALIAGLGTALVPALAIA AVGVVIGLIPALVLLVAGVVAIVSAIVVGLMVITWPLWVPAGIVLVLCGAGIIPVTIGGTLLSGG ALTPFLIAGAVILGLGLLPGIALVVAGLLFGVAAVVALLTLPLVGLAGLILSVAAFVFLGVMALI GSLGLLPAIVAFVVFGLIAAVVTIPVGAVIGLAAGGVIGAVAGAVGALIYVFVIAPRKKKDEEA PADDANAPKEKKMAVQADYALAA
Other gene	ctg1_759	840963	843263	766	MSTPATAQRTRQNILAISLLVLLSLLGGFLLPLLILWVLTFTNFPAFGNSNFIRSLTTVAQVFF CLTTVAGAIWAYSSRLHTQRGDNPHYWSTHFGPGLLSSLSVSGFVTVSLGLPLAATKLYLH GISGDQAFRTEYLTRLTDSPAPYDMAYPNMAPYYPSAWFWFGGRFANLLGIPGWEAFKP WAIISLALMCSLFTAIWCIIIRPELGIIIGMVTALVTTTYGSTEPYAAIVAMALPLTFVLSWHAIR PYRGALTKGRYATGGVAAILGVGACTYTLFTALGAVTLILMALIVTIHRLWVDRRELRQLAG KKVKTPDNTGITPDPVKLRKLRRHIWQPIIRLVVIGLISIAISLIHWGYYLISGLGQPHSRSGSA LHYLPSDSATVLLPMFKSDTVGLLCLIGLIWLIVAIRTSRISQALTIGIIAMYLWTLASMSSVLF GVTLLGFRVALPLTLTFAIAGVCGIFDATHRIISVAVRHVVIDATTTLDNDATPLVLAGQTPRY AGTTDADPNTPALVLPDDRLPGPSLFTSIDPDIRRSYSGPNQPIHVRRTARTIQAVIATITFLV GIGFAQDIPTSLTDEIQMAYTDTDGYGVRADRFPAGISSYYPQVNQALLEGASARQGYPAQ ARDIIILTTEEALLAYYPYWSFQAISPHYANPLGLYEKRNELIEDWALSTSPHDLLRKLQASPF PTPNAFVFRRDGDNYSLLLSEDVYPNQPNVHDYTVSFAANIFHSHCFLVKEIGPLAVVTRTC QTPQEMANA
Other gene	ctg1_760	843276	843788	170	MMTQNRILAAATTALIILTGSSALATADSTPEGPLGPHNPFAKDAKDLSWGNKDIPTGSQIR REMKEGKRPQCTIRLANNEASANCINNTDTPFFAAVSTTCTTKDANGIPTAKPVNASFAIPA HNSGAGIARCQPGTTARGISLTGPVTRQEAELMQKNSTEFTIKAFL
Additional biosynthetic gene	ctg1_761	843855	847466	1203	VTTSTPTPHDATPATQAASTTDDGTSTAPRRTRRTRRRVAIAATITGILGFLLAIASAVLPVR QAVATIDWQQTQQPTSVVSPLVSYLPQNLVVTLPCGSVADAPTNTLLFSTIPANSANAGSE GLQIRRTPDAAGTDSLIITSKSHLLASIAVTALTNQPADCRITTTVRADAGPSDPSVVVDVTG NTFPMPGSNNDGTENNSPVSGELATDSTTRPQVVGFFTTLPRDTRIPGLTAHMVVDTRFS TSPTPLKTALMAIGVLTSLASLLLLWRIDRMSRRYRRDALALREDAGEDPADLTDTNTVGEII DSTSTAAVTSRAASADAAAADAGDSTQCTTQRTARRKRAGKGGRTHAHNHGRTGVWITD ATVTLLLIIWHIFGANTSDDGYLLNMARVADHAGYISNYYRWLGSPESPIGWYYSILQGLTTI STASPFIRIPTLLAGIASWFIISHSILPRLGAAIRTNAVAYWTAGMLFLACWMPLNNGLRPEPI EAVLFIACWALVERAITRGSLLPGALAILAATFALGAGPTGIMCLAILFAGFRSYWQNIRMGIY TLRSAGTPKGLAATGVFAPILATGFLILYTVFADQTLGAMREAIRVRTEIGPNLGWFEDKAR WIALFGISPDGSVTRRFPIFIMLLILGIMIYTIARRTPIPGLSTPVLTRVAGISIGSVLLLALTPTK WTHHFGTFIGIVAVVGAVGAVLMTNRGIGIARNRWLAAAAVTAVTALSFTANNDWWYVSQ FGIHTKHGGQDFPTFLGKAFSTWFLALALLFLLIAALLHFFPSIDRTATTTSNTAIRNTTSWLA

					RQPLMLAVLSTFVVVFSLVSLLGAFIQRYPAYTVGSANLAAIKSHPCSLADQVLVEPNPGR WILQPVDGSNPTEALAGTLRDGFRANGLPDTIDDNQTELNTGDAVKTPDRQNHDVPPFGL PRDVPILGTYRADNQVSAHLVSDWYRLPSQRDDSPLLVLTIAGSLPADAVQLEWTTDPLPT PSTAANTSTTGGNSYKVAGSLTPMDPGPTPSWRNLRVRRPHDLPANATAVRLVVHDTSLA TDKWIAVTPPRVPQLHTLNQVVGTTDAVLIDWSVGLAFPCQRPFNHRNGIAEQPKWRILPN APNVESANAWEDRFGGGPLGWTDLLYYAENVPTYLNHDWLQDWGTLQRFIPRDPTAVPV TVTVTHTEGNGFYTPGTMRDTPPAPKRT
Additional biosynthetic gene	ctg1_762	847715	851167	1150	MGRIVVMSEPAPHTPHHADTTGATTASAHPAQSAADGKKRHTVPALWNLTTARLTAAITGF LGMVLFILAPFLPVNQTTATLTWPQDKSVTSVEAPLIAHSPHSITLDVPFSAVRAMKGQKDG IIASTLPGQSTQATQRALFVRVSERTVDVLSRDTPLLTLRRADVEAARNGNIHIDVDKNGAR ATVTGLNEPVTFTSDDAHLRPITMGIFSDLPLDTPQAAIQGMTITVHIDTRFTTEPTLVKLLAM IFGVLCMILSWIALARIDKIIQPTPFATTGYTRADGTIRSHWRSFTALDAVVYAVLLIWHFIGAN TSDDGYIHTMVRVANQNAGYMANYYRWYGVTESPFGSPYYDILQGFAGANPTSPFVRLPA LICAIITWMLISKAIIPRLGEGLHTRKITTWTAAGVFLIFHMAFNNGMRPEPFVAMMAILTWAL LEKSITTHRMLPATAAVVTAALALSGNPTGLMAVAALAVSTRPLLHIIRERRPQVGLAAQLG PIGASGFAVLTCVFGDHNIGAVLEATRVRGDIGPNMPWHREVLRYLWLSIQTVDGSMSRRL AVFTMLFCLIVVTVVLLRNREIRGALPGPSWRALGITYGTLILMMFSPTKWTHHFGAYAGIAP VLAALAALAIGNTALHSRRNRSLVLGAMGFVLAFSLATVNDWWFISQYHIPWFDKVPQIHG VEFFKIVMVVAVIAFLVAGWQHVTQDFTEMKEPSTAKGRERLRTLAAAPVTVVAWLTVVFY LLSFGKAAVSQYPAYSIAKGNLGVLAGHSCMMADDVLVEPDPNKGMLKPLPRQQWERTP EDGNSAARGPLAGPGTYNFEAWGVPGRVQADPVFVQQGTAGTNYKPEVPPTKQSGQPG GTEGGLDPNLGINGSRAKLPFGLNPKTTPVLGSFQYGAQHSAKLHTAWYSLPKADPEHPVI VFTAAGRLKYLNVNGIERYGQPVYLQYGVTQPDGSVKMLGQVLPYDVGPIPAWRNMRIPR NVIPAEANAVRLVGEDTNLDPDHWVAVTPPRLPHLVTLQHYVGSTDPTLLDWEIALQFPCIR PYTHYAGVTERPKWRITPDRASTVNNTNTWEATEFGGSLGITEGLTDSEEVPTYLKDDWG REWGSLEHLEPYGTRQPGLPKRAEIQYGTEKRSGLWNPGKMNIG
Additional biosynthetic gene	ctg1_763	851164	852711	515	VSMTTAEKLADLRVKMAQSEDPGSERSKAKRTALGRTTPRQRIDSLLDKGSFVEMGRLAK MPGNPMNPYGDGVVTGHGTIDGRPVVVYAHDNTVFGGSVGEAFGKKVCAAMDLAIKIGC PVIGINDSGGARVQDAVTSLAMYSEIGRRQYPLSGLSPQISIMLGKCAGGAVYAPVTTDFVV ATKDSYMFVTGPDVIRSVTGEDITIEELGSAEKQMEYGNVNYVAEDEYEAFSYVRRLLSFLP TSSNDPAPLGRTGLEPEVTDSDLSLNTIIPDSDNASYDVMDVLVKIFDDGDVLEVTPGYGAN VVTAFARVDGRPVGVIANQPKVLSGCLDARAAEKAARHVALCNAYNIPIVFVVDVPGFLPG FEEEKNGVIHRGAKFLAAIVGADVPKVTVVLRKAYGGGYAVMGSKNLGTDINFAWPTARIA VMGAEGAVDLLKRREIEAAGPEEGAKIRQQFIDMYNTFMATPYIAAERGYIDAVIEPADTRL MLRKAFAQLRDKELLTPHRKHEVPLL
Other gene	ctg1_764	852867	853802	311	MWGKTASPRPKKKRTIIIGVLALVVVLVVIAALLWNVVGNRPGIEKEAVNPPGCPQVEVLAV PGTYESAADDDPIHPHFRRNAMLLNVTNPLQQRYSAAQVRVYTVPYVAQFRNPGAIQEAS YNDSRDQGQSRLVAEMNRMHKKCPATQFLLTGFSQGAVIAGDVASEIGNDRGPVPAKNVL GVALLADGRRVNDQGINPGEPLQGKGLEAALGLVGTGLGFLSGSDATMKGRRDDGFGSL NDRTYQLCSVKDIICNATMNPLMLVRKGTNFFSQNPVHGEYNTNPNVVRGMTATQWIVQW MEKLINKALA
Other gene	ctg1_765	853851	854333	160	MRENNRRPLLRSGALTAALLLGLTGITACGNSEVSGTGDALLTDNSRQTSSSSSPSSSTNP ARPSELPEDYPEPSTIMRTAQDKEYLQDLKSQGIVVDGVEDSLIGTGKSLCQGKAETGRLN PVLARAVAGQLAQQGKASQDTQQTTDILTRTAIRHYCR

Additional biosynthetic gene	ctg1_766	85434	85610	586	MLRPLRSVRTTSALIAASLLIATPAVVIPTLGATSHVGPMAGIFAQAAGEPTITKTKWLSPRH VALWIYSPAMQCVRQVQLLLPRSFKANPNLSYPALYLLDGMRASTSYSGWTKYTNIVKTVE STELMVVLPIGGAASFFTNWQKSPEAKNTMQWEKFLATELPNVLRDHWRVNNSAGVAGL SMGGTAAVNLAERHPSLYRFVGSYSGYLDTSSDGMGEAINQAIQEVKPKYSATQMWGTY QSANWQAHDPKLHVQRLAGKSIYVSAGSGNTGPYDKPSQVDGIPQDHAAYALEILAQLTS KTFVSSAQEAGVPVTAKFRPSGTHSWPYWQFEFSQSLPQILKALGRPTPGKVAGNLWYS DSLASFAKSADATVNPKPKQKRKQSKAGSRVLPPSAEVTAELNKRDPKKPIPYKVPQNDPK CQTKGDIRKYLDGHPNSAKAAGHCLTGEYPVPGGRAQNFEHGRVFWSPETGAVLVKGKL NEAYQKMGSSGSVLGLPLDEEHPTHDKRGFYQDFQGGRLYWSFQTGAHWVRGTILQKYK ELGYSSGKLGWPTSDEQRTKTGAVSQFEHGRIEWTAKNNTATYHKK
Additional biosynthetic gene	ctg1_767	85634	85737	341	MTVNKRNKWVASTGATLLAISLVAGMGVDLPTANADNRASLRRDNSANRRCEWDPVKHW VQQCEVWSKAMHRKIRVQIQPAKYGGNHALFLLDGMRARNDWNGWSRDGRAPRIFVNS NVTLVMPVGGETSFYTNWKRKSSTNGQKYTYKWETFLTKELPGYLNRRFGVAYSGNAIGG LSMSGSAAFTLAIYHKPYFKQALSFSGYMNVSDPIMQQALQYAANDAGNYNLQDMWGPY SDPTWRRNDPYVNAAKLRGLKLFMSSATGLPGRYDNPRNGQEAFNTSVGFLLEGVARQQ HILMKNQLEQLGIPCRHIFMANGIHNWGYWRDQLRMAYPYVKAVIG
Other gene	ctg1_768	85798	85808	33	MRWGVFYVALAMGVVLFAVVAVAEEKRIRTLAK
Additional biosynthetic gene	ctg1_769	85830	85933	342	MADRANPSPHAADANSDTADDVMFEETTSAAEMEKEGRKLRPPKNLAEAMWKALRPRQ WVKNVLVLMVPISAGTAVVTDPQVLLQVLYAFITFCLASSSIYLINDAQDVEADRAHPVKRF RPIASGVLPLGVAYAMGIVLMLAAVLLGFFLSGGPLAIVIAVYIVLQLGYCFGLKHQMVLDIVL VSSGFLLRAVAGGVAADVPLSQWFLLVMAFGSIFMASGKRYAEKMLAEQEGRQIRKVLHS YTATYLRFVWTMSATALVLCYSLWAFQQGQMPSAHPGASLWYEISMVPWAIAVMRYAVD VDRGDAGAPEDIALKDHVLQVIALLWLVCIVLAVYVFGG
Core biosynthetic gene	ctg1_770	85940	86434	1646	VKSDLTVVELIAWMREWIANDVSITVSEVNPDQPFEEFGLSSRSILELAGQLENLTGKSINAA VVYQNPTVNKLAVFLLDDSDPAAETHQKVRDRSTVEGADIAIIGVATRFPGDANTPEQYWT LLHDGVDAVTDLPADRYQEFLEDKEVDARVTAAPTRGGFITSEYIRYFDPEFFSISPREAEQ VDPQQRMLLELTYEVFEDAHLPISEQRGHKVGVFIGASNQDYARILESDYSAFHPYSLTGLS PSILANRVSYTFDFRGPSISMDTACSSSLVAVHQAVQALRRGESSLAVAGGINLILAPGTQL AFADLETVLSPDGYIKAFSNDADGISRSDGAGLVLLKRLQDAEADGDHIYAVIKGSAVNQDG RSNGITAPNPDAQMEVLVDAYHDANIVPQDVDYVEAHGTGTILGDPIEALALGKVLGYGRD EEKPLLLGSCKTNFGHMESAAGSASLIKLALAMEKGVIPPMLHFAGPNPYINFEGDHLEPVT ENREWPRYSGKAVAGVSGFGFGGTNAHIVVEEYTGPASPAAVESDESAAADAADTDAPA AADLPALSDAERAQLEAENAMFEELEEPRTYVLPVSGTLATRARQTAKHVADWLEDNRDV DLAAVAHTLSTRSIGRFRKEVVSTDVDSAIAGLLAIAEGTESSDVYSGHHVALEGPVWVYS GFGGQHRKMGKRLYLADKFRGDFGGIFAASFDNVAALIKEESGHDIREIMLTDALNWDTET AQDGIFAIQVALTDTLTYFGCQPAAVIGHSMGEVAAAYAAGGLSLDDAVRVICVRSRLLGEG EATLSEDEQGGMALVEYSAEEIAQLVADNPGTFDTVEPAVYAAPTQITVGGKKPDVQAFVD YATENGKFARMLNVNGAGHTSMVAPLIGELIGEIADIEPLPLHCTLFSSIDKEAVYQVGDTPT TYKYFAKGMRHSVWFTQAVSQACQAGHTCFMEVSAHPVAILSVAAATYAANIPNAKLFYTL RRKEYEPNTLVRAIAQLAAAGHHANLRNLALTQEYADVPRTAFQRKYCWTSARLSGANDS HLPGAHVLLPTGQHVWKAQHKAVKQLEMLAQSAAEYCFGEEGVVTSWQLDGEDVPDAR YITMLNPQEGGATVEIFTQLHGESRPLASATITVGDPNAVSAIVIAKPDADEFGPEDPLAENN TDNEFYDPTTGETVVERLTKIVATNLAYNPEDLTSELPLMELGLDSLMAIRIKNRIQYEFQIPE

					LDLMVVRSSTIGDVAEYISNVRDAQSQGLSGEEASKVSAQKLAEKKADENAAGTPGDGAG ADVPPRDASERLAFSTWATVTGESAGDVMRPLPPISDDTRVKLAERLAERTGSDIDPVLLQ DVKNIEQLANIIRPFLEKEVEGLVRTLRDFPEGQQSKALFLFHPAGGTTAQYSALVDDLPADI PVFGLERVEGPLEERAAEYLPAIKEVQPEGPYTLGGWSLGGALAYEVAKQLIAAGDEVAAIL LIDTVQPSEPDPGTREELHARWDRYASFIQKTYGVNFQVPHDILDAQGEDGMMKLFAQMA ATADLSNTGVSAAVLEHQRASFVDNRMLANLDMETWASVQVPVTLYRAERMHDGAIELEP RYATIDPDGGWGKIVKDLTVVPLKGDHLSVIDEPEVGKIARNITERRESLGL
Additional biosynthetic gene	ctg1_771	864792	866705	637	MSFEVEPLFRGTTIDTSVLTTFPAIAEQRVAQILAGEAEDYPIYRFFDYLRDPNPTPDVLTWI SFTTRAKAIAARLQQVCERGDRVVVMMAQNLDYITGFFAPFFGGQISIPAFSPTEPAHAGYL EAILTDAEPKVILTDAKVAGSVRQFLKNVHVENKPRIIAIEGIEDEMAASWVDPELDPEDIAYL QYSSGSTRRPTAAKISHRAALVSVLQIMRTIGMKPRTRGVNWIPIFHSMSLIYLFAAPLSYTY IDIMEPAAFLQQPSRWVNALQSIDGEDVFSSGPDFAFGLAAARGKPEEGQSLDLSKVVALM NGSESVTQGTVDFFNATFKPYGLHDNAMRASYGMSETTSGAATPEPEVPTFIAKVDGYEL AEGKFVEVAPGEEDKGLSQVGVGGPLLNEYAVIVRTEFDENEEYLGRGSEVPDGTIGEIWI AGPNLASGYWNRPEETAETFDNELVETLDDDLTHTIRFGQKVDRDTKWLRTGDCGAYYN GQLFITGRVKDLVIVDGRNHVATDIEETVMMAGAGKFLPNTIAAFSVNASELLDSASHRTGR VIAPDTKGEQLVVVAEVNPDNEVEDFEETLNAVRTLVARRHGVQLADFRVVEQGAIPRTPT QKIKRLECRNAYESDSLNARF
Other gene	ctg1_772	866924	867967	347	MTNPPRPPPRPDQPGQQGRPTGPSRPSPSSRPAAAGRPGRAPAQRIGGTAGAAGRGDST VRRGDGPTQRPSRPVRPTASRGGAAGRAGQGTAAGAARGTGSNAGRGRGAAWGKIGG AGAAKRGGAGQRGAAQRMSGLRGQAQRGNVQRGSSQRSAARSANAQPTDARGGKRTG LAAFFGTDALANKRTYLIIGAIVLVFIIVIAAIASQRSSSTNAAATVTAITSSTVAGENRTSAPDA EAPAGDKAVTPDLSGLAMDPNSDPNAKAYFRAVKKGGIPVKDGMETALLGMGYTHCQAK MSNDKKTDEIQDVMMDTLVQMTPGVNKAKAKKVIYASSDKYLCPSLQKRQ
Other gene	ctg1_773	868000	868878	292	MANASSEPESHESAAHEVAAHHGGHDAPAADHTVPVDAHTGATAHHDGHGGVPHDLHY VDAHTGASQIHPELGSAAGYRKRSVFDVTWEEWKRAQWFGRGKVPADVPDAFSGASRY FSEEEHLLVNLQRWCRPYPWTVKAAQGLSHFGEHSLGWLALSGVCALSFPNQRKQWGW AALSAFGSHALSVIIKRIVRRHRPSNPAIDVNVTTPSELSFPSSHATSTTSWATSATMITHNP LPLVLSPIMMASRMLAGVHYPTDVVAGAALGAATSLATHKWGPVVLQRLIGQKNS
Other gene	ctg1_774	869003	871153	716	MPTASTSAQFSSFVGERLSRIITPRPGEPLDVRSLYLVEASTNPRRCHAPSRTSLRIPADSE VSFETYFNGLPASYWRRWSTLTSVVLRLHLTGSGRVDTYRSKIDGSRIGITGVSFCGPEGD GATAATTAPHTLDPQVTQCADGSSIVELTLGLEPFEDGGWLWFDITADTPVELLEGGWYAP VAAPTTVTYTDQQTGLTHEVDNANKVVIGIPTFNRPNDAVAALQALASDPDVDAIIEAVLMP DQGTKKVIDHPDFASAAAHFGIGGGVDGATDAAGQPLSVTPDGRFFRFEQGNLGGSGGY SRIMYEARQHTTAPFILYMDDDIAIEPDSILRAVAFARFAKQPMLVGGQMLNLQERSHLHSM GEVINRADFMWTSAPHVHYDHDFSQYPLSAARDHRGAQHTGLADPRWQETADLHRRIDV DYNGWWMCLIPRLVADTIGQPLPLFIKWDDAEFGLRAGAHGFPTVTMPGIAIWHMAWSDK DDAIDWQAYFHLRNRLVVAALTQPEGASINGMLASMAKAVVKHTLCLEYSTVAIQIEAIKDFL RGPDALFDLLPTALPKVAALRREFPDAVVLRSASDLPAPSGGPGGVTEEPRGIRRKVTALA KGLLHSVRPANPAHHSTPQMNLPPIDARWYSLSRLDGATVTTADGRGVVYRQRDLDKAK QLVKDSFAMHKQLREKWAEMQRQYRAAHPRLTSLEAWSEIFGPDTLG
Other gene	ctg1_775	871153	872355	400	MSTQSYDLIVVGSGFFGLTVAERAASQLGKRVLIIEKRSHIGGNAFSQPEPETGIEVHQYGA HLFHTSNERVWDYVTKFTTFTDYQHRVFAMHDGKAYQFPMGLGLICEFFGKYYSPDEART LIKEQASEVKTEEATNLEEKAISLIGRPLYEAFVRDYTAKQWQTDPKELPAANITRLPVRYTF

					NNRYFNDTYEGLPVHGYAAWLRAMINHDLIDVRLNTDWFDVCEDVRAANPDAPVVYTGPL DRYFDYSAGRLGWRTLDFDVEVLPIGDFQGTAVMNYNDADVPYTRIHEFRHFHPERADRY PKDKTVIMKEYSRFAEEDDEPYYPINTPEDREILAAYRELAAAETANQKVLFGGRLGTYQYL DMHMAIASALTMFDNQLAPYWTDGAALKGKGQR
Other gene	ctg1_776	87301 2	87425 9	415	MKRLTTRLLAAAGALTLTVGLINPEAAQAAATMPTPILDPFFNVPAAQYVHKPHGKLLRSRK LPAFLVPGGNATQMVFSTRDTFNRPTYATAVLVKPANFPKNGKVVVFNDFINSLGIGCQPS FSYSSLNPEWNTRSFIAMWYSAIAARYGYALLIPDHEGMKAAYTANILAGHIVLDAVRAMKN TPDFGMQRSWTGMLGYSGGSMVSLWAANLRAEYAPDVHFSAIAIGGTPTDLQYYGVRFG NRPNDAFGLAFASLVGLEREYPKSMRVTPRLSKHGKNKMRVLKNACSPRLLQSLKGESVN TIFNGVSLNPAQERSAFRVLRENSLYYDQRHPTRGTKLYVFGSRTDIGAPIRPLRATIRRYC ATGSRVQYIEVNDPNHVSTAFHNIPAAAQWLFQMSNGGKMRNDCRRIPR
Other gene	ctg1_777	87429 5	87558 1	428	MTTRTTRLKSITAATVTAAVVAGSVVAATPADATPLYDNTPLGSFFKVSQAKLNSVPNGTLI RVRPVSNFVVQGTQVFQYVFKTTNSHGKAVLASSAVMRPVIARPNANVLVYNDFINSLGVQ CQPSFSFNATADWLTTGGRTPNANRKAEIMARNGSLYGIGALAALWGIITIFPDFLGLQSAY GANLLGGHIMLDATKAARQIKALNIPNPKIILSGYSGGAMVAGYAASMAPKYAPNLNIVGLAA GGMPVDMVWMGKALGNNRNAGFGIAMGTMIGMEREYPKRMNVYDRLSKHGKRLVDANR DACTPRMLDALANESATTMFNNVKLREQHNEFAVFAENSLINYKPVPRVPVFIWSSSRDDL VPLRLIKKVTNRWCHTNRNLKIQILDTYVSNHVTNAAVGSWFAFPWVLSRFAGLPPHNTEC
Other gene	ctg1_778	87591 9	87720 2	427	MSALCAVALAAGTAVAVSPDASAAPIYPSKVKDPWFRVPASKYKNLPHGRLIKSRSMKLLQ FPSAVAEQLVFVSTNSHGEKIFATALLLRQPQSFAIPNVPLMVYNHFINSLGTECQPTFSLTT PINQFNNPESLTVAGEISLGLNAALSQGWNILLPDFEGMQGAYGANILAGHIILDATTALVTS RKFNSRHSKVILGGYSGGAMSTLIAASMAPKYAPAPRFVGAIAGGTPADMAWMGLALGNN TNPGFGIVAGTMLGLEREYPNRMNVFARMKPEWQDRLRKGLKDSCVSTMIGSFAGQSMR SVFNDVDIRKQKRELEVVRENSPLYYPLSPKMPLYLYHGNQDIAVPVAKVKKLARRYCSTG TKVTLATMDLSEHLLVGFMGLPSAAAWAQARFAGTPAPNEFCEGREYNILPLGQLQPPKN
Additional biosynthetic gene	ctg1_779	87745 9	87833 4	291	MPADAEFGVGVSPLALGGLGLVATDLDGTLLDSYKRVSPCTQAAIAQLDERHIPLVIATGRP LRWLPPVWEQLPLRPLCICMNGAVIYDSFHHQVLRRAVLKPTVLRDIVQSVHEALPEACFA VERITSEMTPSTGEGPVEEFLIAEDNEHIWYEQNAPRTSLAALAEEPAAKLLVRAVGMSSAE IAEVVCPLVEPVANTTYSVDGGLVEISHPVATKGEALVRILDYWDLSEETLLAFGDMPNDMD MLELAAWGVAVENAHPAVRAAADEVTMSNDEDGVAVVLERFLSA
Other gene	ctg1_780	87832 4	87979 9	491	LLVFRSFSHYSGVMAFEPLYSFFFRVGRILSSVLRHNIIVTGQENIPKKGGALLCINHTGYVD FYFSAMPLLRAKRYPRYMAKQEIFENSVAGPLMRRLGHIPVDRIAGRESMDTAVDQLKKGE LVAIFPEGTMSNSFEIKDMKTGAVRMAQGAGVKIYPTIVWGSHRVYRRDIKPNLERGIPIMIH FCEPMDVAGGDPEELTKQLRSIMQKNLEEVQQRYIEQFGPFEENLPWLPARLGGSAPTLE EAQAADDKIDEDRERANAMMRHSYKAAKEVHTVILHTSDATVDEAAPEAKLSRLTSMRAAA EKLNRAAEELVEAAKVPDGEVAKRKAGAGSLRNMTETVASAAVELERATVLSSRKASRRL GASSDSVTEFATGFRHRTAKSLHAGAEYLSQKAQSLLAYLSDAEEAARKAGEERQIRHAIA REARVEQAAADEVRKANKKMEAAIKHKEREASRALEDAARATEEAVHALEQTVALEDDAD AR
Additional biosynthetic gene	ctg1_781	88003 5	88176 8	577	LSAPSELTLELRESNWDRLGSEQFDILIIGGGSVGAGAALDAATRGLKVAMVEARDLACGS SSRSSKMFHGGLRYLQPNPIPQFGLVAESLHERELSMHTLAPHLVKPLKFIYPLTTPFFERI MMFCGFTLYDRMGGAKTVPMQKHFSRKKILKKIPSLKPNATCGGVQYYDTLVDDARHTMM VTRTAAKYGAVIRTSTQVVDFLKEGDRVVGVTIKDTEDGRTQNVKASAVINATGVWTDEMQ NLAGAKAKFHVHQSKGIHIVVPKDRIKSDSALCFVTEKSVLFVIPWGNYWIVGTTDTSWKLD

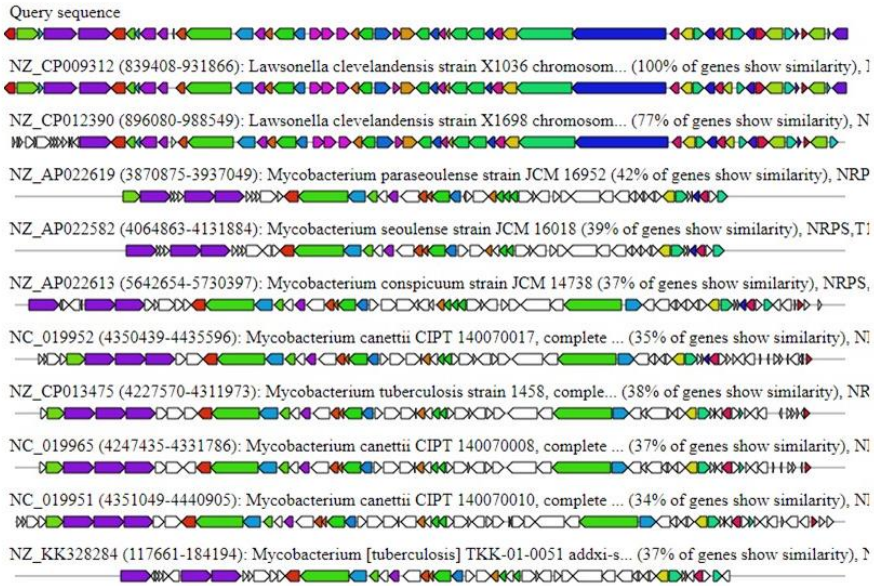
					LAHPAATKTDIDYVLEELNSRLQEPVTYADIVGVYAGLRPLVAQAGQENTEKLARNHEVARL CPGLVSVAGGKYTTYRVIGKDAVEAAIKDVPGQVAKSTTEETPIIGADGYWALTNQLDALAA KYGITEDQVEHLLNRYGSLIFDVLSYAKNDASLKKPITNAEGYLRVEAVYAAAVEGALHLED VITRRTRISIEYDDRGMDSAQEIADLIAPVLGWDDATKEKEVELYRDRVVAELNSQKAITDEE ADALRNQAEEARPAYVDDVDKK
Other gene	ctg1_782	882061	882834	257	MYMINGTELSGMAIFGWEFLGTMVLLLLGNGVCAANSLRTSAAKNSGWLLITFGWGLAVFA GASIAHPTGGHINPAVTLGLCIAGKTPWGAFIFYILGQLLGGMIGALLAWAAFKKQFDANAY NEDGTESGANATIGNVFHTRPAVHSMLWNHVTEIIATFVLMLFILVALWNNDAVNLGTLTYA AVASIIVSIGMSLGSPTGYSLNPARDLGPRFMYAFVLPIKGKGKGEWINAIHPTVGPLIGAAL AAGVYLAIA
Additional biosynthetic gene	ctg1_783	882893	884416	507	VTDKKYIAAIDEGTTSARCIIFDHDGEIVSVGQMEFEQIFPQKGWVEHNPIEIWDCVRRVIGE ALADADLAPSDLVALGITNQRETTIVWDKNTGEPVYNGIVWQDTRTAGICRQLAGDDGIYK YQKKTGLPLSSYSSGPKIKWILDNVDGAREKAENGDLLFGTPDTWLLWHLTGGANPEGDD ATVHATDVTNASRTLLMDITTMQWDEELCDAMGIPMSMLPEIRPSSGDFGKVRGRSNLAD VPITGILGDQQAATFGQLCFQPGEAKNTYGTGLFLLMNTGTEPKFSEHGLITTVCYQIGDDK PVYALEGSVAMGGSLVQWLRDNLKMIPNAASCDRLAETVADNGGVYFVPAFSGLLAPRW RDDARGVIVGLTRYANRGHIARSVLEATAYQTRELVDAMVADSGVKLTELKVDGGMVKSK PLMQFQADLLDVPVIRPQVAETTALGAAYAAGLHVGYWENVDELKANWNEDKRWDPDME SAERDRLFSEWNKAVERTYGWEE
Other gene	ctg1_784	884570	886000	476	MAIRPFKRVAVTAATALSMTALLAGTAQAVPVVQQPEFVTYIVKQDGTTVKTANPRAAYTQ PAQELSGPAKGVVMMAGKQTKPADPKLFKKLAGYNPFQVLPVEGDGPDADELDYSQGYS ADKLFTLYHSDIPQNISYTDVISGYDAFMAKPAVKKANDELVIATNNNASLVQIRRAIEDSDD ENLATMSDALGSNLGKIYRDLYNEGKLVKLQQLVGGNMSRAFNFSNATALEKVAYRNPRP FAVYPERIKYYYEDKDDAKTDDGYKGLIWSYPSGHTNRAYIKAIVLSAVFPEFAAQLLSRGS EVGYNRLVMGVHYPLDTVAGRMTASASMAWHFSDERFFPLIMEAARDTRSRVEKACGTTI MKCWAKGTPYLSEAESIRVYTVRMNYNLPRVGKKGLPMEIPSTASSLLRVKFPQLTKAQRD RILLLSAMDSGYPLDLGGKAGYWQRVNLAEAMSAHVKMLGNGAIQVSFLA
Other gene	ctg1_785	886117	886983	288	MRKDVVMDKNTAPDASQPTASRKTVVGTCLSTYFTLLGRAITTEYAKLRYLASTYIVVACAL IFPIIITCVVSFLQNIASALGVMEQPVESILYWTRVLGITVFMVWGINHVTADVKYGTFDTTYL AVPRWSVVLLAKWISVGSIAGLGTLLVQPLLLWSQKMIFPAITQNWEFSDHTTITFLWAQPV LAFLSVGVSVGLGALFRSAAPAISLLLVWEFVGEAVVSATPKSALAYPAMPFYNAALFVGQ EQQVPPLWSPGWSGVYWLVLAIVLVALGVWRLRARQRTI
Transport-related gene	ctg1_786	886976	887761	261	MIKLVHLSKRYTTTSKDKDNSQRVALQSINTEFKDGEVSYLLGPNGAGKSTLIRLIAGLSAPS EGAVYIDGLAPGSTANPLTSLGVMVDAQAFYPHHTARQHLQWVAWAGGISKKRVPEVLDL VGLDSVADRRISSFSLGMRQRLSIGTTLLGDPQNIVLDEPLNGLDVDGIIWARKLFRGFAAE GRCVVVSTHLLPEVARTADHVIVMGRGEVFADTTLHGLIDDVDPAVASDDAGRLEARFVEL TSAAVEFTSGGGVHA
Other gene	ctg1_787	887799	888266	155	MSLITRSRAATAAVAAATVLSLAVAPAASAAPFMVKGNNPSLADWTAQMKYLLNPKTPDSA LAINLESGKDGVPAAKFFRAMGNRNANWKWEFQGPTKTTGNISTAMFHQGAPGYPTVTK PAKWKKINGNWRISNETLCHYIGAYNIGKNTMFCN
Transport-related gene	ctg1_788	888313	890127	604	MIRELIDIGGTSSADKDFARPFHIFLALAILNGVCQGAALGTMVPVLMSVLQGDMHQAHIWL CVFAAATLVSIIVSVVANGRAFDSSMIIIESMHRRLGQKLINLPLGWFNADATGRASHLAVKG TMFVASAAMDVLVPYIGNIVSPLTLIIVSLIFDWRLGLCLLVGAPLIYGSAKLAMWLNNRADV HVHSAQINTDTRLLEFGRAQVSLRAAGLTGSDYQPLAEAIERQRKAGRGALWASVISMILQ

					NIVVQLLFAAVVSLAVYLAMGGTDPVLMLALIGLLTQFVKPLQVIAEVGTTLRTTDHELEDITE ILNLDPMPEPDPEAATALPAAPADYDIAVQDVHFRYRPDLPEILRGASFQLPAGSMTALVGA SGSGKTTITRLIARFWDVTSGSITIDGKDIRSYTTADLMSMLSLVFQDIYLFDDTLEANIAYGN PQASAADIRAAAERAGVTEIAQRLPDGWESRVGEGGRLLSGGERQRVSIARALLKQAPIVL FDEATASLDASMEKTVQESIAELAQTSTVLVIAHQLSTVQNADNIVVLDDGVVAEQGTHSQL IALGGRYSEFWRRRDAAHGWVLTTSSVTGFDASAERSASLTT
Transport-related gene	ctg1_789	890124	891968	614	MNEKTKAPTGSSLTRLLRPVSAHTAIALVFEVIGAAASVLPFIALTQIAHELVDAYQHGVAPN TDATWDGVFLLLIGIGIQLGFTSLALLITHFADVRLQAHLRLTLADKMGRLPLGWFDNNSSG RVRHLIANDVDALHQLVAHSVVETVSAVITPLVGLIYCFYLDWRLGLIALVPILAYFVTYSVLTI KCERDMMDQIAAGLAVISAAIVDYVNGVSVLKIFNQTGTGFSRFKEASEAFTKRFRELVNPA MRAQSIAVSFMLLPTSALITLGFGVWFVQEGWTTPINLLVVTIIAMTIPSTVYTVAISNTAMKD AVAAAGRITDLLDEPELPEPEPATAQVPTRKDVEFDHVTFSYREGKDVINDVSFTLPAGSIT ALVGPSGAGKSTLGTLLPRFRDVTGGAIRIGGVDLRNMTTDTLYHTVGIVLQDVQMLKISVR DNIRLACPAATDEQVIDAAKVARIHSRIMEFEDGYDTIVGVDALLSGGEAQRISIARTLLSNTP ILVLDEATSATDPESEAEIQQALSRLVAGRTVLVIAHRLSTIAQVDNIIVLDHGSIVESGTHDEL RAQGGMYAHMWDTLHSAESDTLLAQDTSTPHAATTMQHTTQEAQA
Additional biosynthetic gene	ctg1_790	892074	892889	271	MGPMAGALSELFTPAVGSVADSAAGVPTFVFPHAGGSPRFFHMWGSHWQWGPVWGVT YPGRDARLTEPMPAHVQALAAQVGEELAVHLRRHPLPTVRFFGHSLGAVVAYETAIWLLA HRLQWDPARTTVIQLVVSGHNAPHISYPPDYVQVHNKPTEKLVEEMVRVDPRNAEIFAIPEL AALLIPAIREDYRLAETYRSSQTSLFVDEIVVVAGEEDPDCFLEGLEGWKSYCRKWGGVHL QPGGHFYLAEYPESVPGLLARLTGGQGRGVH
Other gene	ctg1_791	892889	894025	378	MTTSPRPHRPRALVVGSTFGLHYSRGLAHPDSPVELAGIVARGSDRSRALADRFGVPFYP LPDLTEPVTSLSAAGLPEVELACVVVRSAVTGGPGDDITRFFLQSGTSVLQEHPVHPSSIVQ HLHTARAATPSQVVAATGVAHPVYAVNPFYPTLPTVQNLVRCVSTLRKQSRIMAVEVRGSI HIFYAILAILADMLPGLPRLILQRENIAQQGVNKQVYNLTLDWDGTPVDIRLYNRMAPLDPDN HSQPLFGCVVECEDGELIFDHMHGPVRWMPRSHYANGTFTHSADVSASMLVGDDEFTDA GAGPFREFTGRELLEELWPRGVMNRVGAVAAATCQQEVFGQRDLALTKLWQEVAAEIPR PVTIDSVPPRQIVW
Other gene	ctg1_792	894025	895521	498	MSPVIAWNGSTGTVGRTALRYFLRRWDDESPAGTAQTVSLELRLTGRSTERLEAVAAEVS AAYPQHKVRVCAGCDAEQDPADWPGAPLTGADLLVNTAGPGYVVTPRLARYCLTHNIAYA DPSGDHSMVAEIMATSDLPATYPHSGQASVPVGETPTAPAGQPADVPAVTVPLILGAGIQP GLSGILVRLLARAAGPGATVELYSGGAQPTTEAALHEFVDSVNSGMLWAGKEWVDGQLR ASCLPESYLAQVAATFSPTAVAHAHCDREVAAAAAAAQAHSVRWHNVSDAPKTTLEIQKMI AGDATVEQVMEAARMDDFGRRHYFRMEGTAVSAGLSDAPLREFHTPLGESMADTMNSS DPVNSSDPVNRADPSEMWHATLECVDSFLVTASVTAIAALYALRRACPSVGAAQPSLNSG VGGCPATDRNTAGAPDPVSPHGAGTPTTAATPVVPAGVCLPVDLPDTDYWWEDIMKDPA TVTTSLRQGDTALDDEEGEL
Core biosynthetic gene	ctg1_793	895521	901631	2036	VNIPELLDELAARRIELWTDNGALKFRAPAGAFTAALKQAVSANKPAIIEALQRGTIIRDDAN ATAPFPLTDVQASYLVGRTNAFEHGGVGCHGYTEFTVEGGYTPAQLTAAWDAVVAAHPM MKCEVHPDGWQRSNPALQVPLQVFDLREERDGSAVAAQRASIAATYRNKVYDITAGQPLI DAVVTVGPADTVLHVSVDLLLTDFLGLSVIFADMEQALRHPDTPPPAPSLTFRDYLQSEKLA MESPAGKRRRERDRQWWTDRLDSLPPAFALSAEARPEVLRSSTITLDGQEIPHYSRHSTS LSPAEWAAFTARAQQRGLTPSAVLMAALGRVLRRYTGLDAGLVTLTVLARQPFAPDVNRIV GDFTSTALLPLGNDRHQSFAANAQQVQVDLFESLDHTAVPGTEVARMIGRHTGQDNFFTPI

					VITSTVGAAADTESEVLIPRVGSGISQTPQVLMDVQLSPQGAGISIDWDTRDGAFSASVLDA AYSDYIALVRALLSDDAAVAEAAWEGDGLARRTIPAIARVASSQPTTGVLGTLHGPIVEQAL AHPDAVAVVDPALVDPGLADSVLVGSMAQEGDESRLAGASARGEVTYRQLLNLALAIGATL RPLQSGEPVALALPAGRDQVAAQLAVLMAGGAYVPLDTAWPEARRAGIVEQLATQAAAAG MDTPLVLDEAAVAAAYQQAVASLPAADSADVPAGAIAPSGLAAVTAQDLPADLLAQVAATA GDPSEVAYIIFTSGSTGTPKGVVIAHEQARTTIDAINDNLGVNAQDAVLAVSRYSFDLAVYNV FGLLGRGGRVVFPSCGTLADPLAWVRDIRLHGITIWNSVPAQLQLVTDLLAETREPLGLSAK QNPNALPLRAALLSGDWIPVTQPLDIQTDAPGLRFFAMGGATEGSIWSNLHEVMVGDEIEA SVPYGRGLRNQGMWILNADQEDCVPGQIGEITIGGDGVAVGYFGDSDRTAAAFFTEPDHG ERAYRTGDHGRYLPNGEIEFLGRVDGQVKIRGHRIELGEIDAALRGQEQVDQAFSLVHQDV HSGAYQVAAVVTPAHDPELAAQREAVFAQQCAEMTATADEFRATVEERALRSLSATLDMI ALRAMAGEVELALQTQSVSELQQDDRLDRWYRTLVERGWQPTGAATTPAPAASVSAAVT TPEPAITLADASPLGDPAVTTEELWRHVDELEERTHYGAGMLQFVRTCTDNIHRLCTGELD AKAVLFPEGNMDVARAVYAENLSSRYLNTVLVAGVKARVRQVVENPELRPLRIMEAGAGV GASTASIVEWLRSVDAFTPDENGVPTVEYIFSDLTTFFLDTARELWPELQYQIFDINRPHTE QGIADASIDLLIGVNVFHNCPNMVQLLTNLHGMLAPGGQLAATESTVPSPVLMSTVEFQEG LSSSAFTDVRGETKTAFMTLVQWQEAIANSPFHTVKHWPARGDIMELGNNHVLWMRRTD TLDTRHLTAEPLIDACRELLPSYMVPATIVITDSIPLTDNGKVDRKALAALCTSYVHQQHSGD VGVAAELNPTEEAIAAVWHDVLSLDPDMHLSPANDFFSLGGDSLLLARCVGRMRSEVPGG DAISWDDLLREVVADPTIGAVATVITTANAGDESATAEAATVGSHGAATDPAVAGSRSAAG TRAVSGSAGGAAVDSRAGTGTPAGAPPAVTPDGHGFLSVRDLPEPKGNCWSVVMEGTG NYAGTDPSEEAILFVHDGTGNLTPYETLFAQLAKQPNRPPVIGIQRQPNDSYLQVAPQQLF ALLAHRYTRNLLDTLRPTKLHVVGYCMGGLICATMVPELAMAGVEIAHYAAISAYRMPFMV TDPILLDFSLARILGIDPRDMGLCFDEASLGEALRTVRATYGRTLPEGCIYEAGDTSVKQAW DNAPATHTERIQLLSDKHPERGGVELIEQVSEVFHHSLQAVGAWEGLPYLGDVHFLRNTG DIDFLPTLGDDMESFWEEYCLGNLRVTSIPGTHFTCMEGENAPGVADLLAEMYDPQVAAR LAAERAAAADEAAPAAARTARATAEGDAQ
Core biosynthetic gene	ctg1_794	901628	911866	3412	VSNLGDNSAQLYASNPLVSPLWEKTQQTPNDIALRFITGTTPDDVDEWTWAEFLHGISTAQ ADLRAAGITAGDRVMMILPASADYLLTFWALITIGAQPVSVPVTPEAPLAAGIPTLRHIAANC GPAALLTTPAKQQELVTHQQSDTTLLPGLTVLTPHYQRPTHPHLPASLDDLAPHPVGPDTV VYVQYSSGSTGKPKGVLDTYRCILWQLDVMRITWNSDTPPVSISWLPLHHDMGIVWGIFDI LCLGGCAGYIPTALFAKNPTIWVNAMSTYKCGMTAAPDTMWRTICGLYQDPARRAGLDFS SLRLAVDGSEPVNTTTQPLLAATFGECGWTSTTLTPMYGLAEGGLAMSGHPGPREAVCVH FDAPTAARGQIVPADLNPNDPAQRTLYSVGDNFFGGDMRIVNPDTCQVLGDCQVGEVWF RGDVAAGYLGNPALTDATFHARTADDDGPYLRTGDYGFLCDGNLYISGRLSNTIIVGGENY FPSDIEELVATANLGFTTAQCCATQPDIDGPWYLVMGEAADIDTLDRTAYSGLLSRAIARSI GKQPATVLWTDNALPRTSSGKIARGEVAQLALQILAKQQGDSSTAHYANAGGRGTSGASN GSDDGAEGADSEDPDGEAGSRNSAAAAATDLTQHPVVVQLADQLSCPPDAVDLRADVSE LGLTSLAVNEFLSWAGHTGHTLDPVEVWENPTVGAWMDLYDLAAAATAQQGDAADGTQG GDAANTSAGTATTSITRPDGAANAAPDAAISTPTDLQKSYLVGADPSQPWGGIDCLAYAEF QPNLGTLTPADFYAALREAVTLVDQEPALHTALPTTETTCYADRPAPDVAVIDLTGVAADQL AAAREELRHVYLHRHFDTPQGENWALAVSTLPDGTTTVHVALSLIAADIVGVGEILARLDAA LQGDTERAFPPHAHHTATGRAAAERAGTHHPAGSITSDAAGTTPNPAAAPIPARGLPLPPD VPTLDTGSTHPNPQVTRFTHAFTTPQWQALSATARTLGTTPAGITLACYDAILRRWSSRDD FLVTLTTINAPQPGHVAERTIAVAHRAHCTLGGPWADTLVAVRTDLRAGIANPTPLGVELRH

					ALASPAGHTGLSPFIFTYSADRPLLAEPLSVLGAPTELRSTTPQVLIDCQVVNVSGHVTLSC DIRTDGIRADVAAQMFQALVDLITLQCAGHPQQLAQRPLIDLIPLPAATRTQREAANSSAPV TTSSLLYTPFRENVQKQPKAVAVVWAPEDQPADKTAGDPLGYNRRGYLTYQELDCMARR LAAKLIPLTRPNQVIGIQLPKGPCQVIATLGVLYAGCTYLPIGLHQPEERLRAIANSAGMEIVL RREDMAELISDELGGRPRPTCHQAIPRVVEANELAYIIYTSGSTGTPKGVAIRHYSALNTCL DVNARNQVTAKDNCLAISALEFDLSVYDIFGMLAAGGTIVCIGDESRRDAFRWADLVKRFR VTIWNTVPGLGEMLYAAALGSQLPSLRRFYFSGDWIAMDLPKRLATISPDATCISMGGATE GSIWSNEYVVTDVDFPADWTSIPYGGPLTGQKYRVVNPNVPGFPDQPDYVPGELWIGGA GVAAGYFGQPELTAERYPMDEQGEHWYRTGDLGEYHTDGLIFFVGRMDTQVKIRGHRVE CGEVEHACRALTGVETAIVVPIRKNSALGAVLVTGATGAKPGSEPGSADGAAAVAAEVQLA DPAHLRAQLATKLPDYMVPAVVMPRESLPHTPNGKIDRRLLATELESYQHEQATGHKGAA AATETPIAGATSAPTVGTAGATVSATAARADGSTIAAENTPGVVGVPQLDSFEQQVVNAWA AVLGEEPEVLATQLAAERAGTQPVNFFALGGDSLAATTVATSLVQAGCQVTVADLFAAPTL EAFIVRARGKLTATPTATQDAAPAAASAVAPAAPTTVAPTAVSEPAAPTATPLSGTSFPLTS LQQAYALGADGVRGITCTTPAVGVIISSGDGSALDPQRFATLIAELTAAWEPLRCVRSGDNE QAVVDAATVPAAAPVLTLRAPTGADLTTDSGVEAALAALRAELSGQRIPLDAVPMVQVVAV EGLTTHVGLSMNYLGLDARSIGVLFNTILDRYEGAADTFPVDPSAQPFYEYVQRELAAGTV GYSADHATGYSADARPQQLPPAPSLPVTTKKWTATSDARFSSLHTQLSVADTRALQTAAA HAGVTVSAVVLQAYGAVIAASSGLDVCGISVPISQRPSTNTAQQGSREVLGNFTELALCTL GTESSVTDVHRTLGHIGDGTLPNSRDIARAGRAAYPVVFTSTTGVAALHRSGTLTPSWVLT RTPGVQLDCQVSMVGDQLELRWDYPVEYVHAQRLHHLFTRFVDQVTAFSADPAPNELDT ELADLAQQLATSQPPSPQQIMAAALEHLDTHTEGIQHTPSRTPLPQYVPLVELWRNRVQDV DVAQAPLTADARRTGTWLADALCGGFDATDIIGHPVLSPQALLVHDPAAASIVTDIAHLLAQ KAEQLGRPVSVVELGSGTGLVRTLIERELTPTGAAGASVDVSWQSVERDAHWRKSATTHN EMPLEALSDIPADVVLAVGSLHRDPRLVRDVAQVPTLREAEVVVVETTELSDAALVSAAVID PTLLESAATVLQNAGQWWMNLTTAGWQPYAMSTPASYICVLQGKREEATAAELCPLHADT AAGSGLPEPELSPEELAAARTVLEAWHSVLPDAASLQVENPANLSADFFALGGDSLAATRV LQQLKQAGYQTLRLVDLFNTPELSELAVLLSKMERVTVDTDGDGAADDSHPATHPLTGVQ QAYLAGRSDAHILGGVASHCYYEFTTTSLDRTAFAQAITTVVDAHDELRALIVDGQATIQPA TPAHILTTVEDPRAATEAETPDPTQQVGMVVRLSPADVPGPVTISIGMDNLILDGASMMLVL QEIDATYRELAAGQPASLPQESLSFGAYLATHQEALDPADITDPAAAHRVAEARAYWSSEV ATLPAAPQIAERAQVVAIREPKIDRVSADVPAELWAAAQQAAKQSRVTAASLVLAAYAVELG QWSGSSDFTVNVTLFDRDMSVPGVNRVVGDFTSLTPVACRVLPGDSLAKVARAIQQTLAS VRDYDAAGALWVQRELLQLTGDPYASMLPVVFTCGLGLVADGVSTTDFSFGTLGRVRSQT PQTLLDLQVHEDVKGLHLTADYVTQALDAQRVQDAISAVAQRICDCAERVHGAQAAADGV GDGAAAAGGVAGAAASITASADQLATSPAANPATVLADQPTADTTPVTAASAADLTGPAAD LYAQIAPLWEAALEGVAVTLDSNFFKEGGDSLRATMVTRQVQDTTGREVDLRILLTNPTLR DYLQAVSAMPAATATEVGELDDYPNSPGSPDRAGAGVVVGTLPHGGDQADSSLADADLE EGTL

smCOG; bacterial specialised metabolite Clusters of Orthologous Genes -; not predicted by antiSMASH



Appendix 109: NRP/type 1 PKS LcA biosynthetic gene cluster and its predicted similar gene clusters according to antiSMASH.

Appendix 110: Summary of genes in BGC possibly coding for NRPS/PKS LcA as predicted by PRISM.

Open read frame	Start	End	Amino acid Sequence
orf_770	859408	864349	MKSDLTVVELIALDDSDPAAETHQKVRDRSTVEGADIGPASPAAVESDESAAADAADTDAPAA ADLPALSDAERAQLEAENAMFEELEEPRTYVLPVSGTLATRARQTAKHVADWLEDNRDVDLA AVAHTLSTRSIGRFRKEVVSTDVDSAIAGLLAIAEGTESSDVYSGHHVALEGPVWVEYEPNTLV RAIAQLAAAGHHANLRNLALTQEYADVPRTAFQRKYCWTSARLSGANDSHLPGAHVLLPTGQ HVWKAQHKAVKQLEMLAQSAAEYCFGEEGVVTSWQLDGEDVPDARYITMLNPQEGGATVEIF TQLHGESRPLASATITVGDPNAVSAIVIAKPDADEFGPEDPLAENNTDNEFYDPTTGETVVESN VRDAQSQGLSGEEASKVSAQKLAEKKADENAAGTPGDGAGADVPPRDASERLAFSTWATVT GESAGDVMRPLPPISDDTRVKLAERLAERTGSDIDPVLLQDVKNIEQLNRMLANLDMETWASV QVPVTLYRAERMHDGAIELEPRYATIDPDGGWGKIVKDLTVVPLKGDHLSVIDEPEVGKIARNIT ERRESLGL
orf_771	864791	866705	MSFEVEPLFRGTTIDTSVLTTFPAIAEQRVAQILAGEAEDYPIYRFFDYDGYELAEGKFVEVAPG EEDKGLSQVGVGGPLLNEYAVIVRTEFDENEEYLGRGSEVPDGTIGEIWIAGPNLASGYWNRP EETAETFDNELVETLDDDLTHTIRFGQKVDRDTKWLRTGDCGAYYNGQLFITGRVKDLVIVDG RNHVATDIEETVMMAGAGKFLPNTIAAFSVNASELLDSASHRTGRVIAPDTKGEQLVVVAEVNP DNEVEDFEETLNAVRTLVARRHGVQLADFRVVEQGAIPRTPTQKIKRLECRNAYESDSLNARF

Appendix 111: Summary of genes in BGC possibly coding for NRP LcA as predicted by PRISM.

Open read frame	Start	End	Amino acid Sequence
orf_790	892073	892889	MGPMAGALSELFTPAVGSVADSAAGVPTFVFPHAGGSPRFFHMWGSHWQWGPVWGVTYPGRDARLTEPMPAHVQALAAQVGEELA VHLRRHPLPTVRFFGHSLGAVVAYETAIWLLAHRLQWDPARTTVIQLVVSGHNAPHISYPPDYVQVHNKPTEKLVEEMVRVDPRNAEIFA IPELAALLIPAIREDYRLAETYSSQTSLFVDEIVVVAGEEDPDCFLEGLEGWKSYCRKWGGVHLQPGGHFYLAEYPESVPGLLARLTGGQ GRGVH
orf_793	895520	901631	MNIPELLDELAARRIELWTDNGALKFRAPAGAFTAALKQAVSANKPAIIEALQRGTIIRDDANATAPFPLTDVQASYLVGRTNAFEHGGVG CHGYTEFTVEGGYTPAQLTAAWDAVVAAHPMMKCEVHPDGWQRSNPALQVPLQVFDLREERDGSAVAAQRASIAATYRNKVYDITAG QPLIDAVVTVGPADTVLHVSVDLLLTDFLGLSVIFADMEQALRHPDTPPPAPSLTFRDYLQSEKLAMESPAGKRRRERDRQWWTDRLDS LPPAFALSAEARPEVLRSSTITLDGQEIPHYSRHSTSLSPAEWAAFTARAQQRGLTPSAVLMAALGRVLRRYTGLDAGLVTLTVLARQPF APDVNRIVGDFTSTALLPLGNDRHQSFAANAQQVQVDLFESLDHTAVPGTEVARMIGRHTGQDNFFTPIVITSTVGAAADTESEVLIPRV GSGISQTPQVLMDVQLSPQGAGISIDWDTRDGAFSASVLDAAYSDYIALVRALLSDDAAVAEAAWEGDGLARRTIPAIARVASSQPTTGV LGTLHGPIVEQALAHPDAVAVVDPALVDPGLADSVLVGSMAQEGDESRLAGASARGEVTYRQLLNLALAIGATLRPLQSGEPVALALPA GRDQVAAQLAVLMAGGAYVPLDTAWPEARRAGIVEQLATQAAAAGMDTPLVLDEAAVAAAYQQAVASLPAADSADVPAGAIAPSGLAA VTAQDLPADLLAQVAATAGDPSEVAYIIFTSGSTGTPKGVVIAHEQARTTIDAINDNLGVNAQDAVLAVSRYSFDLAVYNVFGLLGRGGRV VFPSCGTLADPLAWVRDIRLHGITIWNSVPAQLQLVTDLLAETREPLGLSAKQNPNALPLRAALLSGDWIPVTQPLDIQTDAPGLRFFAMG GATEGSIWSNLHEVMVGDEIEASVPYGRGLRNQGMWILNADQEDCVPGQIGEITIGGDGVAVGYFGDSDRTAAAFFTEPDHGERAYRT GDHGRYLPNGEIEFLGRVDGQVKIRGHRIELGEIDAALRGQEQVDQAFSLVHQDVHSGAYQVAAVVTPAHDPELAAQREAVFAQQCAE MTATADEFRATVEERALRSLSATLDMIALRAMAGEVELALQTQSVSELQQDDRLDRWYRTLVERGWQPTGAATTPAPAASVSAAVTTP EPAITLADASPLGDPAVTTEELWRHVDELEERTHYGAGMLQFVRTCTDNIHRLCTGELDAKAVLFPEGNMDVARAVYAENLSSRYLNTV LVAGVKARVRQVVENPELRPLRIMEAGAGVGASTASIVEWLRSVDAFTPDENGVPTVEYIFSDLTTFFLDTARELWPELQYQIFDINRPHT EQGIADASIDLLIGVNVFHNCPNMVQLLTNLHGMLAPGGQLAATESTVPSPVLMSTVEFQEGLSSSAFTDVRGETKTAFMTLVQWQEAIA NSPFHTVKHWPARGDIMELGNNHVLWMRRTDTLDTRHLTAEPLIDACRELLPSYMVPATIVITDSIPLTDNGKVDRKALAALCTSYVHQQ HSGDVGVAAELNPTEEAIAAVWHDVLSLDPDMHLSPANDFFSLGGDSLLLARCVGRMRSEVPGGDAISWDDLLREVVADPTIGAVATVI TTANAGDESATAEAATVGSHGAATDPAVAGSRSAAGTRAVSGSAGGAAVDSRAGTGTPAGAPPAVTPDGHGFLSVRDLPEPKGNCW SVVMEGTGNYAGTDPSEEAILFVHDGTGNLTPYETLFAQLAKQPNRPPVIGIQRQPNDSYLQVAPQQLFALLAHRYTRNLLDTLRPTKLH VVGYCMGGLICATMVPELAMAGVEIAHYAAISAYRMPFMVTDPILLDFSLARILGIDPRDMGLCFDEASLGEALRTVRATYGRTLPEGCIY EAGDTSVKQAWDNAPATHTERIQLLSDKHPERGGVELIEQVSEVFHHSLQAVGWEGLPYLGDVHFLRNTGDIDFLPTLGDDMESFWEE YCLGNLRVTSIPGTHFTCMEGENAPGVADLLAEMYDPQVAARLAAERAAAADEAAPAAARTARATAEGDAQ Domain Analysis
orf_794	901627	911866	MSNLGDNSAQLYASNPLVSPLWEKTQQTPNDIALRFITGTTPDDVDEWTWAEFLHGISTAQADLRAAGITAGDRVMMILPASADYLLTF WALITIGAQPVSVPVTPEAPLAAGIPTLRHIAANCGPAALLTTPAKQQELVTHQQSDTTLLPGLTVLTPHYQRPTHPHLPASLDDLAPHPV GPDTVVYVQYSSGSTGKPKGVLDTYRCILWQLDVMRITWNSDTPPVSISWLPLHHDMGIVWGIFDILCLGGCAGYIPTALFAKNPTIWVN AMSTYKCGMTAAPDTMWRTICGLYQDPARRAGLDFSSLRLAVDGSEPVNTTTQPLLAATFGECGWTSTTLTPMYGLAEGGLAMSGHP GPREAVCVHFDAPTAARGQIVPADLNPNDPAQRTLYSVGDNFFGGDMRIVNPDTCQVLGDCQVGEVWFRGDVAAGYLGNPALTDATF HARTADDDGPYLRTGDYGFLCDGNLYISGRLSNTIIVGGENYFPSDIEELVATANLGFTTAQCCATQPDIDGPWYLVMGEAADIDTLDRT AYSGLLSRAIARSIGKQPATVLWTDNALPRTSSGKIARGEVAQLALQILAKQQGDSSTAHYANAGGRGTSGASNGSDDGAEGADSEDP DGEAGSRNSAAAAATDLTQHPVVVQLADQLSCPPDAVDLRADVSELGLTSLAVNEFLSWAGHTGHTLDPVEVWENPTVGAWMDLYDL AAAATAQQGDAADGTQGGDAANTSAGTATTSITRPDGAANAAPDAAISTPTDLQKSYLVGADPSQPWGGIDCLAYAEFQPNLGTLTPA DFYAALREAVTLVDQEPALHTALPTTETTCYADRPAPDVAVIDLTGVAADQLAAAREELRHVYLHRHFDTPQGENWALAVSTLPDGTTTV HVALSLIAADIVGVGEILARLDAALQGDTERAFPPHAHHTATGRAAAERAGTHHPAGSITSDAAGTTPNPAAAPIPARGLPLPPDVPTLDT

			GSTHPNPQVTRFTHAFTTPQWQALSATARTLGTTPAGITLACYDAILRRWSSRDDFLVTLTTINAPQPGHVAERTIAVAHRAHCTLGGPW ADTLVAVRTDLRAGIANPTPLGVELRHALASPAGHTGLSPFIFTYSADRPLLAEPLSVLGAPTELRSTTPQVLIDCQVVNVSGHVTLSCDIR TDGIRADVAAQMFQALVDLITLQCAGHPQQLAQRPLIDLIPLPAATRTQREAANSSAPVTTSSLLYTPFRENVQKQPKAVAVVWAPEDQP ADKTAGDPLGYNRRGYLTYQELDCMARRLAAKLIPLTRPNQVIGIQLPKGPCQVIATLGVLYAGCTYLPIGLHQPEERLRAIANSAGMEIVL RREDMAELISDELGGRPRPTCHQAIPRVVEANELAYIIYTSGSTGTPKGVAIRHYSALNTCLDVNARNQVTAKDNCLAISALEFDLSVYDIF GMLAAGGTIVCIGDESRRDAFRWADLVKRFRVTIWNTVPGLGEMLYAAALGSQLPSLRRFYFSGDWIAMDLPKRLATISPDATCISMGG ATEGSIWSNEYVVTDVDFPADWTSIPYGGPLTGQKYRVVNPNVPGFPDQPDYVPGELWIGGAGVAAGYFGQPELTAERYPMDEQGEH WYRTGDLGEYHTDGLIFFVGRMDTQVKIRGHRVECGEVEHACRALTGVETAIVVPIRKNSALGAVLVTGATGAKPGSEPGSADGAAAVA AEVQLADPAHLRAQLATKLPDYMVPAVVMPRESLPHTPNGKIDRRLLATELESYQHEQATGHKGAAAATETPIAGATSAPTVGTAGATV SATAARADGSTIAAENTPGVVGVPQLDSFEQQVVNAWAAVLGEEPEVLATQLAAERAGTQPVNFFALGGDSLAATTVATSLVQAGCQV TVADLFAAPTLEAFIVRARGKLTATPTATQDAAPAAASAVAPAAPTTVAPTAVSEPAAPTATPLSGTSFPLTSLQQAYALGADGVRGITCT TPAVGVIISSGDGSALDPQRFATLIAELTAAWEPLRCVRSGDNEQAVVDAATVPAAAPVLTLRAPTGADLTTDSGVEAALAALRAELSGQ RIPLDAVPMVQVVAVEGLTTHVGLSMNYLGLDARSIGVLFNTILDRYEGAADTFPVDPSAQPFYEYVQRELAAGTVGYSADHATGYSAD ARPQQLPPAPSLPVTTKKWTATSDARFSSLHTQLSVADTRALQTAAAHAGVTVSAVVLQAYGAVIAASSGLDVCGISVPISQRPSTNTAQ QGSREVLGNFTELALCTLGTESSVTDVHRTLGHIGDGTLPNSRDIARAGRAAYPVVFTSTTGVAALHRSGTLTPSWVLTRTPGVQLDCQ VSMVGDQLELRWDYPVEYVHAQRLHHLFTRFVDQVTAFSADPAPNELDTELADLAQQLATSQPPSPQQIMAAALEHLDTHTEGIQHTPS RTPLPQYVPLVELWRNRVQDVDVAQAPLTADARRTGTWLADALCGGFDATDIIGHPVLSPQALLVHDPAAASIVTDIAHLLAQKAEQLGR PVSVVELGSGTGLVRTLIERELTPTGAAGASVDVSWQSVERDAHWRKSATTHNEMPLEALSDIPADVVLAVGSLHRDPRLVRDVAQVPT LREAEVVVVETTELSDAALVSAAVIDPTLLESAATVLQNAGQWWMNLTTAGWQPYAMSTPASYICVLQGKREEATAAELCPLHADTAAG SGLPEPELSPEELAAARTVLEAWHSVLPDAASLQVENPANLSADFFALGGDSLAATRVLQQLKQAGYQTLRLVDLFNTPELSELAVLLSK MERVTVDTDGDGAADDSHPATHPLTGVQQAYLAGRSDAHILGGVASHCYYEFTTTSLDRTAFAQAITTVVDAHDELRALIVDGQATIQPA TPAHILTTVEDPRAATEAETPDPTQQVGMVVRLSPADVPGPVTISIGMDNLILDGASMMLVLQEIDATYRELAAGQPASLPQESLSFGAYL ATHQEALDPADITDPAAAHRVAEARAYWSSEVATLPAAPQIAERAQVVAIREPKIDRVSADVPAELWAAAQQAAKQSRVTAASLVLAAYA VELGQWSGSSDFTVNVTLFDRDMSVPGVNRVVGDFTSLTPVACRVLPGDSLAKVARAIQQTLASVRDYDAAGALWVQRELLQLTGDPY ASMLPVVFTCGLGLVADGVSTTDFSFGTLGRVRSQTPQTLLDLQVHEDVKGLHLTADYVTQALDAQRVQDAISAVAQRICDCAERVHGA QAAADGVGDGAAAAGGVAGAAASITASADQLATSPAANPATVLADQPTADTTPVTAASAADLTGPAADLYAQIAPLWEAALEGVAVTLD SNFFKEGGDSLRATMVTRQVQDTTGREVDLRILLTNPTLRDYLQASAMPAATATEVGELDDYPNSPGSPDRAGAGVVVGTLPHGGDQA DSSLADADLEEGTL

Appendix 112: Genes predicted to be involved in the production of bacteriocin DpA.

***D. pigrum* strain**	**Amino acid sequence**
D. pigrum (accession number MUYF01000003.1)	MSDYRVLLYYQYVDIEDPERFRKEHKALCDQLELKGRILVSDEGLNGTLSGTVENTQKYMDAMHADE RFAEMPFKIDEADEHAFKKMHVRTRPEIVSLNLGEEDVDPNQTTGQHLEPTEFRDALLDEDTIVLDAR NDYEYDLGHFRGAIRPDIRNFRELPDWIRENKEQFMEKKMVVYCTGGIRCEKLSGWLLKEGFEDVAQ LKGGIHNYGNDEETQGELWDGKMYVFDERISVDINRKEKTIIGRDWFDGEPCERYVNCANPYCNKQIL MSEENEHKYLRGCTHECRVHPENRYVKEHNLSTEEVQERLEAIGESLPTFA
D. pigrum ATCC 51524	MSDYRVLLYYQYVDIEDPERFRKEHKALCDQLELKGRILVSDEGLNGTLSGTVENTQKYMDAMHADE RFAEMPFKIDEADEHAFKKMHVRTRPEIVSLNLGEEDVDPNQTTGQHLEPTEFRDALLDEDTIVLDAR NDYEYDLGHFRGAIRPDIRNFRELPDWIRENKEQFMEKKMVVYCTGGIRCEKLSGWLLKEGFEDVAQ LKGGIHNYGTDEETQGELWDGKMYVFDERISVDINRKEKTIIGRDWFDGEPCERYVNCANPYCNKQIL MSEENEHKYLRGCTHECRVHPENRYVKEHNLSTEEVQERLEAIGESLPTFA

Appendix 113: Genes predicted to be involved in the production of bacteriocin DpB.

***D. pigrum* strain**	**Amino acid sequence**
D. pigrum (accession number MUYF01000003.1)	MHSIIICEDNQEQLAYLSMLVQNYIQFHENQFRLVYEDTNPQTTLNYIQQEAIKKGLYLIDINLGADMTG VDLAEAIRDSDIQSKIIFITSEEDQATNILNRHIEPLAYISKNEGLETMQSNLHRALDDAYSRFSEITVRTK DVFSFNFDTLTYQFDLDEVISIEVSGDHRLTLKTITGQYDFFNTLAQIEQDYPTLLRIGRSDMINPVNIKHI DYKRRNVTMVNDEQFTIAASRIIQLRQLYKS
D. pigrum ATCC 51524	MHSIIICEDNQEQLAYLSMLVQNYIQFHENQFRLVYEGTDPQTTLNYIQQEAIKKGLYLIDINLGADMTG VDLAEAIRDSDIQSKIIFITSEEDQATNILNRHIEPLAYITKNERLETMQSNLHRALDDAYSRFSETTVQTK DVFSFNFDTLTYQFDLDEVISIEVSGDHRLTLKTITGQYDFFNTLAQIEQDYPTLLRIGRSDIINPVNIKHI DYKRRNVTMVNDEQFTIAASRIIQLRQLYKS

Appendix 114: Genes predicted to be involved in the production of bacteriocin DpC.

***D. pigrum* strain**	**Amino acid sequence**
D. pigrum (accession number MUYF01000003.1)	MNKHEYNKMIRGELYDPRDPYLRDLMTKQHALMDAFNESGRTNHATQNKLLEKLLGQYHGESSHIR RPFFIDYGLNITIGKHFFANYNCTMLDVAPITIGKNVMFGPNVSLNTPEHPLSAAERNMGLEFARPITIE DNIWIGASVTVVGGVTIGQGAVIAAGAVVTKDVPANTVVGGVPARVLKTIENC
D. pigrum ATCC 51524	MNKHEYNKMIRGELYDPRDPYLRDLMTKQHALMDAFNESDRTDHATQNKLLEELLGQYHGESSHIRR PFFIDYGLNITIGKHFFANYNCTMLDVAPITIGKNVMFGPNVSLNTPEHPLSAAERNMGLEFARPITIED NVWIGASVTVIGGVTIGQGAVIAAGAVVTKDVPANTVVGGVPARVLKIIEN

Appendix 115: Summary of genes in BGC possibly coding for bacteriocin DpD in the whole genome shotgun sequence of *D. pigrum* ATCC 51524 as predicted by antiSMASH.

Type of gene	Locus tag	Location of gene	Amino acid sequence
additional biosynthetic gene	Ctg1_538	567403 - 569058	MGKNFHDKVVYQVYPKSWKDTTGTGMGDLQGVIEKLPYLADLGIDLLWLNPFYKSPQ NDNGYDIADYKSIDPMYGDFDTFDKLVQEASKYNIGLMLDMVLNHVSTEHEWFQRAL NGEEQYQDYFILRQAQPDGSLPTNWESKFGGPAWNQLGDSDYYYLCLYDKTQADLN WHNPEVREALYDVVNFWIDKGVKGFRFDVLNVIGKSQDLKDAADGVGKKEYTDTPIV HEWVRELNKRTFGPYDDIITVGEMSSTNVSNSVQYSHSESNELDMVFSFHHLKVDYK EGSKWTLMDFDFDELKRLLDHWQTGMQAGNGWNALFWNNHDQPRSNSRFADPVN YPYESATMLATTIHLLRGTPYIYQGEEIGMTNPNYESIDQYDDVETHNAYRTLLAQGM NNSEALAIIQVKSRDNSRTPMQWTSEDQAGFTTGQPWLEVASNYKYINTNVTDPSQH IRDYYKQLIRLRKDYAVISEGTYRSIHLDHSNVYSYIREYNNQQLLVVNNFYGKQTAIDV PEDFLKKDARVLIGNYDQHALEQQLILQPYESIAFLLE
Other gene	Ctg1_539	569104 - 569664	MRTLTTKDLVYISVLTAMLCVASLVAIPVGIGIPITLQVFFWLLIPALLKAYRGFLSLALYV LIGLIGIPVFAGGTGGFQAVLSPSFGFLLGSLIIALYIGKVAQKRPSLLTMILHMIVAILILY TIGILYQYFIFNVIAESGGSTTLISLIIANVSSFLPLDILKAILAGIVYDRLIRHTRFKNI
Other gene	Ctg1_540	569836 - 570480	MTPNKENFLKAIYELGGMSKLINNKSLAEYLNVSAAAITDMNTRLVKQSIITYEPYKGVK LTDKGVRIVNQLIRRHRLWEVFLAEKLGYEWDEVHTDADLLEHISSDKLIERLDAFLGH PTVDPHGDTIPTSDGEVIVNQYHALVECKQGESFKVKQVDDDTEFLTYLTDKGIQLNE TYQITEIEPYEGPITLTNNDEENILVSYKAAFRIFGQ
Other gene	Ctg1_541	570755 - 572056	MKLKNILKTASILALLTGPSATMTVDHVYAEEAQDKVADSETNLEDLRQQVRSSVQTIE DLNNELVQIDETIDGIEAEIAQTEEDISQQEAIIQDYFDQAKSRLQNMQLSNVNENAVLS LLEAESFQDIIARVNAVVQLTQANSDQIQSLQAQKDELDQLVQSLETKQTDLSNKKETI DAKKQSVHTEIADLERLIEENREDFETLKEESEIVEICRRAEELVEERQAEQEVQEDAS SNEEAKQVASQSAEETQQPQAEETVAIEENEEPAEPTDDSAAEQSVEEPQVTETEPA EAEQSQPEPETQKQPDQQVQEESQTQTEPEVQEEPIQPQAAVPTTAQYSIDDLEFQ GVINALGKKWTFYSERVLPGGGLNIPGRHTADGFVRDGEGYIVLAASSSVGHGTIIDT PFGSQGKVYDTCASCHAGWFDVYTR
additional biosynthetic gene	allorf_0572098_0572175	572098 - 572175	VRVAMTGSFSYRHICCGVFLINKIM
transport-related gene	Ctg1_542	572760 - 573677	MYYMKNVIEINNLTKSYDNVAVIKSININIVEGTIHALLGKNGAGKTTILKSILGLIKPTSG SIFILGENQLTEKGRNVLKQVGCMIETPGFYPNLTGTENLSIFAKLRELDKQAVQEALTI VNLPYADSKKFKEYSLGMKQRLAIANAIMHDPDILILDEPTNGLDPSGVIEIRELCKEMK RQGKTLLISSHILSEVEQVADEVSIIDEGELVKYINLKKMNENSQVTIHIFTSNFDKVKSV LLEAGVSQHNMINTAKGVRLLSKDMDISEVNKLLTQNDVDIEGIMREKLTLEEQFQRV TETR

Other gene	Ctg1_543	573682 - 574416	MLDLVSTEITKITRMKQMWVYASVLGIYITMMSVYTAQAVGLFDQFIHLYKFSLSYTGF LLLPFILLSLSSSTISDDYRNEVMKNLTVIPISKQKIIAAKLIAIYIVLSVAMVMIWISVCVIG YVTRFRFSMTWLLIFRYLWLCILTSLVIFLSMLPVTLISVATKGNVVITNLIGSMYIIASFF LTNFMNGIIPLATAPHIIWYGSMEGVEVNSNIALMVISIIFYTIIVLFAMNKILKKQEL
Other gene	Ctg1_544	574436 - 575152	MVLEFKKMKRSKLIIVLILIVVLNIGIHYLMGNVKYVGIPYAAEPGWTLQNTLIVGTYYLLL PVFTLIGSASFIIERENSVYINILTIPVRKSHLILNKSKFIWVVSMIFTSSIFVFTLVLEKIHS NILTQEIVLKYLFEYVVHSNGLYVIAMLILSVIIYMDYNMQLAVMVGFVSSFVSVFIEQTA ISYLYPVNALFNISGFKESNLYEYSSSVVILCVIAICAVFIYKKIAQQESW
additional biosynthetic gene	Ctg1_545	575201 - 575428	MKNKVNNWKTPVYEKEEDYDNPAGDIFRELKSDDIDKVMASGTANTYCRCYSGRHS CGRACTITAECPVFTVACC
Other gene	Ctg1_546	575515 - 576258	MKLIDKFKNGLYSFERDIQHDETNNYKSENRLQYWKKFLGVNEREIENILSNGLGINTA NLNELLSENDNFSCKVTETNVLWNQLIHDLQVLSIESIILPEFYIIGDIGQKELPMFYGFH EPFLKLAILRFENYWKNIPGISDNVFNKLLIYLYDQLAEISYRTLILELNIAREENKLAGET SEERYNYFSTQYLSDNYWLILEEYPVMFRLMCEATQKWINNTTRFIDRILSDKDDLEKL LKLREN
Core biosynthetic gene	Ctg1_547	576366 - 578741	MDVTFQSTIKWFNEVTKSHLYSLKIINHTEYGWVEYIPHEECKDYSDFKNYYTELGQLL FLFYLLRGNDIHYENIIAKGKHPVLIDLETLFHNNTSNTSGIDTAADRVNELLENSVRTV GILPNLVWAQNGKNGVDISAISTSENKEIPIEQASITNVNKDNMKVEYKTSTLASQKNN PYIIGEEISLTSYHKYLKKGFIESYTKIKNINKKEIINQVENYKEIYARQILRPTQYYTTLIQI SLHPDFLRSAIDREMLFSKLWIYFDENNSFRKVSEIEFTSLLKNDIPWLISNVSKNNITT KDGSEIESIFKHSSIALVKEKINILGDKDLTLQVELIETALNYDSEYNKAESQRENDRKIIE INDDKLNKNHLDQQLLEISTNIGDYLINQSFIGMNGDVSWIDMNVIGEKANDWNMVPT SMDLYSGLSGIMIYFIFLYKETKQNKYLIMVKRCYKSIINYIKNVRKRTNINSEVMFGGF SGETPIIYALTILEEELGGIFDLDELEKIRSWIFKECKKNISVGNEHDIIIGSSGVIAILLRYY DLTSNDAILEVCQQYAEQIIDNYIEMDNNSIAWIGIASRNALGGFAHGVSGIVWALSKLY SYLPDERYIEVIEKALRYEDYLYSEDDKNWVDRRETEEGIEYNNLSSNMPVAWCHGA SGILLSRASLKKHNLPLSEKRKNKIDEDIEIAVRTTLKNGFGHSHCLCHGDLGNMLILKY ASSELNTKNDIDKRYDIYMSHLISQLKDKWECGIPYKNSPGMMLGLSGIGFGLLSLMN EDLPFILLLE
transport-related gene	Ctg1_548	578813 - 580291	LKKIKFVQQLSESDCGIAALTMVLNYYGCKLHISDLVEKCSISRDGVHLRDLMNVANEY GLKSKAVELNKKENIFSEKIVFPCIAVLSMSHYVVIEKAKRGTIYYLDPECGRIIMDGSE FSKKFTGILLLFFPENIKKQKSQNKIIEILTLGEIKKKFIVGVIFFSFIIQLFVLLMPLFTEYMI DNVINGIALVSHVKLIIGILFAILSYGVFSFVRELLTTLLEIKYISTLKKNVVRKLFYLPLSFF DVRSSGDIVSRINNIDSIQQLLSNIITGVFVDVLTIVISISMMLYISRILSIIMIVYGILLCVLL NIFFKKTDVKNKLSLMKREKTQSYLIELSSNINMMKTSNYGEALYNKWVKNYNEQMHL EFKRQRLLGMYRSLIISYRLLPSVTVLFIGSNFVNQGMLTLGQVMSFLALGNTLLSPLAI LIQNIFDFQYSKNNIDRLSEIITAKSEKNFDGREINSFKNIKFQDVNFSYSGINGKKMLKS
Other gene	Ctg1_549	580366 - 580929	LILSLYNNFEGNILLNNIDKEVYDLKSYRQLFGTVLQDETLLNDTIQKNIDPTHSHTISKI REAAKLACLDEDVMNRMPSKYNTEIGDNGRNLSGGQRQRVAIARALLTNPKVIILDEG

			TSQLDVSTEREIFNNLKKKNITIIAVTHKLSTTTISDCVYVLKKGKIIAHGKHEKLMQNNR YYSDFFR
Other gene	Ctg1_550	581210 - 581617	MLYEVVAVNETGIDGQSYLVDGESYQTSSPTSDEPGTNPEQLMGLSLATCFNATLHA VLKEREIEPKSRVQVTVQLHQDEVSEEYYFTLDLTAAVADIPLDEAETIIQATHKRCPV AKIVGDYKHLTVETVPFEA
additional biosynthetic gene	Ctg1_551	581712 - 582740	MGIIWIIGYYWADLVIGITRVFSKMRQRLKRRRVKPAKQANVYKNVEYKSDYPNNKLDI FIPSHSSHPTRSPEENRIKEQQNTCDNENDMEIRPALYPVIVYFHGGGFAWGDKIDER RYLLEFVKAGYAVVAANYALTPMYKYPVPIIQASEVMDFLVNEGEQYGLDPDNIVLAG LSAGGHLAGQLALAESSSAYADRLGLSQHPDLTIRGIMLNSALIDPNRSNQVGHWLIN WLFSVMGFSYMDMNPRNMKEQMLAEANLLNHINEALPPVYISDGNRVSFTQQATDL VAQLRARDIYVQSNIYTDDYPGFMFHGYEGQLWLREARENVDKQLDFLATIIE
Other gene	Ctg1_552	582882 - 583046	MSNQVVNILLTLVLVYIFSAYIIFLVKSVLYGKEKVYGCKNPTLLQKIKVIFLM
Other gene	Ctg1_553	583361 - 584062	MDKSYIFAVSDVHGSDEQLEALLEYWNPEEEQLVILGDLCDRGPDAQAVFRRAKQLK EEYGAICLRGNHEDMLMKFLDDPADNVAHYYRNGGEITLESFLGDAVSSSTPQELAE ALKDQYPWLIDYLDNLELTYEWGKYFFVHAGMDLTLDDWRESTVRDKIWIRKGFLDV LNETEKTFVFGHTPTMKLRTDGQPDLWVSEDHKIGIDGGAVYGGQLNGVKLSQAGIE ATYQA
Other gene	Ctg1_554	584105 - 586282	VEDKQTFQLIVRELSDYSTKQVRTVLSLLDDGNTVPFIARYRKDQTGALDEVQIREIQQ RSEYIQNLQDRKETVLSAIEEQGELTDELAAEIREATQLQEVEDLYRPYKKKRQTLATK AKEAGLEPLADWLLGFPDASEAEIEAIASDYLNSEYDVETVADALAGAHEIIAEQVADN PEYRKRLRDYTIYNAQITSTVKNEEIDEKNVFQQYYEYAEDYRKIAPHRTLALNRGESE EVLTVKLEVGDERAHRYLYKQILPSLDSSSPVISIVKEAIKDAYKRFIAPSIERELRSTLT EKAETHAIQIFGENLKNLLMQAPLKQKTILGLDPAYRSGCKLAVIDSTGKVLAIKVIYPH TSGETKRNQSKEALKQLITEHAVDVIAIGNGTASRESEQFVAEVVSEMPEKTAFIIVNE AGASVYSASDEARREFPDLAVEERSAISIARRLQDPLAELVKIDPKSIGVGQYQHDVS QKELEGQLDFVVETAVNQVGVNLNTASGALLEHVSGLTKTAANNIVAYRESEGQFTT RDQIKQVKRLGPKSYQQAVGFLRIPDAPEPFDQTGIHPESYTEAEKILAHVKQDKSAIG SDELKHLLEQLPDIQTAEQLEIGLETYRDIVAALIAPGRDARDDMPAPILSTDVLKMDDL REGMQLQGTVRNVVDFGAFVDIGVKEDGLVHISELSDQFVEHPKDVVQVGDIVTVWI KSIDLNRSRIGLTMVQKD
Other gene	Ctg1_555	586307 - 586786	MKLIRTDDSLQAYVEAVSEEWFHHPFRHRATFNSRLRTTGGRYHLQTHHLDFNPKIL DVFGEAVFLGIVKHELCHYHLHIAGRGYQHRDADFRKLLKEVGGLRHTPSIEQKSGVA KRWHYQCSQCGQSYYRKRRVNTASYCCGRCHGKLKLFEQVICEQ
Other gene	Ctg1_556	586872 - 588131	MGLLLKNIKLERCYKEQNNGLVVTETETKDLLIKEGCFAEIADEIAENTPGVQKVIDGN GQLLVPSLRESHIHIDKTYFSGPWQAPTRPEDFSIYTRLNEEKELLPAQLDVAAERAH KVVQHYIQNGHTHIRTHCNIDPQIGLKHVELTLDVLKQYEDQITYEIVAFPQHGLLRNG EEFIQLFEQALSMGVTHIGGLDFATLERDVEASLQLLVKWAKQYDLGIDVHLHDRGAL GVYEIERLLDMMAMYDYQGEVTLSHAFALAQVPAQKKETLFRRLAERNVDISTTVVIS

			PGLTLPIFELDQYGVKVSVGHDSLTDHWSPFGTGDTIQKLNLLSEQFNLIDEKRLSYAL KFGTGGVTPLDSAGRQVWPKVGMPANALLVDAVSSAHLIARRCPISTVISRGQVISEQ SIQQKGAYRG

Appendix 116: Summary of genes in BGC possibly coding for bacteriocin DpD in the whole genome shotgun sequence of *D. pigrum* ATCC 51524 as predicted by BAGEL.

Name of gene	Function	Amino acid sequence length	Amino acid sequence
orf00001	Dihydrofolate synthase/folylpolyglutamate synthase	133	MEQLSEQPFVLVDGAHNEEGVMMLSKSLEQIAPNKKWTLFVGMMADRNYMMISDQMGHIAG KCYVLSPDSERGFNPEQVADHLTESSTPTTVISNVQAIKQYIDTEATDDEHIVIFGSLYLVGDIL TLYKKTL
orf00004	PTS system trehalose-specific EIIBC component	502	MSNFKRDAQSLLKLVGGKDNIQAVTHCATRMRFVLVDPSRADKNQIEELESVKGSFTQAGQF QVIIGNRVSDFFNEFQAVSGIEGTSKDSVKSAGKGNMNWLQQLMANLAEIFSPLIPAIIIGGLIL GFRNVLEGIQIQALGQAIADGAPAFTESGEPIYNTIVEVSPFWNGVNHFLWLPGEAIFHFLPVGI TWAVTRKMGTTQILGIVLGITLVSPQLLNAYGVQEILAGDVQQWDFGFFQLDMIGYQAQVIPA MLAGFTLAYLERFFRKITPEAISMIVVPFFALVPTIFIAHAVIGPIGWQLGKWLANGVQWGLSGS FNWLFGFIFGGLYAPFVVTGLHHMTNAIDLQLVAEYNQTILWPMIAFSNIAQGSAVLAIWWKNR HDEKESAVSLPAAISAYLGVTEPALFGINIKYVYPLVAGMIGSAFAGMYSVATSTAAYTIGVGGL PGILSATNSSYLNFGIAMAIAFFVPFVLVFIFEKYNVLTDQKLEGLPIPKLGS
orf00006	Trehalose-6-phosphate hydrolase	561	MELFRRKDDTMGKNFHDKVVYQVYPKSWKDTTGTGMGDLQGVIEKLPYLADLGIDLLWLNP FYKSPQNDNGYDIADYKSIDPMYGDFDTFDKLVQEASKYNIGLMLDMVLNHVSTEHEWFQRA LNGEEQYQDYFILRQAQPDGSLPTNWESKFGGPAWNQLGDSDYYYLCLYDKTQADLNWHN PEVREALYDVVNFWIDKGVKGFRFDVLNVIGKSQDLKDAADGVGKKEYTDTPIVHEWVRELN KRTFGPYDDIITVGEMSSTNVSNSVQYSHSESNELDMVFSFHHLKVDYKEGSKWTLMDFDFD ELKRLLDHWQTGMQAGNGWNALFWNNHDQPRSNSRFADPVNYPYESATMLATTIHLLRGT PYIYQGEEIGMTNPNYESIDQYDDVETHNAYRTLLAQGMNNSEALAIIQVKSRDNSRTPMQW TSEDQAGFTTGQPWLEVASNYKYINTNVTDPSQHIRDYYKQLIRLRKDYAVISEGTYRSIHLDH SNVYSYIREYNNQQLLVVNNFYGKQTAIDVPEDFLKKDARVLIGNYDQHALEQQLILQPYESIA FLLE
orf00007	Probable biotin transporter BioY	186	MRTLTTKDLVYISVLTAMLCVASLVAIPVGIGIPITLQVFFWLLIPALLKAYRGFLSLALYVLIGLIG IPVFAGGTGGFQAVLSPSFGFLLGSLIIALYIGKVAQKRPSLLTMILHMIVAILILYTIGILYQYFIFN VIAESGGSTTLISLIIANVSSFLPLDILKAILAGIVYDRLIRHTRFKNI
orf00008	Iron-dependent repressor IdeR	214	MTPNKENFLKAIYELGGMSKLINNKSLAEYLNVSAAAITDMNTRLVKQSIITYEPYKGVKLTDK GVRIVNQLIRRHRLWEVFLAEKLGYEWDEVHTDADLLEHISSDKLIERLDAFLGHPTVDPHGD

			TIPTSDGEVIVNQYHALVECKQGESFKVKQVDDDTEFLTYLTDKGIQLNETYQITEIEPYEGPIT LTNNDEENILVSYKAAFRIFGQ
orf00011	-	433	MKLKNILKTASILALLTGPSATMTVDHVYAEEAQDKVADSETNLEDLRQQVRSSVQTIEDLNN ELVQIDETIDGIEAEIAQTEEDISQQEAIIQDYFDQAKSRLQNMQLSNVNENAVLSLLEAESFQD IIARVNAVVQLTQANSDQIQSLQAQKDELDQLVQSLETKQTDLSNKKETIDAKKQSVHTEIADL ERLIEENREDFETLKEESEIVEICRRAEELVEERQAEQEVQEDASSNEEAKQVASQSAEETQQ PQAEETVAIEENEEPAEPTDDSAAEQSVEEPQVTETEPAEAEQSQPEPETQKQPDQQVQEE SQTQTEPEVQEEPIQPQAAVPTTAQYSIDDLEFQGVINALGKKWTFYSERVLPGGGLNIPGRH TADGFVRDGEGYIVLAASSSVGHGTIIDTPFGSQGKVYDTCASCHAGWFDVYTR
ABC	Putative lantibiotic ABC transporter,ATP-binding protein precursor	302	MKNVIEINNLTKSYDNVAVIKSININIVEGTIHALLGKNGAGKTTILKSILGLIKPTSGSIFILGENQL TEKGRNVLKQVGCMIETPGFYPNLTGTENLSIFAKLRELDKQAVQEALTIVNLPYADSKKFKEY SLGMKQRLAIANAIMHDPDILILDEPTNGLDPSGVIEIRELCKEMKRQGKTLLISSHILSEVEQVA DEVSIIDEGELVKYINLKKMNENSQVTIHIFTSNFDKVKSVLLEAGVSQHNMINTAKGVRLLSKD MDISEVNKLLTQNDVDIEGIMREKLTLEEQFQRVTETR
ABC	Putative lantibiotic ABC transporter, permease protein precursor	244	MLDLVSTEITKITRMKQMWVYASVLGIYITMMSVYTAQAVGLFDQFIHLYKFSLSYTGFLLLPFI LLSLSSSTISDDYRNEVMKNLTVIPISKQKIIAAKLIAIYIVLSVAMVMIWISVCVIGYVTRFRFSMT WLLIFRYLWLCILTSLVIFLSMLPVTLISVATKGNVVITNLIGSMYIIASFFLTNFMNGIIPLATAPHI IWYGSMEGVEVNSNIALMVISIIFYTIIVLFAMNKILKKQEL
orf00018	-	238	MVLEFKKMKRSKLIIVLILIVVLNIGIHYLMGNVKYVGIPYAAEPGWTLQNTLIVGTYYLLLPVFTL IGSASFIIERENSVYINILTIPVRKSHLILNKSKFIWVVSMIFTSSIFVFTLVLEKIHSNILTQEIVLKY LFEYVVHSNGLYVIAMLILSVIIYMDYNMQLAVMVGFVSSFVSVFIEQTAISYLYPVNALFNISGF KESNLYEYSSSVVILCVIAICAVFIYKKIAQQESW
42.1;Haloduracin_alpha	42.1;Haloduracin_alpha	81	MRGGNIMKNKVNNWKTPVYEKEEDYDNPAGDIFRELKSDDIDKVMASGTANTYCRCYSGRH SCGRACTITAECPVFTVACC
LanC	MrsM protein OS=Bacillus sp. (strain HIL-Y85/54728) GN=mrsM PE=4 SV=1	247	MKLIDKFKNGLYSFERDIQHDETNNYKSENRLQYWKKFLGVNEREIENILSNGLGINTANLNEL LSENDNFSCKVTETNVLWNQLIHDLQVLSIESIILPEFYIIGDIGQKELPMFYGFHEPFLKLAILRF ENYWKNIPGISDNVFNKLLIYLYDQLAEISYRTLILELNIAREENKLAGETSEERYNYFSTQYLS DNYWLILEEYPVMFRLMCEATQKWINNTTRFIDRILSDKDDLEKLLKLREN
LanM	Lantibiotic mersacidin modifying enzyme	791	MDVTFQSTIKWFNEVTKSHLYSLKIINHTEYGWVEYIPHEECKDYSDFKNYYTELGQLLFLFYL LRGNDIHYENIIAKGKHPVLIDLETLFHNNTSNTSGIDTAADRVNELLENSVRTVGILPNLVWAQ NGKNGVDISAISTSENKEIPIEQASITNVNKDNMKVEYKTSTLASQKNNPYIIGEEISLTSYHKYL KKGFIESYTKIKNINKKEIINQVENYKEIYARQILRPTQYYTTLIQISLHPDFLRSAIDREMLFSKL WIYFDENNSFRKVSEIEFTSLLKNDIPWLISNVSKNNITTKDGSEIESIFKHSSIALVKEKINILGD KDLTLQVELIETALNYDSEYNKAESQRENDRKIIEINDDKLNKNHLDQQLLEISTNIGDYLINQSF IGMNGDVSWIDMNVIGEKANDWNMVPTSMDLYSGLSGIMIYFIFLYKETKQNKYLIMVKRCYK

			SIINYIKNVRKRTNINSEVMFGGFSGETPIIYALTILEEELGGIFDLDELEKIRSWIFKECKKNISV GNEHDIIIGSSGVIAILLRYYDLTSNDAILEVCQQYAEQIIDNYIEMDNNSIAWIGIASRNALGGFA HGVSGIVWALSKLYSYLPDERYIEVIEKALRYEDYLYSEDDKNWVDRRETEEGIEYNNLSSNM PVAWCHGASGILLSRASLKKHNLPLSEKRKNKIDEDIEIAVRTTLKNGFGHSHCLCHGDLGNM LILKYASSELNTKNDIDKRYDIYMSHLISQLKDKWECGIPYKNSPGMMLGLSGIGFGLLSLMNE DLPFILLLE
LanT	MrsT protein	492	LKKIKFVQQLSESDCGIAALTMVLNYYGCKLHISDLVEKCSISRDGVHLRDLMNVANEYGLKSK AVELNKKENIFSEKIVFPCIAVLSMSHYVVIEKAKRGTIYYLDPECGRIIMDGSEFSKKFTGILLLF FPENIKKQKSQNKIIEILTLGEIKKKFIVGVIFFSFIIQLFVLLMPLFTEYMIDNVINGIALVSHVKLII GILFAILSYGVFSFVRELLTTLLEIKYISTLKKNVVRKLFYLPLSFFDVRSSGDIVSRINNIDSIQQL LSNIITGVFVDVLTIVISISMMLYISRILSIIMIVYGILLCVLLNIFFKKTDVKNKLSLMKREKTQSYLI ELSSNINMMKTSNYGEALYNKWVKNYNEQMHLEFKRQRLLGMYRSLIISYRLLPSVTVLFIGS NFVNQGMLTLGQVMSFLALGNTLLSPLAILIQNIFDFQYSKNNIDRLSEIITAKSEKNFDGREINS FKNIKFQDVNFSYSGINGKKMLKS
LanT	MrsT protein	187	LILSLYNNFEGNILLNNIDKEVYDLKSYRQLFGTVLQDETLLNDTIQKNIDPTHSHTISKIREAAKL ACLDEDVMNRMPSKYNTEIGDNGRNLSGGQRQRVAIARALLTNPKVIILDEGTSQLDVSTERE IFNNLKKKNITIIAVTHKLSTTTISDCVYVLKKGKIIAHGKHEKLMQNNRYYSDFFR
orf00024	-	56	MYKIVPYLNKIAHIQNLSFSLITVTNSETYPTFFLPQQNRGPYKKWTEIMTAMTDQ
orf00025	Organic hydroperoxide resistance protein OhrB	135	MLYEVVAVNETGIDGQSYLVDGESYQTSSPTSDEPGTNPEQLMGLSLATCFNATLHAVLKER EIEPKSRVQVTVQLHQDEVSEEYYFTLDLTAAVADIPLDEAETIIQATHKRCPVAKIVGDYKHLT VETVPFEA
orf00027	Lipase 2	342	MGIIWIIGYYWADLVIGITRVFSKMRQRLKRRRVKPAKQANVYKNVEYKSDYPNNKLDIFIPSH SSHPTRSPEENRIKEQQNTCDNENDMEIRPALYPVIVYFHGGGFAWGDKIDERRYLLEFVKA GYAVVAANYALTPMYKYPVPIIQASEVMDFLVNEGEQYGLDPDNIVLAGLSAGGHLAGQLALA ESSSAYADRLGLSQHPDLTIRGIMLNSALIDPNRSNQVGHWLINWLFSVMGFSYMDMNPRNM KEQMLAEANLLNHINEALPPVYISDGNRVSFTQQATDLVAQLRARDIYVQSNIYTDDYPGFMF HGYEGQLWLREARENVDKQLDFLATIIE
orf00029	Serine/threonine-protein phosphatase 1 n=474 RepID=PRP1_ECOLI	233	MDKSYIFAVSDVHGSDEQLEALLEYWNPEEEQLVILGDLCDRGPDAQAVFRRAKQLKEEYGA ICLRGNHEDMLMKFLDDPADNVAHYYRNGGEITLESFLGDAVSSSTPQELAEALKDQYPWLI DYLDNLELTYEWGKYFFVHAGMDLTLDDWRESTVRDKIWIRKGFLDVLNETEKTFVFGHTPT MKLRTDGQPDLWVSEDHKIGIDGGAVYGGQLNGVKLSQAGIEATYQA

orf00031	Uncharacterized protein YdcI	725	VEDKQTFQLIVRELSDYSTKQVRTVLSLLDDGNTVPFIARYRKDQTGALDEVQIREIQQRSEYI QNLQDRKETVLSAIEEQGELTDELAAEIREATQLQEVEDLYRPYKKKRQTLATKAKEAGLEPL ADWLLGFPDASEAEIEAIASDYLNSEYDVETVADALAGAHEIIAEQVADNPEYRKRLRDYTIYN AQITSTVKNEEIDEKNVFQQYYEYAEDYRKIAPHRTLALNRGESEEVLTVKLEVGDERAHRYL YKQILPSLDSSSPVISIVKEAIKDAYKRFIAPSIERELRSTLTEKAETHAIQIFGENLKNLLMQAPL KQKTILGLDPAYRSGCKLAVIDSTGKVLAIKVIYPHTSGETKRNQSKEALKQLITEHAVDVIAIGN GTASRESEQFVAEVVSEMPEKTAFIIVNEAGASVYSASDEARREFPDLAVEERSAISIARRLQD PLAELVKIDPKSIGVGQYQHDVSQKELEGQLDFVVETAVNQVGVNLNTASGALLEHVSGLTKT AANNIVAYRESEGQFTTRDQIKQVKRLGPKSYQQAVGFLRIPDAPEPFDQTGIHPESYTEAEKI LAHVKQDKSAIGSDELKHLLEQLPDIQTAEQLEIGLETYRDIVAALIAPGRDARDDMPAPILSTD VLKMDDLREGMQLQGTVRNVVDFGAFVDIGVKEDGLVHISELSDQFVEHPKDVVQVGDIVTV WIKSIDLNRSRIGLTMVQKD
orf00034	-	87	MMLYSHDAYDNKTARIDYTDNASVSPHRSFVRWMGYAASLQLLSTISENQHRDVDSPDPLY ADDSDKVHASQYLKRLLLQTRPKSWD
orf00036	Uncharacterized protein	419	MGLLLKNIKLERCYKEQNNGLVVTETETKDLLIKEGCFAEIADEIAENTPGVQKVIDGNGQLLV PSLRESHIHIDKTYFSGPWQAPTRPEDFSIYTRLNEEKELLPAQLDVAAERAHKVVQHYIQNG HTHIRTHCNIDPQIGLKHVELTLDVLKQYEDQITYEIVAFPQHGLLRNGEEFIQLFEQALSMGVT HIGGLDFATLERDVEASLQLLVKWAKQYDLGIDVHLHDRGALGVYEIERLLDMMAMYDYQGE VTLSHAFALAQVPAQKKETLFRRLAERNVDISTTVVISPGLTLPIFELDQYGVKVSVGHDSLTD HWSPFGTGDTIQKLNLLSEQFNLIDEKRLSYALKFGTGGVTPLDSAGRQVWPKVGMPANALL VDAVSSAHLIARRCPISTVISRGQVISEQSIQQKGAYRG

Appendix 117: Genes predicted to be involved in the production of bacteriocin DpD.

***D. pigrum* strain**	**Amino acid sequence**
*D. pigrum (*accession number MUYF01000003.1)	MLDLVSTEITKITRMKQMWVYASVLGIYITMMSVYTAQAVGLFDQFIHLYKFSLSYTGFLLLPFILLLSSS TISDDYRNEVMKNLTVIPISKQKIIAAKLIAIYIVLSVAMVMIWISVCVIGYVTRFRFSMTWLLIFRYLWLCIL TSLVIFLSMLPVTLISVATKGNVVITNLIGSMYIIASFFLTNFMNGIIPLATAPHIIWYGSMEGVEVNSNIAL MVISIIFYTIIVLFAMNKILKKQEL
D. pigrum ATCC 51524	MLDLVSTEITKITRMKQMWVYASVLGIYITMMSVYTAQAVGLFDQFIHLYKFSLSYTGFLLLPFILLSLSS STISDDYRNEVMKNLTVIPISKQKIIAAKLIAIYIVLSVAMVMIWISVCVIGYVTRFRFSMTWLLIFRYLWLC ILTSLVIFLSMLPVTLISVATKGNVVITNLIGSMYIIASFFLTNFMNGIIPLATAPHIIWYGSMEGVEVNSNIA LMVISIIFYTIIVLFAMNKILKKQEL

Appendix 118: Genes predicted to be involved in the production of bacteriocin DpE.

***D. pigrum* strain**	**Amino acid sequence**
*D. pigrum (*accession number MUYF01000003.1)	VYKQGKEPFVAVDDVSMTIHEGEIVALIGPNGAGKTTTVSMIGGYLLPTSGDILLEGKSIVKTSSKHKP KIGVVFGGHSGFYGRATLADNLSFFADLVRIPPKEHEQEVARVLKLVDLYDVREKEAHQLSTGMMQ RLHIARAMLGNPSLLLLDEPTTGLDVEIAKEVRDTIKKLAQQGMAILLTSHIMSEIEVLADRIYLIGGGQI KHEGTVADILDLAKVTHIDRPATLEESYLSIAPTLKRGN
D. pigrum ATCC 51524	VYKQGKEPFVAVDDVSMTIHEGEIVALIGPNGAGKTTTVSMIGGYLLPTSGDILLEGKSIVKTSSKHKP KIGVVFGGHSGFYGRATLADNLSFFADLVRIPPKEHEQEVARVLKLVDLYDVREKEAHQLSTGMMQ RLHIARAMLGNPSLLLLDEPTTGLDVEIAKEVRDTIKKLAQQGMAILLTSHIMSEIEVLADRIYLIGGGQI KHEGTVADILDLAKVTHIDRPATLEESYLSIAPTLKRGN

Appendix 119: Summary of genes in BGC possibly coding for bacteriocin DpE in the whole genome shotgun sequence of *D. pigrum* ATCC 51524 as predicted by BAGEL.

Name of gene	Function	Amino acid sequence length	Amino acid sequence
orf00002	-	69	MKLEKIINGYMMIALLLLFIMGRLLDYALTMDFWGAIFSSSTFYHLVALSTYIACMINMKRRGIIDSYW
orf00005	Putative D-galactosamine-6-phosphate deaminase AgaS	387	MFNKSTDELKALDALHTTTEIKQQPDLWRETLAIYRENKERIDTFLDQIKEQHKQINIIFTGAGTSAFVGETIQPYLHGKYRNTGISVQSIPTTSIVSNPEDFLSEEVATILVSFARSGNSPESVATVELAKQIIKDLYQVTITCNADGELAKNAEDDAKNLSLLMPKQANDQGFAMTGAFTAMTLSALLIFDTDADKVAYAEELIPLAENVIARESEIAQVADLDFNRIVYLGSGSLEGLSHEASLKLLELTAGKVATFYESSLGFRHGPKSIVDEQTAVIVFQSTDPYTKQYDNDVLREVYHDKITDHVFAVEQGESNFEGQSILVDKSDMKLPDAYLALPYIVVAQIIALHKAVNIKNGVDNPSPSGTVNRVVQGVIIHDYNRD
orf00007	HTH-type transcriptional repressor YvoA	239	VTTKRVPLYLQLTEKIIDQINDGTYEAGDKLPSERELCHIYDMSRITVRSALSELERDGYVKKFQGKGTFIANTTYQQNLLNVYSFTEETKKMGKTPQTNIVSFELVLADKKYASKLGIRVGDEMYRVVRCRLADQEPLIVETSYLPRYKFAHLTEKDLANSPMYDVFNRAYNIQATRAEEEFSITTLRDHEAELLAEAVGDPAMLVKRTAYDKLEEVIEYTISVINGQKYKYKVELQQ
orf00008	Uracil-DNA glycosylase	231	MNLPVMNDWRPILEEAMQTEAYQQLRAFLKQEYREHTVYPAMEHIWNAFEQTPYEKVKAVILGQDPYHGEGQAHGLSFSVQPGIPIPPSLNNIYKELQSDLGIAPVNHGFLRAWTEEGVLMLNTVLTVRAGEANSHRGKGWEALTDHVIKALNERSTPIVFILWGNQAIAKEAMIDETRHAIIRSSHPSPLAAYRSFFGSQPFSKTNTVLRALEMEPINWELPHHVSSDEV
orf00009	Phosphatase YwpJ	291	MIKLIVSDMDGTLLNEEIELSDENLAAIKDAQAHGIHFAIATGRDYQTGYTIVQERGINCSFFGLNGAIGYDEEGNRLYTKNLKPSTVQTLLHVLDREDVHVNIMTDKGVYSTNYERENEYLRHVLSDINKTLAPERLEEKLDLFLEQHNITYLDNFQELIRRDDEEILKVSAQTTAGEEMLTELKDTLLQAADDIVITASSAMNLEINHKDATKGFAVATYARELGIDHTEVLTIGDNINDLSMLEWAEHGTAMANAAPEAKETAAYETGSNSEHGVAQIINRVLAGEIY
orf00011	Inner membrane protein YjjP n=265 RepID=YJJP_ECOLI	255	MTEATTYQAEEALNICMDIGRLMLSNGAETYRVEDTMHRIATSFKLEYVNVFVVPTAIIMTTKNEVGADVTQLVRVTDRATNLEMVAELNQLSRDLSAQPKSPNEVRAYLYLLQMRIREFPPYLTVLLAAITTGFFPFLFGGSWPDIIPAFLAGGLGEFLFEYVNGSTNITFFAEVVAAFAIGLTAFTLYTLGLGENMNAIIISGVMTLVPGIAITNGIRDLMAGHLLAGVSTLAKALLTAGAIGVGIAVVLTFI
orf00012	Uncharacterized protein HI_0108	155	MFTYYIMPIFSATIATVGFGILYNIPKRTIPASATTSGLGWIVYFICTQVFHLPLFVGTTLASFTIALISQLFAKHYRMPVTIFAIPAIIPLVPGGSAYNSMLAFVTGEILSAMRYLIETFIVAGGLALGLTVNSAIFQVLSPRAIIQQGRRYLP
orf00014	-	42	MSKEEYQNGPSILSQVNDVRAALPIASLPAQLNPILLGLMLI
orf00020	Protein/nucleic acid deglycase HchA	288	MTMTELSKQPQRDRAEHNAYFPSEYSLDAYTSPKTDFDGYKIDTPYEGPTRKILVVASDERYVQMKNGKFFSTGNHPVETLLPMMHLAHAGFDLEVATLSGNSVKIEMWAMPEDDKPVMDFYHEILPKFEQPHKLADILATVTAEDSVYEGVFFPGGHAALVNLPESSDVNQVLNWAMTEKRDIITLCHGPAALLAGASDDQEFLFKDFEMVVFPDALDEGANQEIGYMPGKLKWLLAERLEELGATVLNDEMSGQVHKDRNLLTGDSPLAANALGILAADELLKRLG
186.2;Propionicin_SM1	186.2;Propionicin_SM1	244	MGFAVDDIETHAEEISDKVDISIDVGHNYSNRSSFDVGVEENYRVVDTSSSNEVNLDVLLEGINGISAELDGENVFSIYDSVNGYNSSLSLMLTYIDDNTGEVKQAQVHEPFIENQKISYTIDTSDVNGDLTI

			ATARSVNSFSYYFQSSKWISRSGVISLSLSPKQSWVRTVVHGDNPNASASLKNDSWQTVLNNYS GHSNWRNTSSMRSQYMCHVHWAGYWKTPWNLEPHRTGTAIASNSCNP
ABC	Putative lantibiotic ABC transporter,ATP-binding protein precursor	310	MEGTQWIFSRKGDHMTFQSIQVQDLTKKFGRKTVIDNANFNIAAGEICAVIGKNGAGKTTLFKLLT EQLFPNKGHIEFQGSFERPSIGTLIENPVFFPKFSAYHNLAYFSKQITGSIDKERIQEILDLVELENS RRKFEQFSLGMKQRLGIALALLFKPDLLVLDEPSNGLDPEGVRDIRQILLKVNRERRTTIIVSSHVL TELEEIATDYIILNEGQIVEKVSKDKLIDNMVKTLVIKVDAAKQAATVLNEQFDALEIKIVDDTTLHLS SDDIPSYDINRRLVSKGIQVHSLTIQNETLEDYFFEKVGV
orf00028	-	249	MKNYMKSEFRRLRKKKVGYVVLLLGIALLVGAAFGLDFMSQRELEFPYATNMFYYSSIFVAPNFL LMLTGTLAIVLLGRDRDLISVSIGFGVNRSHIFWGKYFVTLINFLIIGAIFFGVAYASGEMIVPNTEVE HLHRFINNSLNLLPILLSALTVTYVTAILFNSEISAFILILLIYRVINYASNAIIGILPQSEPVFDYLPGTLF SELPANYLTGNVQLEYVHWGINLGIILVFLLLGSLLYRRKSY
orf00029	-	231	MANRKTQFVHFLCLPIINMILFLFIHRQFTGGDVLNPVVAVTSVAVAAITISLSSVSQLLTHDSMRG VDKIMVINRPYSPRYWGNKILTAMVVSWGLMAINLGLLAIVGVDWVVISRIVLASPALLVTGVLLG MLGFFLAWRESNPYFYTNLISAAIPIIYGVIVSVSEYPPVFKVFSQIFPFYNVRDYLQTGQWEHLAI EGIKLVVLIIVTGLVYQLKHKKVVDKKGLLF
orf00031	-	222	LPISSTISIFLLQYLGAYASGTLASSNIWLISGTFGMWASTTTAAGSIGFQRYMGTLQYIVNTRIDDR VSVATAITPASTYGLLAYLVALVMSVILRVGIHGLTIGTVLAIISLWLSALIMSLFIAAFFVFTPNAMTY EELIMIPILLLSGLFSLQLVQLPIFSVFQWLLPLATPIKFLLTEAVEFDILPWLSSLVLWSGLSWWLS SVLLKRANITGQIGGV
ABC	ABC transporter ATP-binding protein NatA	245	VYKQGKEPFVAVDDVSMTIHEGEIVALIGPNGAGKTTTVSMIGGYLLPTSGDILLEGKSIVKTSSKH KPKIGVVFGGHSGFYGRATLADNLSFFADLVRIPPKEHEQEVARVLKLVDLYDVREKEAHQLSTG MMQRLHIARAMLGNPSLLLLDEPTTGLDVEIAKEVRDTIKKLAQQGMAILLTSHIMSEIEVLADRIYL IGGGQIKHEGTVADILDLAKVTHIDRPATLEESYLSIAPTLKRGN
orf00034	Formylaminopyrimidine-binding protein	351	MTTYIKRLMKGITAVLVLLVLAACGQSTQNDADQAADSPADDQTTESVAEGELQDVELTLDWYP NANHVPIYTALKHGYFEEAGLNVTLKMPAEADDPIRLVGANQTDIAVSYPGVLMKARAEDIPVKAF GSLVQRRLDAIMYKEESGIQSPKYLEGKKIGYASDSISEEIIYSMVEEDGGDASKVEMIDVGYDLM PALSTDNVDALISAYMNHEYLLLEDEGYNMGHFEFQDYGIPENQELIFIASDQTIDERSDVLTKFM TALQKGYETAVENPDEAIETLFENEENEYALDKDIEIRSWKEFLIEYMSADGEFGTIDPAQYEEYA KWIYERGAIDSELTGEDLTAPAL
orf00035	Formylaminopyrimidine transport permease protein ThiX	239	MALLGFFEWSVSAGITPHFIIPKPSSVVLTLIEQYELLWRHTLITLLEVAVGLGVSIGLGIPLGILLHY SSWAKRALYPFVLVSQTIPIIALSPIFVMWFGYGLTIKVAIIFLFCFFPLVVSTYDGLKVTDSAYLSLF RNLKASKWQMFKYLQWRMALPSVLSGIKLSVIYALMGATVGEWLGGSDGLGYYIRRTASNLKAD GVFAGIVILSLIGLILFGVVSGLEAVLLHYRNQERNA
ABC	Uncharacterized ABC transporter ATP-binding protein MJ0412	250	MMLELQSVDFSFQHDQEKRVLSEVSLSVAPNQIVSVVGKSGCGKSTLFKLCTTELTPTRGAIRFQ GREIQLGDVAYMPQQDLLLPWQTVLGNVMLPTKLDEANQLTEEHGRRWLEKAGLGDVADQLPH QLSGGMKQRAAFIRTLMTDSDVLLLDEPFGALDYFTQKEMQEWLLTLWMATNKTILVITHNIEEAL YLSDVVYVMQPYRPGQSEQVTPIQIDLPRPREEAVRYSREFIAYKRQLEEAIYEKN
orf00039	-	516	LVDRADKIEQDATQLTKHVVELEGKLHGLDAKVIQELAERKQSIISIAVQIFNIKQELQLSTKDIAELE RKLVELSAELAAEKERLSNVIKRVNEHDERITDLEKRTNAVETVVSRLTGRIAELELTVGLLSDQLK QEVENRKQAVSNLKDQITELKQELRVTTEDRAKLERKLADLEANLEAEKARINDLSQRVNENKANI DVLGHRTDKLESNANHLAERVTELENRVNHFGGALKDEVEQRKQGINEIYAEIKRVKESFTKATS DVREELMKHLDELEKALAEQSNALIQLDIRITIEINRLEEQNNKVDNGSNDTEKLGSDDSVTNSGN DSDDQDTDDVTPPTDKDITDSDEADTDTDTEDGTLTEPETGDNSEETVNPEKTLETDVETGSDS

			DKSPETDKEIDPKVDPEVETDPETNSEVGADPERESDTDEDSQPNVSLEVDNSKGSDSVTDTEE AQPTVTQAGLTGSTSQTQGELLPATATGAWTLGLIGLTSMGLGGVLSIRRKDDEE

-; not mentioned by BAGEL

Appendix 120: Summary of genes in BGC possibly coding for bacteriocin DpG in the whole genome shotgun sequence of *D. pigrum* ATCC 51524 as predicted by antiSMASH.

Type of gene	Locus tag	Location of gene	Amino acid sequence
Additional biosynthetic genes	HMPREF9703_01020	1101976 - 1102482	MKIRHAKQSDAFAVACVLIDSWRETYQGIFPQSLLDNLDQATQTQRMLAAIQMSNM LVLENDAQQLIGFIGGGNNRQTAQFPFIDGELYALYVSLSAQGRGYGAKLVVAFKTL MKELGYKGLLVKCLKENTNAQAFYQHMGFSTIGDQMLTLQESSVVETVLAKSINP
Other genes	HMPREF9703_01021	1102859 - 1103794	MKKKIIYFGALLTISLSGCSIINGSKSEELEEGSFHEIAENSTYTSTEMSLQDLSLESV GGGVVIDNELYLSDEEAGKINVYNESYELIDEIEQDSILSPTLLAGEGESLYIIDIGLKSI LHVDKQTHEIVETIPFPVEIHEPEFLSLVVKDTALFLTYDSNQEKTAHIFKIDLESKQV SEIGELFSGSIHEDEFGDLYALELGEYYKDNDSEGFESGENSLYKVIDNKLVEEVEL PYKLTALDFLLKDDHIITYTYFRHSVVRLDKQGNLIDSLATFDPDYVDYSPVLLEYNH DILVLGRQYDQIFKLSK
Other genes	HMPREF9703_01022	1103925 - 1104953	MNKVLYALESKFRQRLANKESLLLSLVILAIGFIFVIMFVSKWLDAKAEINRIDWLYRD EDVLVITDYTNDLFPDMSPNAERILTDELLVGDHYNMVSGITNHYHNYIPPLQSERD LWQLFDTRSDVIIAGENLKRQHNLAIGDEIHIGESIYHLIGFTLEPLLQNRYVTHFDEV DNQLQTSLRRSEYLINISEVSSQYQSNPTLADRARELEASTRGVQNIISVVTLFLLFI MLMNVSLVLAATLEQVRHADMVRTIFGQGEKLRQLLVLVDTLLLVTVSFHLGVGVY YIIRGSIPVFFYFELTWPVYTISWGLIIIVSSVLAAILSRWQGRRLTIQLLRG
Other genes	HMPREF9703_01023	1104964 - 1106097	MGYQISQVMHQLWREKWKYLFIFAQCVICLTLILMGLNEQHSYAYRRDVLMYDRSH QLTTISSVQGDLLNQTVPDDLLDRQYIIYHKRSLINYLTTDDDIGSIYLVRGNQAFFDT YFPDQPVEKGVAYGDKDVIVSLGKADGQTLTEGIAFYSDKVVIEEKMYTYDSIVPSR ERIPFQAIGSDDIDTSNALFILASDLAHYGEDEGIPVIKTSSSLTDVADLVTEMTHLNP NNRYEAVDLLGHFDSRVADMRSTVQLFSWIGYVAMIMIAFGIIGIMMLIFHKRQKSYII HHLFGATWQILAIQMASELSILMISAFLLSIVAAGLIQPELSTDYFPIVTHGSSVGLTGF IVISIIMINTVWMTYNLSRLKLTNWLI
Transport-related genes	HMPREF9703_01024	1106113 - 1106796	MGGTLIKLSNIHKTFRQGSSEVDIFRQLNFVMKCGESVAIVGQSGLGKTTLLNLIALL DTDYNGQYSLLGEDTSHLTTKELFSLRMAHIGYMCQYHNLLVDMTVMENVQMPLG YAGVSKVERKERSLAIVEKIGLADKASEKAANLSGGQQMRVALARALVSKPDILIAD EPTGNLDASTSQDIMELFQSINSQGMSMVIVTHDLEVAHNCNRIVELRDRQLWEVA K
Other genes	HMPREF9703_01025	1106864 - 1107151	MNDRQLKTKWAQLNWRSRIAYLSIVFVLLFIPFVRIVLETLFGVELNSFLIIVIAWGFS VLANLIAVEWKWLIFAAVVPIELYLFTSRVIELLPL
Other genes	HMPREF9703_01026	1107221 - 1107502	MTISNLTNNWKNMNGLSRTAFVAVGTLVLTIIITVTLDIGFDFNLLSWTQLNDLYFLCL GLCIIAREWKLMISTIIVIRVMFALVVGIGMMIP

Other genes	HMPREF9703_01027	1107540 - 1108157	MFKLVRKFTIVMLVTLIGMFSAAQITTISVHAHENKYNIESSENIETTEEDIAELRSTLG SVEAYIDMDGDQLFNEELARQNNESQEVIEAGLTYNDMILTEQGALTSRSIIDYGNW CGPGDKGSPPIDTLDRQCQKHDKCYEQNGWGNTKCDINFVHNIASNFGSIKGTGA QAYAIAAIMTFAAKVGGTTALAIQYPILAPFLP
Transport-related genes	HMPREF9703_01028	1108646 - 1109530	MNITIPEFSYGTNLILKDISLELTKPQIIGLVAPNGTGKSTFIKLISGHLSAPEVDINLDG HTYETDLIYMRRQIVKMPDQMDLYDELTARDHLRYYASLWEVEKEHVEVVIQTLNM SSYCDNPIKSYSLGMRQRVCFALMLVTKAQYLLLDEVMNGLDPDNVDMISDILFKLK EKGTTMLIASHLLENLDQIADVVYFMKDHQLAFRYEPAISQREMLSVQFSDEVHQK SFIDTFTDIPIQFDNDSLYIDVPLIDEFMLEEIFRWLTTHLDRVGEVKLGTKGCHKLFR ELYDR
Other genes	HMPREF9703_01029	1109565 - 1110668	MKLIQLDWKNFLKKHLMWKFLALACCSMLLSVGYFKLFQEGKIYYELDQAINQVREL EAELKYSEALTDERSDFISRYRSTYNRQRGIYQVDKNISSLPAAIDERIVLREQADEL LGYLADEYLESIDGIQKDKTIIQYLSDHELLISDYASNNIVVQNIVVLLGLGGGLFIYAL LSSELLVSHQLKQTLLEAIPEKLFIRHMSYYGYTVVVYFMLPILAILLVLLGYSLAVNS HVWQHPYLISQGDSYIALANWQIIAWSLLMTLVATALAYQFLFIVTPILQDSIFSALIFI ALVSLPIFFTSAWVNYLPYSHSILYPVFHQQSGSLVIDITKQLGISIGYVLILQLIIWVIH KLRVTKRLLFKL
Other genes	HMPREF9703_01030	1110689 - 1111870	MSGFEWKLLWRQRLFRFKLLTIVMLTCGLFFVMIEDLSERELAIIESNQLSQDYYLAS ANQMMQTYGENNKDVQMMIQLASLAEDQSQALQRKDYRAFVQKRYKFLQAHQQA IEELNWAQSNVLDNFAFEAATTDKVEALGLELYQATAHSLEQLLAQDTISQPEALGK HTTMMVRSLFLGESPQFINFWLAVFLVTLFLVSQTCLEGRKNHQSFTEIIPQSKAVH TIATTVSNAVVVTIFQILLLILVLISYTFVYSFGHLSVTIRRFGSSGFMSLIVVIALCLAML WGLNMLITALTALFNQLFNNQVLTLFVVVLFICFIPMVNTLEVDASWFQYNPFSYVEI GHVVWGVQHYFFTEGSFSLKEFWWSIVFTVTAIYGMLYGVAYFKQRN
Core biosynthetic genes	HMPREF9703_01031	1111892 - 1113046	MGNTEETREKIKLILNNIHDNDYSLNNGNVGKLLLLDIIHEQKYLSTIDYMDQREKVL KYLWHNLKNNPIYPGIFDGISGVGLYFKQRTELTEYEDMIYQYTLNKCSQIIEKNNHL ETFVYYELYGGLIGCGKVLLGEKKYANDLLALTEKLVLETEKFVDKIMHDRVAYFFIE NGIGINATSITYNLGLSHGLAGLLSFLIDFYEENKENRYFDSIIERIKKAIDSLLLIYMSS RKYTDDYWYWNDEITLIWKHKIKEVSTFTWCRGMVGISTLVYKGFKLLGKEENAQN VRENIIQLFNSVKKDVLFENNSICHGSFGIIYTMMNKFNYYNEIFTKELANMMIDVISSI EDYSFLQGELGSFIIALQAENLKKLKIDTFLGY
Transport-related genes	HMPREF9703_01032	1113060 - 1114712	MINQIKKYYDILKNYSTVYMYVYGISTIIIAIFPFFNIFFIRYILNSLTNFTVSDLNVAVMIY LTVSLLQYIVTKINSFASDMMELTLGNQIAINIMESTENLTLFQFEQHDFQNQIRRAM ENSLGVFVSNITIFINTLTTVVTIIISVIYISQWNIFIGITVLIFPFLFYKLYIRINQEHYRVSI KQTEPKKINWYITFLLTQDDAFRENKIFKFSRYLIQKYRKNVGSFIQDTKELFKFDSKI SFVPELANILLVILVIYHLLREAIIGTILVGSIVAMLQMTFQILDSSKELSGNIISLEKNSY YIDELIDVLNISKDKRNDKEVRYNQKIKSITLKNISFFRDDIPVFQNFNLHLERGLYFVI GDNGSGKTTLLKLISGLYKPDSGNIVINNRVCERTSDILKQESSVLFQNFKRYENTLK ENIFFGDYKEKENKDRLDMILNKELLNFIEWESKNNDIQLGSWFEGARDFSGGEWQ

			KIAFGRTLFKEASIYIFDEPNSMIDEKGCKVIDHELRKLAKENIVIVITHKTHLIKKDDNV IHLKKIVT
Core biosynthetic genes	HMPREF970 3_01033	1114725 - 1115741	MINETYYTRTALLSIDKYFNDIIKSEEEIRELVKLPGLVTAIKFTSNNLLEAISNYDKNR TKQSIKTHESLINYASRAALRPTPYGKFASVGRGIFVSENKKENSVPYNLMKTKMYL NINMQWISKLISSLEKNLDIFEVLSLKISPQILFENNSVLVLNNKDANQSKIIELTPLLSY IINLMGNNSMSVQNLIKHILNKYNASREDVIRYLKKLMKEKLLFSNLQPQPPFINSLD RILNFFIKNNLTDKIIYEKLLSLNTIIIRINENNSLYQIDDIRRMMDDILSDFKGDYFHVDT KDCKDTSLLLGVKQKIDQLEQINKYFLYNDYGKFGNQKNC
Core biosynthetic genes	HMPREF970 3_01034	1115704 - 1117647	MIMENSGIKKIVNYFNKYYEMNDEIPLLEFLYSENAMKELVFTGYNEEDEYLSKRINII LNWFNKNSGRSIEINDEILNTLGKYVNNTPVDYTLIVESLDECNENIIINHTLTGCRVE SLMGRFAYMFDNKLRIKYDNNLYLNNHPFIEAEITTLSDKSGKINNLFTSVESSIPKIT CNTVNVNNEIPLNKILVGLSNGEIYLKYKGKILLPILNNMVNLFSGFSRVEKFLMLIYIS VAKPYQSLLNQFISNSNYMPRITYKGMVISKAKYRVLIDDFSKESQKNFNLFKNEFY RYMEEQGLPDHVGIEQFGDVQYLYLKNNRFLHIFYKNLKKYKVLILEELGFEYDCEK HQGYHFSQYVFSICKKNSCNYIETLDEDVLLPKDSVRMIDYSDTVYLNLYTREIYAD EILINLSEMLPKILKEQDFFFIRYWDPIFHLRLRIRNCGEKKFIITEEILKAVEKLGYLDS GVIFRYTTDNFLPEINRYGGEKLYKYVEQYFIAESKLIIQLLSCKEEDIIYISIKCLLLILN ELTKSDLVESKKILKSIIDNTEPIADSTQEVYKKYISDRQNKKVYEVRKNIIECRKILTQ MANLEPKTVSKRKFIGSILHMFVNRIGKNSTISENQMYKLLLRIINRDTYVKLKIR
Other genes	HMPREF970 3_01035	1117824 - 1119032	MSLFELKLLWKKPVFRFKLLLVVLLVGGLWWQFIFKDDHSKESELYTVELDAQISES EKLSLREHYGEDDEQAYLYEQRLNASLKAYDSLIKEDYQTYVEARHRVVDIFLSHM YEQFQLRGGYRFDWAKSFTATEEDNQEFGLKWYRYQEHAFSQLIDQEGLTEDQAL FTTSITMLTELFSVFRSWTFIGRYTWHITLLVTLLVSSHLSLSEKRHQSFLEVLPQYK SREAVKTAIGHGTLVTIFEISMFLVVTFIYGMLNEFGDISATTVTFWSGTWNKDYRIIP VFGVLTLCLLILWFFNLLITGLITLFNVLFKEELITMIIISLAILFLPIVNVLGVEGSWMQY NPFSYVEIGNLVWGVQEHFFTEGAFSFREFGISMSIALVSVYLLSYVVISVIEKV
Other genes	HMPREF970 3_01036	1119126 - 1119356	MRNLFDLNIKSATIKSIGENSSRQNLTVPTTGCNTVMTGYGSGVGGNTGGTRHPNP GCTLRTMICPAPFKNNNGNQ
Core biosynthetic genes	HMPREF970 3_01037	1119932 - 1120138	MKDMFDLNVKSVTLHTVDKQISKNAVGGLGGVNPCIEPPSFGKIEPSTRHTNPGCT LKTMICPTIHRR
Other genes	HMPREF970 3_01038	1120208 - 1120540	MLEEFLTEFRKSKMLQNYKKYLKQVSNNEEASAILDEYRETTALFIERKKNGEDIPD DEMERYTDVTRMLTKKEILKDFIEYDTKVYFFIEELVDGLRDEVKEAFYDSYK
Other genes	HMPREF970 3_01039	1121606 - 1122202	MINLLAMIIAYLLGSIPSGIWIGKLFYGKDIRNYGSGNSGTTNTFRVLGVPAGIAVFIID TAKGAIPVLLPLVLATSVHPLLFGLVAVLGHTFPIFANFKGGKAVATTAGVGLGIYPV FVLTMAAVFATILFISSMVSLSSMLTVGIAILASLYIGDTIFTGTVVVLFLIIVIRHKSNIK RILDGTESTVPFGLRKDK
Other genes	HMPREF970 3_01040	1122433 - 1124436	MAEQGYTDASIKILEGLDAVRKRPGMYIGSTDSRGLHHLVYEIVDNAIDEVLAGYAD QVDVVVHKDGSVSITDNGRGIPVGMHESGIPTMQVIFTVLHAGGKFGGDGAYKTSG GLHGVGASVVNALSEWLTVTVEKGGYRYTQQFVNGGHPEGGVVKKKISHKKSST

			MIRFKPSAEIFSSTAFHYQTIKERLRESAFLVKGTKMTLKDERGDQEADEFLFEDGIQ AFVSYLNEEKDTLSEIAYFEGTAQGIEVEFSFQYNDGYSETLLSFVNNVRTRNGGTH ETGAKSAITKTFNDYARKVDLLKDKDKNLEGSDVREGIAMVLSIRVPEALLQFEGQT KEKLGTPQARGVVDSIISDQLSYFLAENGELSKKLLRKAIKARKAREAARRAKDNSR NGTSKKKRERLLSGKLTPAQSKDADKNELYLVEGDSAGGSAKQGRDRKFQAILPL RGKVINTEKASLQDILKNEEISTMIYTIGAGAGPEFDLADCNYDKVIIMTDADTDGAHI QVLLLTFFYRYMKPLIEAGKVYLALPPLYKVESGRGKKKTIEYAWTDEDLERMLDVV GKPYTIQRYKGLGEMNADQLWETTMDPATRTLIRVTIEDSEKAERRVSTLMGSKVD PRREWIEEHVEFTLAEDGSILETNVGGDHVPDETVDMTGRKDND
Other genes	HMPREF9703_01041	1124437 - 1126908	MAQEQPQQIQELTLENVMSDRFGRYSKYIIQDRALPDVRDGLKPVQRRILYAMYAD NNTYNKAFRKSAKTVGNVIGNYHPHGDSSVYDAMVRMSQDWKMRQPLIDMHGNN GSMDGDPPAAMRYTEARLSKIAGELLSDIDKETVQHVLNFDDTLEEPTVLPASYPNL LVNGATGISAGYATDIPPHNLAEVIDATIHLLKYPNARLETLMDYIQGPDFPTGGIMQ GTKGLKKAYKTGKGKIIVRGKTMIESIRGGKEQIVITEIPYEVNKSKLVRKIDDIRIQKKI DGIADVRDETDREGLRIVVEMKRDVDAEGILTYILKHTDLQVSYNLNMIAIDHRRPQQ VSLAQILNSYLEHKKDVITKRTTFLLNKAEKRQHIVQGLIKAVSILDKLIKTIRSSENKA DAKANIMDEYGFTDAQAEAIVTLQLYRLTNTDITQLQEEADELAGNIAHYQDILSNAS TLDEIVKEELLAVKKAYGSPRLTQVEDKIEKLKVETEVLVTEEQAIVTVTEQGYLKRT SLRSYAASKVEELGKKQDDILLFAEELSTLDHLLIFTNKGNVINRPVHELPDIRWKDV GEHLSRSLSLASGEIIKAVYSYRELSDTAKYVFITRDGYIKQTLENEFEPKRTYKSRS STAIKLKSDADYVVGLYRIESDQHKDIFIATAKGYGLRYQLSEVSTVGANAIGVKSINL KDDDVVVSGIIFEREDPSKEAMVITHRGSVKKLSLASFDEGSRANRGLKIIRQLKTNP HRIQYVIDVDMNEPDPIQLLTDKDKLFAVNPADYTPATRDNNGSFILNTDEDGEITGY RYQLLDIIETNN
Other genes	HMPREF9703_01042	1127146 - 1129395	MNTKVADIKQWDGFKEGKWQESVDVRDFIQNNYDLYEGDESFLAEATDATKQLWE QVMELNRKEFAAGGVLDMDTEVVSTITSHGAAYLNKELETIVGFQTDKPFKRSFQP FGGIRMAELSAEAYGYEIDSEMSHIFRDYRKTHNDGVFSAYTPEMREVRRSGVITG LPDAYGRGRIIGDYRRVALYGLDFLIEDKKREHSTYGNGKMTEAVIRQREELSEQIK SLNELIELGRIYGYDLSRPAETAQEAFQWTYLAYLAAIKQQNGAAMSLGRVSTFLDI YIERDLERGVITEKEAQEFVDHFVMKLRLVKFARTPEYNELFSGDPTWVTEAIAGVG LDGRPLVTKNSFRFLHTLTNLGPAPEPNLTVLWSPDLPQNFKEYSAKVSIESSAVQY ENDEVMRPEYGDDYGIACCVSAMEIGKQMQFFGARANLGKALLYAINGGVDEKSE VQVGPDFGKIESDILDFDEVMEKYDQVLEWLCELYINTLNVIHYMHDKYSYESIMMA LHDTEIHRTMATGIAGFSVAADALSAIKYAKVHVIRNEDGLAVDYKVEGDFPKYGNN DKRVDDLAVDILDRFITKLKKHDTYRAQETTTSILTITSNVVYGKKTGNTPDGRRAGE PFAPGANPLHGRDCQGALASLNSVAKLPYKNGKDGISNTFSIVPGALGKDIDTQKQ NLSLMLDGYAKKGGHHLNVNVLNKETLLDAMENPSEYPQLTIRVSGYAVNFIKLTRE QQMDVISRTFHESM

Appendix 121: Summary of genes in BGC possibly coding for bacteriocin DpG as predicted by BAGEL.

Name of gene	Function	Amino acid sequence length	Amino acid sequence
orf00001	Uncharacterized N-acetyltransferase YuaI	168	MKIRHAKQSDAFAVACVLIDSWRETYQGIFPQSLLDNLDQATQTQRMLAAIQMSNMLVLENDAQQLIGFIGGGNNRQTAQFPFIDGELYALYVSLSAQGRGYGAKLVVAFKTLMKELGYKGLLVKCLKENTNAQAFYQHMGFSTIGDQMLTLQESSVVETVLAKSINP
orf00003	-	311	MKKKIIYFGALLTISLSGCSIINGSKSEELEEGSFHEIAENSTYTSTEMSLQDLSLESVGGGVVIDNELYLSDEEAGKINVYNESYELIDEIEQDSILSPTLLAGEGESLYIIDIGLKSILHVDKQTHEIVETIPFPVEIHEPEFLSLVVKDTALFLTYDSNQEKTAHIFKIDLESKQVSEIGELFSGSIHEDEFGDLYALELGEYYKDNDSEGFESGENSLYKVIDNKLVEEVELPYKLTALDFLLKDDHIITYTYFRHSVVRLDKQGNLIDSLATFDPDYVDYSPVLLEYNHDILVLGRQYDQIFKLSK
orf00004	-	342	MNKVLYALESKFRQRLANKESLLLSLVILAIGFIFVIMFVSKWLDAKAEINRIDWLYRDEDVLVITDYTNDLFPDMSPNAERILTDELLVGDHYNMVSGITNHYHNYIPPLQSERDLWQLFDTRSDVIIAGENLKRQHNLAIGDEIHIGESIYHLIGFTLEPLLQNRYVTHFDEVDNQLQTSLRRSEYLINISEVSSQYQSNPTLADRARELEASTRGVQNIISVVTLFLLFIMLMNVSLVLAATLEQVRHADMVRTIFGQGEKLRQLLVLVDTLLLVTVSFHLGVGVYYIIRGSIPVFFYFELTWPVYTISWGLIIIVSSVLAAILSRWQGRRLTIQLLRG
orf00006	-	377	MGYQISQVMHQLWREKWKYLFIFAQCVICLTLILMGLNEQHSYAYRRDVLMYDRSHQLTTISSVQGDLLNQTVPDDLLDRQYIIYHKRSLINYLTTDDDIGSIYLVRGNQAFFDTYFPDQPVEKGVAYGDKDVIVSLGKADGQTLTEGIAFYSDKVVIEEKMYTYDSIVPSRERIPFQAIGSDDIDTSNALFILASDLAHYGEDEGIPVIKTSSSLTDVADLVTEMTHLNPNNRYEAVDLLGHFDSRVADMRSTVQLFSWIGYVAMIMIAFGIIGIMMLIFHKRQKSYIIHHLFGATWQILAIQMASELSILMISAFLLSIVAAGLIQPELSTDYFPIVTHGSSVGLTGFIVISIIMINTVWMTYNLSRLKLTNWLI
ABC	Uncharacterized ABC transporter ATP-binding protein YknY	227	MGGTLIKLSNIHKTFRQGSSEVDIFRQLNFVMKCGESVAIVGQSGLGKTTLLNLIALLDTDYNGQYSLLGEDTSHLTTKELFSLRMAHIGYMCQYHNLLVDMTVMENVQMPLGYAGVSKVERKERSLAIVEKIGLADKASEKAANLSGGQQMRVALARALVSKPDILIADEPTGNLDASTSQDIMELFQSINSQGMSMVIVTHDLEVAHNCNRIVELRDRQLWEVAK
orf00008	-	95	MNDRQLKTKWAQLNWRSRIAYLSIVFVLLFIPFVRIVLETLFGVELNSFLIIVIAWGFSVLANLIAVEWKWLIFAAVVPIELYLFTSRVIELLPL

orf00009	-	93	MTISNLTNNWKNMNGLSRTAFVAVGTLVLTIIITVTLDIGFDFNLLSWTQLNDLYF LCLGLCIIAREWKLMISTIIVIRVMFALVVGIGMMIP
orf00010	-	226	MLKGDHIQHIDILFSTKGGKLMFKLVRKFTIVMLVTLIGMFSAAQITTISVHAHENK YNIESSENIETTEEDIAELRSTLGSVEAYIDMDGDQLFNEELARQNNESQEVIEA GLTYNDMILTEQGALTSRSIIDYGNWCGPGDKGSPPIDTLDRQCQKHDKCYEQ NGWGNTKCDINFVHNIASNFGSIKGTGAQAYAIAAIMTFAAKVGGTTALAIQYPIL APFLP
ABC	Putative bacteriocin ABC transporter, ATP-binding protein	325	MGKSICDNIANYFIAYIREIISYNLYKGRQQMNITIPEFSYGTNLILKDISLELTKPQI IGLVAPNGTGKSTFIKLISGHLSAPEVDINLDGHTYETDLIYMRRQIVKMPDQMD LYDELTARDHLRYYASLWEVEKEHVEVVIQTLNMSSYCDNPIKSYSLGMRQRV CFALMLVTKAQYLLLDEVMNGLDPDNVDMISDILFKLKEKGTTMLIASHLLENLD QIADVVYFMKDHQLAFRYEPAISQREMLSVQFSDEVHQKSFIDTFTDIPIQFDND SLYIDVPLIDEFMLEEIFRWLTTHLDRVGEVKLGTKGCHKLFRELYDR
orf00012	-	367	MKLIQLDWKNFLKKHLMWKFLALACCSMLLSVGYFKLFQEGKIYYELDQAINQV RELEAELKYSEALTDERSDFISRYRSTYNRQRGIYQVDKNISSLPAAIDERIVLRE QADELLGYLADEYLESIDGIQKDKTIIQYLSDHELLISDYASNNIVVQNIVVLLGLG GGLFIYALLSSELLVSHQLKQTLLEAIPEKLFIRHMSYYGYTVVVYFMLPILAILLV LLGYSLAVNSHVWQHPYLISQGDSYIALANWQIIAWSLLMTLVATALAYQFLFIV TPILQDSIFSALIFIALVSLPIFFTSAWVNYLPYSHSILYPVFHQQSGSLVIDITKQL GISIGYVLILQLIIWVIHKLRVTKRLLFKL
orf00014	-	393	MSGFEWKLLWRQRLFRFKLLTIVMLTCGLFFVMIEDLSERELAIIESNQLSQDYY LASANQMMQTYGENNKDVQMMIQLASLAEDQSQALQRKDYRAFVQKRYKFLQ AHQQAIEELNWAQSNVLDNFAFEAATTDKVEALGLELYQATAHSLEQLLAQDTI SQPEALGKHTTMMVRSLFLGESPQFINFWLAVFLVTLFLVSQTCLEGRKNHQS FTEIIPQSKAVHTIATTVSNAVVVTIFQILLLILVLISYTFVYSFGHLSVTIRRFGSSG FMSLIVVIALCLAMLWGLNMLITALTALFNQLFNNQVLTLFVVVLFICFIPMVNTLE VDASWFQYNPFSYVEIGHVVWGVQHYFFTEGSFSLKEFWWSIVFTVTAIYGML YGVAYFKQRN
LanC	Lantibiotic biosynthesis protein	384	MGNTEETREKIKLILNNIHDNDYSLNNGNVGKLLLLDIIHEQKYLSTIDYMDQREK VLKYLWHNLKNNPIYPGIFDGISGVGLYFKQRTELTEYEDMIYQYTLNKCSQIIEK NNHLETFVYYELYGGLIGCGKVLLGEKKYANDLLALTEKLVLETEKFVDKIMHDR VAYFFIENGIGINATSITYNLGLSHGLAGLLSFLIDFYEENKENRYFDSIIERIKKAI DSLLLIYMSSRKYTDDYWYWNDEITLIWKHKIKEVSTFTWCRGMVGISTLVYKG FKLLGKEENAQNVRENIIQLFNSVKKDVLFENNSICHGSFGIIYTMMNKFNYYNEI FTKELANMMIDVISSIEDYSFLQGELGSFIIALQAENLKKLKIDTFLGY
ABC	Subtilin transport ATP-binding protein spaT	551	MVINQIKKYYDILKNYSTVYMYVYGISTIIIAIFPFFNIFFIRYILNSLTNFTVSDLNVA VMIYLTVSLLQYIVTKINSFASDMMELTLGNQIAININIMESTENLTLFQFEQHDFQN QIRRAMENSLGVFVSNITIFINTLTTVVTIIISVIYISQWNIFIGITVLIFPFLFYKLYIRI NQEHYRVSIKQTEPKKINWYITFLLTQDDAFRENKIFKFSRYLIQKYRKNVGSFIQ DTKELFKFDSKISFVPELANILLVILVIYHLLREAIIGTILVGSIVAMLQMTFQILDSS

			KELSGNIISLEKNSYYIDELIDVLNISKDKRNDKEVRYNQKIKSITLKNISFFRDDIP VFQNFNLHLERGLYFVIGDNGSGKTTLLKLISGLYKPDSGNIVINNRVCERTSDIL KQESSVLFQNFKRYENTLKENIFFGDYKEKENKDRLDMILNKELLNFIEWESKN NDIQLGSWFEGARDFSGGEWQKIAFGRTLFKEASIYIFDEPNSMIDEKGCKVID HELRKLAKENIVIVITHKTHLIKKDDNVIHLKKIVT
orf00017	MutB	338	MINETYYTRTALLSIDKYFNDIIKSEEEIRELVKLPGLVTAIKFTSNNLLEAISNYDK NRTKQSIKTHESLINYASRAALRPTPYGKFASVGRGIFVSENKKENSVPYNLMK TKMYLNINMQWISKLISSLEKNLDIFEVLSLKISPQILFENNSVLVLNNKDANQSKII ELTPLLSYIINLMGNNSMSVQNLIKHILNKYNASREDVIRYLKKLMKEKLLFSNLQ PQPPFINSLDRILNFFIKNNLTDKIIYEKLLSLNTIIIRINENNSLYQIDDIRRMMDDIL SDFKGDYFHVDTKDCKDTSLLLGVKQKIDQLEQINKYFLYNDYGKFGNQKNC
LanB	Lantibiotic biosynthesis protein	627	MNDEIPLLEFLYSENAMKELVFTGYNEEDEYLSKRINIILNWFNKNSGRSIEINDE ILNTLGKYVNNTPVDYTLIVESLDECNENIIINHTLTGCRVESLMGRFAYMFDNKL RIKYDNNLYLNNHPFIEAEITTLSDKSGKINNLFTSVESSIPKITCNTVNVNNEIPL NKILVGLSNGEIYLKYKGKILLPILNNMVNLFSGFSRVEKFLMLIYISVAKPYQSLL NQFISNSNYMPRITYKGMVISKAKYRVLIDDFSKESQKNFNLFKNEFYRYMEEQ GLPDHVGIEQFGDVQYLYLKNNRFLHIFYKNLKKYKVLILEELGFEYDCEKHQGY HFSQYVFSICKKNSCNYIETLDEDVLLPKDSVRMIDYSDTVYLNLYTREIYADEILI NLSEMLPKILKEQDFFFIRYWDPIFHLRLRIRNCGEKKFIITEEILKAVEKLGYLDS GVIFRYTTDNFLPEINRYGGEKLYKYVEQYFIAESKLIIQLLSCKEEDIIYISIKCLLL ILNELTKSDLVESKKILKSIIDNTEPIADSTQEVYKKYISDRQNKKVYEVRKNIIECR KILTQMANLEPKTVSKRKFIGSILHMFVNRIGKNSTISENQMYKLLLRIINRDTYVK LKIR
orf00021	-	409	MLGFVRTMSLFELKLLWKKPVFRFKLLLVVLLVGGLWWQFIFKDDHSKESELYT VELDAQISESEKLSLREHYGEDDEQAYLYEQRLNASLKAYDSLIKEDYQTYVEA RHRVVDIFLSHMYEQFQLRGGYRFDWAKSFTATEEDNQEFGLKWYRYQEHAF SQLIDQEGLTEDQALFTTSITMLTELFSVFRSWTFIGRYTWHITLLVTLLVSSHLS LSEKRHQSFLEVLPQYKSREAVKTAIGHGTLVTIFEISMFLVVTFIYGMLNEFGDI SATTVTFWSGTWNKDYRIIPVFGVLTLCLLILWFFNLLITGLITLFNVLFKEELITMII ISLAILFLPIVNVLGVEGSWMQYNPFSYVEIGNLVWGVQEHFFTEGAFSFREFGI SMSIALVSVYLLSYVVISVIEKV
orf00022	-	76	MRNLFDLNIKSATIKSIGENSSRQNLTVPTTGCNTVMTGYGSGVGGNTGGTRH PNPGCTLRTMICPAPFKNNNGNQ
orf00023	-	43	MKEDDKKVFYSSINRVVTLVCKEIKREYHFDQFKLSTLTEEVN
orf00025	-	77	MSIREEGKIMKDMFDLNVKSVTLHTVDKQISKNAVGGLGGVNPCIEPPSFGKIE PSTRHTNPGCTLKTMICPTIHRR
orf00026	UPF0342 protein SH1117 n=49 RepID=Y1117_STAHJ	110	MLEEFLTEFRKSKMLQNYKKYLKQVSNNEEASAILDEYRETTALFIERKKNGEDI PDDEMERYTDVTRMLTKKEILKDFIEYDTKVYFFIEELVDGLRDEVKEAFYDSYK

orf00027	-	80	MLSSIHNKKGGYHVFLTENKIIILLIISFLIILNAIIIYSVVKNKKEVKIFFIPIIIFHVFTM LGLVYILLFIMFIGVNS
orf00028	-	41	MVESIDAYIIQQEVMGITVESMTVALKDLQTPLVNLVIEHL
orf00030	Glycerol-3-phosphate acyltransferase	198	MINLLAMIIAYLLGSIPSGIWIGKLFYGKDIRNYGSGNSGTTNTFRVLGVPAGIAVF IIDTAKGAIPVLLPLVLATSVHPLLFGLVAVLGHTFPIFANFKGGKAVATTAGVGL GIYPVFVLTMAAVFATILFISSMVSLSSMLTVGIAILASLYIGDTIFTGTVVVLFLIIVI RHKSNIKRILDGTESTVPFGLRKDK

-; not mentioned by BAGEL

Appendix 122: Genes predicted to be involved in the production of bacteriocin DpG.

***D. pigrum* strain**	**Amino acid sequence**
*D. pigrum (*accession number MUYF01000003.1)	MKIRHAKQSDAFAVACVLIDSWRETYQGIFPQSLLDNLDQATQTQRMLTAIKMSNMLVLENDAQQLIGF IGGGNNRQTAQFPFIDGELYALYVSLSAQGRGYGAKLVEAFETLMKELGYKGLLVKCLKENVNAQAFY QHMGFSTIGDQMLTLQESSVAETVLAKSINP
D. pigrum ATCC 51524	MKIRHAKQSDAFAVACVLIDSWRETYQGIFPQSLLDNLDQATQTQRMLAAIQMSNMLVLENDAQQLIGF IGGGNNRQTAQFPFIDGELYALYVSLSAQGRGYGAKLVVAFKTLMKELGYKGLLVKCLKENTNAQAFY QHMGFSTIGDQMLTLQESSVVETVLAKSINP

Appendix 123: Summary of genes in BGC possibly coding for polyketide DpA in the whole genome shotgun sequence of *D. pigrum* ATCC 51524 as predicted by antiSMASH.

Type of gene	Locus tag	Location of gene	Amino acid sequence
Other gene	HMPREF9703_00316	354007	MFIKTGDKVQVITGKEKGKQGTVLKAIPRENRVIVEGLNIAKKHTRPSMESEGGIVETEAPIHVSNVQLVDPKSGEPTRVGFRFEDGKKVRYAKKTGEAI
Other gene	HMPREF9703_00317	354343	MIQTESRMKVADNSGAREVLVIKVLGGSGAKTANIGDEVVVTVKHATPGGVVKKGDVARAVIVRSKSGLRRKDGSYIKFDENACVIVREDKAPRGTRIFGPVARELRDNDFMRIISLAPEVL
Other gene	HMPREF9703_00318	354752 - 355015	MTERKERKQYVGKVVSDKMDKTITVEIATQKQHKKYKKRMKYSTKLKAHDEKNIAKEGDIVRIMETRPLSKEKRFRLVEVVEEAVVL
Other gene	HMPREF9703_00319	355052 - 355246	MKANELRELSHQELKDKEKEFKDELFNLRFQLATGQLEDTSRIKKVRQNIARVKTVLRQAELAQ
Other gene	HMPREF9703_00320	355236 - 355667	MLVPKRVKYRREHRGKMRGEAKGGKEIAYGQYGLQSLDSKWITNRQIESARIAMTRYMKRGGKVWIKIFPHKSVTAKGIGVRMGSGKGAPEKWVAPVKRGKIMFEVGGVSEEVAHEALRLASMKLPVRTRVVTREESGDAHEG
Other gene	HMPREF9703_00321	355671 - 356327	MGQKVNPHGLRVGVIQDWDAKWYADSDFSDKLHEDLAVRELIAKDLEEASVSRVEIERAANRINVSIHTAKPGMVIGKGGSEVDALRNKLSNLTNKRVHVNIIEVKKPDMDATLVAKSIAEQLENRISFRRAQKQAIQRALRAGAEGVRTQVAGRLNGADMARTESFSEGTVPLHTIRADIDYANVEADTTFGKIGVKVWIYKGEVLPEIEEDQKGGK
Other gene	HMPREF9703_00322	356334 - 356690	MASSRTEAHATARMVRIAPRKVRLVVDQIRGKDAAEAISILRFTNRGAAEAVEKVLKSAIANAEHNFDMNIENLVVSEAYANEGPTLKRFRPRAKGAASRINKRTSHITVVVSEKKEG
Other gene	HMPREF9703_00323	356712 - 356990	MARSLKKGPFVDEHLMKKVQAMDSDNKRVIKTWSRRSTIFPNFVGHTIAVYDGRKHVPVYVQEDMVGHKLGEFAPTRTFKGHSKTEKVTKKF
Other gene	HMPREF9703_00324	357037 - 357882	MAIKKYKATSNGRRNMTTSSQSDITTNKPEKSLLVSQKRGSGRNNAGKITVRHKGGGHKHKYRLVDFKRRKDGIRGIVKTIEYDPNRSANISLIQYEDGTKAYILAPKGIKVGQEIYSGPDADIKPGNALALKDIPVGTVVHNIELKPGKGGQLVRSAGASAQVLGKEGKYVLVKLPSTENRLILAECRATIGTIGNEQHELVRIGKAGRNRWKGIRPTVRGSVMNPNDHPHGGGEGRAPIGMPSPVSPWGKPTLGKKTRKGKKHSDKLIVRRRRTKKRKK
Other gene	HMPREF9703_00325	357924 - 358214	MQAQDIILRPVITEQSMADMELDKYTFEVDTRATKSQVKRAVKELFDVDVEKVNIMNTKPKPKRMGRYVGYTKKKRKAIVTLKPDSKEIEIFQTEE
Other gene	HMPREF9703_00326	358214 - 358837	MPKVTVFNQQGDANGEVSLNNDIFGIEPNENVLFDAIIMQRASQRQGTHAVKNRSAVRGGGRKPWRQKGTGRARHGSNTSPIWRGGGVIFGPTPRSYSYKLPKKVRRLAILSALSQKALDDEIIVIDELNFDQPKTKDFQHMLDHIGVERKALVVLEKENEFAKLSARNIEGVKVVAPDNVSVLDVVAHDDLILTKTALEAVEEALQ
Other gene	HMPREF9703_00327	358866 - 359498	MTKGILGKKVGMTQVFTETGELVPVTVIEAKPNVVLQVKTVETDGYNAVQLGFDDKRNVLSNQPEQGHVKKADTSPKRFIREIRDAELGDVEVGSEITVETFKQGDIIDVTGTSKGKGFQGVIKRHNQSRGPETHGSRYHRRPGSMGQAADPARVFKGKKLPGRMGGQTTTIQNLEIVRVDADKNVILVKGNVPGPKKSMVEIRSAKKAD

Other gene	HMPREF9703_00328	359542 - 359850	MAKQKIRIRLKAYEHRVLDQSATKIVETAKRTGAEVAGPVPLPTERKLFTIIRSPHKYKDSREQFEMLTHKRLVDILNPTPKTVDALTKLDLPSGVDIEIKL
Additional biosynthetic gene	HMPREF9703_00329	360294 - 361079	MKAYDIVKTARDINRISTRDVIQALCQEKFIEQFGDRLSGNDGAIIGGVGLANNTPLTIIGIEKGRTVEENIQCNFGSASPAGYRKASRLLKQANKFNRPVLTLINTPGAYAAPESEEAGIGEAIARNLLIMSELTVPTLAIILGEGGSGGAIALALADEVWMMELSVYAILSPEGFASILWKDAKRAPEAAELMKLTSKDLSELAIVDRVIKEVDNNGHSIDKTIILSQLTRQINEKMTQLAEQPVHERLKKRQERFRKF
Additional biosynthetic gene	HMPREF9703_00330	360294 - 361079	MGLFRKRKGIKLNKILEEKDQERLAHVPDDLVERCPSCRKMLLHKQIEADCCCPNCGYHMHFAAYDRIEWLVDVGSFSEWEAALQTTNPLDFPRYAEKMAQQREKTGLNEAVITGQALINQQAVAIGVMDSRYIMASMGTVVGEKITRLFERATAQRLPVVLYIASGGARMQEGILSLMQMAKISQAVQQHHQAGLFYLPILTHPTTGGVTASFAMQGDIILAEPDATIGFAGKRVIEQTIKAKLPLDFQTAERVLETGFIDRIVPRERQKSVIQTLIQIHAGDGVNESL
Additional biosynthetic gene	HMPREF9703_00331	361961 - 363325	MFKKVLVANRGEVAVRIIRVLREMGISSVAIYSSADKQALHTELADEAICIGPAKTLDSYGNPVAVISAALQMDCDAIHPGYGFLSEKSEFVDLCEEVGLTFIGPSSHVINQMGNKQHARETMRAAGVPCTPGSDGLVQTVEEAEQVAEVIGYPLMVKAADGGGGKGMRRVADASELAHKFSAAMMEAQAVYGNKDVYIEKIIAPAKHIEVQLLADTHGNVIHLGERDCSLQRNNQKVIEMAPATFLDASVREALCDAAVTAAKAIGYTNAGTIEFLVDEQQQFYFMEMNTRLQVEHSVTEMITGIDIVREQINIAMGQPLAITQEDVTFNGMAIECRLNAEDPKQSFRPASGYIERLILPSGGMGLRVESGVYPNYSLPPYYDSMIAKIIVHKPSRQETFKTLQRALVEVVVEGLVTNIELLEELAYSEEVLADQYHTKWLEDEFLPAWMMTD
Additional biosynthetic gene	HMPREF9703_00332	363327 - 363761	MSVLTAQEVMEIIPNRFPIFFIDAVDELIPGKRIVCRKNVTINEHVFQGHFPGEPVLPGVYIVEAMAQAGSIPLLKQEGFEGKTGYLGGLNKVKFRQKVVPGDVLRIEVDIIKQKKNAGIGRGRAFVGDKKVAEADMTFIIGAN
Additional biosynthetic gene	HMPREF9703_00333	363762 - 364256	MNYDQLLALVDKLDQSSLAYIRYEHDGSKVELSKEVPHSQPVSMPAETSSVSSPAIREATPTVPSVVSDDKLHNETSSDVSESVGVEVLSPMVGVAYLQPAPDKDPYVQVGDRVEAGDVVVIIEAMKIMTEIKAECSGIVVDILVANEELVEYDQPLIRIKEDN
Core biosynthetic gene	HMPREF9703_00334	364258 - 365511	MGQADKLNRVVVTGLGTISPLGNNVDEFWKKVRANESGIAPITKFDASEVGVHVAGEVKDFDPTLTMDRKEYKRMDLFCQYGIAASVEAVKMSGYDIAANASRVGTLISSGIGGLIEIENGIRKMIDKGPKRIPPLFVPLTIGNMAAGNISMKLGAKGISMDIVTACASSTNSIGEAFLKIQAGFLDACLAGGCEGTINEIGIGGFNALTALSTNEDPTKASRPFDKDRDGFVMGEGAGVLFLESLDSAQERGAHILAEIVGYGATSDAYHMTAPVPDGSGAGEAIKMALASAHITPEQVSYINAHGTSTPTNDSGETTAIKYALGDAAYNIPVSSSKGHFGHLLGAAGGIEAVTCVKALEDGFIPATLGLETSDEACDLDYVPQTGREADLQYVLSNSLGFGGHNAILCFKRWEGK
Core biosynthetic gene	HMPREF9703_00335	365511 - 366233	MSQVCVVTGSSRGIGLAIAKQLADDGHQVVLNSRSPLKPEVLEQFADAKLEVGTIVGDVSDFADAERMISAVKEQYGRIDVLVNNAGITRDGLVMRMKEEDFDQVIATNLKGCFNMARHTTPIMLKQRSGTIINVSSVSGIMGNAGQVNYAASKAGVIGLTKSLARELASRSITVNTIAPGFIETDMTAEMSERVTEAMLGEIPLKRFGQAEEVASAVKFLMENRYVTGQTIEVNGGLHI
Core biosynthetic gene	HMPREF9703_00336	366230 - 367177	MKVAVIFNGQGAQFEGMGLDFREHFPEARAVFNQASEATGLDMVQLVSEDFSKLRQTKYAQPAIGTVSLAIWASIREILPQVDYMAGLSLGEYAALMASGIFTVSDGLRLLFERGQVMSDVCKDIAEDEPMQMLAVIGMPREVVEHLVEDLPQTYLANFNSPEQIIL

			AGPKSSLKLFNQASKAAGYRKGLPLKVEGPFHTPLMAAACQPLEALLDSYELQPG CAPVISNTTVEPHDLETLKSTLVRHLIEPVQWEQTIDWLIQAEVTHLIQIGPGQTLQK LLKAHDQAPLCLAISQVEDVSEIEKFLNENKGEKE
Other gene	HMPREF9703 _00337	367226 - 368197	MKTAITELLGIRYPIIQGAMAWVADADLASAVSNAGGLGIVGTGHDSREVVKEKIDK MKELTDQPFAVNAMLLNPHIEEVIDYIIEESGVKIVTTGAGNPSQYMKRFQEAGIRVI PVVASVALAKRMERIGADAIVIEGMEAGGHIGRSTTLTLLPQVVEAVDIPVIAAGGF GNGESLAAALMLGAEAIQVGTRFVVSKESNAHQNFKQKILKANDIATVVTGQITGH PVRVLRNQLTTDYLELERLQTSEDEPDFSEMEKLGKGALRRAVVDGDIKQGSMMA GQIAGLVKKEETVAEMIQDYIDGAQQAYRRCQACFTQDDA
Core biosynthetic gene	HMPREF9703 _00338	368220 - 368444	MVFETVKDIIVEQLGIDESEVTKETDLENGLDADSLDIFQIISDIEDEYDITIDTDLNLQ TVGELVDYVEQLIG
Core biosynthetic gene	HMPREF9703 _00339	368485 - 369429	MSKVVATGQYLPANRLTNEQLIAQAGIDSSDEWIVQRTGVKARRFAEGESVADLAT EAARAIIANVGPDVKEDIRLIIVASMSSGNPTPSIANQVQANLGIAEAWGFDISGACS GFTMAVDIAERMSRTYESGYVLVIGADKMSQILDLSDRTSSIIFGDGAGGLLIACDG AGLPGYRSQLVALEDTKDAITLDKTVPDRQYLTMLGRDVFNFVVRAVIPGLAKFIDE LNGSYDYVLSHQANDRLLDVMSQKLGIDRQQIPANIAEVANTSSASIPILLDDLVQA GTITLSDQQQIVMVGFGGGLAYGINYFKL
Regulatory gene	HMPREF9703 _00340	369435 - 369896	MDHQTMNEVNEYLVMIFNEVLSIEEEAISRSEFSDLSVKEMHTIEAIGLSGDLNSAQ VAKRLDITPGTLTVSIQNLVKKGYVTRVKSETDRRVVKLALTKRGKLIYRLHHKFHM RMVEESLTGFDPEEAQVLIRGLRNLHEFLNDLKNQQRKE
Other gene	HMPREF9703 _00341	370214 - 371188	MTQSEKKHKKIILIGNGAVGSAYAYALVNQNIGQELGIIDLDPFKVEGDVMDLNSAL AFTAPKDIYVADYSDCADADLVVFTAGAGQKPGETRLDLIHKNLKITKTIVDQVMAS GFNGIFLVASNPVDILSYAVHKFSGMPAHKVVGSGTSLDSARFRIELAKRLNIDARN VHGYIIGEHGDTEFPVWSHANIAGLQITEWLETNPLEDEGELLEIFEAVRDQAYHIIE RKGATYYGIGAALAKITQAIFNNENSVLPLSVLLEGEYGHDDIYIGTPAIINDRGVKR AIEIPLNDYEQGKMNHSINTLRKFIQDAREKAPELGL
Other gene	HMPREF9703 _00342	371731 - 372999	MNIIDELQWRGAINQQTDEDGLKELTEKQSIALYCGVDPTGDSMHIGHLIPFIILKRF QMAGHKPVVLIGGGTGSIGDPSGRNSERQLQTKETVQHNVDKLSAQMRKLFKAG EAESGIRLTNNADWLSPLSFLDFLRDYGKEFNINTMLSKDVVASRLEQGISFAEFSY QIIQSIDFLHLYRHEDVQLQIGGSDQWGNITAGLDLIRKKEGHEAEAYGLTIPLLLKS DGTKFGKTAGGAIWLDPEKTSPYEFYQFWINQQDADVVKYLKYFTFLDQTEIAKLE QSLTDEPWKRQAHKALAEEMTRFVHGQAALDEAIQITEALFSGEIADLTAEQIEETF GDVPTSEATRGEHDIVEFLADITGICSSRREAREFIEQGAITISGEKVTDLDYTISSTD TFDDRFIIVRRGKKKYYRVNLTD
Other gene	HMPREF9703 _00343	373019 - 373438	MRLWHETLISDLPRQQLLGQHRECCALRGKGWNRPHATVQYVFDYSPYKLYQYH QLIMEEMKSRTYQPDERWEDPLYRGKACAPYRELEPVTPTKPIYPEHNATYLAECL ENLADKGIKLSVRMKQSEKIVDNKRGYPI
Other gene	HMPREF9703 _00344	373602 - 374657	MSKTNSILNLIGIQDENTKLIEKNPIKTSGPNEIIYATLTYQPQGCCKCGIKNDGSIIIK YGTKTSRITLSGLEKNRYLVLKKQRFQCKACGQTFIAQTSVVDKYCYIANQVKRKS LDLSRETICKTTNARLCHISEATVQRICNAEALKYKGYGQRLPSNIGIDEFRYRNQT MAFDYIDNDSGNILSILPNRRIKDIKDHFYNRYNLKERKKVQTVTTDLNAGYIHMIPE LFPNATIILDRFHLVQLMVRALDRTRIKVMNRLKKGNNTDQKRYRRLKRYGKMLKL

			KSKDISFTDYKLYYLFGYFTAQMIVDKLLAYDELLKRTWEVYQCFMRAVDQRDGVA MSTLIHTN
Other gene	HMPREF9703 _00345	375614 - 376216	MSDHYYTKEPTSTSNPEQFKEVVIGQTLSLTSDAGVFSRGQMDFGTRTLIEAVQTD KKCSGPLLDMGCGYGVVGIALAAANPERTIHMVDINERAVALAKQNAEVNAVENVII QQSSLFENIVERSFSAVVSNPPIRAGKQVVHGILEQSYDYLQPGGKLYIVIQKKQGA PSAKKKMQAVFGNVERIALEKGYWVLMAEK
Other gene	HMPREF9703 _00346	376362 - 376922	MTKTQRMIYISLLSAQAVILGLVENSIPFPFAFAPGAKLGVANLIVIIAIYTLPLKDSFK VLMMRLLMTTLLGGTLSTFMYSFMGALLSYLSMLLVKTLGCSQVSVIGVSATGGML HNLGQLLVASFIAQTWTVLMYLPVLSFFGILSGMAIGIAANYLMEHVHTVRRFRVEQ EQADQAKHSSRLS
Other gene	HMPREF9703 _00347	377028 - 377186	MRKIRSKQFNITDFHEKNLKPIKICIIHKILQKADIGKKIICLISIVIMIYL
Other gene	HMPREF9703 _00348	377227 - 377556	MIEDRDFLMRQIKQMIQNLGKILDRNTLRELLALSEDDMSNEELDSLYLMGQVDIIA AEKQLTPQHLSVDLGIDIDRLEDLFKGQAFLTEPEATNVQSFLETHDNSEPT
Other gene	HMPREF9703 _00349	377652 - 378605	MSTEPIFLTPVLQDKIWGGKKLQTEFGFDLPSDTVGEAWVISTHPHGESTVSSPEQ YAGMGLNELYHEYPELFGAQQPDTFPLLTKILDAKEDLSVQVHPDDEYGQEHEGE LGKTECWFVISADEDATIIYGHNARTEEEFRHLVEAGEWDELLREVPVKAGDFFYV PHGTIHAIKGGITILETQQNSDTTYRVYDYNRTDKDGNTRELHLEDSIAVSNIPHIDP AIQQQEDRVGMSAITHYLTNEYFSVYRWQIRDQLVVDLTGDYTLVTVLDGQGMIEI DQESYPVEKAQSFIIPHGVSSVTLSGDLDMIASNPE
Additional biosynthetic gene	HMPREF9703 _00350	378752 - 379756	MNTLNWGIIGTGGIAEDFAAGFQDTGITCHGVSSRSLEKAEAFKEAFGLHAAYGDY HELLQDEVIDIVYIATPHHTHAQIAADALQAGKHTVVEKPIVTQADDLAQLKQLAEAQ GVYLFEAMTTHYRPVYQTIRELIDQKDLGALKMIQVNFGSFKPKDAGYFFQKDLAG GALLDIGVYALNFVMEFLSTPPTEMKTLGYINEQFGVDESAGIVLKNDQNELATITM TFHAKLPKRGVIAFEGGYFEIDNYPRADRVTFTTPEGKTEHYQAGQSSQALQEEMI AITNLIASGTPQPHLTKTTHVIELMDAIRQEWNLTYPFDNHTLDAPNREKA

Appendix 124: Genes predicted to be involved in the production of polyketide DpA.

***D. pigrum* strain**	**Amino acid sequence**

*D. pigrum (*accession number MUYF01000003.1)	MKVAVIFNGQGAQFEGMGLDFREHFPEARAVFNQASEATGLDMVQLVSEDFSKLRQTKYAQPAIGTVSLAIWA SIRETLPQVDYMAGLSLGEYAALMASGIFTVSDGLRLLFERGQVMSDVCEEIAEDEPMQMLAVIGMSREVVEHL VEDLPQTYLANFNSPEQIILAGPKSNLKLFNQAAKAAGYRKGLPLKVEGPFHTPLMAAACQPLEVLLDSYELQPG CAPVISNTTVEPHDLETLKSTLVRHLIEPVQWEQTIDWLIQAKVTHLIQIGPGQTLQKLLKAHDQAPLCLASQVED VSEIEKFLNENKGEKE
D. pigrum ATCC 51524	MKVAVIFNGQGAQFEGMGLDFREHFPEARAVFNQASEATGLDMVQLVSEDFSKLRQTKYAQPAIGTVSLAIWA SIREILPQVDYMAGLSLGEYAALMASGIFTVSDGLRLLFERGQVMSDVCKDIAEDEPMQMLAVIGMPREVVEHLV EDLPQTYLANFNSPEQIILAGPKSSLKLFNQASKAAGYRKGLPLKVEGPFHTPLMAAACQPLEALLDSYELQPGC APVISNTTVEPHDLETLKSTLVRHLIEPVQWEQTIDWLIQAEVTHLIQIGPGQTLQKLLKAHDQAPLCLASQVEDV SEIEKFLNENKGEKE

Appendix 125: Summary of genes in BGC possibly coding for bacteriocin DpF as predicted by BAGEL.

orf	Biosynthetic assembly	Domain	Start of sequence	End of sequence	Sequence
orf_334	KS	Ketosynthase	364257	365511	MGQADKLNRVVVTGLGTISPLGNNVDEFWKKVRANESGIAPITKFDASEVGVHVAGEVKDFDPTLTMDRKEYKRMDLFCQYGIAASVEAVKMSGYDIAANASRVGTLISSGIGGLIEIENGIRKMIDKGPKRIPPLFVPLTIGNMAAGNISMKLGAKGISMDIVTACASSTNSIGEAFLKIQAGFLDACLAGGCEGTINEIGIGGFNALTALSTNEDPTKASRPFDKDRDGFVMGEGAGVLFLESLDSAQERGAHILAEIVGYGATSDAYHMTAPVPDGSGAGEAIKMALASAHITPEQVSYINAHGTSTPTNDSGETTAIKYALGDAAYNIPVSSSKGHFGHLLGAAGGIEAVTCVKALEDGFIPATLGLETSDEACDLDYVPQTGREADLQYVLSNSLGFGGHNAILCFKRWGK
orf_336	Mal	Acyltransferase	366229	367177	MKVAVIFNGQGAQFEGMGLDFREHFPEARAVFNQASEATGLDMVQLVSEDFSKLRQTKYAQPAIGTVSLAIWASIREILPQVDYMAGLSLGEYAALMASGIFTVSDGLRLLFERGQVMSDVCKDIAEDEPMQMLAVIGMPREVVEHLVEDLPQTYLANFNSPEQIILAGPKSSLKLFNQASKAAGYRKGLPLKVEGPFHTPLMAAACQPLEALLDSYELQPGCAPVISNTTVEPHDLETLKSTLVRHLIEPVQWEQTIDWLIQAEVTHLIQIGPGQTLQKLLKAHDQAPLCLASQVEDVSEIEKFLNENKGEKE
orf_338	T	Thiolation	368219	368444	MVFETVKDIIVEQLGIDESEVTKETDLENGLDADSLDIFQIISDIEDEYDITIDTDLNLQTVGELVDYEQLIG