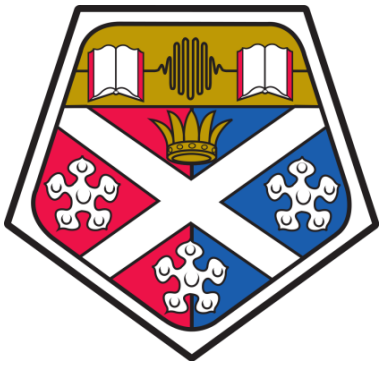


University of Strathclyde

Glasgow, United Kingdom

Strathclyde Institute of Pharmacy and Biomedical Sciences

Heavy Metal Inducible Antimicrobial Activity of *Streptomyces* spp. Isolated from the Leadhills and Wanlockhead Lead Mines in Scotland, UK

By

Talal Sabhan Salih

Thesis presented in fulfilment of the requirement for the degree of

Doctor of Philosophy (PhD)

2017

Declaration of Authenticity and Author's Rights

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed:

Date:

Acknowledgements

I would firstly like to sincerely thank Dr. Nicholas Tucker, my supervisor, for his excellent supervision and for looking after me with great support and patience during the hardest of my PhD study. I can't thank you enough Dr. Nick for everything you have done to help me complete this adventure over the last four years.

I am also thankful to Dr. Paul Hoskisson, Dr. Paul Herron, Dr. Iain Hunter, Dr. Arnaud Javelle Dr. Katherine Duncan and Dr. Gareth Westrop for their invaluable advice, support and encouragement.

I am very thankful for the Iraqi government/Ministry of Higher Education and Scientific Research/University of Mosul for awarding me a scholarship allowing me such an opportunity to gain knowledge and experience as well as achieve my life's dream to study for a PhD in the United Kingdom.

I am grateful to the past and present members of the microbiology group for their friendship, help and all the stimulating discussions, Leena, Alison, Kirsty, Sara, Sarah, Cezar, Florence, Emilio, James, Andy, Gaetan, Anna, Lis, Rebecca, Tiago, Gabriel, John, Walid and Ainsley. Profound thanks to Jana, who was an infinite source of knowledge and encouragement throughout the study (Vielen Dank Dr. Jana).

Loving thanks also go to my wife Hind. She has always been standing beside me and is responsible for so many my dreams coming to life. And to my wonderful daughters Reem and Yasameen, for always bringing smiles to my face and for helping me remember what is important in life. I would especially like to thank my parents for their decades of love, teaching me ethics and determination. I am greatly indebted to them.

Abstract

There is a real and urgent need for new antimicrobial compounds to combat the rise of multidrug resistant infections. Whole genome sequencing and genome mining have revealed that *Streptomyces* spp. possess a range of biosynthetic gene clusters for secondary metabolites, however many of them are silent and cryptic under conventional laboratory conditions. It has also been shown that triggering the expression of these cryptic gene clusters in *Streptomyces* might yield novel antibiotics. Here, we isolated over 100 *Streptomyces* strains from sediments contaminated with heavy metals from a former industrial site in Scotland, United Kingdom. These strains were assayed for heavy metal dependent antimicrobial activity. We have used a combination of genomics, bioinformatics and transcriptomics to investigate these novel strains, providing a phylogenetic context and molecular evidence towards the discovery of the gene clusters responsible for antibiotic biosynthesis.

Our findings highlight the potential of using heavy metals for activation of silent biosynthetic metabolite gene clusters in *Streptomyces* isolated from extreme environments for natural product discovery. These findings are supported by the fact that these compounds are produced only in the presence of sub-inhibitory concentrations of heavy metals but not in the absence of metal induction. Whole genome sequencing has enabled us to investigate the taxonomy and biosynthetic capacity of the strains whilst transcriptomic analysis and bioinformatics tools are allowing us to investigate the changes in the transcript levels of these genes that occur under metal-inducing conditions.

Table of Contents

List of Symbols and Abbreviations

%:	percentage
°C:	Celsius
ACT:	Actinorhodin
Ah:	Silver
ANI:	Average Nucleotide Identity
antiSMASH:	antibiotics & Secondary Metabolite Analysis SHell
ATCC:	American Type Culture Collection
AU:	Adenine Uracil
BacMet:	Antibacterial Biocide & Metal Resistance Genes
BGCs:	Biosynthetic Gene Clusters
BLAST:	Nucleotide Basic Local Alignment
bp:	base pair
cDNA:	Complementary DNA
CDS:	Coding Sequencing
CF:	Cluster Finder
Co:	Cobalt
Conc.:	Concentration
Cu:	Copper
CV:	Coefficient of Variation
DDH:	DNA-DNA Hybridisation
DNA:	Deoxyribonucleic Acid
dNTP:	Deoxynucleotide Triphosphate
DSMZ:	Deutsche Sammlung von Mikroorganismen und Zellkulturen (German Collection of Microorganisms and Cell Cultures GmbH)
EDTA :	Ethylenediaminetetraacetic acid
E-Gel:	Electrophoresis System Gel
embl:	European Molecular Biology Laboratory
Fe:	Ferrous
FPKM:	Fragments Per Kilobase of transcript per Million fragmentmapped

g:	gram
GB:	Gigabyte
gbk:	Genbank
GC:	Guinine Cytocine
GlcNAc:	N- acetylglucosamine
HCl:	Hydrochloric Acid
HDAC:	Histone Deacetylase
Hg:	Mercury
HSP:	Highest point on the shoulders
***is*DDH:**	*in silico* DNA-DNA Hybridisation
ISP:	International *Streptomyces* Project
ISPs:	Ion Sphere Particles
L:	litre
LB:	Luria Bertani
leBIBI:	Bioinformatics Bacterial Identification Tool
LPSN:	List of Prokaryotic names with Standing in Nomenclature
Mbp:	Million Base Pair
MEGA:	Molecular Evolutionary Genetics Analysis
mg:	milligram
ml:	millilitre
MLST:	Multilocus Sequence Typing
mM:	millimolar
MM:	Minimum Medium
MRSA:	Methicillin-resistant *Staphylococcus aureus*
MTC:	Maximum Tolerance Concentration
MUSCLE:	Multiple Sequence Comparison by Log Expectation
NaBu:	Sodium Butyrate
NaOH:	Sodium Hydroxide
NCBI:	National Centre for Biotechnology Information
ng:	nanogram
NGS:	Next Generation Sequencing

Ni:	Nickel
NJ:	Neighbour Joining
NRPs:	Nonribosomal Peptides
OD:	Optical Density
OSMAC:	One Strain−Many Compounds
OT2:	OneTouch 2
Pb:	Lead
PBS:	Phosphate Buffered Saline
PCA:	Principal Component Analysis
PCR:	Polymerase Chain Reaction
pg:	Picagram
PGM:	Personal Genome Machine
PKs:	Polyketides
PM:	Phenotype Microarray
pmol:	Picamolar
RAST:	Rapid Annotation Subsystem Technology
RIN:	RNA Integrity Number
RiPPs:	Ribosomally synthesized and Post-translationally modified Peptides
RNA:	Ribonucleic acid
rRNA:	ribosomal Ribonucleic Acid
SAM:	S-adenosylmethionine
SDS:	Sodium Dodecyl Sulphate
SMs:	Secondary Metabolites
SSU:	Small Subunit
VRSA:	Vancomycin-resistant *Staphylococcus aureus*
w/v:	weight/volume
WGS:	Whole Genome Sequencing
WHO:	World Health Organisation
Zn:	Zink
µl:	microliter

Chapter 1:

Introduction

1. Introduction

There is a rapid and global spread of severe antibiotic-resistant microbial pathogens, such as Gram-positive methicillin-resistant *Staphylococcus aureus* (MRSA), vancomycin-resistant *S. aureus* (VRSA) (Fraise, 2002; Woodford and Livermore, 2009), and vancomycin-tolerant *Enterococcus* strains (Akins & Rybak , 2001; Saribas & Bagdatli, 2004), in addition to *Pseudomonas aeruginosa*, one of the most difficult bacteria to treat, due to its distinctive nature for antibiotic resistance (Overhage *et al.*, 2008). The current emergence in the number of new pathogens which are multi-drug resistance has received increasing attention. These pathogens also include the Gram-negative New Delhi metallo-beta lactmase (NMD-1), carbapenemase producing Enterobacteriaceae expressing enzymes such as KPC-2 (*Klebsiella pneumoniae* carbapenemase-2) (Brink *et al.*, 2012; Bushnell *et al.*, 2013; Munoz-Price *et al.*, 2-13) and *Acinetobacter baumannii* strains which have been reported resistant to all known antibiotics that have been reported (Pleg *et al.*, 2008). The most recent issue is the emergence to MCR-1 (*Escherichia coli* strain SHP45) which has been found to be resistant of the last group of antibiotics, polymyxins, by plasmid mediated resistance (Liu *et al.*, 2016). The World Health Organisation (WHO) has expressed serious concerns and raised awareness of emerging antibiotic resistant bacteria (WHO, 2001; Leung *et al.*, 2011). The WHO has just published its first ever list of antibiotic-resistant "priority pathogens" which includes 12 families of bacteria that pose the greatest threat to human health (WHO, 2017). This highlights that much effort is needed to address the growing global resistance to antimicrobial medicines and that there is urgent demand to seek new (novel) chemical compounds of high medical value.

The family Actinomycetes is well-known as a rich source of bioactive chemical compounds with great potential for medical and pharmaceutical applications. Genera of this family such as *Streptomyces, Saccharopolyspora and Salinispora* were underestimated for years based on traditional isolation approaches. However, more modern techniques, such as, genome sequencing (Bentley *et al.*, 2002; Nett *et al.*, 2009; Harrison and Studholme, 2014;) and genome mining using *in silico* bioinformatics tools

(Medema *et al.*, 2011; Blin *et al.*, 2013; Cimermancic *et al.*, 2014; Weber *et al.*, 2015) of some members of these genera have shown that there are many silent gene clusters. These genes may encode candidates for novel drugs and antimicrobials.

1.1 Actinomycetes family

Actinomycetales, commonly called actinomycetes, are Gram positive or Gram variable bacteria that belong to the phylum Actinobacteria which is one of the largest phyla in the domain Bacteria, as deduced from its branching pattern which has been based on the 16S rRNA phylogenetic tree (Goodfellow, 2012). Actinomycetes have relatively standard nutritional requirements. They are chemo-organotrophs, possess a fermentative carbohydrate metabolism and grow as aerobes, facultative anaerobes or obligate anaerobes (Goodfellow *et al.*, 1983). The major constituents of their cell wall are glucosamine, muramic acid, glutamic acid, and alanine (Lechevalier, *et al.*, 1966), and wall teichoic acids are widespread among their members (Streshinskaya *et al.*, 2002). The GC content of the DNA of actinomycetes is generally in the range of 65-75%, although certain thermophilic actinomycetes have a low GC content (44-54%) (Queener & Day, 1986). Since they are producing a wide range of industrially important compounds, they are of special interest in research. They are a widely distributed group of microorganisms in nature which primarily inhabit soils, sediments and water (Takizawa *et al.*, 1993). Some grow well at neutral pH, however others are acidophiles, alkaliphiles, halophiles or thermophiles (Goodfellow, 2012). Moreover, some can tolerate and grow in extremely high levels of heavy metals (Amoroso *et al.*, 1998; Ibrahim *et al.*, 2011; Moraga *et al.*, 2013).

1.2 The family of Streptomycetaceae

The family streptomycetaceae was proposed in 1943 by Waksman and Henrici. Originally, this family had only two genera: the genus *Streptomyces* and the genus *Micromonospora* (Stackebrandt *et al.*, 1997). However, *Streptacidiphilus*; an additional genus to the family; has been established by Kim *et al.* (2003), which included acidophilic species isolated from low pH soils and litter. Accordingly, the family now has three genera: the genus *Streptomyces*, the genus *Kitasatospora* as well as the genus *Streptacidiphilus* as indicated in the 2nd edition of the Bergey's manual of systematic bacteriology (Kampfer, 2012). The family Streptomycetaceae has been composed of 614 taxa based on 16S rRNA gene sequences (Labeda *et al.*, 2012), where *Streptomyces*, *Kitasatospora* and *Streptacidiphilus* contain 583, 23 and 8 species respectively.

Streptomyces species have a complex life cycle and tend to grow slowly and produce vegetative hyphae, 0.5-2.0 µm in diameter, which form an extensively branched mycelium with rarely fragments. The hyphae mature to form chains of three or more non-motile spores, in which uni-genomic spores germinate to produce a multigenomic substrate mycelium of branching hyphae which give rise to aerial hyphae and finally to spores. A few species produce spores on substrate mycelium (HPA, 2012). As nutrients start to deplete, the aerial hyphae develop marking the beginning of the morphological differentiation. Ultimately, sporulation takes place, leading to septation of the hyphae into chains of spores that mature and disperse (Piette *et al.* 2005).

Secondary metabolism occurs during the late stages of vegetative growth (Bérdy 2005). The expression of different specialised metabolites is influenced by environmental factors like the presence or lack of certain nutrients or the metabolism of other microorganisms in the same niches. Antibiotic metabolites in particular are expressed as a way to compete and defend against adjacent bacteria (Goodfellow *et al.*, 1983; Hopwood, 2007). Production of these antibiotics is encoded by genes distributed in large clusters within the chromosome or on plasmids. The presence of many biosynthetic clusters means members of the same species can lead to the biosynthesis of different antimicrobial products (Liu *et al.* 2013).

1.3 Actinomycetes Infections and diseases

Actinomycetoma is a subcutaneous tissue disease, commonly affecting the foot (Dunne *et al.*, 1998). The disease is endemic in tropical, subtropical, and temperate regions of the world. Although the highest incidence occurs in Asia, India, Yemen, and Pakistan, it is frequent in the countries of Sudan, Mauritania, and Senegal in Africa. It is also present in most of North, Central, and South America, with the highest incidence in Mexico and Venezuela (Welsh *et al.*, 2012; Mattioni *et al.*, 2013). Actinomycetoma is most commonly caused by the genus *Actinomadura*, followed by *Nocardia* and *Streptomyces* species (Hazra *et al.*, 1998). Among the genus *Streptomyces*, *S. sudanensis* (Quintana *et al.*, 2008) and *S. somaliensis* (Fahal and Sabaa, 2010) have been reported the most common etiologic agent causing human infections. Recently, the first case of human cervicofacial actinomycetoma caused by *Streptomyces griseus* was reported (Chander *et al.*, 2013). However, the sequenced genomes of this bacterium should allow identification and diagnosis of the infectious agent, and ensure targeted, rapid treatment regimes (Kirby *et al.*, 2012)

In comparison to other bacteria, it has been found that a few members of actinomycetes are plant pathogens. Nevertheless, they appear to be major pathogens of certain crops in particular areas and under special conditions (Locci, 1994). For instance, *Streptomyces scabies* is one of the streptomycetes that cause common scab symptoms on potatoes and other root crops (Joshi *et al.*, 2007).

1.4 Antibiotic Production in *Streptomyces*

Actinomycetes are the major antibiotic producers in the pharmaceutical industries. Two-thirds of natural antibiotics have been isolated from actinomycete bacteria (Newman *et al.*, 2003). Among actinomycetes, members of the genus *Streptomyces* are exploited for the production of commercially significant and bioactive molecules (Goodfellow, 1988; Bérdy, 2005). Over 7600 bioactive metabolites derived from

Streptomyces have been recognised (Bérdy, 2005). These metabolites include antibiotics, immunomodulators, anticancer drugs, antiviral drugs, herbicides, and insecticides (Bibb, 2005; Hopwood, 2007, Table 1.1).

There have been several approaches and many studies have been carried out to maximise secondary metabolite yields in actinomycetes by means of genetic manipulation of biosynthetic pathways. Improving oxytetracycline production in *Streptomyces rimosus* M4018 by metabolic engineering of the G6PDH gene in the pentose phosphate pathway; improving antibiotic production in *Streptomyces coelicolor* A3 by the engineering of N-acetylglucosamine metabolism; and increasing production of antitumor mithramycins in *Streptomyces argillaceus* by engineering precursor metabolite (Tang *et al.*, 2011; Świątek *et al.*, 2012; Zabala *et al.*, 2013) are good examples of such approaches. Studies have revealed that N- acetylglucosamine (GlcNAc), the monomer of chitin and constituent of bacterial cell wall peptidoglycan, and the related amino acid glutamate are highly preferred carbon and nitrogen sources for streptomycetes (van Wezel et al, 2006; Świątek *et al.*, 2012). Rigali *et al.* (2008) have shown that under famine conditions, GlcNAc addition stimulated antibiotic production in *S. clavuligerus, S. collinus, S. griseus, S. hygroscopicus* and *S. venezuelae* grown on MM agar (5 mM GlcNAc or higher) by functioning as an allosteric effector of the pleiotropic transcriptional repressor DasR, which controls the GlcNAc transport and metabolism as well as antibiotic production.

A study by Moore *et al.*, 2012 has shown that adding sodium butyrate to MSF agar induced expression of secondary metabolic gene clusters in *S. coelicolor* A3(2) strain M145 by stimulating actinorhodin (ACT) pigment production and antibiotic production in *Pseudonocardia, Saccharopolyspora* and *Amycolatopsis* as non-model actinomycetes. Sodium butyrate belongs to a group of histone deacetylase inhibitors (HDAC) which have a vital role in controlling and regulating gene expression in eukaryotes by antagonising the regulatory acetylation of histone proteins (Munshi *et al.*, 2009). Thus, the study shows the potential use of HDAC inhibitors to increase antibiotics yield or trigger the silent genes in *S. coelicolor* as well as in other bacteria.

A few studies have focused on exploiting metal stress for new antibiotics and secondary metabolites production from *Streptomyces*. It has been shown that new secondary metabolites with potential medical uses might be discovered by triggering cryptic gene clusters in *Streptomyces* by utilisation of heavy metals (Amoroso et al, 2013). Haferburg *et al.* (2009) investigated the potential effects of sub-inhibitory nickel and cadmium concentrations on new metabolite production. The researchers found that two out of ten stains isolated from heavy metal contaminated areas could produce new metabolites after being exposed to the above two mentioned heavy metals. Moreover, one of these two strains, *S. tendae* F4, produced a new class of isoflavones (Ueberschaar *et al.*, 2011), which have many health benefits, including protection against cancer.

Table 1.1: bioactive compounds produced by *Streptomyces* and their applications.

Bioactive compound	Strains	Application
Actinomycin	*S. antibioticus*	Anticancer
Adriamycin (doxorubicin)	*S. peucetius*	Anticancer
Amphotericin (natamycin)	*S. nataensis*	Antifungal
Avermectin	*S. avermitilis*	Antiparasitic (worms and insects)
Avoparcin	*S. candidus*	Growth promotion of farm animals
Bialaphos	*S. hygroscopicus*	Herbicide
Bleomycin	*S. verticillus*	Anticancer
Candicidin	*S. griseus*	Antifungal
Chloramphenicol	*S. venezuelae*	Antibacterial
Chlortetracycline	*S. aureofaciens*	Antibacterial
Clavulanic acid	*S. clavuligerus*	Antibacterial
Fosfomycin	*S.wedmorensis*	Antibacterial
Kanamycin	*S. kanamyceticus*	Antibacterial
Kasugamycin	*S. kasugaensis*	Antifungal (Herbicide)
Lincomycin	*S. lincolnensis*	Antibacterial
Mitomycin	*S. caespitosus*	Anticancer
Monensin	*S. cinnamonensis*	Growth promotion of farm animals
Neomycin	*S. fradiae*	Antibacterial
Nikkomycin	*S. tendae*	Antifungal, insecticide
Novobiocin	*S. niveus*	Antibacterial
Nystatin	*S. noursei*	Antifungal
Oleandomycin	*S. antibioticus*	Antibacterial
Oxytetracycline	*S. rimosus*	Antibacteria
Polyoxin	*S. cacoi*	Antifungal (Herbicide)
Pristinamycin	*S. pristinaespiralis*	Antibacterial

Rapamycin	*S. hygroscopicus*	Immunosuppression
Streptothricin	*S. lavendulae*	Growth promotion of farm animals
Streptomycin	*S. griseus*	Antibacterial
Tacrolimus (FK506)	*S. hygroscopicus*	Antibacterial
Tetracycline	*S. aureofaciens*	Antibacterial
Thienamycin	*S. cattleya*	Antibacterial
Tylosin	*S. fradiae*	Growth promotion of farm animals
Virginiamycin	*S. virginiae*	Growth promotion of farm animals

Note. Table adopted from Hopwood (2007).

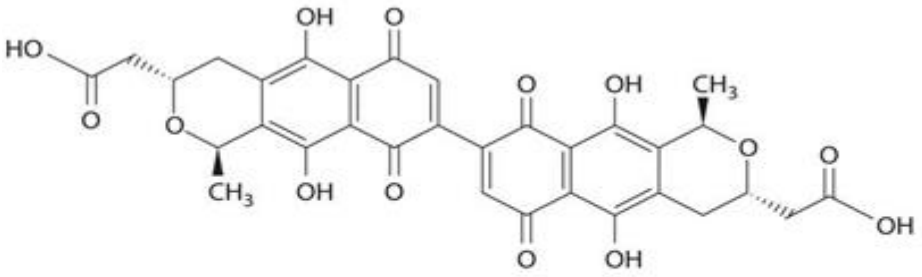
1.5 Bioactive Pigments from *Streptomyces*

Thousands of microorganisms producing bioactive compounds have been isolated and identified. However, some are producing commercially active pigmented molecules that have biological functions. *Streptomyces* is distinctly observed to produce a variety of extra- and intracellular pigments with potent biological activities (Table 1.2).

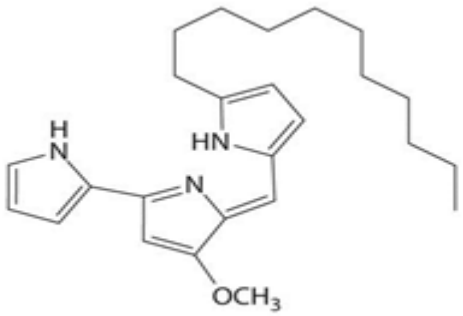
Actinorhodin (ACT) (Figure 1.1) belongs to the benzoisochromanequinone (BIQ) antibiotic family, a class of aromatic polyketides (Liu *et al.*, 2013). ACT is the best model compound for studying polyketide antibiotics synthesis (Figure 1.2) as its entire biosynthetic gene (*act*) cluster has been cloned (Okamoto *et al.*, 2009).

Table 1.2: biologically active pigmented compounds produced by *Streptomyces.*

Pigment	**Activity**	***Streptomyces* strain**	**Reference**
Actinorhodin	Antibiotic	*Streptomyces coelicolor*	Hopwood (2007)
Undecylprodiginines	Antibiotic	*Streptomyces coelicolor*	Hopwood (2007)
Undecylprodigiosin	Anticancer	*Streptomyces rubber*	Soliev *et al.* (2011)
5,10-dihydrophencomycin methyl ester	Antibiotic	*Streptomyces sp*.	Soliev *et al.* (2011)
Fridamycin D, Himalomycin A, Himalomycin B	Antibacterial	*Streptomyces sp.* B6921	Soliev *et al.* (2011)
Chinikomycin A and Chinikomycin B, Manumycin A	Anticancer	*Streptomyces sp*. M045	Soliev *et al.* (2011)
Melanin	Antioxidant	*Streptomyces sp*.	Dharmik & Gomashe (2013)



Actinorhodin (ACT)



Undecylprodiginines (RED)

Figure 1.1: Actinorhodin (ACT) and Undecylprodiginines (RED) pigments produced by *Streptomyces coelicolor* A3 (2).

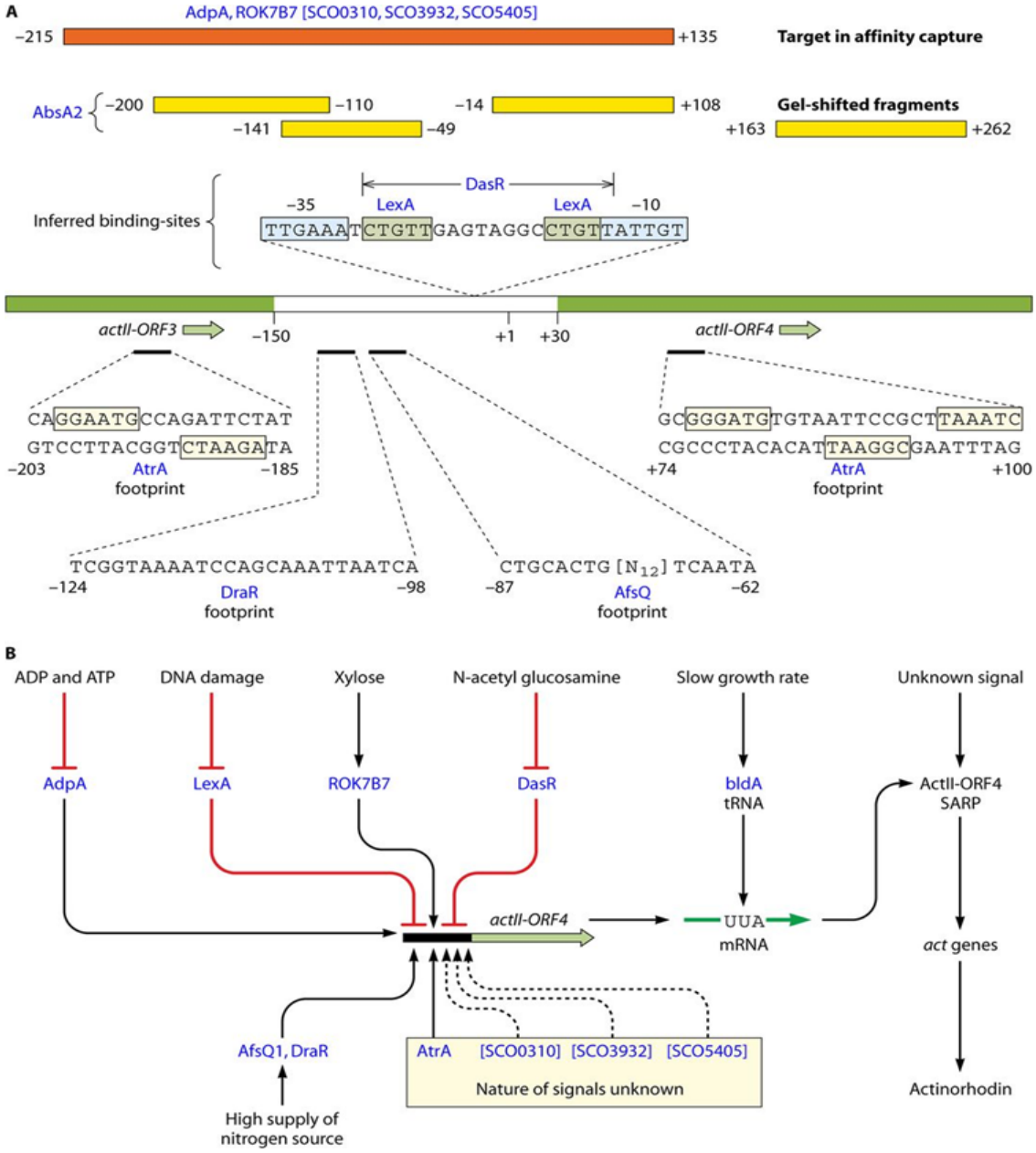


Figure 1.2: (A) Regulatory proteins of the promoter region of *actII-ORF4*, coding sequences shown in green and the noncoding intergenic region in white. Diverse regulators given in red and gel-shifted fragments given in yellow. (B) Nature of signal inputs influencing Actinorhodin production. The figure here is reproduced from Liu *et al.*, 2013.

Streptomyces also synthesise and produce melanin, which is considered an important criterion for morphological and taxonomical studies. Melanin positive *Streptomyces* strains can biosynthesise tyrosine compounds into L-Dopaquinone (Figure 1.3) and the latter is converted into melanin by the fermentative oxidation of tyrosinases enzymes (Faccio *et al.*, 2012). The diffusible pigment of melanin can be clearly observed on tyrosine-containing media. There are three main types of melanin with slight difference in colour: eumelanins, black to brown; pheomelanins, yellow to red and the dark colour allomelanins (Plonka & Grabacka, 2006).

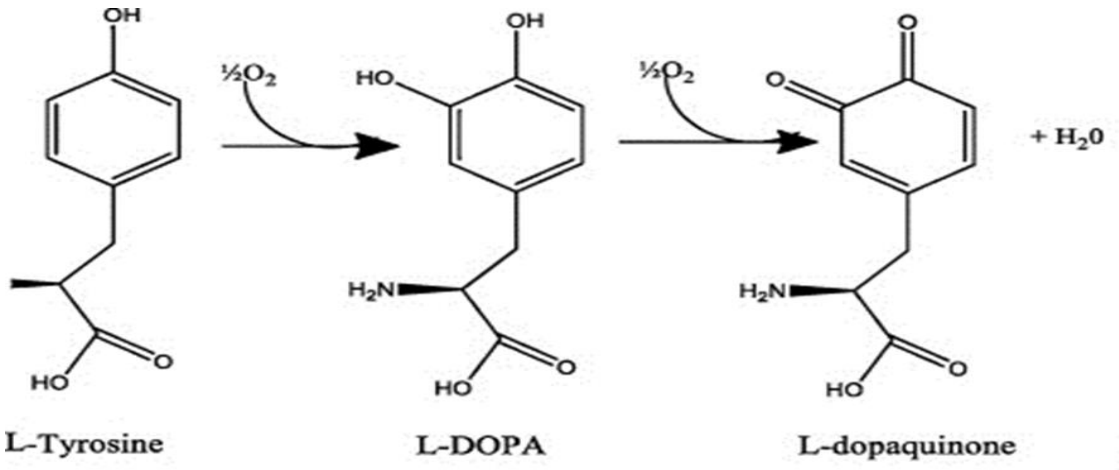


Figure 1.3: Biosynthesis of L -dopaquinone by oxidation of tyrosine with tyrosinase (Faccio *et al.*, 2012).

1.6 New *Streptomyces* strains and new bioactive compounds

Since the rate of discovery of novel antibiotics has decreased in recent years due to the failure to discover new metabolites from existing environments, actinomycetales from uncommon habitats require to be screened for new bioactive compounds. In particular, actinomycetes strains from adverse environments such as extreme temperatures, pH and those high in heavy metal or salt concentration need to be studied as they may be a source of novel antibiotics. The environment can affect the metabolism and antibiotic production of a microbe and therefore, extreme environments are considered to be the most diverse habitats for microbial isolation (Gunatilaka, 2006). Moreover, it has been estimated that less than 1% of the microbial world has so far been cultivated and characterised (Molinari, 2007). This means that the vast majority of microbes remain to be discovered. However, research in this area has intensified in order to isolate new strains of actinomycetes, notably streptomycetes, and several research studies have recently been reported. For instance, a new *Streptomyces* sp. exhibiting activity against methicillin-resistant *Staphylococcus aureus* isolated from a new environment, tropical rainforest, was investigated by Higginbotham and Murphy (2010). A novel *Streptomyces* spp. and secondary metabolite were obtained from a hypersaline estuary, the Laguna Madre (Texas) in the USA (Espinoza *et al.*, 2013). There was also a new thermophilic strain of *Streptomyces* spp. designated PAL114, producing antimicrobial compounds, isolated from a Saharan soil in Ghardaïa, Algeria (Aouiche *et al.*, 2013). Metal-tolerant *Streptomyces* isolates showed the ability to grow in the presence of boron compounds in the Province of Salta in Argentina (Moraga *et al.*, 2013). Additionally, a novel *Streptomyces graminifolii* sp. nov., was isolated from bamboo (Sasa borealis litter (Lee & Whang, 2014). Furthermore, two novel strains of *Streptomyces* were isolated from marine sediments in Thailand (Phongsopitanun *et al.*, 2014). Streptanoate, a new anticancer butanoate produced by *Streptomyces* sp. DC3 (Noomnual *et al.*, 2016); Iminimycin A, the new iminium metabolite was isolated from *Streptomyces griseus* OS-3601 (Nakashima *et al.*, 2016). Potent microbial compounds against Gram negative bacteria produced by *Streptomyces cyaneofuscatus*, which was isolated from fresh water sediment (2017); and just recently, searchers at the University of East Anglia in

the UK have identified a new potential antibiotic in fighting MRSA and VRE produced by a new species of *Streptomyces* (proposed name *S. formicae*) (Qin *et al.*, 2017).

1.7 Typing of *Streptomyces* based on 16S rRNA gene

Many genomic regions other than 16S rRNA have also been used in an attempt to determine the phylogenetic relationships among bacteria. For instance, the entire-genome analysis has been used. However, because the genomes vary in size, and gene duplication, gene transfer, gene deletion, gene fusion, and genes splitting are common; it was observed that this method is quite complicated, and that the phylogenetic trees based on entire-genome analysis are similar (Bansal & Meyer, 2002). Although the data presented by Rong *et al.* (2009) shows good congruence of the *Streptomyces albidoflavus* group using partial sequences of the five housekeeping genes and all of the species that they believed to be later heterotypic synonyms of *Streptomyces albidoflavus*, including *Streptomyces sampsonii*, *Streptomyces champavatii*, *Streptomyces odorifer*, *Streptomyces felleus*, *Streptomyces griseus* subsp. *solvifaciens*, *Streptomyces limosus*, *Streptomyces coelicolor*, *Streptomyces canescens* and *Streptomyces globisporus* subsp. *caucasicus*, are also located in Clade 112 of the study by Labeda *et al.* (2012) which has 99% bootstrap support (Labeda *et al.*, 2012). The data presented by Guo *et al.* (2008) and Rong and Huang *et al.* (2010) for the *S. griseus* group has good agreement with the MLST and the 16S rRNA gene phylogeny study by Labeda *et al.* (2012). However, their MLST studies of the phylogenetic tree of many strains is lacking a clear resolution in the segment of 16S rRNA, resulting in practically no bootstrap support for many of the branches present between clade 34 and clade 35 region (Labeda *et al.*, 2012). Nevertheless, 16S rRNA has been demonstrated to be a robust molecular tool for positioning and clustering of the streptomycetaceae family and illustrating their diversities within the phylogenetic tree (Chun & Goodfellow, 1995; Labeda & Kroppensted, 2000; Kumar *et al.*, 2007; Labeda *et al.*, 2012).

1.8 Phenotype Microarray (PM) using Biolog Omnilog system

The PM technique is a high throughput system that uses tetrazolium redox dye chemistry to produce a colour change in response to microorganism respiration. Tetrazolium, is a colourless water-soluble molecule that is readily reduced to form a purple colour of formazan in the presence of NADH as an electron donor (Bochner *et al*2001; Bochner, 2009, Figure 1.4). The purple colour can be quantified and kinetic information can be provided using the Biolog OmniLog instrument which uses a CCD camera to record the reduction of a tetrazolium dye at various intervals throughout the user-defined incubation period and subsequently the intensity of colour change is proportional to bacterial growth (Shea *et al.*, 2012).

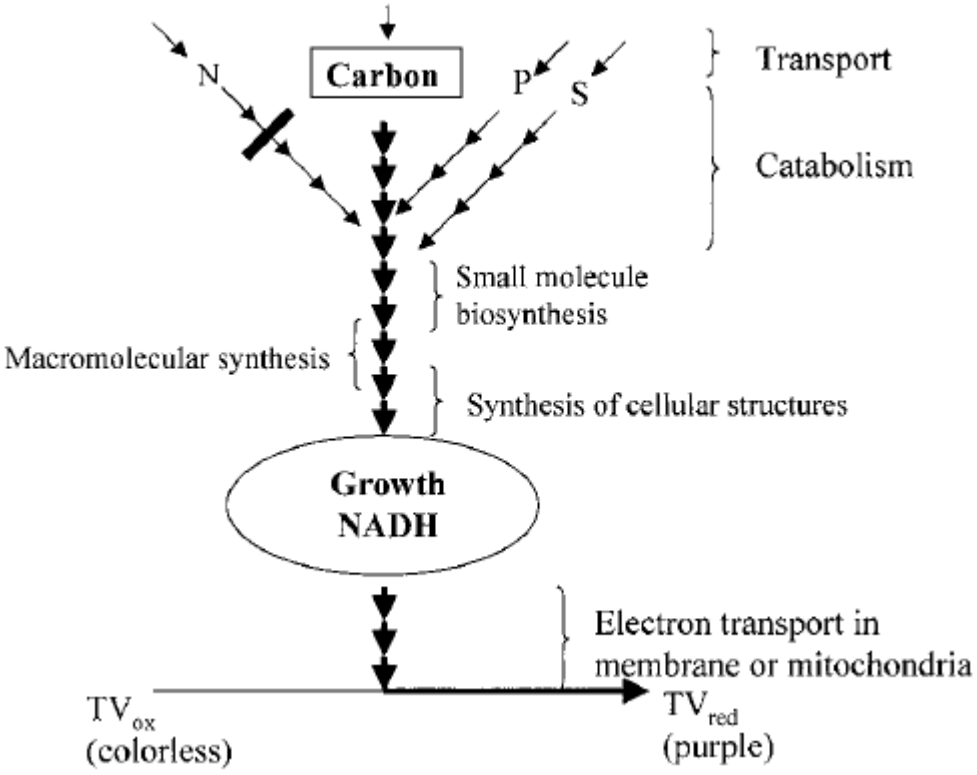


Figure 1.14: Schematic diagram of reduction of tetrazolium violet (TV) by cellular electron transport chain (Bochner *et al.* 2001). When a carbon source is metabolised, NADH is produced and tetrazolium dye is reduced and as a result it produces a purple colour.

Nearly, 2000 bacterial phenotypes can be efficiently identified on twenty preconfigured 96-well plates in a single experiment. Biolog plates PM1 to PM10, give profiles in utilising carbon, nitrogen, sulphur and phosphorus sources as nutritional chemical signals and behaving in different environmental chemical signals such osmotic and pH effects that stimulate cells; whereas plates PM11 to PM20, give profiles of bacterial behaviour in different concentrations of 240 chemicals that can potentially inhibit cells. The conditions and chemicals pattern in each well of the PM1 to PM20 plates can be obtained from Biolog, Hayward, CA (http://www.biolog.com/).

In the recent years, Phenotype Microarray (PM) technology has been applied in discrimination of closely related bacterial strains. In the study conducted by Mukherjee (2008), the utilization of *N-acetyl*-D-galactosamine (Aga) and D-galactosamine (Gam) was exploited to differentiate 194 *Escherichia coli* O157:H7 isolates from those of EDL933 and Sakai closely related strains from the 2006 spinach outbreak. In another paper, Kauko *et al.* (2010) show that serotypes of *Salmonella enterica* were differentiated based on utilising of 18 substrates from biolog plates. Furthermore, in more recent studies phenotypic microarrays were applied to characterise bacteria, such as *Acinetobacter* strains, *Mycobacterium tuberculosis* and *Mycobacterium bovis*, *Cronobacter sakazakii* SP291, *sarR* mutant in *Staphylococcus aureus*, *Rhizobium leguminosarum* bv. *Trifolii*, *Clostridium perfringens* strains, *Rhodococcus* sp. BCP1 and *Rhodococcus opacus* R7, *Clostridium difficile* and *Clostridium sordellii*, *Mycobacterium avium* subsp. *hominissuis* isolates, *Mycobacterium ulcerans* (Mara *et al.*, 2012; Khatri *et al.*, 2013; Yan *et al.*, 2013; Lucas & Manna 2013; Mazur *et al.*, 2013; Park & Rafii 2014; Orro *et al.*, 2015; Scaria *et al.*, 2015; Sanchini *et al.*, 2017; Zingue *et al.*, 2017). However, at the time of writing this chapter there are no published studies focusing on metabolic properties of *Streptomyces* strains evaluated by phenotypic microarray methods.

1.9 Heavy metal resistance in *Streptomyces*

In addition to the production of many varieties of secondary metabolites as a part of their metabolism, there is evidence that the genus *Streptomyces* has adapted to survive under extreme and harsh conditions. Soils and sediments contaminated with high concentration of heavy metals such as nickel (Ni), copper (Cu), Ferrous (Fe), Cobalt (Co), silver (Ag), zinc (Zn), lead (Pb) and mercury (Hg) are among rich metal habitats where a variety of different *Streptomyces* strains have been isolated. Although some of these elements are important for microorganisms since they are incorporated into enzymes and cofactors, higher concentrations are toxic and have adverse effects on enzymes and DNA in bacterial cells (Lopez-Maury *et al.*, 2002; Abou-Shanab *et al.*, 2007). Streptomycetes apply various systems of resistance mechanisms in response to heavy metals. These mechanisms include possessing nickel and iron-containing enzymes which are superoxide dismutase regulated by nickel metal (Kim *et al.*, 1998). Other resistance mechanisms encoded by resistance loci are located on plasmids (Ravel *et al.*, 1998) and include the adsorption and retaining of heavy metal cations to the bacterial cell walls (Schmidt *et al.*, 2005). Recent studies have shown that the mechanisms linked to heavy metal resistance include: 1) biosorption, in which metals precipitate to the cell surface and cell wall components due to ion exchange; 2) metal reduction, in which heavy metals become insoluble and unavailable by bacterial reductase activity; 3) extracellular chelaters, in which metal binding to polysacharides or protein prevent the cell toxicity by reducing the bioavailability of these metals; 4) efflux transport systems, in which heavy metal resistance is due to efflux protein pumps; 5) intracellular sequestration, in which metals accumulate within the cytosol of bacterial cells; 6) bio-mineralisation, in which crystals can be formed from metals driven by microorganisms; 7) superoxide dismutase, in which heavy metals can be tolerated by superoxide dismutase enzyme activity (Schütze & Kothe, 2012).

1.10 Lead (Pb) resistance in *Streptomyces*

Lead pollution of soils, sediments and water result in harsh conditions for microbial growth and survival because of its toxic effects. Pb toxicity comes from replacing essential metal ions such as calcium and zinc from their local binding sites and therefore inhibiting the enzyme activities (Nies, 1999; Naik & Dubey, 2013). Moreover, lead metal causes damage to nucleic acids and proteins, disruption of cell membrane functions, and alteration of the enzyme specificities (Asmub *et al.*, 2000; Hartwig *et al.*, 2003). Although lead is very toxic even in low concentrations, many environmental microorganisms have developed strategies (Figure 1.5) that enable them to tolerate and survive lead exposure (Jarosławiecka & Piotrowska-Seget, 2014).

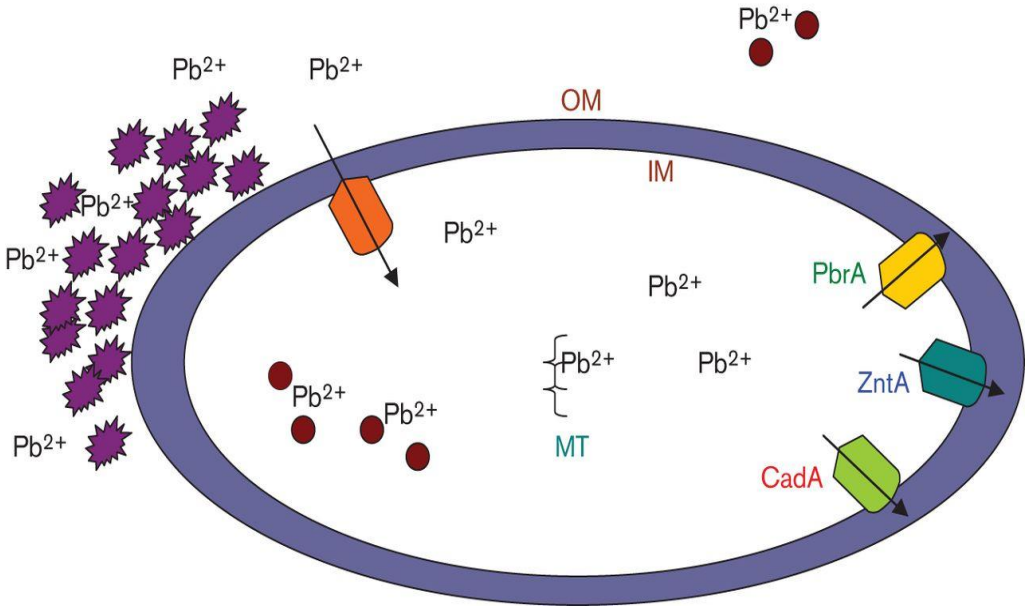


Figure 1.5: Mechanisms of microbial cell protection against lead ((Pb (II)) toxicity. Pb(II) could be kept away from the micro-organisms through its precipitation as insoluble phosphates outside the cell (circles), adsorption on extracellular polysaccharides (violet stars) or by polymers naturally occurring in the cell wall. After entering the cell through essential metal transporters, Pb (II) can be further inactivated by binding to the metallothioneins (MT), sequestered as insoluble phosphates or removed from the cell via transporters such as CadA, ZntA or PbrA. OM, outer membrane; IM, inner membrane (Jarosławiecka & Piotrowska-Seget, 2014).

Little is known about the tolerance of *Streptomyces* strains towards the heavy metal lead and whether the lead might activate the cryptic gene necessary for inducing secondary metabolites in *Streptomyces*. However, in a study carried out by Guo *et al.* (2009), *Streptomyces plumbiresistens,* a novel strain that could grow in a high concentration (4.0 mM) of lead was isolated from lead-polluted soil in north-west China. Despite very limited studies conducted by Haferburg *et al.* (2009) to address the possibility of enhancing the silent gene clusters in *Streptomyces* by induction with heavy metals to produce secondary metabolites of biological activity, to date there are no such studies on lead metal.

1.11 Strategies to activate cryptic and silent biosynthetic gene clusters in *Streptomyces*

Sequencing of the complete genomes of *Streptomyces coelicolor*, *Streptomyces avermitilis* and *Streptomyces griseus* revealed that many biosynthetic gene clusters encode to produce more than three secondary metabolites, whereas they contain more than twenty gene cluster that are predicted to encode for secondary metabolites (Bentley, 2002; Liu *et al.*, 2013; Tanaka, 2013). These facts suggest that *Streptomyces* has silent or cryptic gene clusters that might be responsible for the production of many novel bioactive natural products. These silent or cryptic biosynthetic gene clusters remain inactive for unknown reasons (Nett *et al.*, 2009; Seyedsayamdost, 2014), and particulary when *Streptomyces* cultured under common laboratory conditions (Haferburg *et al.*, 2009; Chiang *et al.*, 2011; Liu *et al.*, 2013, Amoroso *et al.*, 2013).

Many methods and strategies (Figure 1.6) have been devised to activate these cryptic biosynthetic pathways (Liu *et al.*, 2013). Since cryptic gene clusters encode specific functions not expressed under any known factors, mutational events or other genetic mechanisms like recombination or insertion are needed to activate these cryptic gene clusters (Tamburini & Mastromei, 2000). On the other hand, silent gene clusters may be activated by environmental signals (Chiang *et al.*, 2011; Aigle & Corre, 2012; Haferburg

& Kothe, 2013). The one strain-many compounds (OSMCA) method is one of these strategies that exploit the benefits of modification in culture media compositions or conditions for enhancing the secondary metabolites expression of the silent clusters (Bode *et al.*, 2002). At the same time, several studies have shown that when some microorganisms are co-cultured with *Streptomyces*, biosynthetic gene clusters were induced and new bioactive metabolites were isolated (Onaka *et al.*, 2011; Chiang *et al.*, 2011; Seyedsayamdost *et al.*, 2012). Another study by Moore *et al.* (2012) has suggested that HDAC inhibitors can be used to activate biosynthetic pathways of silent gene clusters in *Streptomyces*. A recent study conducted by Haferburg & Kothe (2013) showed that the biosynthetic pattern of secondary metabolites can be altered and new potentially therapeutic compounds might be discovered through activation of silent gene clusters in *Streptomyces* by heavy metal induction.

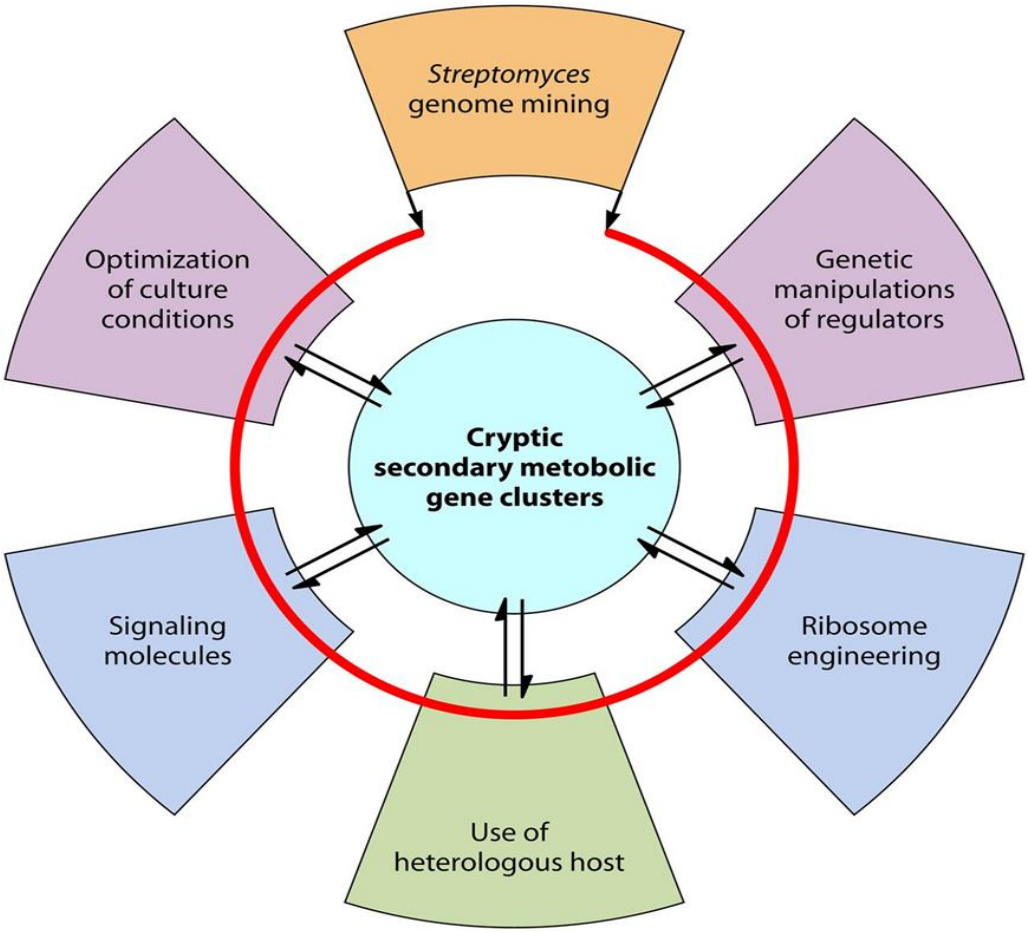


Figure 1.6: Strategies for the activation of cryptic secondary metabolic gene clusters in *Streptomyces*. The red line indicates the sequenced *Streptomyces* linear chromosome. Trapezoidal blocks with different colours represent different methods based on gene clusters to activate the possible expression of clusters (adopted from Liu *et al.*, 2013).

1.12 Aims of the present work

The objective of the present study is to isolate and characterise novel strains of Streptomycetes from such unexplored or under-explored extreme environments, as novel sources for new bioactive compounds.

The hypothesis of this work is that Actinomycetes in an unexplored habitat may hold promising novel strains, or even species that produce new biologically active metabolites. In addition, bacterial strains isolated from heavy metal contaminated environments may produce unique metabolites with potential medical applications via activation of their silent gene clusters by heavy metal induction. We hypothesise that the evolutionary link between antibiotic and heavy metal resistance can be exploited to discover new secondary metabolites.

The specific aims of this thesis are the following:

1. Isolation and cultivation of promising Actinomycetes strains that exhibit strong bioactivity against tested pathogens under heavy metals induction (Chapter 3).

2. Carry out Whole Genome Sequencing (WGS) using Ion Torrent PGM Technology and perform *de novo* assemblies for candidates on the basis of their distinctive properties (Chapter 4).

3. Bioinformatics and Computational High throughput descriptions of novel isolate (isolate_99) with the closest related strains based on 16S and 23S rRNA genes, *in Silico* DNA-DNA Hybridisation (DDH), Average Nucleotide Identity (ANI), and Phenotypic Microarray (PM) experiments (Chapter 5).

4. Perform genome mining of isolate_99 using comprehensive bioinformatics tools to identify the putative Biosynthetic Gene Clusters (BGCs) which may be potential sources for novel natural products that have not been observed under conventional laboratory conditions (Chapter 6).

5. Study the transcriptomic response through the RNA-Seq analysis of Biosynthetic Gene Clusters of isolate_99 under lead metal stress to identify the mechanism of action and the molecular target of each gene cluster (Chapter 7).

Chapter 2:

Materials and Methods

2.1 Biological materials

A total of 65 environmental samples originated from the heavy metal contaminated sites of Leadhills and Wanlockhead villages in South Lanarkshire, Scotland, United Kingdom were collected in March, 2014. Samples were transported immediately to the Dr. James and Mrs. Fitri Hay Laboratories at the University of Strathclyde, Glasgow, UK, stored at 4 °C and processed within one day of collection. Soils were air-dried at 70 °C for 15 minutes in a hot air oven and then cooled to room temperature. This helps in decreasing the population of gram negative bacteria. Soil samples were crushed, mixed thoroughly and large debris were removed to obtain fine soil particles, these samples were used for isolation of *Streptomyces*.

2.2 Bacterial strains and organisms

Table 2.1 contains a list of all *Streptomyces* strains and Table 2.2 contains a list of all other bacterial and fungal used in this study.

Table 2.1 *Streptomyces* references strains used in this study.

Strain	**Source / Reference / Sequence reference**
Isolates 1-120	This Study
S. coelicolor M145	Strathclyde stock cultures
S. albus	DSMZ 40313
S. avermitilis	DSMZ 46492
S. venezuelae	DSMZ 40230
S. rochei	DSMZ 40913
S. turgidiscabies	DSMZ 41838
S. graminilatus	DSMZ 102005
S. xylophagus	NZ_JNWO00000000.1
S. prunicolor	NZ_BARF00000000.1
S. aurantiacus	NZ_LIPP00000000.1

S. bottropensis	ATCC 25435
S. scabiei	NC_013929.1
S. sviceus	NZ_ABJJ00000000.2
S. *coelicolor* A3(2)	NC_003888.3
S. ghanaensis	NZ_ABYA00000000.1

Table 2.2 Other bacterial strains and organisms used in this study.

Strain	**Reference / source**
Escherichia coli	ATCC 25922
Staphylococcus aureus	ATCC 43300
Klebsiella pneumoniae	ATCC 700603
Acinetobacter baumannii	ATCC 19606
Pseudomonas aeruginosa	ATCC 27853
Enterococcus faecalis	ATCC 51299
Candida albicans	Strathclyde stock cultures

2.3 Media used for cultivation of bacterial strains

All media used for growth of bacterial strains were prepared by combining the relevant ingredients, as listed below, followed by autoclaving for 15 minutes at 121 °C (unless otherwise stated) and then allowed to cool slowly in water bath to 55 °C. For the preparation of agar plates, approximately 25 ml of molten agar was poured into Petri dish and allowed to set. Plates were poured and set within a laminar hood for around 30 minutes, to maintain sterility and allow any excess liquid to evaporate. Prepared media were used immediately or stored in fridge at 4 °C.

Humic Acid-Vitamin Agar (Hayakawa and Monomura, 1987)

Humic acid	1.0 g*
Na_2HPO_4	0.5 g
KCl	1.71 g
$MgSO_4$.7H2O	0.05 g
$FeSO_4$.7H2O	0.01 g
CaCO3	0.02 g
B-vitamins **	
Cycloheximide	50 mg
Nystatin (added in this study)	30 mg
Agar	18 g
Distilled water	1000 ml
pH	7.2

* Dissolved in 10 ml of 0.2 N NaOH

** 0.5 mg each of thiamine-HCl, riboflavin, niacin, pyridoxin-HCl, inositol, Ca-pantothenate, *p*-aminobenzoic acid, and 0.25 mg of biotin. B-vitamins and cycloheximide were sterilised by filter membrane and added to the autoclaved media.

Maltose Yeast-extract Malt-extract Agar (MYM) (Shepherd *et al.*, 2010)

Maltose	4.0 g
Yeast extract	4.0 g
Malt extract	10.0 g
Agar	18.0 g
Distilled water	1000 ml
pH	7.3

Manitol Soya Flour Agar (MS) (Kieser *et al.*, 2000)

Mannitol	20.0 g
Soya flour	20.0 g
Agar	20.0 g
Tap water	1000 ml

Glucose Yeast-extract Malt-extract Agar (ISP-2 / GLM) (Shirling and Gottlieb, 1966)

Glucose	4.0 g
Yeast Extract	4.0 g
Malt Extract	10.0 g
Agar	20.0 g
Distilled water	1000 ml
pH	7.3

Oatmeal Agar (ISP-3) (Shirling and Gottlieb, 1966)

Oatmeal	20.0 g
Trace Salts Solution (see solutions)	1.0 ml
Agar	20.0 g
Distilled water	1000 ml
pH	7.2

ISP-4 agar (Shirling and Gottlieb, 1966)

Soluble starch	10.0 g
Dipotassium phosphate	1.0 g
Magnesium sulphate	1.0 g
Ammonium sulphate	2.0 g
Calcium carbonate	1.0 g
Ferrous sulphate	1.0 mg
Magnesium chloride	1.0 mg
Zinc sulphate	1.0 mg
Agar	20.0 g
Distilled water	1000 ml
pH	7.2

ISP-5 (Glycerol Asparagine Agar Base) (Shirling and Gottlieb, 1966)

L-Asparagine	1.0 g
Dipotassium phosphate	1.0 g
Glycerol	10 ml
Trace Salts Solution (see solutions)	1.0 ml
Agar	20.0 g
Distilled water	1000 ml
pH	7.2

ISP-6 (Shirling and Gottlieb, 1966)

Peptone	15.0 g
Proteose Peptone	5.0 g
Ferric Ammonium Citrate	0.5 g
Dipotassium Phosphate	1.0 g

Sodium Thiosulfate	0.08 g
Yeast Extract	1.0 g
Agar	20.0 g
Distilled water	1000 ml
pH	7.3

ISP-7 (Tyrosine medium) (Shirling and Gottlieb, 1966)

Glycerol	15.0 g
L-Tyrosine	0.5 g
L-Asparagine	1.0 g
K_2HPO_4	0.5 g
$MgSO_4$. $7H_2O$	0.5 g
NaCl	0.5 g
$FeSO_4.7H_2O$	10.0 mg
Trace Elements Solution Ho-Le (see solutions)	1.0 ml
Agar	20.0 g
Distilled water	1000 ml
pH	7.3

Minimal Medium (MM) Agar (Kieser *et al.*, 2000)

L-aspargine	0.5 g
K_2HPO_4	0.5 g
$MgSO_4.7H_2O$	0.2 g
$FeSO_4.7H_2O$	0.01 g
Glucose (added after autoclaving)	10.0 g

Agar	16.0 g
Distilled water	1000 ml
pH	7.0

R2YE Medium (Shepherd *et al.*, 2010)

Media A (Prepared in a single autoclavable 1L bottle with a magnetic stirring bar)

Sucrose	103.0 g
K_2SO_4	0.25 g
$MgCl_2.6H_2O$	10.12 g
Glucose	10.0 g
Casamino acids	0.1 g
Distilled water	800.0 ml
Yeast extract	5.0 g
Agar	15.0 g

Media B (each solution was prepared in a separate bottle, sterilised by autoclaving)

Trace element solution	2.0 ml
TES buffer (5.73 %, w/v)	100.0 ml
KH_2PO_4 (0.5%, w/v)	10.0 ml
$CaCl_2.2H_2O$ (3.68%, w/v)	80.0 ml
L-proline (20%, w/v)	15.0 ml
1M NaOH	5.0 ml

Trace element solution (1 L)

$ZnCl_2$	40.0 mg
$FeCl_3.6H_2O$	200.0 mg
$CuCl_2.2H_2O$	10.0 mg
$MnCl_2.4H_2O$	10.0 mg

$Na_2B_4O_7.10H_2O$	10.0 mg
$(NH4)_6Mo_7O_{24}.4H_2O$	10.0 mg

The solution was made in dH_2O, sterilised by autoclaving and stored at room temperature or at 4 °C for a long time period.

Luria Bertani (LB)

Tryptone	10.0 g
Yeast extract	5.0 g
Sodium chloride	5.0 g
Glucose	1.0 g
Agar	16.0 g
Distilled water	1000 ml
pH	7.0

Tryptone Soya Broth (TSB) (Kieser *et al.*, 2000)

Tryptone soya broth powder	30.0 g
Distilled water	1000 ml

2.4 Buffers and Solutions

Trace Salts Solution

$FeSO_4 .7H_2O$	0.1 g
$MnCl_2 .4H_2O$	0.1 g
$ZnSO4 .7H_2O$	0.1 g
Distilled water	100.0 ml

The solution was sterilised by filter membrane.

Metal Ion Cocktail

$ZnSO_4.7H_20$	68 mg
$FeCL_2.6H_20$	135 mg
$MnCl_2.4H_2O$	99 mg
$CaCl_2.2H_2O$	74 mg
Distilled water	100 ml

Ingredients were filtered sterilised

Phosphate Buffered Saline (PBS) (Harisha, 2007)

NaCl	8.0 g
KCl	0.2 g
Na_2HPO_4	1.44 g
KH_2PO_4	0.24 g
Distilled water	1000 ml
pH	7.2

Ingredients were autoclaved for 15 minutes at 121 °C.

Lysozyme buffer

Sucrose	0.3 M
Tris	25 mM
EDTA	25 mM
Biochemistry grade water	as required
pH	8.0

Ingredients were dissolved and autoclaved for 15 minutes at 121 °C. Prepared buffer was tightly-sealed, stored at room temperature and protected from direct light until use.

1 x TE buffer (Harisha, 2007)

1M Tris solution	1.0 ml
0.5 M EDTA	0.2 ml
Distilled water	98.8 ml
pH	7.5

Ingredients were dissolved and autoclaved for 15 minutes at 121 °C. Prepared buffer was tightly-sealed, stored at room temperature and protected from direct light until use.

6X loading dye (Cold Spring Harb Protocols, 2007)

Glycerol	30.0 ml
Bromophenol blue	0.25 g
Xylene cyanol FF	0.25 g
Distilled water	70 ml

The mixture was stored at 4°C in dark bottle.

1X TAE Buffer (Harisha, 2007)

Tris-acetate (Tris base)	4.84 g
Acetic acid	1.14 ml
Na_2 EDTA	0.74 g
Distilled water	1000 ml

The buffer was stored at room temperature.

2.5 Microbiology Methods

2.5.1 Isolation

For each collected sample, the soil (3 g) was added to a falcon tube (50 ml) containing 27 ml of PBS (Harisha, 2007) adjusted to pH 8.0. The tubes were incubated in an orbital shaker incubator (Thermo Scientific, Iowa, USA) at 30 °C with shaking at 220 rpm for 30 min. Mixtures were allowed to settle at room temperature for 15 min. A series of dilutions of the suspension from 10^{-1} to 10^{-3} were aseptically prepared using PBS in a total volume of 10 ml.

An aliquot of 100 µl of each dilution was pipetted and spread evenly over the surface of HAV agar as a selective medium. The B vitamins including thiamine-HCl, riboflavin, niacin, pyrodoxin-HCl, inositol, capantothenate, p-aminobenzoic acid (0.5 mg/l for each) and biotin (0.25 mg/l) were added to the autoclaved medium. The antifungals cycloheximide (50 mg/l) and nystatin (30 mg/l) were supplemented to inhibit development of invasive fungi for the three media. The plates were incubated at 30 °C in aerobic condition, and monitored over three weeks.

2.5.2 Purification

All screened samples are suspected to be inhabited with actinomycetes and the isolates showed actinomycetes like characteristic morphology on selective media and under a stereoscopic microscope were transferred onto fresh medium of MS medium without adding antibiotic. Subculture was repeated on fresh medium of MS until pure isolates were obtained from each visible growth. Colonies were slow growing, aerobic, chalky, heaped, folded and with aerial and substrate mycelia of different colours.

2.5.3 Long term storage of *Streptomyces*

Strains of *Streptomyces* were stored as spore glycerol stocks at – 80 °C following the protocol of Shepherd *et al.* (2010) with a slight modification.

Fresh ISP-2 or MYM agar plates inoculated with *Streptomyces* strains were incubated at 30 °C for one to two weeks. After a full growth was observed, 4 to 5 ml of sterile distilled water was added to the plate. The spores were completely resuspended into the distilled water by gently scraping the surface of the culture with a sterile loop. Then after, the resuspended spores were transferred by pipette to a Falcon tube, and the final volume was made up to 35 ml with distilled water. To break up the spore chains, the tubes were vigorously mixed on a vortex mixer (Vortex-2 Genie, USA). Subsequently, by using a sterilised syringe, the spore suspension was passed through sterile cotton wool, and the filtered spores were collected in a new Falcon tubes. The residual spores attached to the cotton filter were washed with an additional distilled water. The filtered spores were then centrifuged for 10 minutes at 4500 rpm. After centrifugation the supernatant was discarded and the spores were resuspended in 1 ml of distilled water. 1 ml of 50% (v/v) glycerol was added to the spores suspension, mixed and an aliquot of 1 ml of the mixture was pipetted and transferred to sterile 1.5 ml Eppendorf tube and stored at – 80 °C.

2.5.4 Bioactive secondary metabolites induction assay

Different gradual concentrations of the following heavy metals; $CuSO_4$, $CoCl_2$, $ZnCl_2$ and $Pb(CH_3COO)_2$, were prepared using distilled water. The solutions were sterilised through filter membranes and added immediately to autoclaved medium. To determine the maximum tolerance concentration (MTC) the following concentrations were prepared: $CuCl_2$ (1.5, 2, 2.5, 3, 3.5 and 4 mM/L), $CoCl_2$ (1.5, 2, 2.5, 3, 3.5 and 4 mM/L), $ZnCl_2$ (20, 25, 30, 35, 40, 45, 47.5 and 50 mM/L), Pb (2, 2.5, 3, 3.5 and 4 mM/L). Finally, different concentrations of N-acetylglucosamine (GlcNAc) ranging from from 5 to 50 mM and

different concentration of sodium butyrate (NaBu) ranging from 25 to 200 mM were tested. The isolates were inoculated from fresh cultures. After the incubation for two weeks at 30 °C, plates were observed for heavy metals resistance. The highest concentration in which the isolates could grow was considered as the MTC. The qualitative parameter of heavy metals tolerance was estimated by the microbial growth. *Streptomyces coelicolor* which is known not to carry heavy metals resistance (Schmidt *et al.*, 2005) was used as a control, and MM agar plates without metal supplements were used for microbial growth comparison. Finally, plugs of each *Streptomyces* cultures grown on different concentrations of inducers plus a control plug containing only agar, were taken under sterile conditions using a cork borer, and placed over LB agar containing the test pathogens.

2.5.5 Antibiotic production screening assay

To discover the potential production of antibiotic compounds of the isolated Streptomyces strains utilising different media, all isolates were grown on ISP-2, ISP-3, ISP-4, ISP-5, ISP-6 and ISP-7 for three weeks at 30 °C. The isolates were tested against ESKAPE pathogens: *Escherichia coli, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa* and *Enterococcus faecalis*. The pathogenic strains were grown overnight in 5 ml of LB liquid medium at 37°C. Cultures were diluted to a final OD_{600} of 0.08 - 1.3, then an aliquot of 100 µl of each tested microorganism was pipetted and evenly distributed on Petri dishes containing 25 ml of LB agar. Then after, plugs of each *Streptomyces* cultures grown on different ISPs media plus a control plug containing only agar, were taken under sterile conditions using a cork borer, and placed over the LB agar plates containing the test pathogen. The antibiotic production was evaluated by measuring the inhibition zone between the target microorganisms and the plugs of *Streptomyces* cultures, after incubation at 37 °C for 24 h.

2.5.6 Phenotypic Microarray (PM) using Omnilog

PM plates PM1, PM2, PM3, PM9, PM10, Inoculating Fluid-0a (IF-0a) and Inculcating Fluid-10b (IF-10b) and Tetrazolium Dye D were obtained from Technopath, UK. To prepare the inoculum for PM plates, one colony of each isolate_99, *S. turgidiscabies* and *S. graminilatus* strains was grown in 5 ml TSB medium at 30 °C for two days. The growth was centrifuged at 4 °C and 3500 rpm for 10 min to harvest the mycelium. The supernatant was poured off and the mycelium were resuspended in 1 ml of sterile distilled water, centrifuged at 4 °C and 3500 rpm for 10 min to wash the mycelium. This was repeated one more time in distilled water and one time in IF-0a or IF-10b solution depending on the PM plate to be used. The cell suspension was serial diluted to obtain an OD_{450} of 0.4. The cell suspension was further diluted 1 in 10 in the IF solution of interest so the final OD_{450} was 0.04. The final inoculating solution was prepared based on the type and the total number of plates used.

For three PM1 or three PM2 plates: 30 ml of IF-0a at 1.2X, 360 µl of 0.5 M $MgCl_2$, 360 µl of 0.5 M Na_2SO_4, 360 µl of 1.5 M NH_4Cl (pH 7), 360 µl of 0.6 M NaH_2PO_4 , 360 µl of Metal Ion Cocktail (See above), 360 µl of H_2O and 360 µl of Dye D at 100X. Each well of PM plate was filled with 100 µl of the Inoculation Fluid. The plates were loaded onto the Omnilog within 15 min and incubated at 30 °C for 72 hours. For PM3 plates, IF-10b was used instead of IF-0a, 360 µl of 2.5 M glucose was added, ammonium chloride was excluded and the rest as in PM1 and PM2 plates. For PM9 and PM10 plates, IF-0a was used, 360 µl of 2.5 M glucose was added, H2O was excluded and the rest as in PM1 and PM2 plates.

The raw data obtained from the Omnilog experiments were analysed using the following programs, DSE_OKA data v1.1.1.15 was used to convert the raw data in DE5 format to .OKA format, the Kinetic v1.3 program was utilised to change the format from .OKA to .DLB and finally the parametric comparison was performed using the parametric v.1.3 program.

2.5.7 Independent confirmatory testing of Phenotypic Microarray (PM) data

Minimal Medium (MM) agar supplemented with varied selected carbon (20 mM) or nitrogen (20 mM) sources was used to confirm the PM analysis results for some of the key findings. A single colony of each tested strains previously grown on IPS-2 medium was streaked on MM medium. Plates were scored after 5 days of incubation at 30 °C.

2.6 Molecular Biology Methods

2.6.1 DNA extraction

Genomic DNA was isolated from the strains according to the protocol described by Kieser *et al.* (2000) with some modifications. Isolates were grown in 20 ml of TSB in 50 ml falcon tubes for 7 days at 30 °C with shaking at 220 rpm. Cells were harvested by centrifugation for 5 minutes at 4000 rpm (Thermo Scientific, Megafuge 40R, USA) and the pellet was resuspended in 500 µl of lysosyme buffer and transferred to a 1.5-ml Eppendorf tube. The cells were centrifuged at 10,000 rpm (Thermo Fisher, Germany) for 30 second and the pellet was resusprnded again in 500 µl of lysosyme buffer. Following this step, 25 µl lysozyme solution (50 mg ml-1) and 3 µl RNase A (10 mg ml-1) were added and incubated at 37 °C for 60 minutes with occasional mixing. Then 10 µl of 10% SDS solution was added to each tube and incubated at 55 °C for 15 minutes, and the samples were then transferred to a 37 °C water bath for 30 minutes (to allow RNase A to work). This was followed by the addition of 5µl proteinase K (50 mg ml^{-1}) and incubation at 55 °C for 15 minutes. Subsequently, SDS was precipitated by adding 200 µl of 250 mM KCl. Protein and lipids were removed twice by the addition of 500 µl of an equal volume of phenol-chloroform, and the phases were mixed by inversion for 1 minute and centrifuged at 12,500 rpm for 5 minutes. The supernatant was transferred, using a sterile cut tip, to a new eppendorf tube. Thereafter, a measured amount of the aqueous phase was transferred to a new tube, and ddH_2O and 3 M NaAc (pH = 5.2) to a final concentration of 0.3 M NaAc was added. The DNA was precipitated by adding 0.54

volume isopropanol. The DNA was picked out of the mix and washed with ice-cold ethanol (70%) in a new tube, and centrifuged at 12,500 rpm for 5 minutes. After centrifugation, the supernatant was removed, dried, and redissolved in 200 µl TE buffer and incubated at 55 °C for a few hours and/or overnight in the fridge. The concentration and purity of extracted nucleic acids were determined using the Thermo Scientific NanoDrop UV-Vis spectrophotometer 2000c (Thermo Scientific, USA) at 260/280 nm and 260/230 nm. The extracted DNA was stored at -20 °C.

2.6.2 Polymerase Chain Reaction (PCR):

2.6.2.1 16S rRNA gene amplification

The 16S rRNA gene was amplified by PCR using a set of universal primers: 27F (5´- AGA GTT TGA TCC TGG CTC AG - 3´), 1492R (5´- TAC GGC TAC CTT GTT ACG ACT T - 3´) (Chun & Goodfellow , 1995; Labeda & Kroppensted; 2000; Goodfellow & Kumar 2007), 1522R (5´- AAG GAG GTG ATC CAG CCG CA - 3´) (Magarvey *et al.*, 2004; Das *et al.*, 2010; Bouras *et al.*, 2013) and 1525R (5´- AAG GAG GTG ATC CAG CC - 3´) (Zhu *et al.*, 2007). For the PCR assay and for each sample, a total volume of 50 µl mixture in a 0.2 ml reaction tube was performed using the (Bio Rad T100TM, USA) thermal cycler. The reaction mixture was prepared as follows: 10 µl of 5x My *Taq* reaction buffer, containing dNTPs, $MgCl_2$ and enhancers, 0.5 µl of My *Taq* DNA polymerase (Bioline, UK), 1 µl of each forward and reverse primer, 1 µl of purified DNA and 36.5 µl ddH_2O. The amplification cycle was performed according to the following programme: an initial denaturation step at 95 °C for 5 min, followed by 30 amplification cycles of 95 °C for 30 sec for denaturation, 55 °C for 60 sec for primer annealing, and 72 °C for 60 sec for primer extension, and a final extension step was held at 72 °C for 10 min and cooled to 12 °C.

2.6.2.2 Control reactions

For each primer pair combination tested against extracted samples DNA: one negative control without DNA template and one positive control (*Streptomyces coelicolor* M145 obtained from the reference stocks of Dr. James and Mrs. Fitri Hay Laboratories, Strathclyde Institute of Pharmacy and Biomedical Sciences) were included.

2.6.2.3 Agarose gel electrophoresis

Agarose gel electrophoresis was used to visualise DNA from the PCR reactions. From the total 50 µl PCR product 20 µl was pippeted and mixed with 5 µl 6x loading dye and 10 µl of this mixture was loaded on an agarose gel (1%, stained with 3 µl ethidium promide (10 mg/ml)). The gel was run at 80V for 45 min in a 1X TAE buffer. To determine the size of the PCR product, 5 µl of HyperLadder 1kb (Bioline, UK) as a DNA marker was loaded on the gel. The PCR bands were confirmed and visualised with ultraviolet illumination (Syngene, UK) and the images were exported using GeneSnap software.

2.6.2.4 16S rRNA gene sequencing

Amplified PCR products that showed a single clear band under the UV transilluminator (Uplan, CA, USA) were purified using the ISOLATE II PCR and Gel kit for DNA extraction from agarose gel (Bioline, UK) following the manufacturer's instructions. The purified fragments were quantified with the NanoDrop UV-Vis spectrophotometer 2000c according to sample submission guide and shipped to Germany (Eurofins MWG, Ebersberg, Germany) to be sequenced. Sequencing was carried out using an ABI 3730 XL automatic DNA sequencer. The same primers as used for PCR amplification were used for this purpose.

2.6.3 Whole Genome Sequencing (WGS)

The next generation sequencing using Ion Torrent technologies involves different steps including: DNA fragmentation; adaptor ligation, 400 bp size selection, PCR amplification of adaptor ligated DNA, 400 bp library quantification and qualification, Template-Positive Ion Sphere Particles preparation using the Ion OneTouch 2 Instrument (OT2), Template-Positive Ion Sphere Particles enrichment and loading the Chip and starting the sequencing. All Ion PGM protocols and sequencing user guides are provided by Ion Torrent, Life Technologies. However, below are the most notable records that applied to our samples.

2.6.3.1 Library preparation and DNA fragments size selection

For library preparation, we followed the NEBNext Fast DNA Fragmentation & Library Prep Set for Ion Torrent (New England Biolabs, Inc., UK). The bacterial genomes were enzymatically fragmented so that 400 base pair fragments could be selected to be run on the PGM sequencer. Each fragmentation reaction consisted of around 800 ng of genome DNA. 1 µl $MgCl_2$ (200 mM) was added to each reaction as recommended for high GC DNA content. After fragmentation, each genomic DNA was barcoded using Ion Xpress Barcode Adapters (Life Technologies), ligated and cleaned up. For 400bp DNA fragments size selection, the optimal size read for the PGM, the adapter ligated samples were run on the E-Gel Size select-2% (Invitrogen). The 50 bp DNA ladder (Invitrogen) was used as a reference. The run time was about 18 minutes to get the desired fragments size. The size-selected, barcode adapter ligated DNA fragments then underwent for 6 cycles of PCR amplification. The final PCR products were purified using Agencourt AMPure XP Beads (Beckman Coulter).

2.6.3.2 DNA fragments validation

The quantity and the quality of the amplified DNA libraries were assessed using both: Qubit 2.0 Fluorometer (Invitrogen) and High Sensitivity DNA Bioanalyzer 2100 (Agilent Technologies) respectively. The concentration (in µg/ml) and the length (in bp) of each library were used to calculate the final DNA concentration required for pooling of each sample.

2.6.3.3 Templated and Enrichment of positive Ion Sphere Particles

To prepare template-positive Ion Sphere Particles (ISPs), the Ion PGM Template OT2 400 protocol (catalog number: 4479878) was followed The pooled library for each sample then underwent emulsion PCR on the Ion OneTouch-2 (Life Technologies) system for 7 h and enriched for 45 min on the Ion OneTouch ES instrument (Life Technologies) using Dynabeads MyOne Streptavidin C1 (Invitrogen). The enriched, template-positive ISPs are then loaded onto an Ion 316v2 chip following a protocol developed in our lab (appendix A) and subsequently sequenced using 300 sequencing cycles (1200 flows) on the Ion Personal Genome Machine (Life Technologies) following the Ion PGM Sequencing 400 protocol.

2.6.4 RNA Sequencing

2.6.4.1 Total RNA isolation from isolate_99

Total RNA was extracted using an adaptation of the ISOLATE II RNA-Mini Kit (Bioline). Before the extraction, isolate_99 was grown on cellophane discs (previously cut and boiled for 15 min in dH_2O on the heater, then placed between pre-wetted Whatman No.1 paper and autoclaved) laid on top of both MM agar plates and MM agar plates supplemented with 1mM Pb for 7 days at 30 °C. Once the bioactive metabolite was

produced by screening its activity against *E. coli*, biomass was harvested by scraping the surface of the cellophane with a sterile razor blade and transferred to an RNase-free Eppendorf tube, and this was used as starting material. Immediately, 1 ml of RNAprotect (Qiagen) was added to the biomass, vortexed for 5 seconds and incubated for 5 min at room temperature.

After RNA protect treatment, the pellet was resuspended in 200 µl of TE buffer containing 15 mg/ml of lysozyme and 40 µl of 10 mg/ml of proteinase K. The samples were incubated with rotation for 1 hour at room temperature, 350 µl of RLT buffer containing β-mercaptoethanol (10 µl per ml of RLT) and vortexed vigorously. The samples were transferred to a new Eppendorf tube containing 250 µl of Zirconia beads (Ambion) and vortexed for 10 min at room temperatures. The tubes were centrifuged for 5 min at 4 °C, the supernatant for each sample was removed and combined with 350 µl of 100% EtOH. Finally, the samples were transferred to a RNA Mini spin column and the RNA Mini protocol followed until the elution step. For elution, 25 µl of RNase free water was added to the spin column and incubated at room temperature for 2 min. The column was centrifuged at maximum speed for 2 min to elute the RNA. The elution step was repeated one more time to give a total of 50 µl of RNA.

2.6.4.2 Removal of genomic DNA from RNA samples

Residual DNA was removed by performing DNAse digestion in solution followed by RNA cleanup. The RNA sample was treated with DNase I reagents (Ambion, Life Technologies) as specified by the manufacturer. Briefly, 5.5 µl of 10X DNase buffer and 4 µl of DNAseI were added to each sample and incubated at 37 °C for 30 min, then 10 µl of DNAse inactivation reagent were added into RNA sample, vortexed and incubated for 2 min at room temperature. The mixtures were centrifuged for 1 min at maximum speed and the supernatant was collected and transferred into a new tube.

2.6.4.3 Control of DNA contamination

To check for genomic DNA contamination, the isolated RNA samples were used as templates for a standard PCR using *hrdB* primers. None of the samples showed a PCR product apart from the positive control with genomic DNA (isolate_99) as a template.

2.6.4.4 Quantification and qualification of isolated RNA

A Qubit 2.0 fluorometer was used to measure the RNA concentration in each sample by applying RNA HS Assay Kit (Life Technologies catalog number Q32852). In short, 1 µl of RNA sample was mixed with 198 µl RNA HS solution and 1 µl RNA HS reagent in a 0.5 ml Qubit assay tube. The sample and the 2 standards provided with the kit were incubated for 2 min at room temperature before the measurement.

For RNA quality, Bioanalyzer 2100 (Agilent) using Agilent RNA 600 Pico Kit assay (catalog number 5067-1513) was used. This assay can measure the RNA Integrity as an indicator for RNA quality following the manufacturer's instructions.

2.6.4.5 Ribosomal RNA depletion from total RNA samples

Ribosomal RNA was removed from total RNA samples using the Ribo-Zero Magnetic Kit for Gram positive bacteria (Illumina, catalog number MRZGP126). The kit uses biotinylated capture probes to hybridise to the rRNA, the rRNA-probe hybrid is bound to magnetic beads and removed by a magnet. The remaining solution contains the desired mRNA in solution. The manufacturer's instructions were followed and the rRNA-Depleted samples were stored at -80 °C.

2.6.4.6 cDNA synthesis and RNA sequencing library preparation

Protocol, materials and reagents for whole transcriptome barcoded libraries preparation were carried out using user guide in Ion Total RNA-Seq Kit v2 (Life Technologies catalog number 4475936) which is compatible for use with Ion Personal

Machine (PGM) system. The RNA library preparation protocol involves different steps including: RNA fragmentation using RNase III, purification of the fragmented RNA, yield and size distribution assessment of the fragmented RNA, hybridisation and ligation of the RNA, reverse transcription (RT), the cDNA purification, the cDNA amplification, purification the amplified cDNA and yield and size distribution assessment of the amplified DNA.

2.6.4.7 Sequencing

For sequencing of final RNA libraries, Templated and Enrichment of positive Ion Sphere Particles steps from Whole Genome Sequencing (WGS) section were followed after pooling the barcoded whole transcriptome libraries and determining the library dilution required for template preparation for each sample.

2.7 Bioinformatics Methods

2.7.1 Bioinformatics analysis tools

2.7.1.1 BLASTn to search for sequences

The Nucleotide Basic Local Alignment Search Tool (BLASTn) program (http://blast.st-va.ncbi.nlm.nih.gov/Blast.cgi, Altschul *et al.*, 1990) was used to search for homology to the input template sequence against entire sequences which are available on the sequence databases. The default was 100 hits and Nucleotide collection (nr/nt) database was chosen, and the program selection was set to highly similar sequences (megablast). Most closely related homologues which had ≥98% sequence identity (Cole *et al.*, 2009), and had query coverage 100% with an E value $<10^{-3}$ and met the criteria of aligning (Hall, 2011) were identified, downloaded and saved in FASTA format. If the coding sequencing (CDS) files were in the wrong orientation, the orientation of the sequence was corrected.

2.7.1.2 MUSCLE for aligning the gene sequences

The alignment of the data set sequences of the 16S rRNA gene was performed with Multiple Sequence Comparison by Log Expectation (MUSCLE) software program (Edgar, 2004). All gaps were deleted by aligning the codon sequences rather than aligning the DNA sequences (Hall, 2011). The excess sequences were trimmed manually; and duplicate sequences were eliminated by computing pairwise distances. Reliability of the alignment was preliminarily estimated by determining the average of amino acid identity with p-distance default (Thompson et al, 1999). For more accuracy, the alignment reliability was evaluated with GUIDANCE program which is a multiple sequence alignments web-based server (http://guidance.tau.ac.il/) using FASTA format. Moreover, GUIDANCE was used to detect the sufficiency of the outgroup sequences with which to root a tree. The PRANK method (Loytynoja & Goldman 2008) was chosen to estimate a base alignment. The paired alignments that had a high score (i.e., near 1.0) were considered to be reliable (Penn *et al.*, 2010) while unreliable sequences were removed and the trimmed multiple sequence alignments were saved in FASTA format.

2.7.1.3 leBIBI^{QBPP} tool for species identification based on 16S rRNA gene

The Quick Bioinformatics Phylogeny of Prokaryotes leBIBI^{QBPP} tool (Falndrois *et al.,* 2015) was used to identify the isolates based on their query sequences by using a proximate Maximum Likelihood (ML) approach.

2.7.1.4 MEGA 6.0 for Phylogenetic analysis

The program MEGA (Molecular Evolutionary Genetics Analysis) version 6.0 (Tamura *et al.*, 2013) employing the distance algorithmic Neighbour Joining (NJ) approach (Saitou & Nie 1987) was used to infer phylogenetic relationships and produce a phylogenetic tree from the data.. The suitability of the sequences for a NJ tree was determined with the average pairwise Junkes-Canter (JC) distance. The reliability of the resulting tree

topologies was estimated by 1000 bootstrap resamplings and branches that have less than 50% support were removed by choosing Condensed Tree default (Hall, 2011).

2.7.1.5 FastQC tool

FastQC (Andrews *et al.*, 2011) provides a high quality control for high throughput raw sequence data. It provides a modular set of analyses which can be used to give a quick impression of whether the raw sequence data has any issues before doing any further analysis.

2.7.1.6 SPAdes 3.5 Genome Assembler

Spades 3.5 (Nurk *et al.*, 2013) was used to perform *de novo* assemblies for sequenced genomes. The loaded raw data were in fastq format. The settings were: run assembly with read error correction, suggesting enabling single-cell mode, careful correction, *k*-mer values: 22,33,55,77 and library type: single reeds. The outputs were 5 files of SPAdes contigs (fasta), SPAdes contigs (stats), SPAdes scaffolds (fasta), SPAdes scaffold (stats) and SPAdes log.

2.7.1.7 QUAST 2.2 tool

To evaluate genome assemblies QUAST was used (Gurevich *et al.*, 2013). The input data was in fasta format. The tool accepts multiple assemblies, thus was suitable for comparison.

2.7.1.8 ResFinder - 2.1 tool

ResFinder (Zankari *et al.* 2012) can identify acquired antimicrobial resistance genes in complete or partial sequenced isolates of bacteria. The settings were 50% identity threshold, 60% minimal length and contigs as type of reads.

2.7.1.9 PHAge Search Tool (PHAST)

PHAST (Zhou *et al.*, 2011) was used to identify prophage sequences within bacterial genomes. The input data was in fasta format.

2.7.1.10 RAST for Bacterial Genome Rapid Automatic Annotation

The genomic contigs belonging to the strains used in this study were submitted in Rapid Annotation Subsystem Technology (RAST) (Aziz *et al.*, 2008). RAST is designed to rapidly call and annotate the genes of a complete or essentially complete prokaryotic genome. The input data was in fasta format. The settings were: Taxonomic ID = 1883, Domain = bacteria, Genus = *Streptomyces*, Species = undefined, Genetic code = 11, RAST annotation scheme = RASTtk, Gene caller = RAST, FIGfam version = Release 70. The output data was in Genbank (.gbk), EMBL (.embl) and GFF (.gff) formats.

2.7.1.11 antiSMASH (antibiotics & Secondary Metabolite Analysis SHell) tool

antiSMASH 3.0 tool (Blin *et al.*, 2013) was used to search for secondary metabolite biosynthesis gene clusters in sequenced genes. The input data was in GenBank format.

The settings were: search for Gene Cluster Blast analysis, Known Gene Cluster Blast analysis, Subcluster Blast analysis, smCOG analysis for functional prediction, Active site finder, Whole-genome PFAM analysis and Enzyme Commission (EC) number prediction.

2.7.1.12 ClusterFinder tool

ClusterFinder tool (Cimermancic *et al.*, 2014) was used to detect and extract putative biosynthetic gene clusters of unknown types in sequenced genomes. The input data was in GenBank format. The settings are same as in AntiSMASH 3.0 tool.

2.7.1.13 REALPHY 1.10 tool

REALPHY tool (Bertels *et al.*, 2014) was used to infer phylogenetic tree of isolate_99 with those of closely related *Streptomyces* strains based on whole genome sequences. The input data was in fasta format. The settings were minimum read length 200, minimum per base coverage 10, minimum PHRED quality 20 and the query mapped to each of the references via bowtie2 algorithm (Trapnell *et al.*, 2012).

2.7.1.14 Genome to Genome Distance Calculator (GGDC) tool

GGDC tool (Meier-Kolthoff *et al.*, 2013) was used to measure *in silico* DNA-DNA hybridisation (*is*DDH) values between isolate_99 and the most closely related strains based on whole genome sequences data. This tool supports both complete and draft genomes. The input data was in fasta format and GGDC 2 BLAST + alignment method was chosed for finding intergenomic matches.

2.7.1.15 Average Nucleotide Identity (ANI) (Goris et al, 2007)

ANI tool was used to estimate the average nucleotide identity between isolate_99 and the most closely related strains using both ANIb based on BLAST+ and ANIm based on MuMer options. This tool supports both complete and draft genomes. The input data was in fasta format. The settings were minimum length 200, minimum identity 50 and minimum alignment 50.

2.7.1.16 Proteome Comparison tool

Proteome comparison tool (Wattam *et al.*, 2014) was utilised to perform proteome comparison identity between isolate_99 and the most closely related strains. The tool runs a BLASTP comparison for CDSs identity which represented in colour intensities.

2.7.1.17 SEED viewer tool

SEED viewer tool (Overbeek *et al.*, 2014) was used for functional comparisons between isolate_99 and *S. turgidiscabies* and *S. graminilatus* genomes via defining genes associated with a functional role in a bacterial genome as a subsystem.

2.7.1.18 CLC Bio Genomics Workbench 8.0

CLC software version 8.0 (Qiagen Bioinformatics) was utilised for RNA-Seq data analysis. The RNA-Seq raw data in fastq format were imported into CLC Genomics workbench and analysed by using the tools: Trim, RNA Seq Analysis, and Differential expression analysis using the EGDE algorithm (Robinson *et al.*, 2010). The default settings were selected and the isolate_99 genome sequence was used as a track (reference) in a Genbank (gbk) format.

2.7.1.19 TopHat, Cufflinks and CummeRbund tools

TopHat and Cufflinks tools (Trapnell *et al.*, 2009; Trapnell *et al.*, 2012) were used for differential gene and transcript expression analysis of RNA-Seq experiment. The raw data files were downloaded in BAM format from the Ion Torrent Server and were uploaded onto the Galaxy platform (Afgan *et al.*, 2016). CummeRbund tool (Trapnell *et al.*, 2012) was utilised for visualising RNA-Seq analysis results.

2.7.2 Databases

2.7.2.1 GenBank database

The GenBank, National Centre for Biotechnology Information (NCBI) nr/nt database was searched with the query sequences of the 16S rRNA gene obtained from sequencing analysis using the Nucleotide Basic Local Alignment Search Tool (BLASTn) program

(http://blast.st-va.ncbi.nlm.nih.gov/Blast.cgi). Also, it was used to download complete or draft genome sequences which used in this study.

2.7.2.2 BLAST (Local Alignment Search Tool)

BLAST is an online software that searches for homology to the input template sequence against entire sequences which are available on the sequence database (Altschul *et al.*, 1990) , such as Genbank (www.ncbi.nlm.nih.gov/genbank/), which maintains at the National Centre for Biotechnology Information (NCBI) in the United States; the European Molecular Biology Laboratories (EMBL) (www.ebi.ac.uk/embl/), which maintains at the European Bioinformatics Institute (EBI) ; and the DNA Data of Japan (DDBJ) (www.ddbj.nig.ac.jp/), which maintains at the National Institute of Genetics (NIG) in Japan. The tool can find similarities in nucleotide database thorough BLASTN algorithm or protein database through BLASTP algorithm using nucleotide or protein query respectively. The tool also has the power for finding similarities genes and protein-coding genes through forms of the algorithms, BLASTX, TBLASTN, and TBLASTX because they identify and translating amino acid similarities in nucleotide sequences (Korf, 2003).

2.7.2.3 BacMet (Antibacterial Biocide & Metal Resistance Genes) database

The BacMet is a robust manually curated database for identifying of biocide and metal-resistance genes in proteins and DNA sequences through BLASTP and BLASTN respectively (Pal *et al.*, 2014). The database maintains experimentally verified resistance genes along with similar sequences genes predicted in public databases that have conserved function. It contains 704 experimentally confirmed resistance genes and 40556 predicted resistance genes. The database has two subdatabases: BacMet Experimentally Confirmed database and BacMet Predicted database. BacMet is maintained at the University of Gothenburg, Sweden.

2.7.2.4 Silva database

The Silva database (Quast *et al.*, 2013) was searched with the query sequences of the 16S and 23S rRNA genes obtained from sequencing analysis. The SILVA database is developed and maintained by the Microbial Genomics and Bioinformatics Research Group in Bremen, Germany, in cooperation with the Department of Microbiology at the Technical University Munich and the Ribocon GmbH.

2.7.2.5 List of Prokaryotic names with Standing in Nomenclature (LPSN) database

LPSN (http://www.bacterio.net) database was used to know the list of species in the genus of *Streptomyces*, which currently contains 793 validly described species, along with their 16S rRNA sequences. The database lists the names of prokaryotes (Bacteria and Archaea) that have been validly published in the *International Journal of Systematic and Evolutionary Microbiology* directly or by inclusion in a Validation List, under the Rules of International Code of Nomenclature of Bacteria (Parte, 2013).

Chapter 3:

Isolation, Identification, Lead (Pb) - metal Induction and Bioassay against ESKAPE Pathogens

3.1 Isolation

Our interest is focused on microorganisms belonging to the actinomycetaceae family, and specifically to the *Streptomyces* genus, the members of which have demonstrated important antimicrobial activity. Therefore, during the course of screening of new isolates for antimicrobial activities, many uninteresting microorganisms could be excluded due to their dissimilarities to actinomycete characteristics; thus, only representative isolates were finally selected for further purification. After a three-week incubation on HAV agar medium, bacterial colonies were isolated. Colonies were slow growing, aerobic, chalky, heaped, folded and with aerial and substrate mycelia of different morphologies. The isolated microorganisms were further subcultured, purified and maintained on MYM and MS media

3.2 Morphological Identification

To elucidate their growth characteristics, the candidates were grown on ISP-2 - ISP-7 media. It was observed that some of the isolates produced diffusible pigments in the surrounding medium; others produced reverse side pigments. Some isolates represented various colony and cell morphologies, and growth characteristics (Appendix A) on the different ISPs media. These properties are among many others recommended by Shirling and Gottlieb (1972) for the description of *Streptomyces*. Moreover, a wide variety of isolates possess an earthy odour which is known as geosmin (Pollak and Berger, 1996). However, the amount of growth, aerial mass colour, reverse side colour and soluble pigment production of any of the isolates appeared to be influenced by the medium composition, growth conditions, and age of culture. Pigment production has been widely used to identify and classify *Streptomyces* species (Kämpfer, 2012). It was recognised that morphological features, in particular aerial growth and soluble pigment colours of *Streptomyces* members are different from other actinomycetes. From the results, it was observed that biosynthesis of pigments related to the medium

composition used. Consequently, it has been suggested to grow *Streptomyces* on different culture media (Shirling & Gottlieb, 1966; Goodfellow *et al.*, 1983; Kieser *et al.*, 2000). Each medium used in our experiments has a different nitrogen and/or carbon source and this might have an effect on metabolic pathways of pigment excretion.

3.3 Stereoscopic Microscope Identification

Microscopically, it was observed that the morphology of the colony and aerial hyphal mycelium on HAV agar varied depending on the isolated strain (Figure. 3.1), which was used among other features to propose differences between the isolates. Microscopic examination is of high value in terms of *Streptomyces* determination and differentiation (Kämpfer, 2006). However, due to the high number of species (many novel) continuing to be isolated, it is highly recommended to identify, classify and discriminate the members of the genus *Streptomyces* on the basis of sequence analysis of 16S rRNA/DNA gene besides cultural and conventional methods.

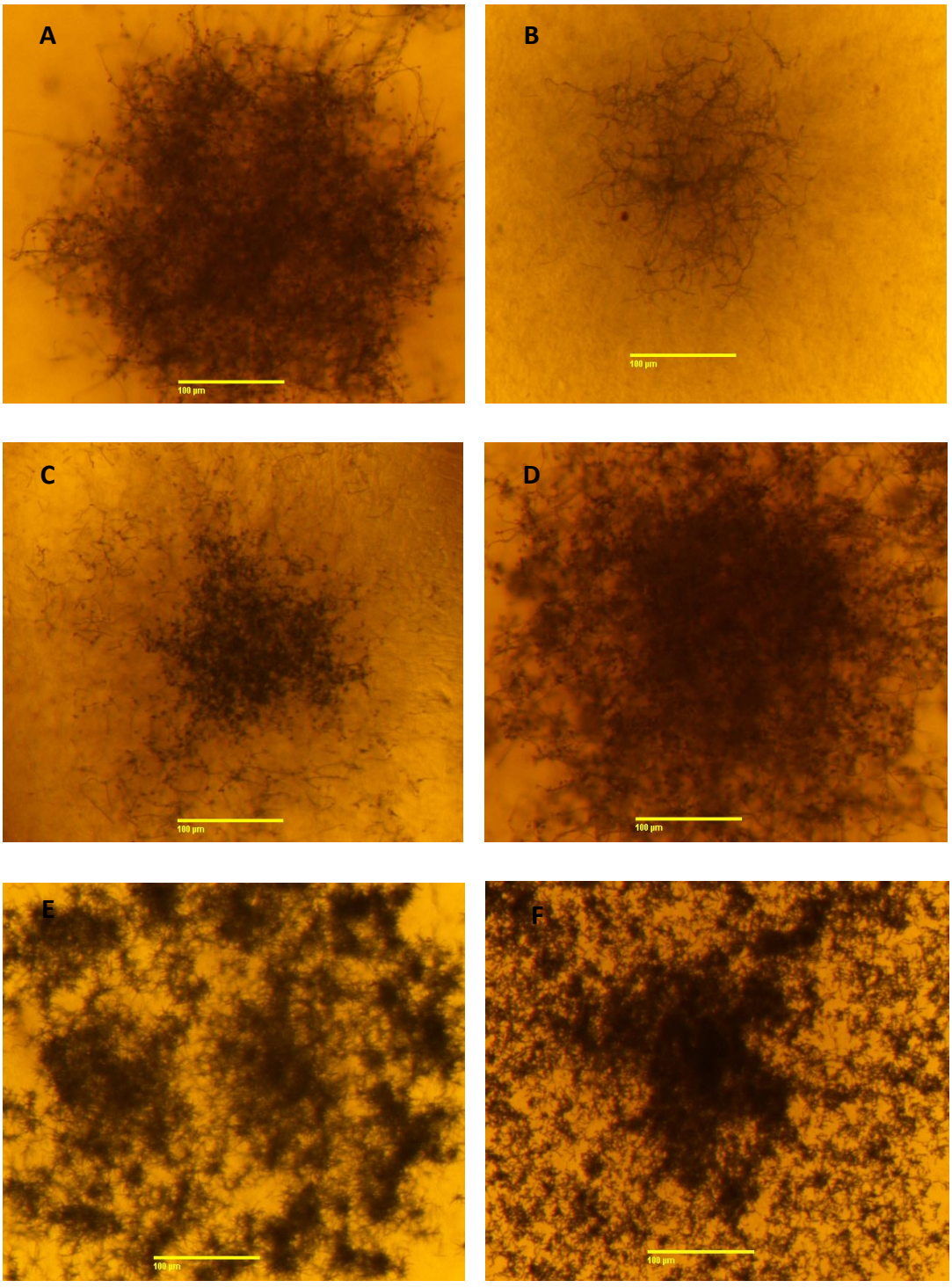


Figure 3.1 Example of six colonies (A-F) under a stereoscopic microscope (Nikon SMZ1500) showing the growth pattern and phenotype characteristics of some of isolates on HAV medium at 30°C for 7 days. Each isolate has different aerial hyphae phenotypic feature.

3.4 Molecular Identification

3.4.1 16S rRNA gene identification

In order to confirm that the isolated species belong to the Actinomycetaceae family, the 16S rDNA gene was amplified, sequenced and analysed. After gel electrophoresis, the size of the PCR fragment bands of 16S rRNA were 1500 bp (Figure 3.2). To find the most similar strains, the 16S rRNA sequences of all isolates were subjected to leBIBI ^{QBPP} system (Falndrois *et al.*, 2015) (Table 3.1). Since the universal primers used in 16S rRNA gene amplification were successfully amplified the control DNA (*Streptomyces coelicolor* M145) and no amplification occurred in the negative control (no DNA template), and produced PCR fragments of 1500bp in size (Figure 3.2), therefore, it is highly recommended to use these primers in detecting the genus of *Streptomyces* from pure cultures. The 16S rRNA is distributed and conserved among all bacteria (Chakravorty, 2007), for this reason it was chosen as the target gene for PCR amplification. Unlike other bioinformatics tools, BIBI promotes a phylogenetic approach to find the most similar strain. By convention, phylogeny is applied to confirm the 97% similarity among of very closely related sequences or to resolve marginal situations (Keller *et al.*, 2010).

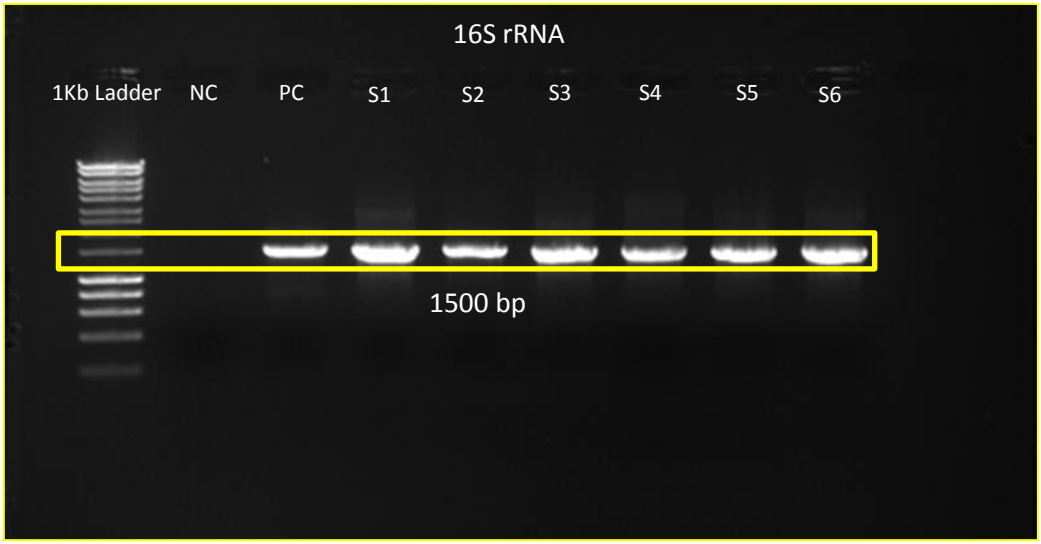


Figure 3.2: Agarose gel electrophoresis of PCR fragments obtained by amplification of the 16S rRNA gene from pure cultures of six isolates as an example (S1-S6). The bands from left are: 1Kb Hyperladder, negative control (NC), positive control (PC, *S. coelicolor* M145) and six isolates in order (S1-S6). All bands are 1500 bp in length as compared with the Hyperladder marker.

Briefly, a BLAST search (Camacho *et al.*, 2009) is applied by most of other alignment-based nucleotide sequence identification tools to give several of the closest hits that match to the query sequence. This, however does not mean that the closest BLAST hit is the nearest phylogenetic neighbour (Koski & Golding, 2001), and this may lead to strain misidentification (Keller *et al.*, 2010). However, to overcome this problem and for a more accurate identification, BIBI combines a BLAST search with the alignment of resulting similar sequences, and proposes an identification of the species of origin of the input sequence through phylogenetic reconstruction (Flandrois *et al.*, 2015). To improve the accuracy of query sequence identification, the references SSU-rDNA databases used by this system contain tags for sequences of Type strains (Table 3.1). This database is less susceptible to be contaminated by erroneous species identifications (Flandrois *et al.,* 2015).

Table 3.1: Sequences with the most closely related strain to each of the query sequence as determined by producing approximate Maximum Likelihood (ML) phylogenetic approach using leBIBI ^{QBPP} tool (Falndrois *et al.*, 2015).

Isolate No./ Query No.	Length of query sequence	Quality of the BLAST analysis	Biodiversity level	Closest strain
9	1028	Passed with 98.8% of the query length	Actinobacteria	*Streptomyces ederensis* T EU594481
11	1437	Passed with 96.2% of the query length	Actinobacteria	*Streptomyces umbrinus* T AB184305
12	1397	Passed with 99.2% of the query length	Streptomycetaceae	*Streptomyces subrutilus* T AB184372
13	1353	Passed with 99.2% of the query length	Actinobacteria	*Streptomyces yanii* T KF772677
14	1352	Passed with 99.4% of the query length	Streptomycetaceae	*Streptomyces sp*. T AF401982
15	1382	Passed with 97.1% of the query length	Actinobacteria	*Streptomyces turgidiscabies* T AB026221
16	1396	Passed with 98.9% of the query length	Actinobacteria-Pseudonocardiales	*Lentzea albida* T AB006176
17	1407	Passed with 97% of the query length	Actinobacteria	*Streptomyces turgidiscabies* T AF361782
18	1359	Passed with 98.7% of the query length	Streptomycetaceae	*Streptomyces anulatus* T AB184644
20	1359	Passed with 99.1% of the query length	Actinobacteria	*Streptomyces turgidiscabies* T AF361782
21	1133	Passed with 99.6% of the query length	Streptomycetaceae	*Streptomyces lydicus* T KF712383
22	1352	Passed with 99.7% of the query length	Streptomycetaceae	*Streptomyces subrutilus* T AB184372
24	1411	Passed with 98.2% of the query length	Streptomycetales	*Streptomyces albolongus* T AY999756
27	1376	Passed with 97.5% of the query length	Streptomycetaceae	*Streptomyces subrutilus* T AB184372
28	1368	Passed with 98.5% of the query length	Actinobacteria	*Streptomyces turgidiscabies* T AF361782
30	1399	Passed with 99.5% of the query length	Actinobacteria	*Streptomyces sp*. T JX983201
31	1305	Passed with 99.3% of the query length	Streptomycetales	*Streptomyces herbaricolor* T AB184212

32	1400	Passed with 97% of the query length	Streptomycetaceae	*Streptomyces rishiriensis* T AB184383
33	1358	Passed with 99.1% of the query length	Streptomycetaceae	*Streptomyces humidus* T AB184213
36	1394	Passed with 98.9% of the query length	Actinobacteria-Pseudonocardiales	*Lentzea violacea* T AJ242633
37	1354	Passed with 99.1% of the query length	Streptomycetaceae	*Streptomyces flavogriseus* T KF991636
38	1398	Passed with 99.2% of the query length	Streptomycetaceae	*Streptomyces subrutilus* T AB184372
39	1394	Passed with 98.9% of the query length	Actinobacteria-Pseudonocardiales	*Lentzea albida T AB006176*
40	1417	Passed with 99.3% of the query length	Streptomycetaceae	*Streptomyces avidinii* T AB184395
41	1452	Passed with 98.4% of the query length	Streptomycetaceae	*Streptomyces sp*. T KJ604921
42	1377	Passed with 99.4% of the query length	Streptomycetaceae	*Streptomyces goshikiensis* T EF178693
43	1027	Passed with 98.8% of the query length	Actinobacteria	*Streptomyces aureus* T KU198669
47	1353	Passed with 99.2% of the query length	Actinobacteria	*Streptomyces sanglieri* T AB249945
49	1390	Passed with 98.7% of the query length	Micromonosporaceae	*Micromonospora aurantiaca* T CP002162
50	1390	Passed with 98.7% of the query length	Micromonosporaceae	*Micromonospora matsumotoense* T AF152109
51	1418	Passed with 99.6% of the query length	Streptomycetaceae	*Streptomyces subrutilus* T AB184372
52	1390	Passed with 98.7% of the query length	Micromonosporaceae	*Micromonospora aurantiaca* T CP002162
53	1390	Passed with 98.8% of the query length	Pseudonocardiales	*Lentzea violacea* T AJ242633
54	1396	Passed with 99.7% of the query length	Pseudonocardiales	*Lentzea violacea* T AJ242633
55	1398	Passed with 96.5% of the query length	Streptomycetaceae	*Streptomyces humidus* T AB184213
56	1352	Passed with 100% of the query length	Streptomycetaceae	*Streptomyces mauvecolor* T AB184532
57	1000	Passed with 99.5% of the query length	Actinobacteria	*Streptomyces mirabilis* T AB184412
58	1413	Passed with 100% of the query length	Streptomycetaceae	*Streptomyces goshikiensis* T EF178693
59	1399	Passed with 99.3% of the query length	Streptomycetaceae	*Streptomyces_prunicolor* T AB184294
60	1352	Passed with 98.7% of the query length	Streptomycetaceae	*Streptomyces avidinii* T AB184395

63	1037	Passed with 97.8% of the query length	Streptomycetaceae	*Streptomyces_lydicus* T Y15507
64	1396	Passed with 99.2% of the query length	Streptomycetaceae	*Streptomyces cirratus* T AY999794
69	1352	Passed with 99.7% of the query length	Streptomycetaceae	*Streptomyces pratensis* T JQ824045
70	1365	Passed with 99.8% of the query length	Streptomycetaceae	*Streptomyces subrutilus* T AB184372
71	1028	Passed with 99.3% of the query length	Streptomycetaceae	*Streptomyces aureus* T KU198669
73	1214	Passed with 100% of the query length	Actinobacteria	*Streptomyces albus* T AY99987
74	1430	Passed with 99.5% of the query length	Streptomycetaceae	*Streptomyces avidinii* T AB184395
75	1405	Passed with 100% of the query length	Streptomycetaceae	*Streptomyces goshikiensis* T EF178693
76	1364	Passed with 98.5% of the query length	Streptomycetales	*Streptomyces paucisporeus* T AY876943
77	1359	Passed with 98.5% of the query length	Actinobacteria	*Streptomyces turgidiscabie* T AF361782
78	1359	Passed with 98.7% of the query length	Streptomycetaceae	*Streptomyces humidus* T AB184213
79	1431	Passed with 97.7% of the query length	Streptomycetaceae	*Streptomyces_sp*. T AB856295
80	1344	Passed with 98.8% of the query length	Streptomycetaceae	*Streptomyces goshikiensis* T EF178693
81	1433	Passed with 99.4% of the query length	Streptomycetaceae	*Streptomyces_sp*. T AF401982
82	1395	Passed with 99.8% of the query length	Streptomycetaceae	*Streptomyces avidinii* T AB184395
83	1404	Passed with 97.7% of the query length	Actinobacteria	*Streptomyces turgidiscabies* T AF361782
85	1533	Passed with 99.1% of the query length	Streptomycetaceae	*Streptomyces_humidus* T AB184213
87	1397	Passed with 98.0% of the query length	Actinobacteria-Propionibacteriales	*Kribbella albertanoniae* T KC283016
88	1390	Passed with 98.9% of the query length	Actinobacteria-Pseudonocardiales	*Lentzea violacea* T AJ242633
89	1020	Passed with 98.9% of the query length	Streptomycetaceae	*Streptomyces_praecox* T AY999853
90	1116	Passed with 97.7% of the query length	Streptomycetaceae	*Streptomyces olivochromogenes* T AY094370
92	1514	Passed with 99.8% of the query length	Streptomycetaceae	*Streptomyces olivochromogenes* T AB184737
93	1351	Passed with 99.7% of the query length	Streptomycetaceae	*Streptomyces flavogriseus* T KF991636

94	1352	Passed with 99.4% of the query length	Streptomycetaceae	*Streptomyces_subrutilus* T AB184372
96	1353	Passed with 99.2% of the query length	Streptomycetaceae	*Streptomyces flavogriseus* T KF991636
99	1533	Passed with 99.7% of the query length	Streptomycetaceae	*Streptomyces turgidiscabies* T AB026221
100	1209	Passed with 98.3% of the query length	Streptomycetaceae	*Streptomyces_canus* T AY999775
101	1435	Passed with 99.5% of the query length	Streptomycetaceae	*Streptomyces subrutilus* T AB184372
102	1399	Passed with 98.4% of the query length	Streptomycetaceae	*Streptomyces aureus* T KU198669
103	1377	Passed with 97.1% of the query length	Streptomycetaceae	*Streptomyces subrutilus* T AB184372
104	1019	Passed with 99.6% of the query length	Streptomycetaceae	*Streptomyces praecox* T AY999853
105	1353	Passed with 99.7% of the query length	Streptomycetaceae	*Streptomyces_subrutilus* T AB184372
106	1373	Passed with 99.4% of the query length	Streptomycetaceae	*Streptomyces flavogriseus* T KF991636
108	1403	Passed with 97.5% of the query length	Streptomycetaceae	*Streptomyces geldanamycininus* T DQ334781
109	1393	Passed with 99.3% of the query length	Streptomycetaceae	*Streptomyces cirratus* T AY999794
110	1430	Passed with 97.5% of the query length	Actinobacteria	*Streptomyces_graminilatus* T HQ268006
111	1028	Passed with 99.3% of the query length	Streptomycetacea	*Streptomyces olivochromogenes* T AY094370
112	1377	Passed with 97.6% of the query length	Streptomycetaceae	*Streptomyces cirratus* T AY999794
113	1352	Passed with 99.2% of the query length	Streptomycetaceae	*Streptomyces_flavogriseus* T KF991636
114	1357	Passed with 99.4% of the query length	Actinobacteria	*Streptomyces thermocarboxydus* T AB249926
115	1032	Passed with 98.1% of the query length	Actinobacteria	*Streptomyces glomeroaurantiacus* T AB249983
116	1403	Passed with 97.7% of the query length	Bacteria	*Nocardioides albus* T AF004988
117	1199	Passed with 99.4% of the query length	Streptomycetacea	*Streptomyces aureus* T KU198669
118	1352	Passed with 99.7% of the query length	Streptomycetacea	*Streptomyces flavofuscus* T AB249935
119	1397	Passed with 99.2% of the query length	Streptomycetaceae	*Streptomyces griseochromogenes* T AJ310923
120	1263	Passed with 99.2% of the query length	Actinobacteria	*Streptomyces bangladeshensis* T AY750056

3.4.2 16S rRNA gene phylogenetic tree

An unrooted phylogenetic tree of all isolates was categorised into five main clades on the basis of their evolutionary distances calculated through the neighbour joining method (Figure 3.3). The first clade (indicated by red circles) was further divided into three sub clades, the first sub clade was further divided into isolates No. 54, 53, 88, 36, 39 and 16; the second sub clade was also divided into isolates No. 50, 52 and 49 leaving the third sub clade with two isolates No. 116 and No. 87. Table 3.1 shows the closet strain for each query sequence determined by BIBI database, the isolates No: 54, 53, 88, 36, 39, 16, 50, 52, 49, 116, and 87 are most closely related to the reference type strains: *L. violacea* T AJ242633, *L. violacea* T AJ242633, *L. violacea* T AJ242633, *L. violacea* T AJ242633, *L. albida* T AB006176*, L. albida* T AB006176, *M. matsumotoense* T AF152109, *M. aurantiaca* T CP002162, *M. aurantiaca* T CP002162, *N. albus* T AF004988 and *K. albertanoniae* T KC283016 respectively. Based on these results, although these isolates mentioned above belong to the same clade (Figure 3.3), they belong to different genera of actinobacteria (Table 3.1). Isolates No. 54 and 53 are close to each other than isolates No. 88 and 36 despite their belonging to the same reference type strain (*L. violacea* T AJ242633). Similarly, isolates No. 52 and 49 (*M. aurantiaca* T CP002162) are more close together than isolate No. 50 (*M. matsumotoense* T AF152109) although they belong to the same sub clade. Conversly, isolates No. 116 (*N. albus* T AF004988) and 87 (*K. albertanoniae* T KC283016) belong to different genera although they are clustered together in the same sub clade. However, because of the highly conserved nature of the 16S rRNA gene, it is not ideal to infer the relationships of the individual clades with each other (Labeda *et al.*, 2012) therefore, genes that exhibit higher levels of variations from the 16S rRNA gene, and notably the housekeeping and/or 23S rRNA are a valuable alternative for inferring the phylogenetic tree and distinguishing between the members of each clad (Han *et al.* 2012; Labeda *et al.*, 2012; Rong & Huang 2012; Labeda *et al.*, 2012).

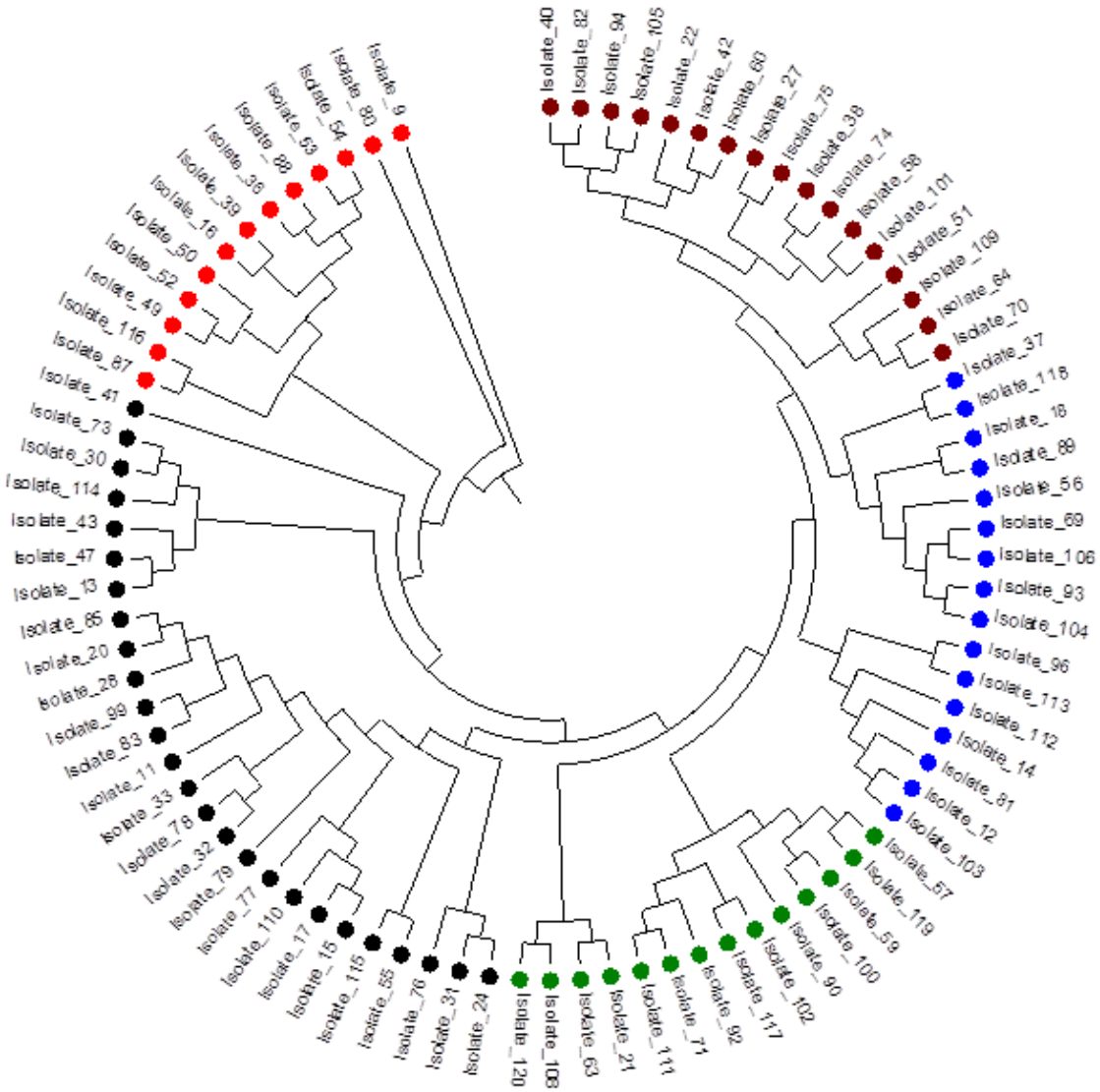


Figure 3.3: Unrooted Neighbour-Joining (NJ) phylogenetic tree based on almost complete 16S rRNA gene sequences (1000 X BOOTSTRAP). Circular tree showing the relationship between all strains isolated in this study using MEGA6 software (Tamura *et al.*, 2013). Clades 1, 2, 3, 4 and 5 are indicated by red, black, green, blue and brown circles respectively.

3.5 Lead (Pb) – metal induction of bioactive secondary metabolites

The information available for exploiting heavy metals for induction of silent gene clusters in actinomycetes and in particular *Streptomyces* to produce bioactive metabolites is very scarce. However, a study carried out by Haferburg *et al.* (2009) showed that compounds extracted from some *Streptomyces* strains grown in the presence of nickel and cadmium metals displayed antimicrobial activity against *Escherichia coli*, *Mycobacterium smegmatis*, *Staph. aureus* and *Candida albicans*. These metal-inducible *Streptomyces* strains were isolated by the authors from the former uranium mining area in Thuringia, Germany.

The Leadhills and Wanlockhead (the highest village in Scotland) sites were chosen to collect our samples in this study. Many minerals had been mined there for many centuries. The lead was mined in the area as early as Roman times (Mines Research Group Committee, 1978).

During the course of screening the isolated *Streptomyces* strains for heavy metal resistance, two of these isolates could be shown to produce bioactive metabolite (s) when they were grown in presence of lead. These unknown compounds inhibited the growth of the neighbouring cultures (Figure 3.4). These two promising isolates, Isolate_85 and Isolate_99, were further investigated and showed an inhibition effect against *C. albicans* and *E. coli* pathogens respectively. To determine the maximum concentration of lead metal that allowed growth of the candidates and gave the maximum effect against tested microorganisms, the antibiotic effect was evaluated by measuring the zones of inhibition when they were grown on MM agar supplemented with different concentrations of lead ranging from 0.5 to 2.0 mM (Table 3.2). Additionally, eight out of all isolates were also found to be induced to produce bioactive secondary metabolites when they were grown on minimal medium supplemented with 1-2 mM Pb (Table 3.3).

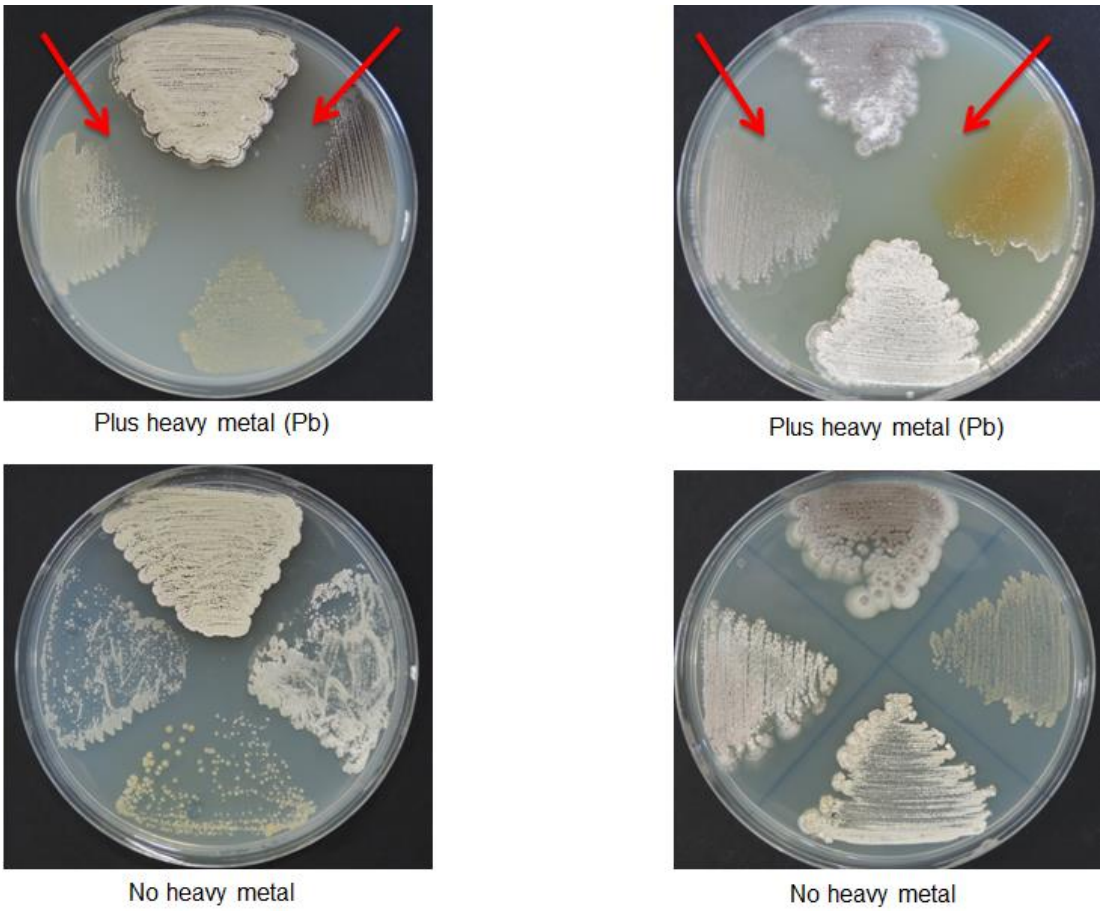


Figure 3.4: Initial evidence for metal induced antibiotic production (areas near to the red arrows. Isolate_85 (left-top) and Isolate_99 (right-top) partially inhibited the growth of the neighbouring *Streptomyces* strains on MM agar supplemented with 2.0 mM of Pb. MM agar without Pb (bottom) for both isolates was used as a control.

Table 3.2: This table shows isolate_ 85 and isolate_99 activity inhibition zones against *Candida albicans* and *E. coli* respectively when they were grown on MM agar supplemented with different concentration of Pb ranging from 0.5 to 2.0 mM. MM agar without isolates (last column) but same concentrations of Pb was used as a control.

Isolate_85		
Pb concentration (mM)	*Candida albicans* / Inhibition zone (mm)	Control / Inhibition zone (mm)
0.0	0.0	0.0
0.5	0.0	0.0
1.0	**17**	0.0
1.5	**17**	0.0
2.0	**21**	0.0
Isolate_99		
Pb concentration (mM)	*E. coli* / Inhibition zone (mm)	Control / Inhibition zone (mm)
0.0	0.0	0.0
0.5	0.0	0.0
1.0	**12**	0.0
1.5	**18**	0.0
2.0	**11**	0.0

Table 3.3: Isolates that showed lead (Pb) - metal induction of bioactive secondary metabolites.

Isolate No.	**Test microorganisms**	**Inhibition zones (mm) control / Pb**	**Source**
99	*E. coli*	0 / 18	Leadhills / sediment
85	*C. albicans*	0 / 21	Leadhills / sediment
73	*C. albicans*	0 / 17	Wanlockhead / sediment
22	*C. albicans*	0 / 15	Wanlockhead / sediment
42	*C. albicans*	0 / 15	Wanlockhead / sediment
58	*C. albicans*	0 / 14	Leadhills / sediment
76	*Staph. aureus*	0 / 18	Wanlockhead / Moss
15	*Staph. aureus*	0 / 13	Leadhills / riverside

Isolate_85 and isolate_99 were further investigated to determine if they were also able to be induced with previously identified antibiotic inducers. These two isolates were grown on MM supplemented with different concentrations of N-acetylglucosamine (GlcNAc) as carbon and nitrogen sources ranging from 5 to 50 mM. Also, these isolates were grown on MM supplemented with different concentrations of sodium butyrate (NaBu) ranging from 25 to 200 mM. These two experiments did not agree with both Rigali *et al* (2008), who have shown that GlcNAc stimulated antibiotic production of *Streptomyces* under famine conditions of optimum induction at 10 mM, and recently Moore *et al* (2012) who found that adding NaBu to MS medium stimulated antibiotic production in some non-*Streptomyces* microorganisms and enhanced ACT pigment production in *S. coelicolor* A3(2) as a *Streptomyces* model of optimal induction at 100 mM (Figure 3.5). In a parallel experiment, sodium acetate ($C_2H_3NaO_2$) was added to MM medium with the same concentrations of lead (II) acetate [$Pb(CH^3COO)_2$] that induced isolate_85 and isolate_99. This experiment was performed to see whether the acetate has an effect of induction of biosynthesis because of low pH. It has been shown that sodium acetate had no effect to stimulate these two strains for bioactive molecule production (Figure 3.6 A).

Gene cluster induction method	Bioactivity Effect	Reference
Isolate/MM Agar	—	—
Pb	+	Present study
GlcNAc	—	Rigali *et al.* (2008)
NaBu	—	Moore *et al.* (2012) *Pseudonocardia, Saccharopolyspora* & *S. coelicolor*

Figure 3.5: Comparison of Pb with other well-known antibiotic production methods. Pb clearly triggers antibiotic production against *Candida albicans* (left) and *E. coli* (right) whereas GlcNAc and NaBu had no effect. Pb alone had no effect on *C. albicans* and *E. coli* growth at the concentrations used in this study.MM + 1.5 mM and MM + 2.0 mM of Pb without strains were used as control.

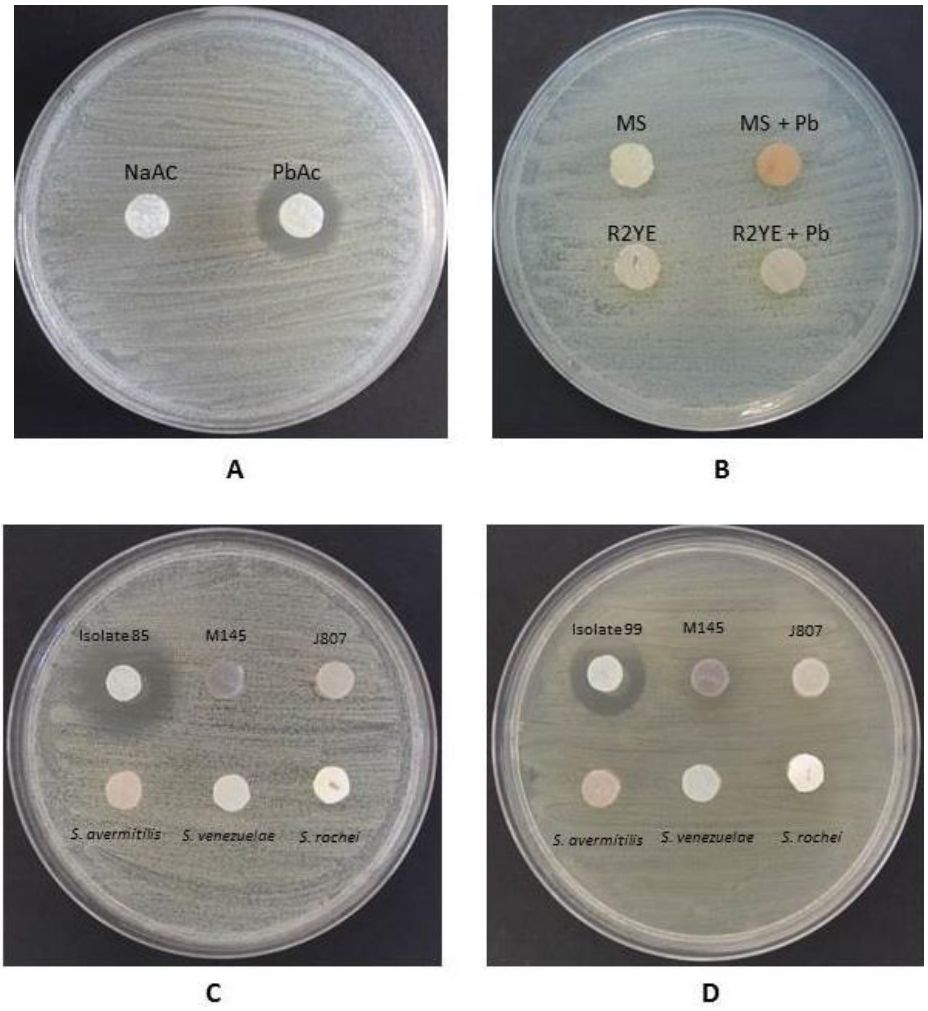


Figure 3.6: (A): comparison of adding lead acetate (PbAc) and sodium acetate (NaAc) to induced isolate_85 for active metabolites against *C. albicans*. NaAc had no effect to stimulate this strain for bioactive molecules production. (B): comparison of Pb induction in MM medium versus complex media. MS and MS plus Pb, R2YE and R2YE plus Pb were conducted to induce isolate_85 for active metabolites against *C. albicans*. Isolate_85 and could not produce bioactive metabolites neither without nor with adding Pb. (C) and (D): *Streptomyces* reference: *S. coelicolor* M145, *S. albus* J807, *S. avermitilis*, *S. venezuelae* and *S. rochei*. None of the reference strains had effects on *C. albicans* (C) and *E. coli* (D) growth at the concentrations used in this study. Agar plugs of each *Streptomyces* cultures grown on different concentrations of inducers plus a control plug containing only agar medium, were taken under sterile conditions using a cork borer, and placed over LB agar containing the test pathogens.

3.5.1 Pb induction in MM medium versus complex media

In order to examine the effect of complex media with high nutrient levels on stimulation of these isolates to produce bioactive metabolites under lead stress, MS agar and R2YE media were used. R2YE, the most rich and complex medium, is widely used in experiments for antibiotics and bioactive metabolites production from *Streptomyces* strains (Kieser *et al.*, 2000; Rigali *et al.*, 2008; Shepherd *et al.*, 2010). From the results (Figure 3.6 B), it was observed that isolate_85 and Isolate_99 could not produce bioactive metabolites either with or without adding the inducer.

3.5.2 Bioactivity of isolates versus reference strains

Here, we carried out an experiment to compare the same conditions of lead induction with some *Streptomyces* reference strains: *S. coelicolor* M145*, S. albus J*807*, S. avermitilis, S. venezuelae and S. rochei.* There is no evidence that these *Streptomyces* species were originally isolated from heavy metal contaminated areas. From the results (Figure 3.6 C & D), none of these reference microorganisms were able to be affected by the metal induction to produce active metabolites that inhibits the growth of tested microorganisms.

3. 6 Bioassay against ESKAPE pathogens

It was worthwhile to test these actinomycetes for their antimicrobial activity against ESKAPE pathogens without adding any activators. The ESKAPE pathogens are emerging pathogens of concern which represent clinically-relevant bacteria from both nosocomial and environmental sources (Rice, 2008). These pathogens belong to the group known as multi-drug resistant bacteria that exhibit resistance towards β-lactamases, vancomycin, methicillin, aminoglycosides, macrolides and fluoroquinolones antibiotics (Pendleton *et al.*, 2013). For these reasons, the discovery of novel antibiotics with new mechanism of action that are able to antagonise the growth of these pathogens would have important implications for the treatment of infectious diseases and medicine as a whole.

It is well known that antimicrobial compounds derived from *Streptomyces* have less effect on Gram-negative than Gram-positive bacteria. The double cell membranes present in Gram-negative bacteria play an important role to protect these organisms against antibiotics (Gupta, 2011). Although the antimicrobial activity of compounds derived from *Streptomyces* strains against Gram-positive bacteria has been reported in several studies, only a few have reported activity against Gram-negative bacteria (Higginbotham & Murphy, 2010; Bouras *et al.*, 2013; Aouiche *et al.*, 2014).

As illustrated in Table 3.4, some of isolates were found to be biologically active against Gram-negative and some against the Gram-positive ESKAPE pathogens. Interestingly, Isolate_56 was able to produce an active metabolite which has strong activity against *E. coli, K. pneumoniae* and *P. aeruginosa.* This unknown metabolite (s) was produced when this *Streptomyces* strain was grown on ISP-2, ISP-3, ISP-4 and ISP-6 against both *E. coli* and *K. pneumoniae,* and only on ISP-6 against *P. aeruginosa.*

Table 3.4: Overview of bioassay using agar plug technique of actinomycete isolates grown on ISP media that exhibited antimicrobial activity against ESKAPE pathogens

Escherichia coli ATCC 25922		*Staphylococcus aureus* ATCC 43300		*Klebsiella pneumoniae* ATCC 700603		*Acinetobacter baumannii* ATCC 19606		*Pseudomonas aeruginosa* ATCC 27853		*Enterococcus faecalis* ATCC 51299	
Isolate No.	ISP-Medium No. (Type)	Isolate No.	ISP-Medium No. (Type)	Isolate No.	ISP-Medium No. (Type)	Isolate No.	ISP-Medium No. (Type)	Isolate No.	ISP-Medium No. (Type)	Isolate No.	ISP-Medium No. (Type)
56	2, 3, 4 & 6	9	3, 4, & 5	56	2, 3, 4 & 6	17	4	56	6	11	5
		11	3 & 4			69	4	63	6	17	2
		19	4			96	4	118	6	38	5
		20	5							50	5
		31	3							70	5
		50	6							73	2
		54	4							76	5
		59	4 & 5							80	5
		63	2 & 5							100	2
		69	4							103	5
		73	2, 3, 4 & 5							108	2

		76	2, 3, 4 & 5								
		96	5								
		100	2, 3, 4 & 5								
		108	2, 3, 4 & 5								
		110	4								
		115	4								

Similarly, isolate 63 and isolate 118 were also found to inhibit the growth of *P. aeruginosa* when they were grown on ISP-6 medium. The growth of *A. baumannii* was inhibited by isolates 17, 69 and 96 when they were grown on ISP-4. The number of isolates that exhibited inhibitory activity against Gram-positive ESKAPE bacteria was much higher than against Gram-negative pathogens. Seventeen and eleven isolates were able to inhibit the growth of *Staph. aureus* and *E. faecalis* respectively.

The *Streptomyces* isolates also displayed a different range of activity based on the ISP medium they were grown on. As shown in Figure 3.7, a higher number of strains presented activity against Gram-positive ESKAPE bacteria when grown on ISP-4 (12 isolates), ISP-5 (8 isolates), ISP-3 (7 isolates) and ISP-2 (3 isolates) against *Staph. aureus*, while ISP-5 (7 isolates) and ISP-2 (4 isolates) against *E. faecalis*. At the same time no strains grown on ISP-6 and/or ISP-7 were active against *Staph. aureus* and/or *E. faecalis*. Conversely, five isolates were found to be active against Gram-negative ESKAPE pathogens which were grown on ISP-6, one isolate against *E. coli*, one isolate against *K. pneumoniae* and three isolates against *P. aeruginosa. A. baumannii* was more sensitive to ISP-4 than other media. None of the isolates were active against ESKAPE pathogens when grown on ISP-7. From these results, the isolates were found to be stimulated to produce secondary metabolites after small variation in culture media composition. Amino acids are described as potential inducers of secondary metabolites under the One Strain−Many Compounds (OSMAC) approach (Höfs *et al.*, 2000; Bode *et al.*, 2002; Christian *et al.*, 2005; Chai *et al.*, 2012). In this regard, isolates 17, 69 and 96 were able to produce a bioactive metabolite which inhibited *A. baumannii* when they were propagated on ISP-4 medium which contains starch and ammonium sulphate as sources for carbon and nitrogen respectively. Similarly, the growth of *P. aeruginosa* was inhibited by isolates 56, 63 and 118 when they were grown on ISP-6 medium which contains glycerol and L-Asparagine as carbon and nitrogen sources respectively. From this it can be deduced that the carbon and nitrogen sources used in ISP-4 and ISP-6 media seemed to play an important role in induction of the silent gene clusters in these promising isolates. In a study performed by Rateb *et al.* (2011) on *Streptomyces sp*. strain C34, three new compounds which exhibited potent activity against Gram-positive bacteria were identified by application

of an OSMAC approach within the same strain using variation of the cultivation media. However, for evaluation of secondary metabolite production and to avoid the complexity of the media compositions, biosynthesis of secondary metabolites of this strain has to be screened on minimal medium supplemented with different carbon and/or nitrogen sources separately.

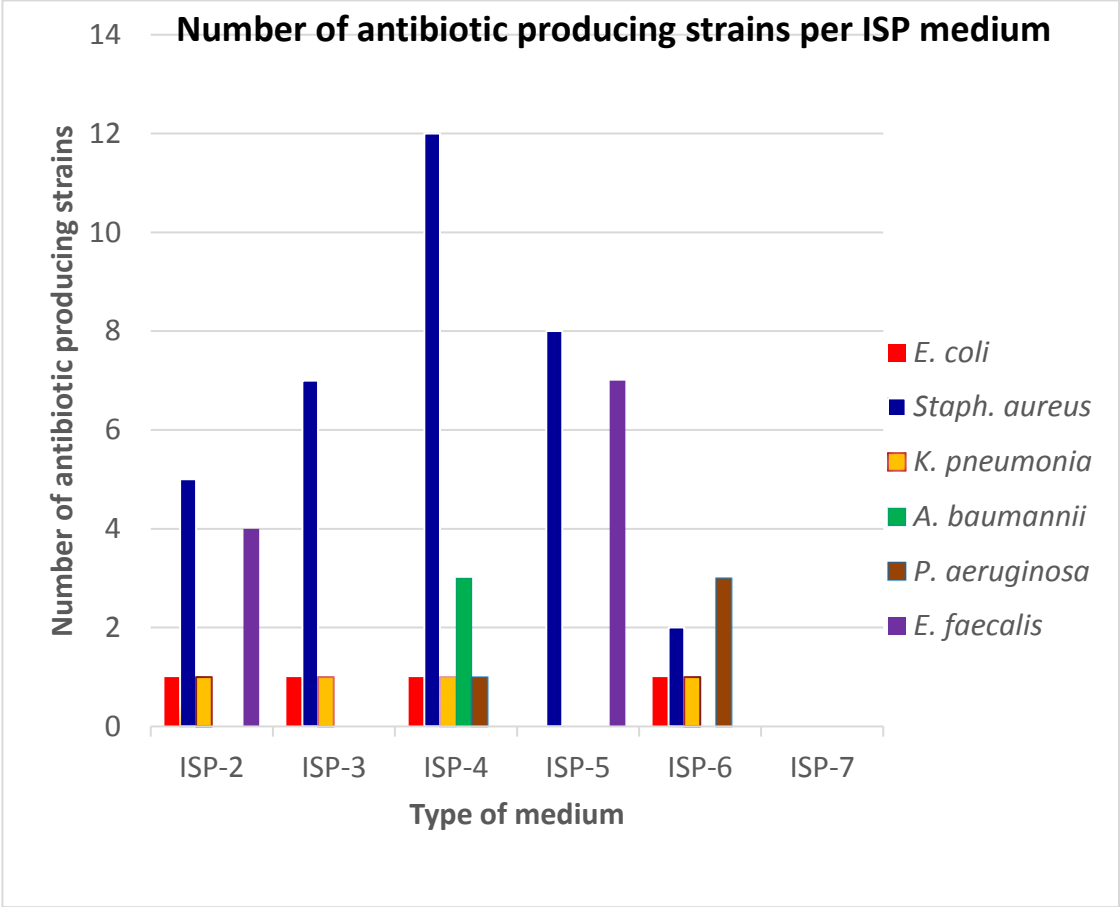


Figure 3.7: Number of actinomycetes strains exhibiting antibiotic activity against ESKAPE pathogens per ISP medium.

Chapter 4:

Whole Genome Sequencing (WGS) of Isolates 56, 85, 92, 99 and 118 using Next Generation Sequencing (NGS) Ion Torrent PGM Technology

4.1 Next generation sequencing system technologies

DNA sequencing is an experimental process of determining the order of nucleotides in a given DNA molecule. Precise knowledge about the DNA sequence of entire genomes is paramount in identifying genes, transcripts, and proteins downstream, and consequently in elucidating biochemical processes taking place in a given organism (Herzyk, 2014). Different sequence technologies have been devised during the last decades. Sanger sequencing with the use of chain-terminating dideoxynucleotide triphosphates (ddNTPs) that caused base-specific termination of primed DNA synthesis (Sanger & Coulson, 1977), was the first and the gold standard method in DNA sequencing.

In the beginning of this century, several sequence platforms have been devised, which differ from each other in terms of read length, data produced, and data quality. These systems are referred to as next generation sequencing (NGS) technologies, which employ massively parallel sequencing of millions of DNA fragments simultaneously (Gupta & Gupta 2014). The most popularly used platforms are the pyrosequencing technology by Roche 454 (Margulies *et al.*, 2005), the reversible terminator technology by Illumina (Turcatti *et al.*, 2008), the oligo ligation technology by SOLiD (Valouev *et al.* 2008), the ion semiconductor technology by Ion Torrent (Rusk, 2011) and the single molecule real-time technology by PacBio (Mardis, 2011).

4.2 Ion Torrent PGM Sequence Technology

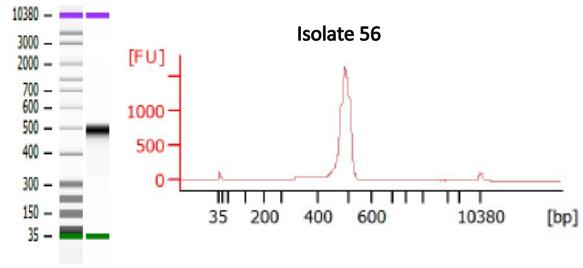
The Ion Torrent sequencer uses a novel detection approach based on ion detection system (Rusk, 2011), in which the information comes from releasing a hydrogen ion (H^+) when a nucleotide is incorporated into the DNA molecules by the polymerase. Ion Torrent, with its Ion Personal Genome Machine (PGM™) sequencer, recognises whether the nucleotide is added or not by shifting in pH that scales with the number of nucleotides incorporated (Rothberg *et al.*, 2011). The built-in hypersensitive

sensors in the ion-semiconductor sequencing chip are designed to distinguish and detect the changes in pH (chemical event) as a result of a nucleotide incorporation into a DNA strand and release of a hydrogen ion, and convert into voltage (digital event). Each time the chip is flooded with one nucleotide after another, if a nucleotide is not a match for a particular template, no voltage change will be detected and no base will be called for that template. On the other hand, if there are two identical bases on the DNA strand, the voltage is doubled, and the chip records two identical base calls (Rothberg *et al.*, 2011; Gupta & Gupta 2014).

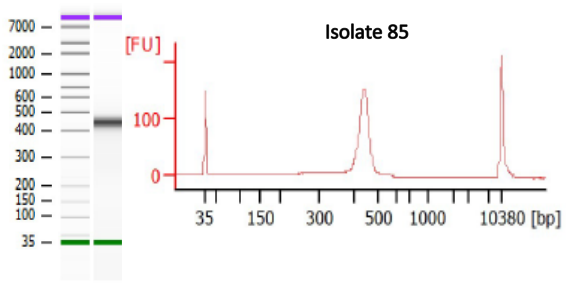
4.3 Library size distribution using Agilent high sensitivity DNA Chip

Following size selection of the libraries and after PCR amplification of adaptor ligated DNA, the amplified libraries were validated with Agilent Bioanalyzer 2100 using an Agilent high sensitivity DNA chip. This measures DNA size fragments distribution in base pairs (bp), concentration in pg/µl and molarity in pmol/l which are required to calculate the DNA input for pooling the final libraries for whole genome sequencing. Figure 4.1 below shows the output from the Agilent Bioanalyzer 2100, in these five samples, the middle peak represents the DNA fragment size distribution, concentration and molarity of DNA in each sample. The other two peaks represent DNA markers.

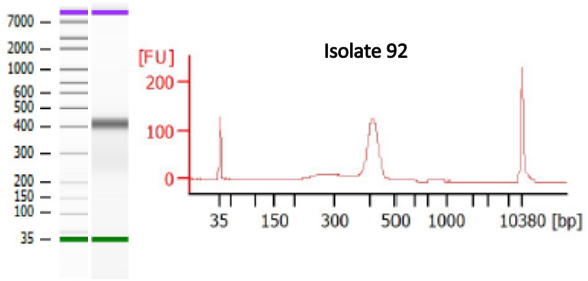
The isolates 56, 85, 92, 99 and 118 were selected for whole genome sequencing on the basis of their distinctive properties. Isolates 85 and 99 secreted unknown bioactive compounds affecting the growth of the neighbouring *Streptomyces spp*. cultures and exhibited a very strong activity against C. *albicans and E.coli* respectively when they were grown on MM agar supplemented with lead metal (discussed in chapter 3). Isolates 56 and 118 were candidates of *Streptomyces* strains that showed bioactivity against Gram negative ESKAPE pathogens (discussed in Chapter 3). Isolate 92 showed a strong resistance towards zinc (Zn) metal. It was found to grow in up to 55 mM $ZnCl_2$.



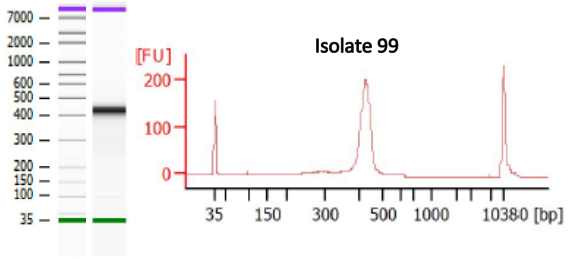
Region table for isolate_56		
Average library size (pb)	Conc. [pg/µl]	Molarity [pmol/l]
490	7,480.26	23,961.8



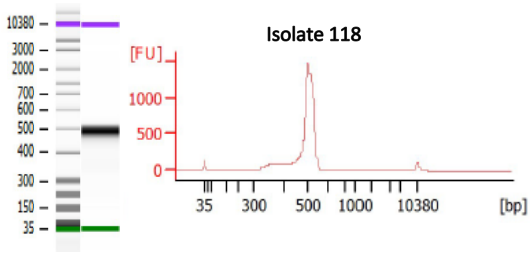
Region table for isolate_85		
Average library size (pb)	Conc. [pg/µl]	Molarity [pmol/l]
445	350.66	1,268.1



Region table for isolate_92		
Average library size (pb)	Conc. [pg/µl]	Molarity [pmol/l]
415	403.17	1,641.5



Region table for isolate_99		
Average library size (pb)	Conc. [pg/µl]	Molarity [pmol/l]
430	453.45	1,711.6



Region table for isolate_118		
Average library size (pb)	Conc. [pg/µl]	Molarity [pmol/l]
511	7,863.78	25,154.8

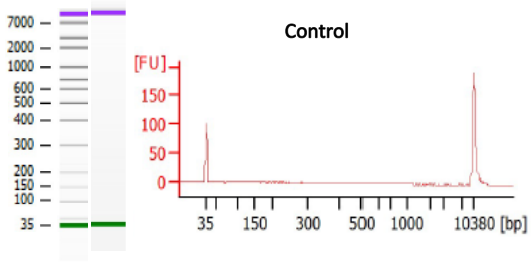


Figure 4.1: Output from the Agilent Bioanalyzer 2100. In these five samples, high quality DNA libraries were obtained. The middle peak represents the DNA fragment size distribution. Isolates 56, 85, 92, 99 and 118 had libraries sizes 490, 445, 415, 430 and 511 respectively. Also, it was possible to measure concentration and molarity of DNA in each sample. The other two peaks represent DNA markers. In the control sample no peak was detected.

4.4 Bead loading percentage

The loading percentage of bead sequencing chip is the percentage of wells that contain templated Ion Sphere Particles (ISPs). This % value considers only the potentially addressable chip wells and is a result of the software well classification step. A pseudo colour image of the Ion chip is generated showing percentage loading across the physical surface. The ISPs loading percentage for isolates 85, 92 and 99 was 91% (Figure 4.2), whereas the ISPs loading percentage of isolate 56 and 118 was 75% (Figure 4.3). Both obtained percentages are highly accepted as the minimum threshold of loading is ≥30%. However, the loading percentage might be affected by many factors. From the results, it has been seen that inserted libraries larger than 470bp (maximum recommended length of the 400pb sequencing) for downstream steps resulted in low loading percentage as in isolate 56 and 118 where the input libraries for these two strains were 490 and 511pb respectively (Figure 4.1). This means that those two libraries did not amplify sufficiently on to the ISPs during template preparation and may result in low loading percentage (Figure 4.3) compared to loading percentage of isolates 85, 92 and 99 (Figure 4.2).

Low templated ISPs can also affect the loading percentage. According to the user guide by Life Technologies (Pub. no. MAN0007218), the optimal template ISPs percentage between 10-30% of that recovered from Ion One Touch-2 (OT2) instrument. Although our samples fall within the recommended range values, 16% and 27% for isolates 85, 92 and 99 and isolates 56 and 118, samples with low percentage templated ISPs (<10%) can lead to low loading density on the Ion chip and consequently can lead to lower key signal and lower throughput, while samples with high template ISPs (>30%) can lead to increase number of filtered reads and lower throughput.

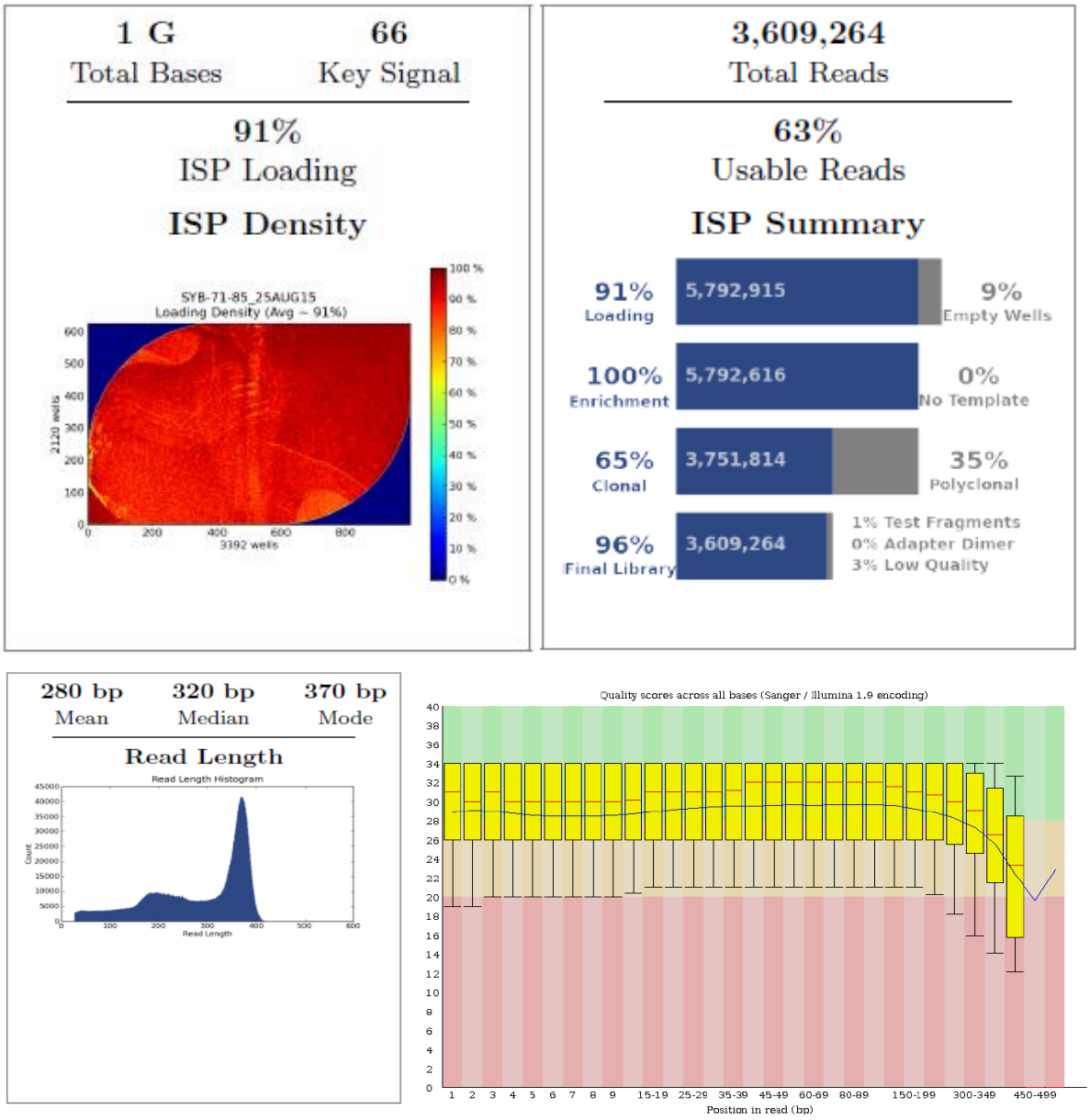


Figure 4.2: Summary report of the whole genome sequencing run of isolates 85, 92 & 99: the ISP density percentage along with total bases and key signal; total reads along with enrichment, clonal and polyclonal and final library values; the trimmed read length; and the range value plot for accuracy and quality scores across all bases at each position in the FastQC file.

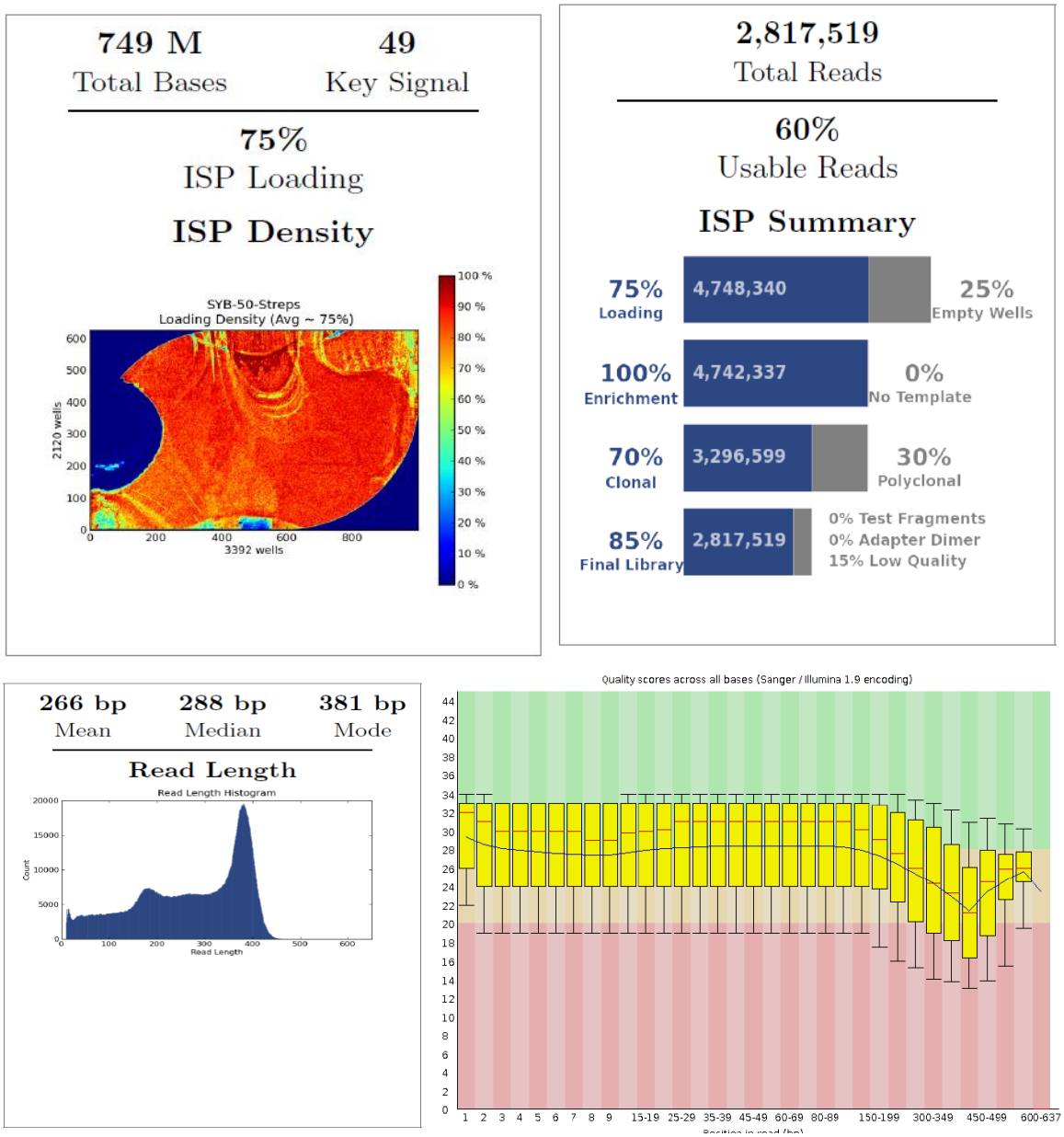


Figure 4.3: Summary report of the whole genome sequencing run of isolates 56 & 118: the ISP density percentage along with total bases and key signal; total reads along with enrichment, clonal and polyclonal and final library values; the trimmed read length; and the range value plot for accuracy and quality scores across all bases at each position in the FastQC file.

We had been facing some issues regarding the low loading percentage since we switched from using Ion PGM 400bp sequencing kit to Ion PGM Hi-Q 400bp sequencing kit in our lab using the Ion PGM weighted chip bucket protocol (Pub. no. MAN0007517). To overcome this issue, we have developed our loading protocol (Appendix B) with help of Dr. Jana K. Schniete, a research associate at University of Strathclyde, that fits with Ion PGM Hi-Q 400bp sequencing kit and Ion 316v2/318v2 chip types. Applying this developed protocol, the loading percentage was higher (>70%) each time. The main different points of this developed protocol are to keep the chip in the bucket and tilt it at a 45 degree angle and the pipette tip needs to be at 90 degrees to the chip while loading the templated ISPs and perform five in and out cycles of slowly pipetting the sample in and out of the chip ten times.

4.5 Total bases, total reads and read length

The throughput specifications of chip 316v2 type which is used in this study for 400bp are 600 Mb – 1GB of total bases and 2-3 millions of total reads. The total bases of isolates 85, 92 and 99 and isolates 56 and 118 were 1 GB and 749 Mb respectively, whereas 3,609,264 total reads were for isolates 85, 92 and 99 and 2,817,519 total reads were for isolates 56 and 118 (Figure 4.2 and Figure 4.3). It can be seen from the read length histograms (Figure 4.2 and Figure 4.3), the median read length of isolates 85, 92 and 99 is 320bp whereas the read length of isolates 56 and 118 is 288bp.

It is obvious, therefore read lengths which were generated from inserting the recommended library size and size distribution will lead to high loading percentage as in case of isolates 85, 92 and 99 (Figure 4.1). Consequently, the final total bases of the sequencing run seems to depend on the loading percentage density, total number of reads as well as the median read length.

4.6 FastQC

FastQC provides quality control (QC) analysis on biological raw data generated from high-throughput sequencing (Andrews *et al.*, 2011). The FastQC report showed that per base sequence quality score for isolates 85, 92 and 99 was over 20 Phred quality score (1 in 100 probability that the base is called wrong) for up to 370bp. Reads falling below the minimum accepted value were trimmed (Figure 4.2). Similarly, the quality score for isolates 56 and 118 was over 20 for up to 349bp (Figure 4.3). The higher the score the better the base call. If the median value, which is represented by the central red line, occurs on green areas these are very good quality calls, if the line occurs on orange areas these are reasonable quality calls whereas calls of poor quality occur on red areas.

4.7 *De novo* assembly

A wide variety of different bioinformatics tools are available to deal with and analyse high throughput read length data that is generated by NGS sequencing (Hernandez *et al.*, 2008; Ekblom & Wolf, 2014). To perform *de novo* assembling, large amounts of high quality reads are required. However, it is possible now with Ion Torrent technology to work and generate read length of around 400bp besides short reads of 200bp. Such a long read length aims to overcome the short read limitations of *de novo* assembly without the use of a reference genome. As mentioned previously, *Streptomyces* bacteria have high GC DNA content. GC-rich fragments can lower assembly completeness and significantly reduce the performance of *de novo* assembly strategy (Chen *et al.*, 2013; Nurk *et al.*, 2013). For this reason, the SPAdes assembler (Nurk *et al.*, 2013) was used to perform *de novo* assemblies on our isolates. This assembly algorithm supports read outputs generated by Ion Torrent technology to overcome assembling datasets with high GC content by suggesting enabling single-cell mode and using different *k-mer* lengths of 21,33,55,77. However, the accuracy of assembly of GC-rich samples can be also improved by increasing the coverage depths of the total amount of NGS data. In this regard, if we assume that the average length

of *Streptomyces* genome is around 8Mbp, thus the obtained coverage depths of isolates 56, 85, 92, 99 and 118 were 37%, 42%, 39%, 44% and 35% respectively, which should be high enough to minimise the size of unsequenced regions and sufficient to get good coverage for *de novo* draft genome assemblies. Desai *et al.* (2013) and Sims *et al.* (2014) showed that 35X–50X data obtained from NGS platforms is sufficient to get good coverage of small genomes such as bacteria and yeast.

4.8 General overview of the sequenced genomes

After the assemblies, it was possible to have a general overview of the sequenced genomes. The general characteristics from the genomes and their quality checks are illustrated in (Table 4.1).

4.8.1 Genome size and number of contigs

The genome sizes of the five isolates are quite big for a bacterium. It can be seen that genome sizes for isolates 56, 85, 92, 99 and 118 were 8.9 Mbp, 10.8 Mbp, 7.1 Mpb, 9.2 Mbp and 8.5 Mbp respectively (Tabl 4.1). Nevertheless, values between 8.7 Mbp and 11.9 Mbp are in the range of previous observations for genomes belonging to the streptomycetes genus which have the largest genomes among bacteria (Harrison and Studholme, 2014). Quast software (Gurevich *et al.*, 2013) was used, in order to assess the quality of the sequences after the *de novo* assembling (Table 4.1). To increase the quality of assembled sequences, only contig lengths above 1000bp were involved for the final number of contigs for each isolate. Accordingly, these sequences were assembled to 788, 233, 464, 362 and 382 contigs for isolates 56, 85, 92, 99 and 118 respectively. In terms of number of contigs, the results from the GenBank database where all complete and draft genomes of *Streptomyces* bacteria are available strongly corroborate the results obtained in the our study. However, fewer but longer contigs are preferable for other downstream analysis, in particularly for closing and finishing bacterial genomes.

4.8.2 N50 value

The N50 value is the major quality parameter in the analysis of the assemblies (Mäkinen, 2012). Isolate 85 shows very good quality (Table 4.1). It's N50 value is high when compared to the other three assemblies from 92, 99 and 118 which still show a satisfactory assembly quality, whereas isolate 56 has poorer quality. Obviously, the N50 value of isolate 85 was three to four times higher than other isolates as a result of reduction in the number of contigs (233) and consequently the number of largest contigs was increased (66 contigs >= 50000bp) with a largest contig of 404158bp.

Table 4.1: Quality assessment of the final assembly of the five isolates performed by QUAST 2.2 (Gurevich *et al.*, 2013).

Attribute	Isolate_56	Isolate_85	Isolate_92	Isolate_99	Isolate_118
# contigs	788	233	464	362	382
# contigs (>= 0 bp)	788	233	464	362	382
# contigs (>= 1000 bp)	788	209	411	333	382
# contigs (>= 5000 bp)	492	165	270	251	284
# contigs (>= 10000 bp)	297	143	194	207	228
# contigs (>= 25000 bp)	85	118	93	116	121
# contigs (>= 50000 bp)	13	66	27	57	45
Largest contig	136584	404158	207880	221638	135713
Total length (pb)	8,9477,88	10,841,945	7,156,780	9,280,527	8,524,242
N50	18635	124964	34889	58666	40407
GC (%)	71.42	70.03	70.01	69.73	71.25
Mismatches:					
# N's	0.0	0.0	0.0	0.0	0.0
# N's per 100 kbp	0.0	0.0	0.0	0.0	0.0

4.8.3 GC % content

From the current sequenced bacterial genomes, it can be seen that GC content in bacteria ranges from 15% to about 85% (Land *et al.*, 2015). Free living bacteria that inhabit the soil tend to have on average higher GC content and larger genome size than associated bacteria, such as obligatory pathogens and symbionts (Rocha and Danchin, 2002; Land *et al.*, 2015). The GC content bias among bacterial genomes is unclear; however Rocha and Danchin (2002) suggested that this might result from competition for metabolic resources. Based on previously sequenced complete and draft genome analysis of *Streptomyces* strains, the GC % varies from 66 – 74% (Kampfer, 2012; Labeda *et al.*, 2012; Aouiche *et al.*, 2013; Harrison and Studholme, 2014). Hence, the values presented from the five isolates of 71.42% for 56, 70.03 % for 85, 70.01% for 92, 69.73% for 99 and 71.25% for 118 (Table 4.1), corroborate the previous observations.

4.8.4 *In silico* antimicrobial resistance gene sequences

Identification of acquired antibiotic resistance genes was carried out using ResFinder tool (Zankari *et al.* 2012). The *ole(C)* resistance gene, which confers macrolide, lincosamide and streptogramin B resistance, was found in all of the five isolates (Table 4.2). Additionally, Chloram-phenicol resistance gene, *cmlV*, was only found in isolates 56 & 118 as well. *Ole(C)* and *cmlV* resistance genes were also found in pangenomes of the fifteen *Streptomyces* strains with identity of 89.7% and 98.55% respectively (Table 4.3). No antimicrobial resistance genes were detected in the core of the fifteen finished genome. These results reveal that antimicrobial resistance genes are not within the essential conserved genes of *Streptomyces* genome but they are occurred much frequently in the dispensable region of the genome, that contributes to the diversity characteristics, environmental adaptation, uniqueness and antimicrobial resistance of the bacterial species (Tettelin *et al.*, 2008).

Table 4.2: Identification of acquired antimicrobial resistance genes in the genome of the five isolates using ResFinder-2.1 tool (Zankari *et al.* 2012).

Isolate No.	Resistance gene	% Identity	Query/HSP length	Predicted phenotype	Accession number
56	*ole(C)*	84.02	676 / 978	Macrolide, Lincosamide and Streptogramin B resistance	L06249
	cmlV	81.14	1108 / 1311	Phenicol resistance	U09991
85	*ole(C)*	85.37	943 / 978	Macrolide, Lincosamide and Streptogramin B resistance	L06249
92	*ole(C)*	86.21	957 / 978	Macrolide, Lincosamide and Streptogramin B resistance	L06249
99	*ole(C)*	85.56	942 / 978	Macrolide, Lincosamide and Streptogramin B resistance	L06249
118	*ole(C)*	83.59	591 / 978	Macrolide, Lincosamide and Streptogramin B resistance	L06249
	cmlV	80.47	1111 / 1311	Phenicol resistance	U09991

HSP length: is the length of the alignment between the best matching resistance gene and the corresponding sequence in the genome.

Table 4.3: Identification of acquired antimicrobial resistance genes in the pangenome of the fifteen *Streptomyces* species (finished genomes) using ResFinder-2.1 tool (Zankari *et al.* 2012).

Resistance gene	% Identity	Query/HSP length	Predicted phenotype	Accession number
Aminoglycoside				
aac(3)-VIIa	80.88	617 / 867	Aminoglycoside resistance	M22999
aac(6')-Isa	81.45	415 / 474	Aminoglycoside resistance	AB116646
aph(6)-Ia	100.00	924 / 924	Aminoglycoside resistance	AY971801
aac(3)-Xa	99.65	855 / 855	Aminoglycoside resistance	AB028210
aph(4)-Ib	99.90	999 / 999	Aminoglycoside resistance Alternate name; aph(7'')-Ia	X03615
aph(3'')-Ia	99.63	819 / 819	Aminoglycoside resistance	M16482
MLS - Macrolide, Lincosamide and Streptogramin B				
erm(O)	99.62	783 / 783	Macrolide resistance	M74717
ole(C)	89.70	942 / 978	Macrolide, Lincosamide and Streptogramin B resistance	L06249
Phenicol				
cmlV	98.55	1311 / 1311	Phenicol resistance	U09991
Tetracycline				
otr(C)	84.76	722 / 1056	Tetracycline resistance	AY509111
Glycopeptide				

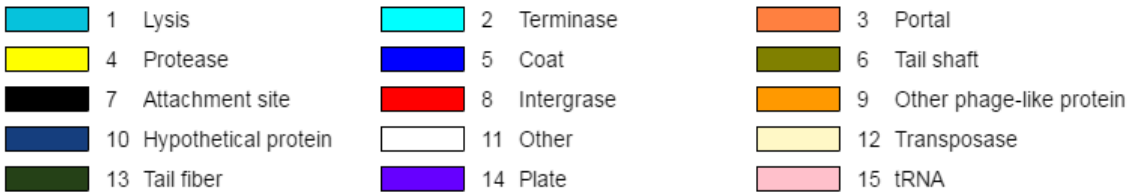
anX-Sc	100.00	609 / 609	Vancomycin resistance (Glycopeptid resistance) VanSc vancomycin resistance operon, (VanSc, VanH-Sc and VanX-Sc)	AL939117
VanH-Sc	100.00	1014 / 1014	Vancomycin resistance (Glycopeptid resistance) VanSc vancomycin resistance operon, (VanSc, VanH-Sc and VanX-Sc)	AL939117
ddlA2-Sc	100.00	1041 / 1041	Vancomycin resistance	AL939117

HSP length: is the length of the alignment between the best matching resistance gene and the corresponding sequence in the genome.

4.8.5 Prophage sequences

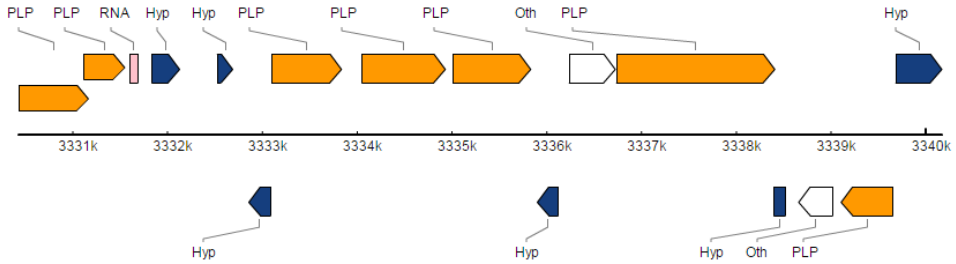
Prophage sequences, which are inserted into the genomes of the bacterial strains, were detected using PHAST (PHAge Search Tool) (Zhou *et al.*, 2011). These analyses showed that all five *Streptomyces* isolates have phage sequences. In isolate 56 there are two prophages with 16 CDSs and 16 CDSs placed in the regions between 3.33 Mbp and 3.34 Mpb and 3.84 Mbp and 3.85 Mbp with lengths of 9.7 Kbp and 11.9 Kbp respectively (Figure 4.4). One prophage was detected in isolate 85 with 12 CDSs located from 7.96 Mpb to 7.97 Mbp with length of 9.4 Kbp. In isolate 92 there are 13 CDSs with a length of 11.3 Kbp. Also, in isolates 99 and 118 there is one prophage region for each of these isolates as well, with 16 CDSs from 1.854 Mbp to 1.866 Mbp with a region length of 11.5 Kpb and 25 CDSs from 2.24 Mbp to 2.26 Mpb with a region length of 2.3 Kbp respectively. Isolates 92 and 99 were also shown to have coding sequences for transposases (Figure 4.4). Recent findings suggest that phages and transposons act as vehicles for antibiotic resistance genes (Balcazar, 2014; Darmon and Leach, 2014), hence the presence of these sequences within *Streptomyces* genomes might be considered to play a major role in resistance to antibiotics and heavy metals.

Identified phage elements



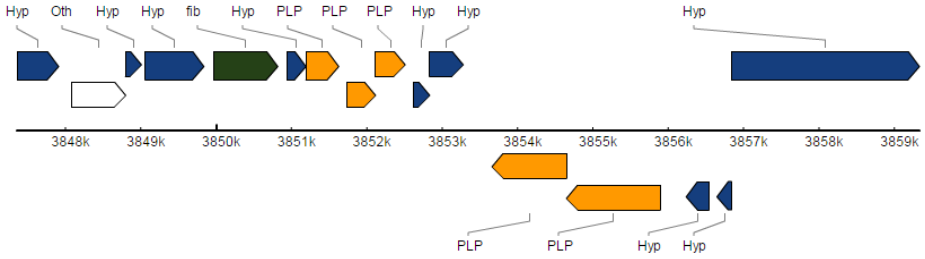
Prophage region: 1
Number of CDS: 16
Location: from 3330428 to 3340167 (9740 bps)
Predicted status: incomplete prophage
GC content: 70.42%

Isolate 56



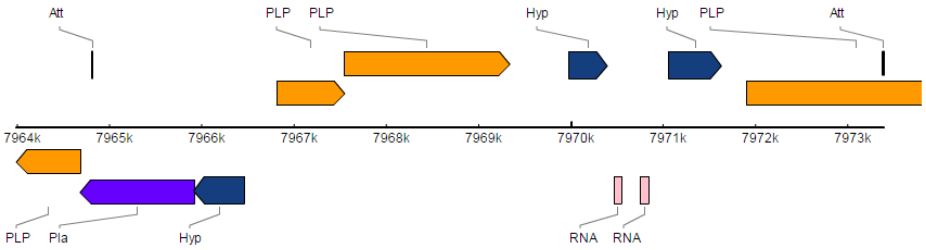
Prophage region: 2
Number of CDS: 16
Location: from 3847357 to 3859333 (11977 bps)
Predicted status: incomplete prophage
GC content: 69.37%

Isolate 56



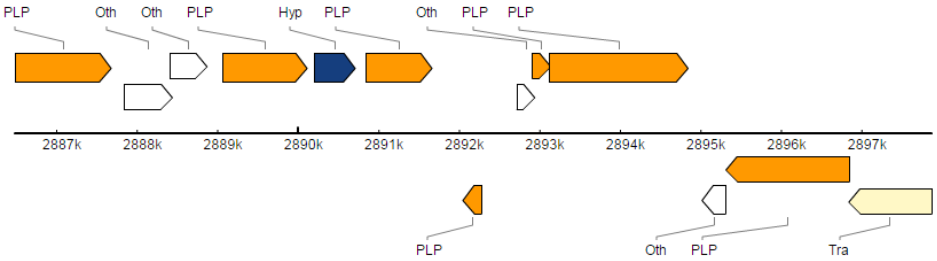
Prophage region: 1
Number of CDS: 12
Location: from 7963989 to 7973390 (9402 bps)
Predicted status: incomplete prophage
GC content: 70.93%

Isolate 85



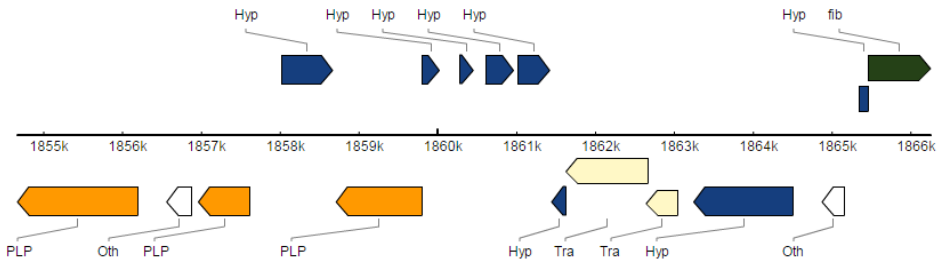
Prophage region: 1
Number of CDS: 13
Location: from 2886479 to 2897865 (11387 bps)
Predicted status: incomplete prophage
GC content: 68.44%

Isolate 92



Prophage region: 1
Number of CDS: 16
Location: from 1854667 to 1866248 (11582 bps)
Predicted status: incomplete prophage
GC content: 68.51%

Isolate 99



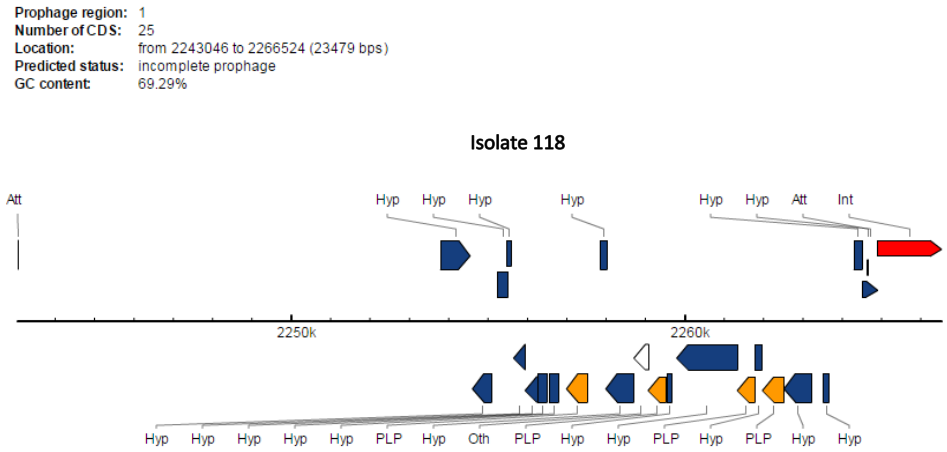


Figure 4.4: CDS in isolates 56, 85, 92, 99 & 118 retrieved by PHAST (Zhou *et al.*, 2011).

Chapter 5:

Genome Computational and Phenotypic Microarray Characterisations of Isolate_99, a novel *Streptomyces* strain, Compared with the Closest Related Strains: *Streptomyces turgidiscabies and Streptomyces graminilatus*

5.1 16S and 23S rRNA genes

The bacterial 16S ribosomal RNA (rRNA) gene is the most conserved DNA that shows considerable sequence diversity among different bacteria (Chakravorty, 2007). There are a number of reasons that favour the extensive utilisation of the 16S rRNA to study bacterial phylogeny and taxonomy. These reasons include firstly, it is universally distributed in bacteria and archaea, thus the lineage among all bacteria can be measured (Woese *et al.* 1985); secondly, it is conserved over time, which enabled the design of suitable primers for PCR amplification; thirdly, it's size (1,500bp) is suitable for sequencing and provides enough information for analysis (Patel, 2001; Mizrahi-Man *et al.*, 2013). The sequencing of the 16S rRNA for a large number of bacterial strains has been determined. Over 198 million sequences have been deposited in Genbank (https://www.ncbi.nlm.nih.gov/), the largest database of gene sequencing, of which more than 13 million are of the 16S rRNA gene at time of writing this chapter. The sequencing of the whole 1,500 bp region is important and usually required to discriminate between particular taxa or strains, or describing a novel species. (Sacchi *et al.*, 2002). Therefore, it is necessary to consider whether the gene should be sequenced in its entirety or whether the more commonly reported shorter sequence can provide sufficient comparable information (Clarridge, 2004). The 23S rRNA region has also been used, rather than 16S rRNA, in an attempt to determine the phylogenetic relationships among bacteria, but it is rarely used because of the lack of established broad-range sequencing primers, limited updated databases and the difficulty of sequencing of the whole region (around 3,000 bp) with early sequencing technology (Pie *et al.*, 2009). Nevertheless, the Silva database (Quast *et al.*, 2013), is a comprehensive web resource that actively incorporates small (16S/18S) and large (23S/28S) subunit ribosomal RNA sequences for all domains of life (bacteria, archaea and eukarya). It was used for quality checked and aligned both sequenced regions of 16S and 23S rRNA. The Silva database contains over 6 million high quality full length sequences of small and large rRNA genes at the time of writing this chapter.

5.2 Phylogenetic relationship of isolate_99 with the closely related strains

Phylogenetic analysis of isolate_99 sequence with 16S rRNA of the highest 30 similar strains of *Streptomyces* was carried out. The phylogenetic tree obtained by applying the neighbour joining method is illustrated in Figure 5.1. It can be seen that *S. graminilatus* and *S. turgidiscabies* are the most closely related strains to isolate_99 with identity of 98.6% and 98.25% respectively. On the other hand, the phylogenetic tree obtained by applying the neighbour joining method of the highest 22 similar strains to isolate_99 based on 23S rRNA sequence demonstrated that *S. turgidiscabies, S. xylophagus and S. prunicolor* strains are the most close strains to isolate_99 (Figure 5.2) with identity of 97%, 96.4% and 96% respectively. Surprisingly, although *S. graminilatus* was the most closely related strain to isolate_99 based on 16S rRNA gene, it failed to appear in the 23S rRNA tree as in the 16S rRNA. It was also possible to exhibit the relationship of isolate_99 and the most closely related strains based on both 16S and 23S rRNA genes using high-throughput whole genome sequencing (WGS) reads. It has been observed that isolate-99 was closely related to *S. turgidiscabies* and *S. graminilatus* based on WGS reads (Figure 5.3).

Surprisingly, while the trees based on both 16S rRNA and WGS reads indicated that isolate_99 is closely related to *S. turgidiscabies* and *S. graminilatus* (Figures 5.1 & 5.3), the tree based on 23S rRNA indicated that isolate_99 is closely related only to *S. turgidiscabies* (Figure 5.2) but not to *S. graminilatus* that appeared on 16S rRNA and WGS trees. This is hard to ignore, considering that 23S rRNA gene of *S. graminilatus* was not found in both Silva and Genbank databases when we tried to retrieve and download it manually.

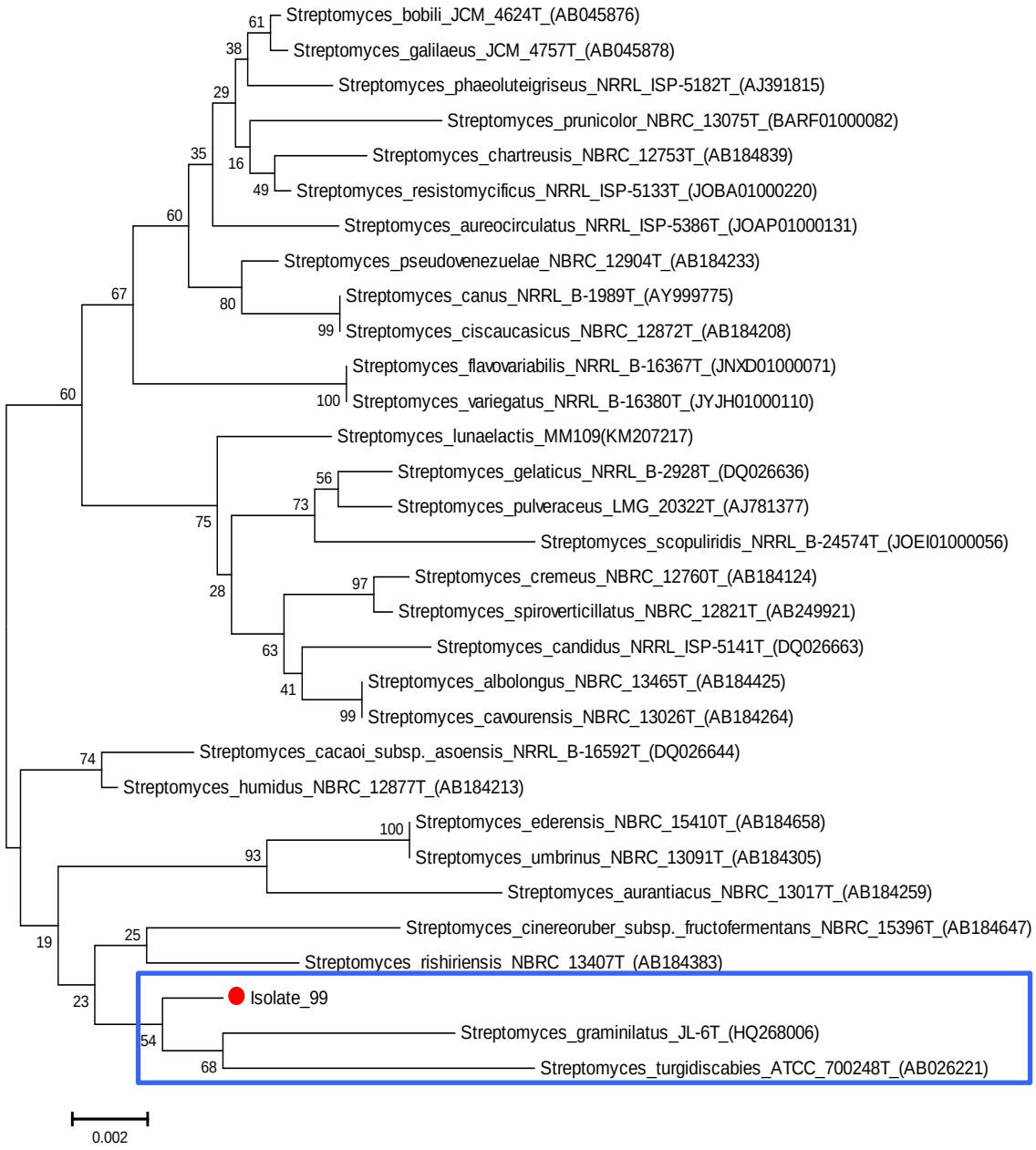


Figure 5.1: Neighbour-Joining (NJ) phylogenetic tree based on 16S rRNA (1000X BOOTSTRAP) showing the relationship between strain Isolate_99 (indicated by red circle) and closely related species in the genus *Streptomyces* that obtained from Silva database. Numbers appearing at nodes indicate the bootstrap percentage. Accession numbers are given in parentheses.

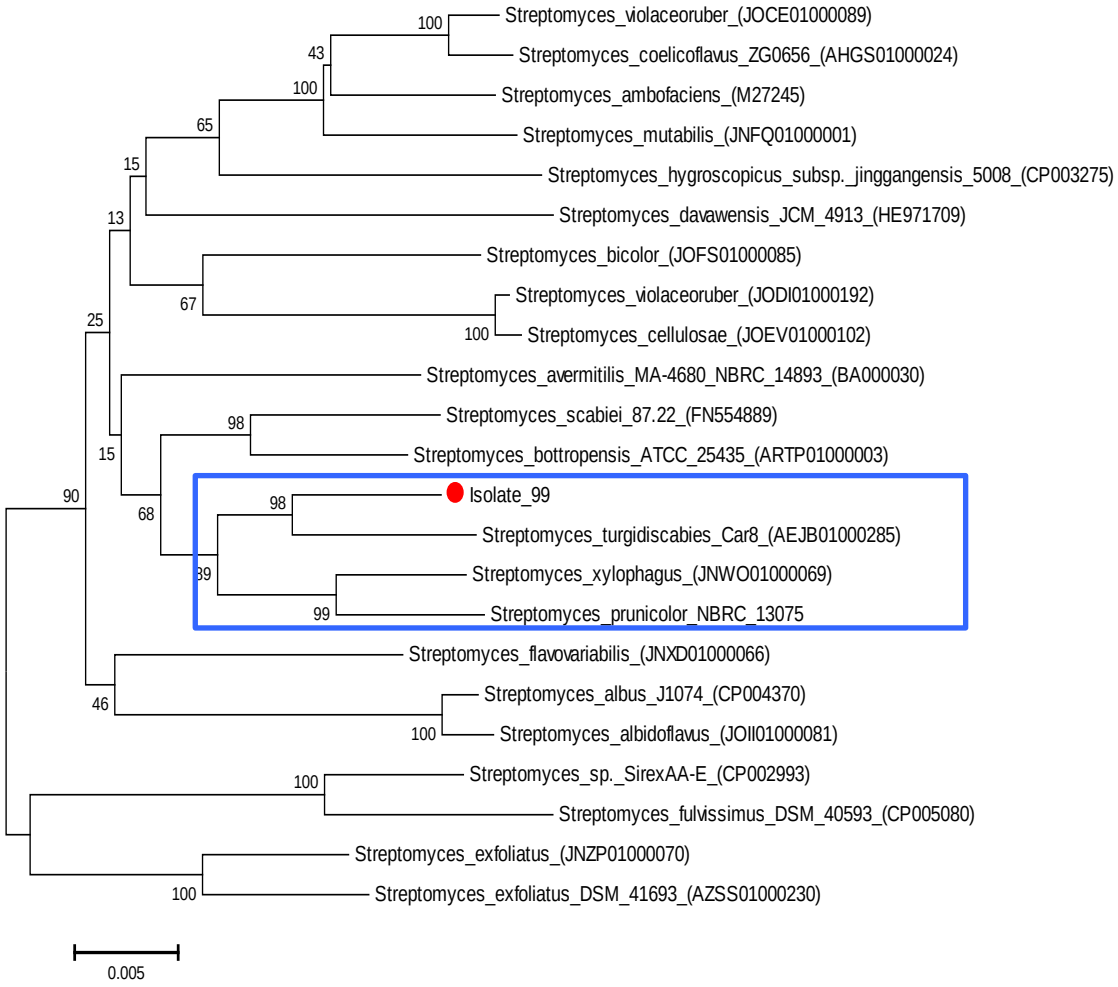


Figure 5.2: Neighbour-Joining (NJ) phylogenetic tree based on 23S rRNA (1000X BOOTSTRAP) showing the relationship between strain Isolate_99 (indicated by red circle) and closely related species in the genus *Streptomyces* that obtained from Silva database. Numbers appearing at nodes indicate the bootstrap percentage. Accession numbers are given in parentheses.

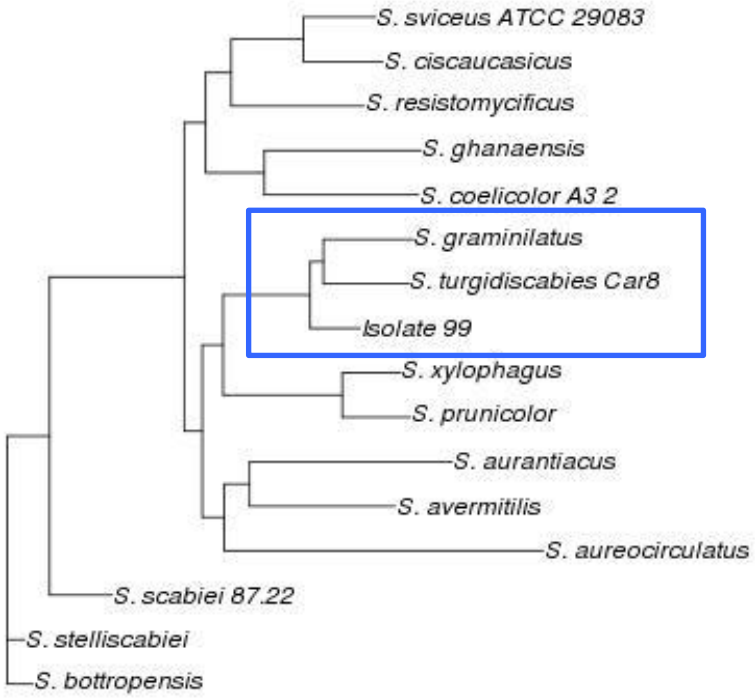


Figure 5.3: Neighbour-Joining (NJ) phylogenetic tree based on whole genome sequence data (assembled contigs) showing the relationship between strain Isolate_99 and closely related species in the genus *Streptomyces* that were obtained from NCBI database.

However, the 23S rRNA gene of *S. graminilatus* was obtained directly from its draft genome using RNAmmer (Lagesen *et al.*, 2007), added and aligned manually to construct a new 23S rRNA phylogenetic tree. Interestingly, *S. graminilatus* has appeared again to be the closely related to isolate_99 based on 23S rRNA (Figure 5.4) as well as 16S rRNA and WGS based trees. This might be explain why 23S rRNA gene has not been widely implemented for both phylogenetic analysis and taxonomic classification of bacteria and the sequence databases not been constantly updated accordingly.

In the case of *Streptomyces* species, genes other than 16S rRNA have also been used for studying phylogenetic relationships among strains. Guo *et al.* (2008), Rong *et al.* (2009) and Rong and Huang (2010) have used MLST (Multilocus Sequencing Typing) of five housekeeping genes, *atpD*, *gyrB*, *recA* , *rpoB*, and *trpB* to investigate the phylogeny of *Streptomyces*. However, their MLST studies of the phylogenetic tree were particularly useful for accurately determining phylogenetic relationships between the clades of *Streptomyces* genus rather than between closely related strains within this genus (Labeda *et al.*, 2012).

As the bacterial genomes vary in size, and gene duplication, gene transfer, gene deletion, gene fusion, and genes splitting are common, it was observed that using the whole genome sequence for phylogenetic analysis is quite complicated, and that the phylogenetic trees based on whole genome analysis are similar (Bansal & Meyer, 2002). Nevertheless, several studies have been attempted to construct phylogenetic analysis of some bacterial genera (Foster *et al.*, 2009; Rodriguez *et al.*, 2012) including *Streptomyces* genus (Alam *et al.*, 2010). Here, we utilised a REALPHY bioinformatics tool (Bertels *et al.*, 2014) to infer a phylogenetic tree of isolate_99 with those of closely related *Streptomyces* strains based on whole genome sequences. Unlike other traditional bioinformatics methods for reconstructing phylogenetic trees from whole genome sequence data which are required the high through put sequence reads to be assembled to contigs, annotated, identified and aligned to orthologous coding regions (Luo *et al.*, 2011; Rodriguez *et al.*, 2012). This is time consuming and requires a sophisticated combination of bioinformatics

methods. REALPHY directly maps raw sequencing reads to the genome sequence of a single reference, extracts and concatenats homologous sites into a multiple sequence alignment from which the phylogenetic tree is reconstructed. Applying this method, mapping biases from mapping to a single reference can be avoided by implementing a procedure for merging alignments obtained by mapping to multiple reference genomes into a single non-redundant alignment and accurate phylogenies can be reconstructed (Bertels *et al.*, 2014).

From the results, it is very obvious that isolate_99 is the most closely related to both *S. graminilatus* and *S. turgidiscabies*. There is an excellent correlation between the phylogeny of the three trees constructed based on 16S rRNA, 23S rRNA and WGS. However, this might highlight the strength of ribosomal genes and WGS reads in identifying relationships between closely related species in *Streptomyces* genus. Thus, because of the limited availability of 23S rRNA gene on the sequence databases, and that nearly 294 genomes of *Streptomyces* have been sequenced and deposited in NCBI databases yet, it is likely that the 16S rRNA gene is an ideal choice to infer the phylogenetic relationships between closely related species in the genus of *Streptomyces*, which currently contains 793 validly described species along with their 16S rRNA sequences according to list of prokaryotic names with standing nomenclature (LPSN; http://www.bacterio.net/streptomyces.html; Parte, 2013).

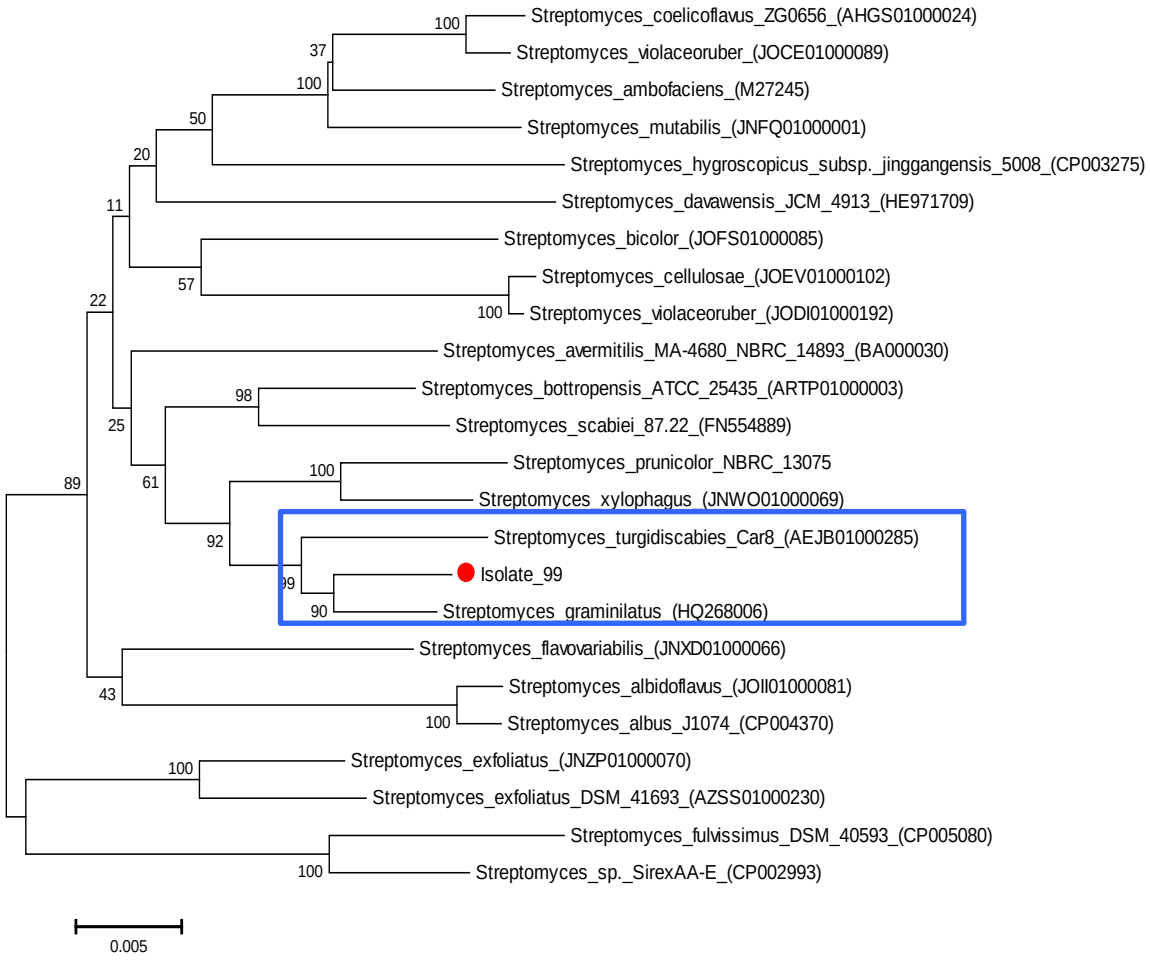


Figure 5.4: Neighbour-Joining (NJ) phylogenetic tree based on 23S rRNA (1000X BOOTSTRAP) showing the relationship between strain Isolate_99 (indicated by red circle) and closely related species in the genus *Streptomyces* that obtained from Silva database after adding *Streptomyces graminilatus* 23S rRNA gene. Numbers appearing at nodes indicate the bootstrap percentage. Accession numbers are given in parentheses.

5.3 Genome similarity using Average Nucleotide Identity (ANI) and *in silico* DNA-DNA Hybridisation (*is*DDH) of isolate_99 with the closely related strains

Over the last few decades, traditional DNA–DNA hybridisation (DDH), which indirectly measures the degree of genetic similarity between two genome sequences, has been the gold standard method for bacterial species demarcation as it provides a clear and objective numerical threshold for a species boundary, for which 70% DDH was suggested and is widely used (Wayne *et al.*, 1987). This conventional method has been reported to be labour-intensive, error prone, the procedures are time consuming and data produced cannot be compared with other labs (Richter & Rossello-Mora, 2009; Auch *et al.*, 2010; Kim *et al.*, 2014). However, in the era of genome sequencing and bioinformatics, it is now generally accepted that genome sequencing has the potential to be a routine approach of measuring genetic similarity between closely related species as a replacement for traditional DDH. New robust bioinformatics tools and pipelines have been developed for correlating the wet lab-based DNA-DNA hybridisation values with digital DDH-like similarity based on computational comparisons of whole genome sequences. Average Nucleotide Identity (ANI) has been developed by the Tiedje group (2005), and in *silico* DNA-DNA Hybridisation (*is*DDH) by means of genome-to-genome sequence comparison developed by the Göker group (2010). These two methods have been most widely used as possible next generation gold standard approaches for measuring the degree of similarity between the genomes of two strains based on whole genome sequencing data (Konstantinidis & Tiedje, 2005; Richter & Rossello-Mora, 2009; Auch *et al.*, 2010; Kim *et al.*, 2014; Varghese *et al.*, 2015). It is now generally accepted that with ANI among conserved and shared genes a threshold of 95 to 96% equates to a conventional DDH threshold of 70% (Goris *et al.*, 2007; Richter & Rossello-Mora, 2009). On the other hand, the *is*DDH method produces values that can potentially mimic the wet-lab DDH boundary of 70% (Auch *et al.*, 2010; Meier-Kolthoff *et al.*, 2013).

Using the values above, the degree of genomic distanse between isolate_99 and closely related species *of S. graminilatus* and *S. turgidiscabies* based on both 16S rRNA (Figure 5.1) and 23S rRNA (Figure 5.4) genes; *S. auranticalis* based on 16S rRNA gene (Figure 5.1); *S. xylophagus*, *S. prunicolor* and *S. bottropensis* based on 23S rRNA gene (Figure 5.4); and *S. avermitilis, S. scabiei, S. svicceus, S. coelicolor* and *S. ghanaensis* based on higher scores obtained from RAST annotation (Figure 5.5) were determined. In general, all the results were below 95% (the species ANI cutoff value) (Table 5.1) and below 70% (the species DDH cutoff value) (Table 5.2).

The results have shown that the ANI values between isolate_99 and the most related species reached 88.7% and 91.2% at most when compared *to S. turgidiscabies* and 87.4% and 90.9% when compared to *S. graminilatus* using ANIb (BLAST algorithm) and ANIm (MUMmer algorithm) respectively, whereas those between isolate_99 and each of the other strains were below 80.7% and 86.3% using ANIb and ANIm methods respectively (Table 5.1). It can be seen from Table 5.1 that the ANI values calculated using the MUMmer algorithm (Kurtz *et al.*, 2004) are higher than those ANI values calculated with the BLAST algorithm (Altschul *et al.*, 1997). Nevertheless, when two genomes are closely related, ANIb and ANIm match ANI values very closely, as it can be seen that the ANI value between *S. albus*_J1074 and *S. sp*. GBA_94-10 was 95.27% and 96.18% using ANIb and ANIm respectively (Table 5.1). S. sp. GBA_94-10 is one of the six sequenced strains of the same species of *S. albus* (Seipke, 2015). These two strains were used for validating the ANI values. Also, it was noticed that when the relatedness between two genomes decreased, the disparity between the ANIb and ANIm values increased (Table 5.1). However, in both algorithms, the query genome is spliced in 1020 nucleotide stretches and each of them blasted against the reference genome to calculate the ANI (Goris *et al.*, 2007). Additionally, MUMmer was shown to be a more efficient algorithm for large DNA sequences, much faster and more precise (Richter & Rossello-Mora, 2009).

Conversely, regarding the *is*DDH values, isolate_99 had an *in silico* DDH 40.8% and 39.2% with *S. turgidiscabies* and *S. graminilatus* respectively (Table 5.2), that clustered most closely together in the trees based on both 16S rRNA gene (Figure

5.1) and 23S rRNA gene (Figure 5.4), whereas those between isolate_99 and each of other strains were below 25% (Table 5.2). It was also possible to estimate if the two strains belong to distinct or related subspecies by conducting comparisons with subspecies boundaries >79% (same subspecies). To our knowledge, the *is*DDH bioinformatics tool (Auch *et al.*, 2010; Meier-Kolthoff *et al.*, 2013), is the only tool for estimating subspecies boundaries for either distinct or same subspecies by applying formula of probability that DDH > 79% (same subspecies). This formula is more useful when trying to discriminate between two genomes sharing >70% DDH value. However, because of none of subjected genomes showed >70% *is*DDH value with isolate_99, it was not worthwhile to consider their values. But, because a value of 74.6% *is*DDH was obtained between *S. albus*_J1074 and S*. sp*. GBA_94-10 trains, which used for validating the DDH threshold of 70%, the probability that DDH > 79% (same subspecies) was considered despite it was rejected as the value was 37.5% (Table 5.2) which is below 79% threshold and therefore S. albus_J1074 and S. sp. GBA_94-10 belong to district subspecies.

Interestingly, although isolate_99 was the most closely related strain to *S. graminilatus* and *S. turgidiscabies* based on 16S rRNA gene with 98.6% and 98.2% identity respectively (Figure 5.1), it appeared to be that *S. turgidiscabies* is the first most closely related based on both the ANI (ANIb 88.7% and ANIm 91.2%) and *is*DDH (40.8%), whereas *S. graminilatus* was the second most related strain based on both the ANI (ANIb 87.4% and ANIm 90.9%) and *is*DDH (39.2%). It has been reported that 97% 16S rRNA gene sequence similarity corresponded to 70% DDH (Stackebrandt and Goebel, 1994), and that DDH is only required when 16S rRNA gene sequence similarity between two strains is > 97% (Tindall *et al.*, 2010). Even though, thresholds over 98.7% have also been applied (Stackebrandt and Ebers, 2006) as it is hard to differentiate two species using 16S rRNA gene sequences alone, as some species share at least 98% 16S rRNA gene sequence similarity, 95% of ANI, and 70% *is*DDH (Konstantinidis & Tiedje, 2005; Thompson *et al.*, 2013; Kim *et al.*, 2014). Thus, although isolate_99 had shared more than 98% 16S rRNA gene sequence similarity with both *S. turgidiscabies* and *S. graminilatus* strains, it is clear that isolate_99 belongs to a new *Streptomyces* strain where *in silico* DNA-DNA hybridisation (*is*DDH)

value cmparing with the genomic sequence of *S. turgidiscabies*, the closely related strain, was lower than 41% with 3.34% probability of being the same species (Table 5.2). This *is*DDH value is even lower than the controversial DDH values >60–70% (Richter & Rossello-Mora, 2009) to show enough dissimilarity to their closest relative species. Moreover, the Average Nucleotide Identity (ANI) values of isolate_99 compared to *S. turgidiscabies* were lower than the threshold of 95% (ANIb 88.7% and ANIm 91.2%), showing a strong evidence that isolate_99 is a novel species of genus *Streptomyces*.

Table 5.1*: in silico* pairwise genome comparison of Average Nucleotide Identity (ANI) between isolate_99 and the most closely related strains using ANIb and ANIm bioinformatics tools.

Query genome	Reference genome	Genome size (Mpb) / GC (%)	ANIb based on BLAST+	ANIm based on MuMer
			Above cutoff (> 95%)	
			Below cutoff (< 95%)	
Isolate 99	*S. turgidiscabieis*	10.8 / 69.9	88.76	91.28
Isolate 99	*S. graminilatus*	9.6 / 70.1	87.46	90.96
Isolate 99	*S. xylophagus*	11.3 / 69.8	80.86	86.11
Isolate 99	*S. prunicolor*	11.7 / 69.7	80.47	86.01
Isolate 99	*S. aurantiacus*	8.7 / 72.0	77.61	85.31
Isolate 99	*S. bottropensis*	8.9 / 71.2	80.29	86.03
Isolate 99	*S. avermitilis*	9.1 / 70.7	80.97	86.29
Isolate 99	*S. scabiei*	10.4 / 70.6	80.77	86.37
Isolate 99	*S. sviceus*	9.3 / 68.5	81.24	86.07
Isolate 99	*S. coelicolor* A3	8.6 / 72.0	79.23	85.70
Isolate 99	*S. ghanaensis*	8.5 / 69.8	79.77	85.95
Isolate 99	Isolate 99	9.2 / 69.7	100.00	100.00

Table 5.2: *in silico* Whole Genome DNA-DNA hybridisation (*is*DDH) between Isolate_99 and the most closely related strains using GGDH bioinformatics tool.

Query genome	Reference genome	DDH estimate (GLM-based)	Probability that DDH > 70% (same species)	Probability that DDH > 79% (same subspiecies)
Isolate 99	*S. turgidiscabieis*	40.80% +/- 2.52	3.34%	0.83%
Isolate 99	*S. graminilatus*	39.20% +/- 2.6	2.23%	0.58%
Isolate 99	*S. xylophagus*	25.60% +/- 2.41	0.01%	
Isolate 99	*S. prunicolor*	25.40% +/- 2.41	0.01%	0.01%
Isolate 99	*S. aurantiacus*	25.50% +/- 2.40	0.01%	0.01%
Isolate 99	*S. bottropensis*	25.00% +/- 2.40	0.01%	0.01%
Isolate 99	*S. avermitilis*	25.80% +/- 2.41	0.01%	0.01%
Isolate 99	*S. scabiei*	24.80% +/- 2.40	0.01%	0.01%
Isolate 99	*S. sviceus*	25.80% +/- 2.3	0.01%	0.01%

Isolate 99	*S. coelicolor* A3 (2)	24.50% +/- 2.3	0.01%	0.0%
Isolate 99	*S. ghanaensis*	25.20% +/- 2.3	0.01%	0.01%
Isolate 99	Isolate 99	100 +/- 0.0	98.3%	80.34%
S. albus_J1074	*S*. sp. GBA_94-10	74.60% +/- 3.33	87.80%	37.51%

5.4 Proteome based comparison using the proteome comparison bioinformatics tool

Isolate_99 and the closely related *Streptomyces* strains*: S. avermitils* (score: 533), *S. scabiei* (score: 417), *S. sviceus* (score: 408), *S. coelicolor* (score: 355) and *S. ghanaensis* (score: 341) based on scores from their annotation using RAST (Aziz *et al.*, 2008); and *S. turgidiscabies* and *S. graminilatus* based on 16S and 23S rRNA genes; *S. xylophagus* and *S. prunicolor* based on 23S rRNA gene were compared. The proteome comparison bioinformatics tool in PATRIC resources (Wattam *et al.*, 2014) was utilised. The tool runs a BLASTP comparison for CDS identity which is represented in colour intensities, where the blue colour represents the highest protein sequence similarity and red represents the lowest. Obviously, figure 5.5 (left) shows the low homology between isolate_99 and the five closest strains from RAST, with very little CDS similarity 70-80%. However, these five strains had also low ANI (ANIm, in this case): *S. avermitilis* (86.2%), *S. scabiei* (86.3%), *S. sviceus* (86%), *S. coelicolor* A3 (2) (85.7%) and *S. ghanaensis* (85.9%) (Table 5.1) and *is*DDH: *S. avermitilis* (25.8%), *S. scabiei* (24.8%), *S. sviceus* (25.8%), *S. coelicolor* (24.5%) and *S. ghanaensis* (25.2%) (Table 5.2) values when they compared with isolate_99. Moreover, *S. turgidiscabies* and *S. graminilatus* strains have more similarity to isolate_99 than the other with CDSs similarity 90-95% (Figure 5.5, right) because its track has more proteins that are oranges and greens than reds and oranges. Also, these two strains had higher ANI and *is*DDH values than the others as discussed above. Nevertheless, isolate_99 appears different from *S. turgidiscabies* and *S. graminilatus* because its track has many more proteins that are blues than yellows and greens.

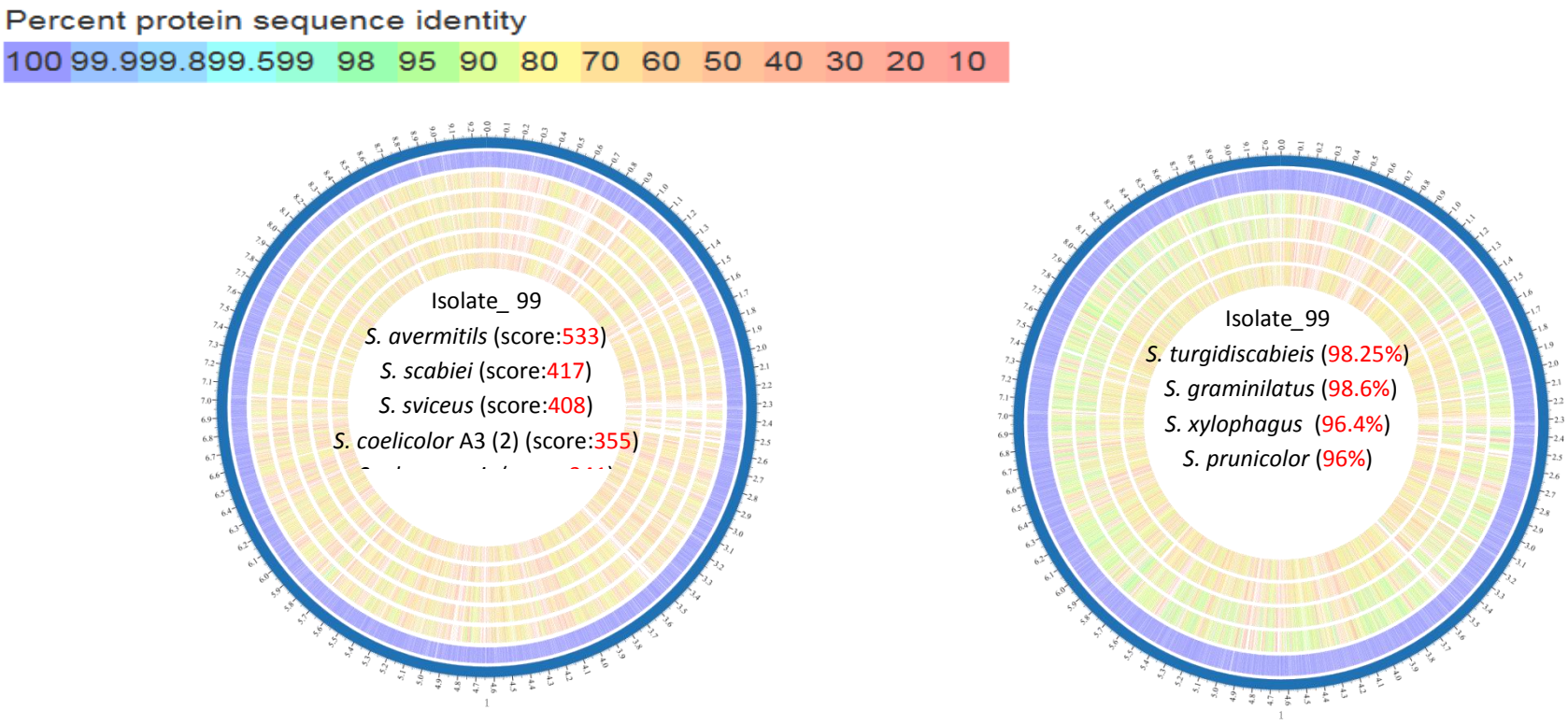


Figure 5.5: Proteome comparison of Isolate_99 against the closest neighbours based on RAST (left), 16S and 23S rRNA (right). Track order is from the outside: Isolate_99, *S. avermitilis* MA-4680 (score:533), *S. scabiei* 87.22 (score:417), *S. sviceus* ATCC 29083 (score:408) *S. coelicolor* A3(2) (score:355), *S. ghanaensis* (score:341) (left) and : Isolate_99, *S. turgidiscabieis* (16S 98.25%), *S. graminilatus* (16S 98.6%), *S. xylophagus* (23S 96.4%) and *S. prunicolor* (96%) (right). Colour key is used to estimate the similarity of each protein over the chromosome. It looks like isolate_99 strain is not similar to the other strains because its track has more proteins that are blues than oranges, yellows and greens.

5.5 Genome function comparative analysis with RAST and SEED – viewer

For the purpose of this analysis, the contigs of isolate_99, *S.turgidiscabies* and *S. graminilatus* were annotated using the Rapid Annotation System Technology (RAST) (Aziz *et al.*, 2008). A total of 11,872, 9640 and 8258 protein encoding genes (CDSs) were predicted respectively. The CDSs were classified by their metabolic function and compared to classified CDSs found in the others. According to the RAST analysis, isolate_99 had 22% subsystem coverage of all known metabolic processes with 2,555 CDSs; *S. turgidiscabies* had 30% subsystem coverage of all known metabolic processes with 2,875 CDSs and *S. graminilatus* had 29% subsystem coverage with 2,328 CDSs (Table 5.3). However, these values indicate that despite isolate_99 having lower subsystem coverage of all known metabolic process than *S. turgidiscabies* and *S. graminilatus*, it still has more genes to be identified. Additionally, the SEED subsystems annotation engine, which defines genes associated with a functional role in a bacterial genome as a subsystem (Overbeek *et al.*, 2014) was used for functional comparisons between isolate_99 and *S. turgidiscabies* and *S. graminilatus* genomes. The annotated ORFs of isolate_99, *S. turgidiscabies* and *S. graminilatus* were classified into categories as shown in figures 5.6, 5.7 and 5.8 respectively. As a result of this classification, 362, 446 and 420 subsystems were found in isolate_99, *S. turgidiscabies* and *S. graminilatus* strains, respectively. However, comparison of the genes found in the metabolic subsystems categories was overall similar with exception of the abundance of the subsystem categories in each strain. In isolate_99, subsystem categories representing the metabolism of carbohydrates, amino acids, proteins, and cofactors, vitamins, prosthetic groups, or pigments, are the most abundant, and they account for 710, 737, 288 and 306 respectively (Figure 5.5).

Table 5.3: Subsystem* coverage of Isolate_99*, S. turgidiscabieis and S. graminilatus* in the RAST annotation (Aziz *et al.,* 2008).

Subsystem coverage	**Isolate_99**	***S. turgidiscabieis***	***S. graminilatus***
in subsystem (%)	26%	30%	29%
non-hypothetical	2462	2748	2235
Hypothetical	93	127	93
not in subsystem (%)	78%	70%	71%
non-hypothetical	4182	3347	2756
hypothetical	5135	3532	3174

*Rapid Annotations based on Subsystem Technology, uses a "Highest Confidence First" assignment propagation strategy based on manually curated subsystems and subsystem-based protein families that automatically guarantees a high degree of assignment consistency.

A total of 356 subsystems participate in the metabolism of fatty acids, lipids, and isoprenoids. 71 subsystems are involved in metabolism of aromatic compounds; of these 30 are annotated and predicted to catalyse the oxidation of organic compounds with industrial and environmental relevance, such as salicylate, quinate, biphenyl and p-Hydroxybezoate (Denef *et al.*, 2004). Similarly, subsystem categories representing the metabolism of carbohydrates, amino acids, cofactors, vitamins, prosthetic groups, or pigments, proteins and metabolism of aromatic compounds are the most abundant in *S. turgidiscabies* (Figure 5.7) and *S. graminilatus* (Figure 5.8) strains. However, like most *Streptomyces*, these genes are responsible for the basic life-sustaining needs of the bacterial cells (Lau *et al.*, 2014). It is very obvious that the virulence, disease and defence subsystem category identified by RAST is more abundant in isolate_99 than in *S. turgidiscabies* and *S. graminilatus*, and they account for 93, 67 and 52 respectively (Figures 5.6, 5.7 and 5.8). Under this category, RAST detected 67 subcategories in isolate_99 which are involved in resistance to antibiotics and toxic compounds including heavy metal resistance genes. However, the presence of high numbers of these resistance genes might explain the survival of isolate_99 in such an extreme environment such as the site contaminated with heavy metals from where it was isolated.

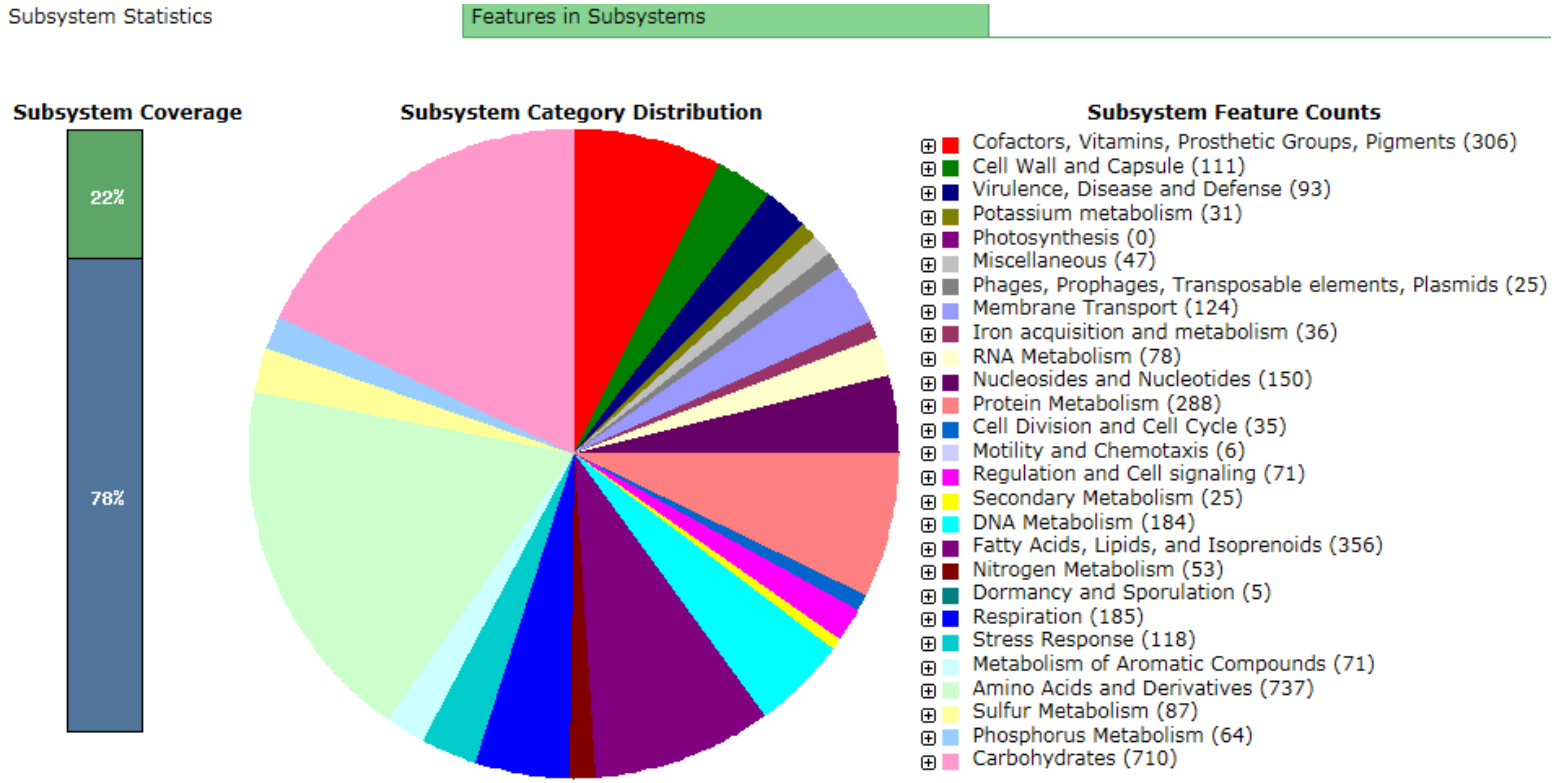


Figure 5.6: Subsystem category distribution statistics for isolate_99. The draft genome was annotated using the Rapid Annotation System Technology (RAST) server. The pie chart showed the count of each subsystem feature and the subsystem coverage was displayed using SEED viewer. The green bar of the subsystem coverage corresponds to the percentage of the proteins included in the subsystems while the blue bar corresponds to the percentage of the proteins that are not included in the subsystems.

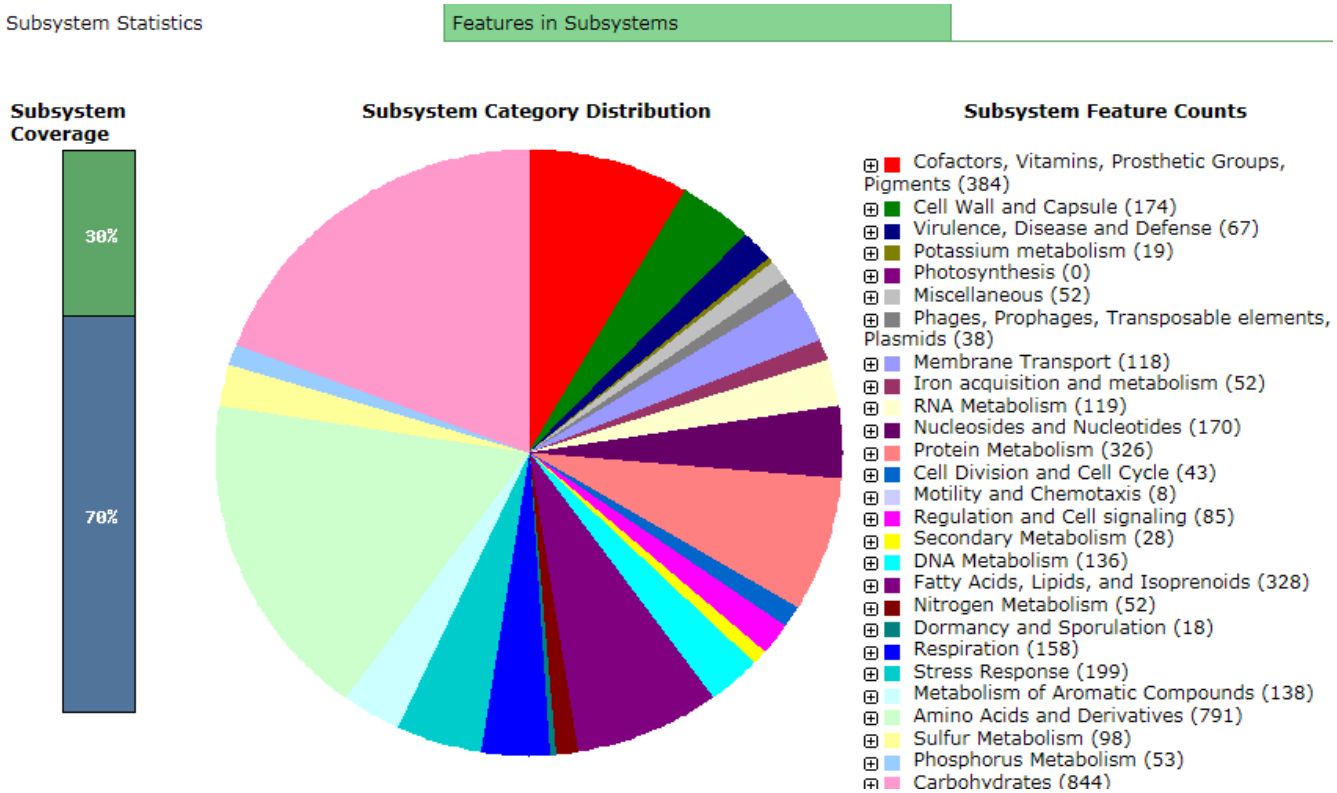


Figure 5.7: Subsystem category distribution statistics for *S.turgidiscabies*. The draft genome was annotated using the Rapid Annotation System Technology (RAST) server. The pie chart showed the count of each subsystem feature and the subsystem coverage was displayed using SEED viewer. The green bar of the subsystem coverage corresponds to the percentage of the proteins included in the subsystems while the blue bar corresponds to the percentage of the proteins that are not included in the subsystems.

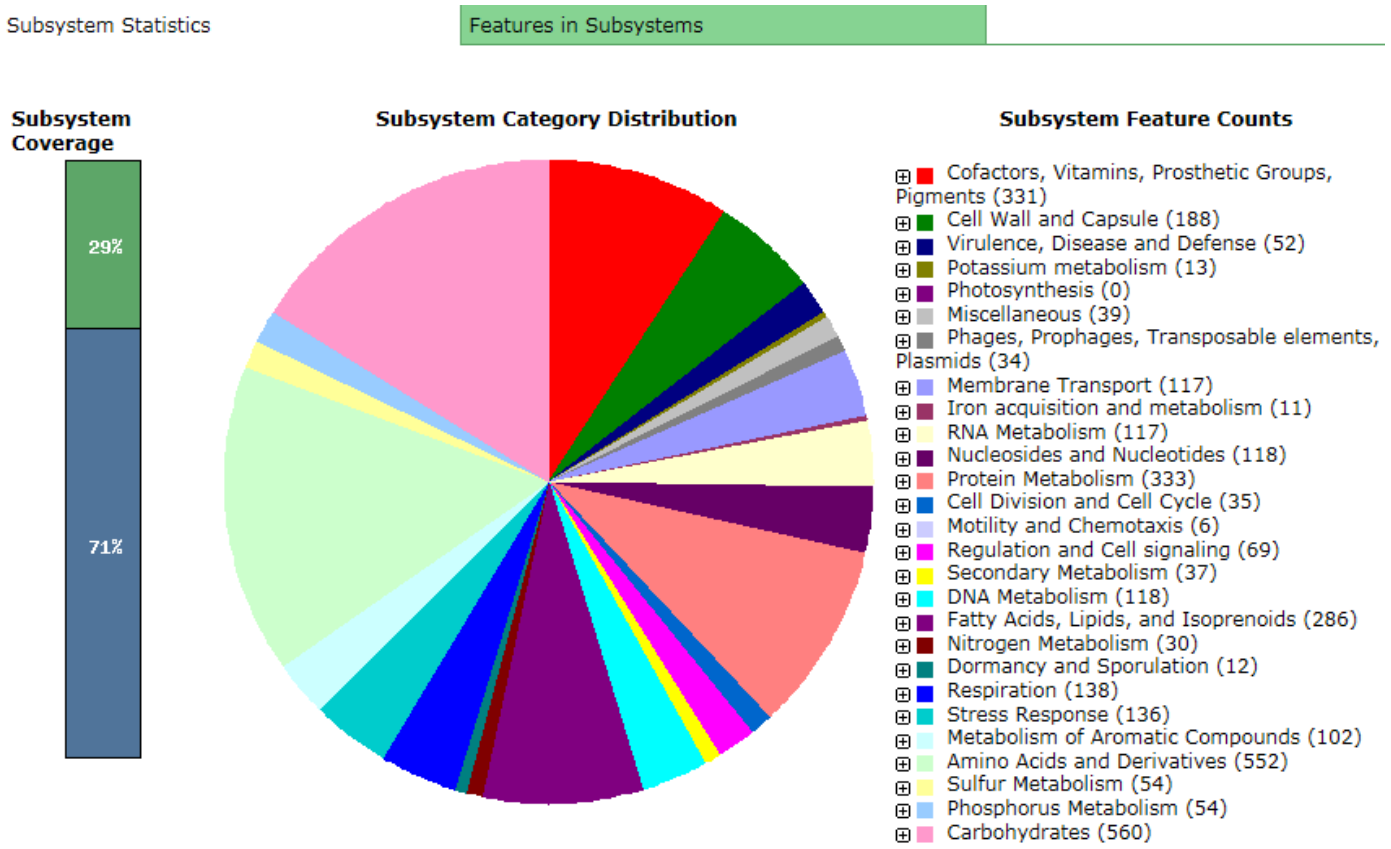


Figure 5.8: Subsystem category distribution statistics for *S. graminilatus*. The draft genome was annotated using the Rapid Annotation System Technology (RAST) server. The pie chart showed the count of each subsystem feature and the subsystem coverage was displayed using SEED viewer. The green bar of the subsystem coverage corresponds to the percentage of the proteins included in the subsystems while the blue bar corresponds to the percentage of the proteins that are not included in the subsystems.

5.6 Phenotypic microarrays identify variant metabolic pathways in isolate_99

We compared the metabolic capabilities of isolate_99 with those of the closely related strains: *S. turgidiscabies* and *S. graminilatus*, with a phenotypic microarray (PM, Bochner *et al.* 2001). Various stimulating sources of nutrients were tested including 190 different carbon sources (PM1, PM2) and 95 nitrogen sources (PM3). Osmotic and pH tolerance in terms of growth (PM9, PM10) were also investigated. In total, 477 different growth conditions were assayed to compare between isolate_99 and *S. turgidiscabies* and *S. graminilatus*. Generally, phenotypic differences based on the bacteria's ability to utilise carbon and nitrogen were noted. Furthermore, growth differences in response to osmolytes and different pH growth environments between the strains were also observed.

5.6.1 Carbon (C) supplements utilisation

When isolate_99 was compared with *S. turgidiscabies*, differences in growth rate between the two strains were observed based on their ability to utilise a range of carbon sources (Figures 5.9 & 5.10). Both strains had shown metabolic advantages over each other. However, isolate_99 grew better than *S. turgidiscabies* when D-alanine, dulcitol, D-aspartic acid, a-hydroxy butyric acid, D-malic acid, β-D-allose, L-sorbose, D-tagatose, 5-keto-D-gluconic acid, L-isoleucine, L-leucine, L-lysine, L-methionine and L-ornithine were tested. On the other hand, *S. turgidiscabies* grew better than isolate_99 when D-saccharic acid, D-serine, formic acid, D-galactonic acid, tyramine, γ-cyclodextrine, arbutine, sedoheptulasan, bytyric acid, sorbic acid, L-homoserine and a-keto valeric acid were added. Also, it was noted that caproic acid (Figure 5.10, E2) was only utilised by *S. turgidiscabies* as a carbon source. Therefore, with the exception of the above mentioned carbon sources, the carbon use pathways for the rest of the carbon sources were similar for isolate_99 and *S. turgidiscabies* (Figures 5.9 & 5.10). Furthermore, isolate_99 grew better than *S. graminilatus* when D-alanine, D-gluconic acid, D-manitol, L-rhamnose, D-melibiose, thymidine, D-glucose aminic acid, maltotriose, adenosine, propionic acid, inosine, glycyle, L-glutamic acid

D-malic acid, γ-cylodextrin, pectin, β-D-allose, D-arabitol, salicin, L-sorbose, D-tagatose, butyric acid, melibionic acid, oxalomalic acid, L-isoleucine, L-leucine, L-lysine, methionine, L-phenyl alanine and L-pyroglutamic acide were supplied. Morover, *S. graminilatus* responded better than isolate_99 when L-arabinose, L-proline, D-xylose, formic acid, 2-Deoxy adenosine, tyramine, inuline, sedoheptulasan, γ-amino butyric acid, a-keto valeric acid, quanic acid and sorbic acid were supplied. However, *S. graminilatus* was found to grow significantly better than isolate_99 in well C5 supplemented with tween 20. By contrast, there were little or no differences in growth rates when other carbon sources were compared between isolate_99 and *S. graminilatus* (Figures 5.11 & 5.12).

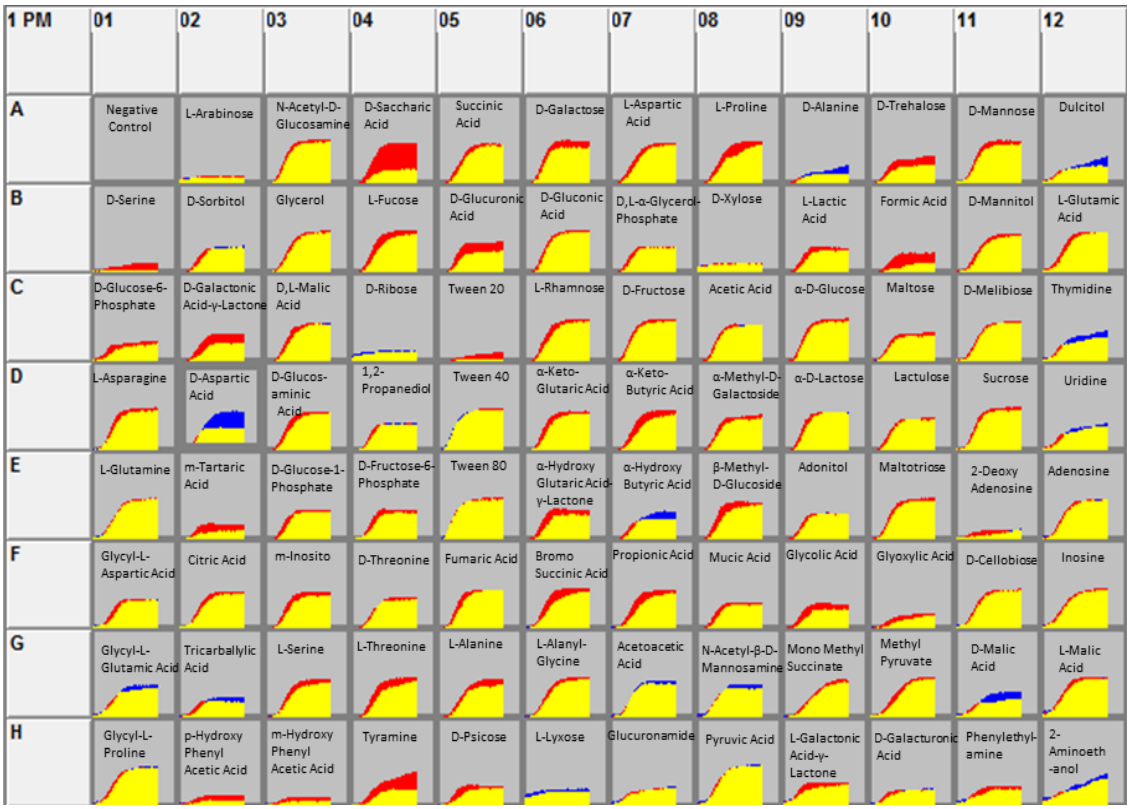


Figure 5.9: Parametric BioLog graph depicting carbon utilisation (PM1) by isolate_99 vs. *S. turgidiscabies*. Isolate_99 is recorded as a blue tracing and *S. turgidiscabies* as a red. Areas of overlap (no change) of the two kinetic curves are coloured yellow, whereas differences are highlighted as patches of blue or red. Apart from differences in respiration with D-alanine (A9), dulcitol (A12), a-hydroxy butyric acid (E7) and D-malic acid (G11) by isolate_99, and D-serine (B1), formic acid (B10), D-galactonic acid (C2), tyramine (H4) by *S. turgidiscabies* as carbon sources, differences with D-saccharic acid (A4) and D-aspartic acid (D2) were most notable in this assay.

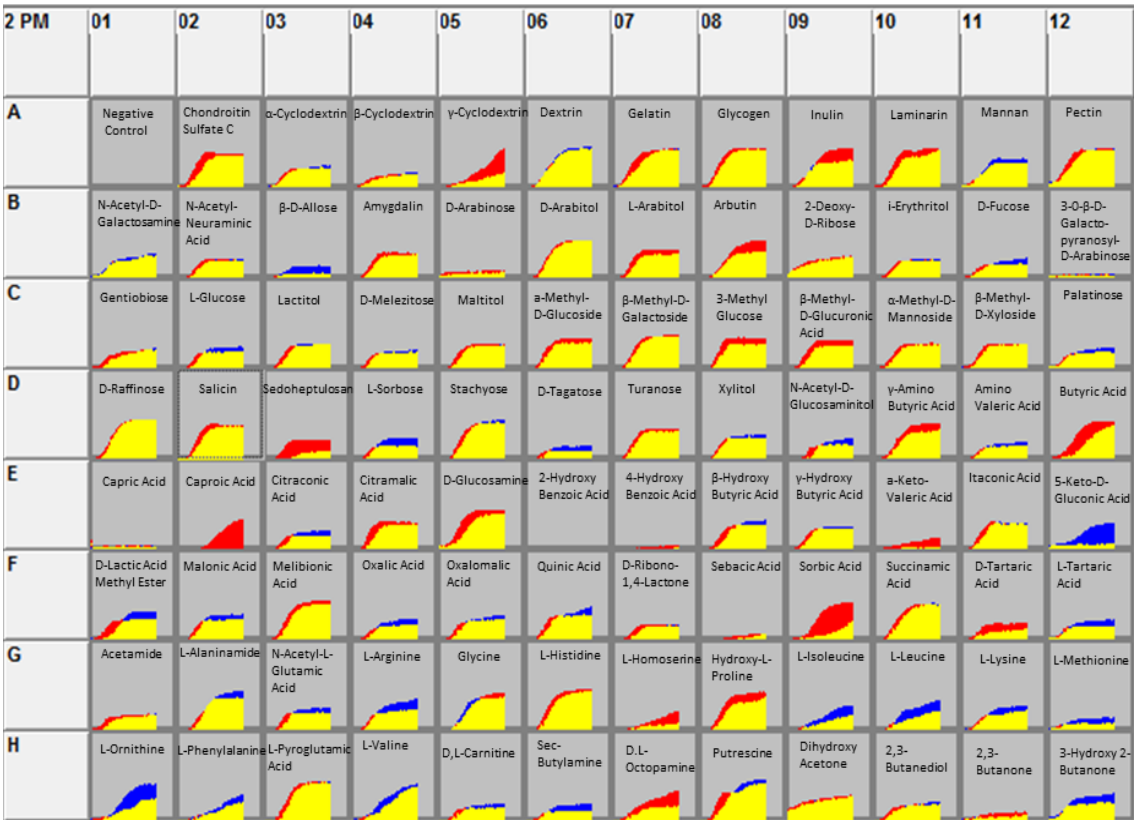


Figure 5.10: Parametric BioLog graph depicting carbon utilisation (PM2) by isolate_99 vs. *S. turgidiscabies*. Isolate_99 is recorded as a blue tracing and *S. turgidiscabies* as a red. Areas of overlap (no change) of the two kinetic curves are coloured yellow, whereas differences are highlighted as patches of blue or red. Apart from differences in respiration with β-D-allose (B3), L-sorbose (D4), D-tagatose (D6), L-isoleucine (G9), L-leucine (G10), L-lysine (G11), L-methionine (G12) by isolate_99 and arbutine (B8), butyric acid (D12) and L-homoserine (G7) by *S. turgidiscabies* as carbon sources, differences with γ-cyclodextrine (A5), sedoheptolosan (D3), caproic acid (E2), 5-keto-D-gluconic acid (E12) and sorbic acid (F9) were most notable in this assay.

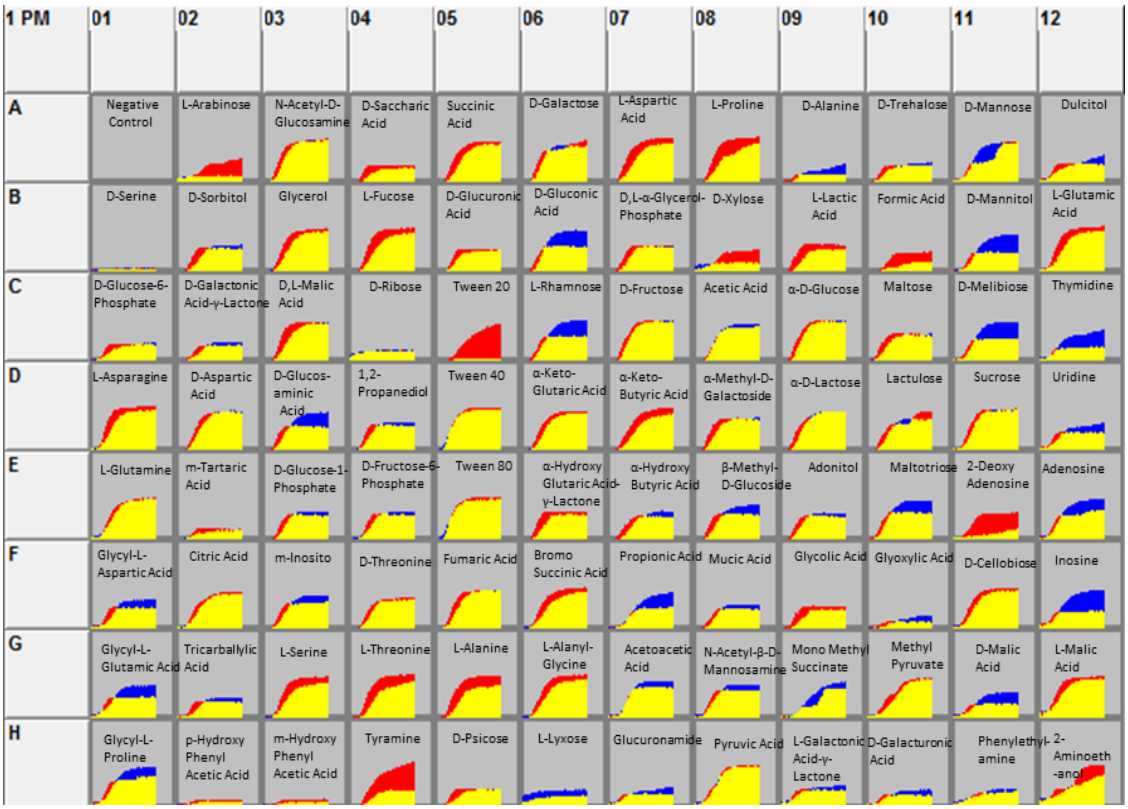


Figure 5.11: Parametric BioLog graph depicting carbon utilisation (PM1) by isolate_99 vs. *S. graminilatus*. Isolate_99 is recorded as a blue tracing and *S. graminilatus* as a red. Areas of overlap (no change) of the two kinetic curves are coloured yellow, whereas differences are highlighted as patches of blue or red. Apart from differences in respiration with D-alanine (A10), thymidine (C12), D-glucoseaminic acid (D3), maltotriose (E10), adenosine (E12), propionic acid (F7) by isolate_99, and L-proline (A8), D-xylose (B8) and formic acid (B10) by *S. graminilatus* as carbon sources, differences with L- arabinose (A2), D-gluconic acid (B6), D-mannitol (B11), Tween 20 (C5), L-rhamnose (C6), D-melibiose (C11), 2-deoxy adenosine (E11), Inosine (F12) and tyramine (H4) were most notable in this assay.

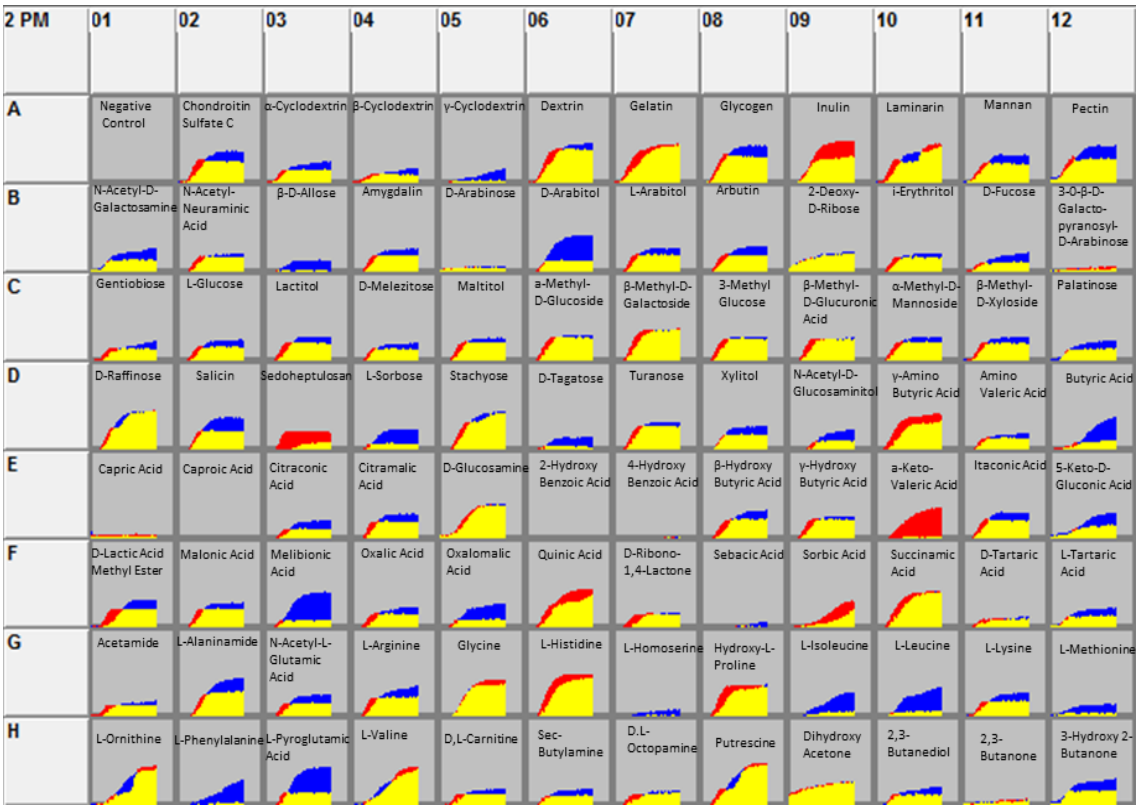


Figure 5.12: Parametric BioLog graph depicting carbon utilisation (PM2) by isolate_99 vs. *S. graminilatus*. Isolate_99 is recorded as a blue tracing and *S. graminilatus* as a red. Areas of overlap (no change) of the two kinetic curves are coloured yellow, whereas differences are highlighted as patches of blue or red. Apart from differences in respiration with γ-cylodextrin (A5), pectin (A12), salicin (D2), D-tagatose (D6), oxalomalic acid (F5), L-lysine (G11), L-methionine (G12) by isolate_99, and γ-amino butyric acid (D10), quanic acid (F6) and sorbic acid (F9) by *S. graminilatus* as carbon sources, differences with inulin (A9), β-D-allose (B3), D-arabitol (B6), sedoheptulasan (D3), L-sorbose (D4), butyric acid (D12), a-keto valeric acid (E10), melibionic acid (F3), L-isoleucine (G9), L-leucine (G10), L-phenylalanine (H2) and L-pyroglutamic acid (H3) were most notable in this assay.

5.6.2 Nitrogen (N) supplements utilisation

Phenotypic microarrays for the nitrogen utilisation generated various positive growth advantages for isolate_99 over both *S. turgidiscabies* and *S. graminilatus* strains (Figures 5.13 & 5.14). Isolate_99 showed better growth than *S. turgidiscabies* in presence of the following N sources: biuret, D-glutamic acid, N-acetyl-L-glutamic acid, methyl amine, N-amylamine, N-butylamine, ethylamine, ethylenediamine, histamine, acetamide, D-glucoseamine, amino-N-caproic acid and a-amino-N-valeric acide, whereas *S. turgidiscabies* showed growth advantages than isolate_99 on only two N sources: L-tyrosine and N-acetyl-D-galactose amine. However, there were little or no differences in growth rates when other nitrogen sources were compared between isolate_99 and *S. turgidiscabies* (Figures 5.13). On the other hand, isolate_99 showed superior growth rate than *S. graminilatus* for utilisation of the following N sources: biuret, L-methionine, L-phenylalanine, D-alanine, D-asparagine, D-glutamic acid, D-lysine, D-serine, D-valine, L-ornithine, N-acetyl-L-glutamic acid, L-pyroglutamic acid, methylamine, N-amylamine, N-butylamine, ethylamine, ethylenediamine, histamine, acetamide, D-glucoseamine, cytosine, thymine, thymidine, uracil, amino-N-caproic acid, amino-N-valeric, a-amino-N-valeric acid and Gly-Glu, whereas a growth advantage was observed for *S. graminilatus* in the presence of parabanic acid only (Figure 5.14). No growth advantages between isolate_99 and *S. graminilatus* were observed for the rest of N sources.

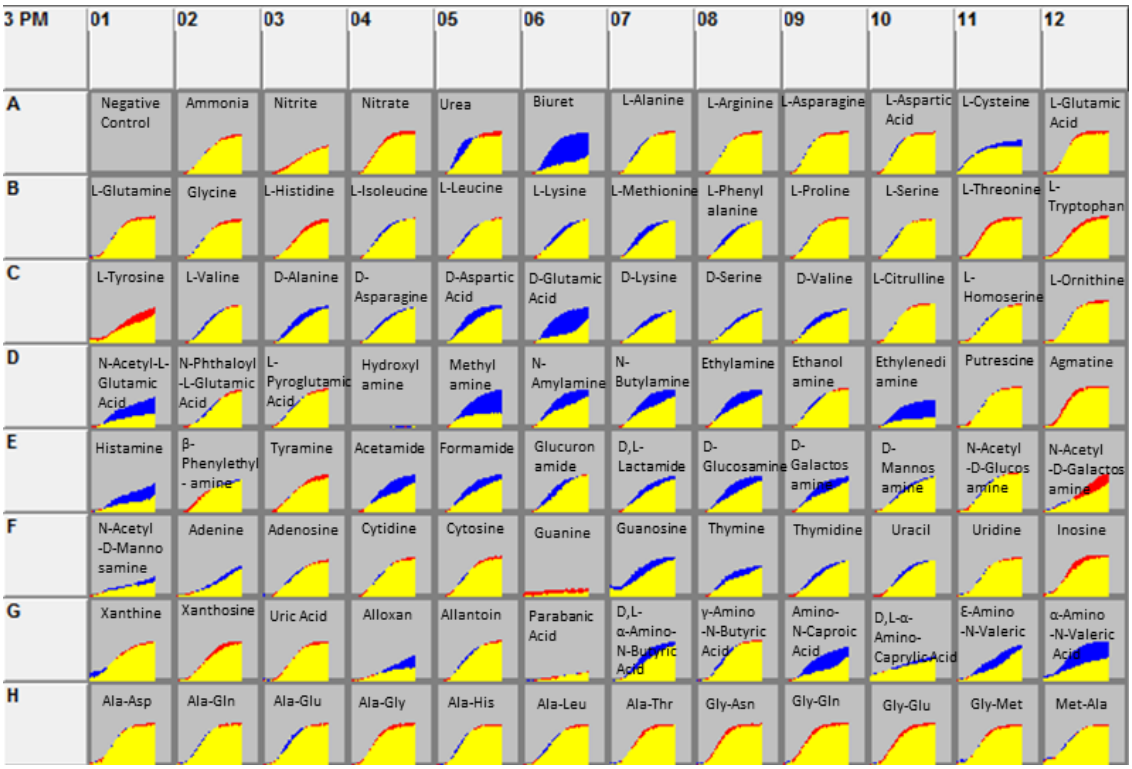


Figure 5.13: Parametric BioLog graph depicting nitrogen utilisation (PM3) by isolate_99 vs. *S. turgidiscabies*. Isolate_99 is recorded as a blue tracing and *S. turgidiscabies* as a red. Areas of overlap (no change) of the two kinetic curves are coloured yellow, whereas differences are highlighted as patches of blue or red. Nitrogen sources more rabidly metabolised by isolate_99 are biuret (A6), D-glutamic acid (C6), N-acetyl-L-glutamic acid (D1), methyl amine (D5), N-amylamine (D6), N-butylamine (D7), ethylamine (D8), ethylenediamine (D10), histamine (E1), acetamide (E4), D-glucoseamine (E9), amino-N-caproic acid (G9) and a-amino-N-valeric acid (G12), and L-tyrosine (C1) and N-acetyl-D-galactose amine (E12) by *S. turgidiscabies* in this assay.

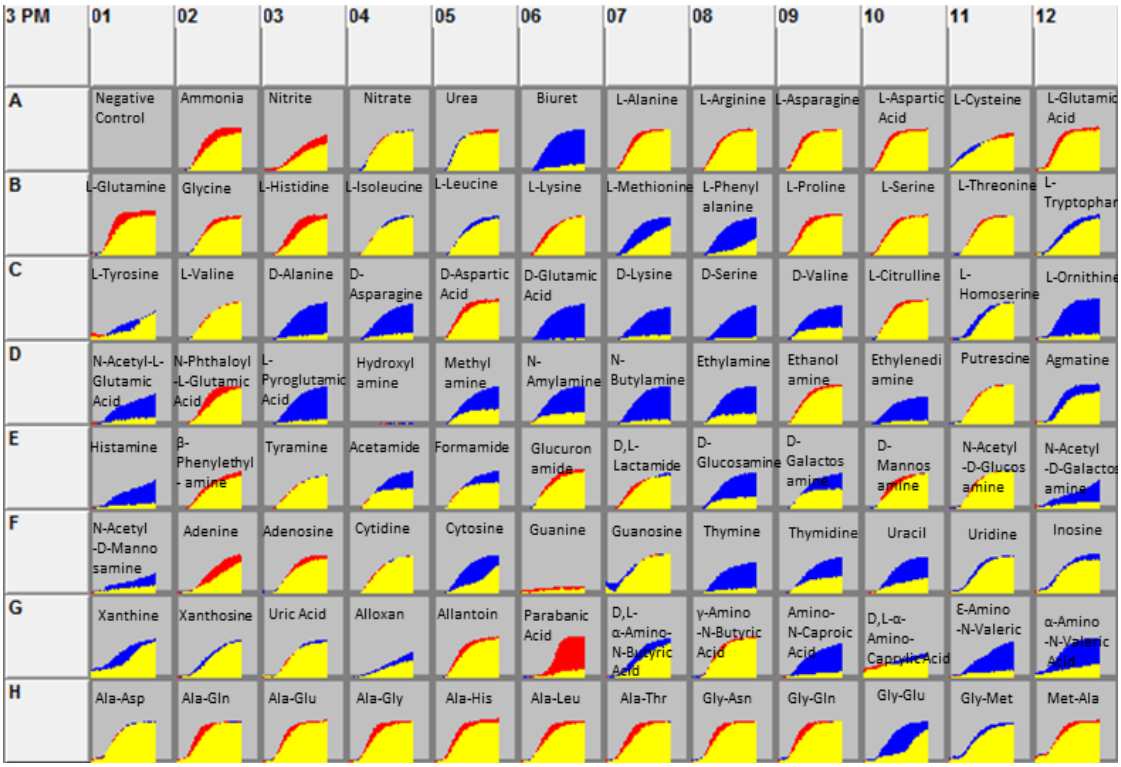


Figure 5.14: Parametric BioLog graph depicting nitrogen utilisation (PM3) by isolate_99 vs. *S. graminilatus.* Isolate_99 is recorded as a blue tracing and *S. graminilatus* as a red. Areas of overlap (no change) of the two kinetic curves are coloured yellow, whereas differences are highlighted as patches of blue or red. Nitrogen sources more rabidly metabolised by isolate_99 arebiuret (A6), L-methionine (B7), L-phenylalanine (B8), D-alanine (C3), D-asparagine (C4), D-glutamic acid (C6), D-lysine (C7), D-serine (C8), D-valine (C9), L-ornithine (C12), N-acetyl-L-glutamic acid (D1), L-pyroglutamic acid (D3), methylamine (D5), N-amylamine (D6), N-butylamine (D7), ethylamine (D8), ethylenediamine (D10), histamine (E1), acetamide (E4), D-glucoseamine (E8), cytosine (F5), thymine (F8), thymidine (F9), uracil (F10), amino-N-caproic acid (G9), amino-N-valeric (G11), a-amino-N-valeric acid (G12) and Gly-Glu (H10) , whereas parabanic acid (G6) was most notable by *S. graminilatus*.

The obtained carbon and nitrogen patterns of phenotypes suggested that isolate_99 is different from both *S. turgidiscabies and S. graminilatus* strains. Using data from PM1 and PM2 (carbon sources), a number of phenotypic differences based on their ability to utilise a range of carbon sources were observed (Figures 5.9, 5.10, 5.11 & 5.12). The strains show obvious discrimination between carbon sources. It can be noted that isolate_99 and *S. graminilatus* couldn't utilise caproic acid as a carbon source(Figure 5.9, E2 & 5.10, E2) whereas *S. turgidiscabies* was active and showed good growth rate on caproic acid (Figure 5.15, right). Caproic acid is a saturated fatty acid that has six carbons and one carboxylic group ($C_6H_{12}O_2$). It has been used in diverse industrial applications such as perfumes, medicine, food additives, lubricating grease, tobacco flavour, rubber, and dyes. Caproic acid is also used as a precursor for the biosynthesis of hexyl esters, which can be converted to hexanol by esterification and hydrogenation reaction (Cheon *et al.*, 2014). Interestingly, Tween 20 supported growth of *S. graminilatus* (Figure 5.15, left) as a source of carbon and energy but did not support growth of both isolate_99 and *S. turgidiscabies* (Figure 5.8, C5). It is possible that Tween 20 restricted growth of the isolate_99 and *S. turgidiscabies* due to generating lauric acid (C12:0), which is among saturated fatty acids that has been reported to have toxicity against Gram-positive bacteria than Gram-negative bacteria (Kitahara *et al.*, 2004; Rouse *et al.*, 2005; Nakatsuji *et al.*, 2009). Conversely, a very good growth rate was observed on Tween 40 and Tween 80 compounds by the three strains (Figures 5.10, D5, E5 & 5.10, D5, E5). The metabolism of these two substrates involves the β-oxidation pathway where oleic acid is produced, which is non-toxic for bacterial growth and has been reported to support growth of a wide range of bacterial species including *Streptomyces* (Slijkhuis *et al.*, 1984; Skrivanova *et al.*, 2005; Orro *et al.*, 2015). Furthermore, in order to independently confirm the Biolog respiration results, the ability of isolate_99, *S. turgidiscabies* and *S. graminilatus* to grow on minimal medium in presence of caproic acid and Tween 20 as sole carbon sources was tested. However, the results were concordant with the phenotype microarray data (Figure 5. 16).

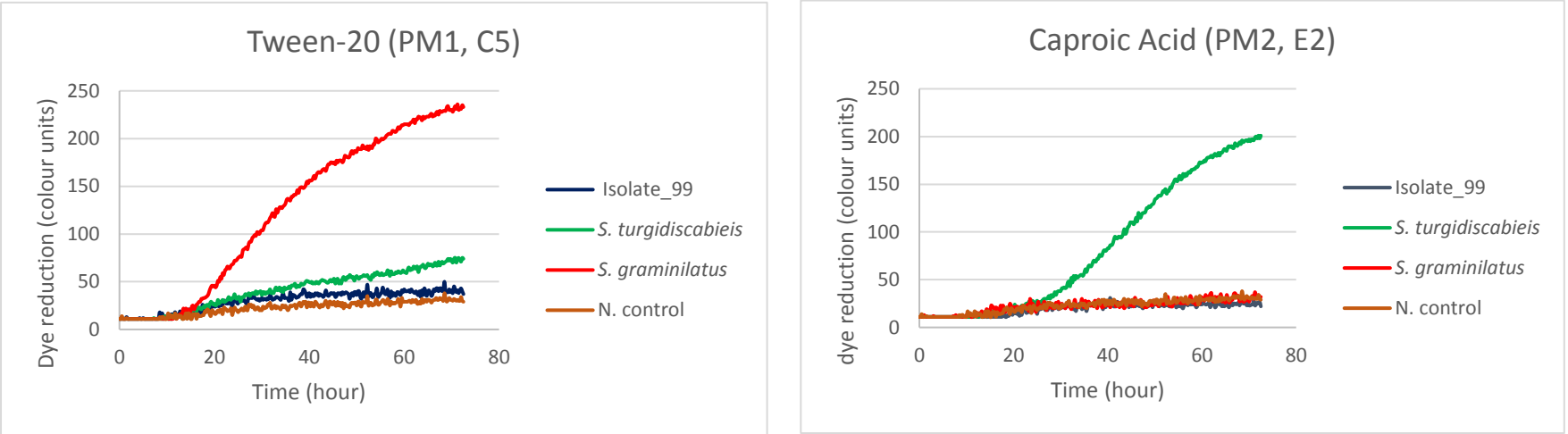


Figure 5.15: Examples of sigmoidal curves of colour development measured in colour units which resulted from tertrazolium dye reduction during 72h of incubation for the substrates Tween-20 (left) and caproinic acid (right). Tween-20 is more rapidly metabolised by *S. graminilatus* (shown in red) than isolate_99 and (shown in blue) and *S. turgidiscabies* (shown in green). Similarly, caproic acid is most notable metabolised only by *S. turgidiscabies*.

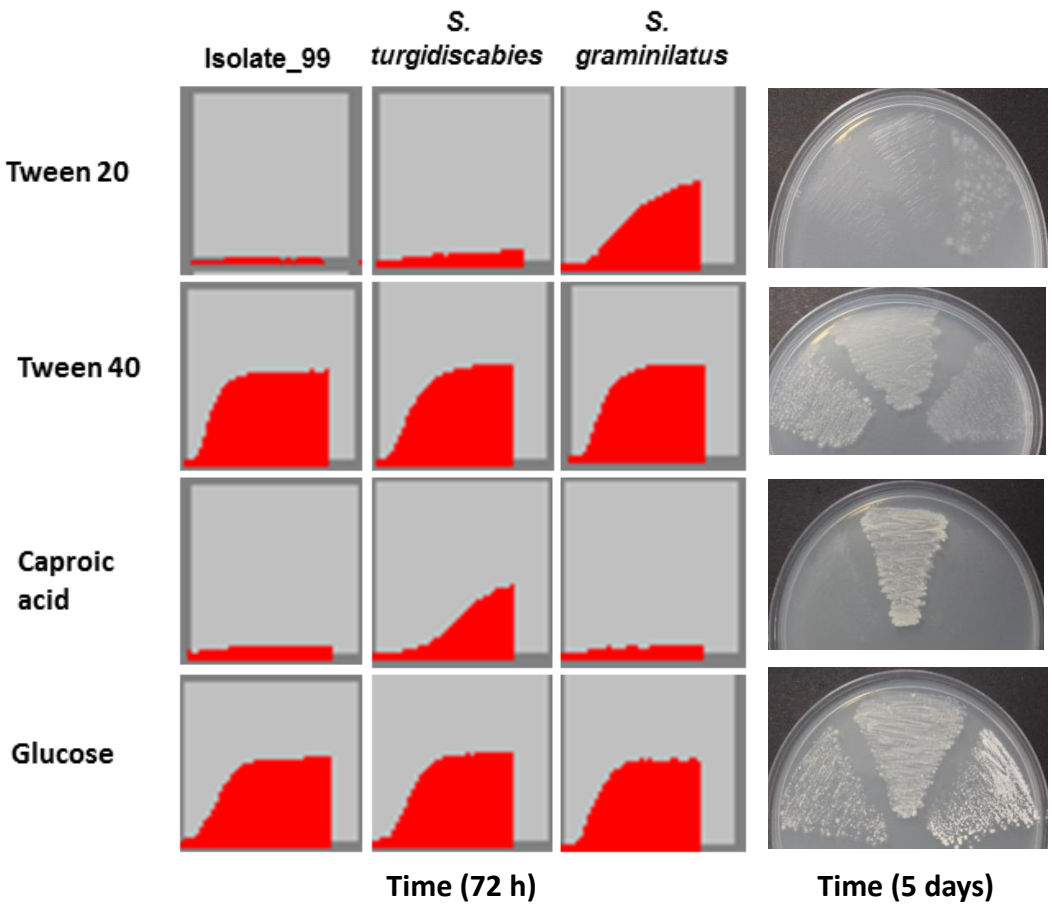


Figure 5.16: PM plots (left) of the utilisation of Tween 20, Tween 40, Caproic acid and Glucose by isolate_99, *S. turgidiscabies* and *S. graminilatus*. Data derived from the increase in intensity of the colour of the reduced tertrazolium dye, measured in colour units, are plotted against time (72 h) using the PM software from Biolog. The three strains were streaked on MM agar plates (right) with Tween 20 (20 mM), Tween 40 (20 mM) and Caproic acid (20 mM) as carbon sources and incubated at 30 °C for 5 days. Glucose (20 mM) was used as a positive control.

Data from the phenotypic microarray (PM3), showed significant differences in nitrogen utilisation when isolate_99, *S. turgidiscabies* and *S. graminilatus* were compared (Figures 5.13 & 5.14). It is very obvious that isolate_99 had a better growth rate during utilisation a wide range of nitrogen sources compared to strains in particularly, *S. graminilatus* (Figure 5.14). This difference can be associated with the great difference between the two strains in terms of genes predicted to be involved in amino acids of subsystem category identified by RAST. They are more abundant in isolate_99 than in *S. graminilatus*, and they account for 737 and 552 respectively (Figures 5.6 & 5.8).

It was possible to measure the metabolic activity of the best used carbon and nitrogen substrates of the three strains by calculating the maximum signal (score) that can be extracted from the metabolic activity curves under the conditions provided in each well of phenotypic microarray plate. The intensity of the purple colour developed from a redox dye due the energy production and cell respiration in each well can reflect the amount of the substrate which is metabolised by the tested strains. Accordingly, the most intensively metabolised carbon and nitrogen substrates by isolate_99 were: N-acetyl-D-glucoseamine (GlcNAc) (score:240) from the carbohydrates guild, Tween 40 (score:236) from the polymers guild, sucrose (score:235) from the carbohydrates guild, D- glucose (score:232) from the carbohydrates guild and L-rhamnose (score:232) from the carbohydrates guild as carbon sources; and biuret (score:252) from other the miscellaneous guild, L-glutamic acid (score:252), L-ornithine (score:251), L-asparagine (score:250), and L-alanine (score:249) as nitrogen sources from the amino acid guild. On the other hand, both *S. turgidiscabies* and *S. graminilatus* strains showed different metabolic activities and intensities of carbon and nitrogen substrates. Amongst carbon sources: GlcNAc (score:270), sucrose (score:267), D-mannose (score:266), D-galactose (score:264) and L-rhamnose (score:264) from the carbohydrates panel, and L-proline (score:278) from the amino acids panel, L-glutamic acid (score:276) from the amino acids panel, inulin (score:273) from the polymers guild,L-fucose (score:268) from the carbohydrates panel and GlcNAc (score:265) from the carbohydrates panel; also, amongst nitrogen sources: L-glutamic acid (score:288) from the amino acids panel,

nitrite (score:287) from the inorganic panel, L-ornithine (score:283), L-arginine (score:283) from the amino acids panel and L-alanine (score:282) from the amino acids panel, and ammonia (score:285) from the inorganic panel, L-glutamic acid (score:285), L-alanine (score:276), L-arginine (score:276) and L-asparagine (score:275) from the amino acids panel were the most metabolised by *S. turgidiscabies* and *S. graminilatus* respectively.

However, although many nutritionally rich and minimal media support the growth of most known strains of *Streptomyces* (Kieser *et al.*, 2000; Hopwood, 2007), some of these strains vary significantly in terms of the source, chemical form, and amount of nutrients they need. Moreover, knowledge on the most nutrients preferred by new *Streptomyces* isolates may provide information for formulating and designing new nutritional media. Thus, the intensity of metabolic activity of carbon and nitrogen sources obtained from the Biolog assays can allow for selecting only relevant carbon and nitrogen sources when formulating media for isolate_99. The carbon and nitrogen sources that produced the highest amount of growth rate (intensity) across all substances tested were, remarkably, GlcNAc and glutamate respectively for the three strains. Moreover, studies have revealed that GlcNAc, the monomer of chitin and constituent of bacterial cell wall peptidoglycan, and the related amino acid glutamate are highly preferred carbon and nitrogen sources for Streptomycetes (van Wezel et al, 2006; Swiatek *et al.*, 2012). Additionally, Rigali *et al.* (2008) have shown that under famine conditions, GlcNAc addition stimulated antibiotic production in *S. clavuligerus, S. collinus, S. griseus, S. hygroscopicus* and *S. venezuelae* grown on MM agar (5 mM GlcNAc or higher) by functioning as an allosteric effector of the pleiotropic transcriptional repressor *DasR* regulon, which controls the GlcNAc transport and metabolism as well as antibiotic production. Nonetheless, to verify findings from the Biolog scale and to obtain optimal production of metabolites, larger scale fermentations along with investigating a possible relation between nitrogen, carbon and other nutrient sources are needed.

However, results obtained from carbon and nitrogen patterns can be used to discriminate between isolate_99 and the closely related *S. turgidiscabies* and *S.*

graminilatus strains. To our knowledge, there have been no published studies on the nutrients utilisation of *Streptomyces* using the Phenotype Microarrays. Therefore, we could not compare this study against other *Streptomyces* strains; in particular to those of complete published genomes. Furthermore, our phenotype array results also reveal some key phenotype differences particularly for carbon and nitrogen sources utilisation that might be enough to discriminate adequately between two species.

5.6.3 Osmolyte tolerance and survival in different pH conditions

The effects of different osmolytes and pH ranges on the three strains were tested by using plates PM9 and PM10. In response to the presence of osmolytes, isolate_99 could not tolerate high concentrations of NaCl, urea, sodium lactate and ammonium sulphate compared with both *S. turgidiscabies* and *S. graminilatus* strains (Figures 5.17 & 5.18). Isolate_99 could tolerate 2% NaCl, 2% urea, 1% sodium lactate and moderate growth on 100 mM ammonium sulphate. In contrast, *S. turgidiscabies* had a growth advantage over isolate_99 and could tolerate 4% NaCl, 3% urea, 4% sodium lactate and 100 mM sodium nitrate, whereas, *S. graminilatus* could tolerate 2% NaCl, 5% urea, 3% sodium lactate and 100 mM ammonium sulphate. Further, the three isolates were able to grow in up to 20% ethylene glycol, 200 mM sodium phosphate and 100 mM sodium nitrate, also no or very little growth was observed on sodium formate, sodium benzoate and sodium nitrite. It was also noted that isolate_99 and *S. turgidiscabies* grew well in the precence of 3% potassium chloride, but *S. graminilatus* could not.

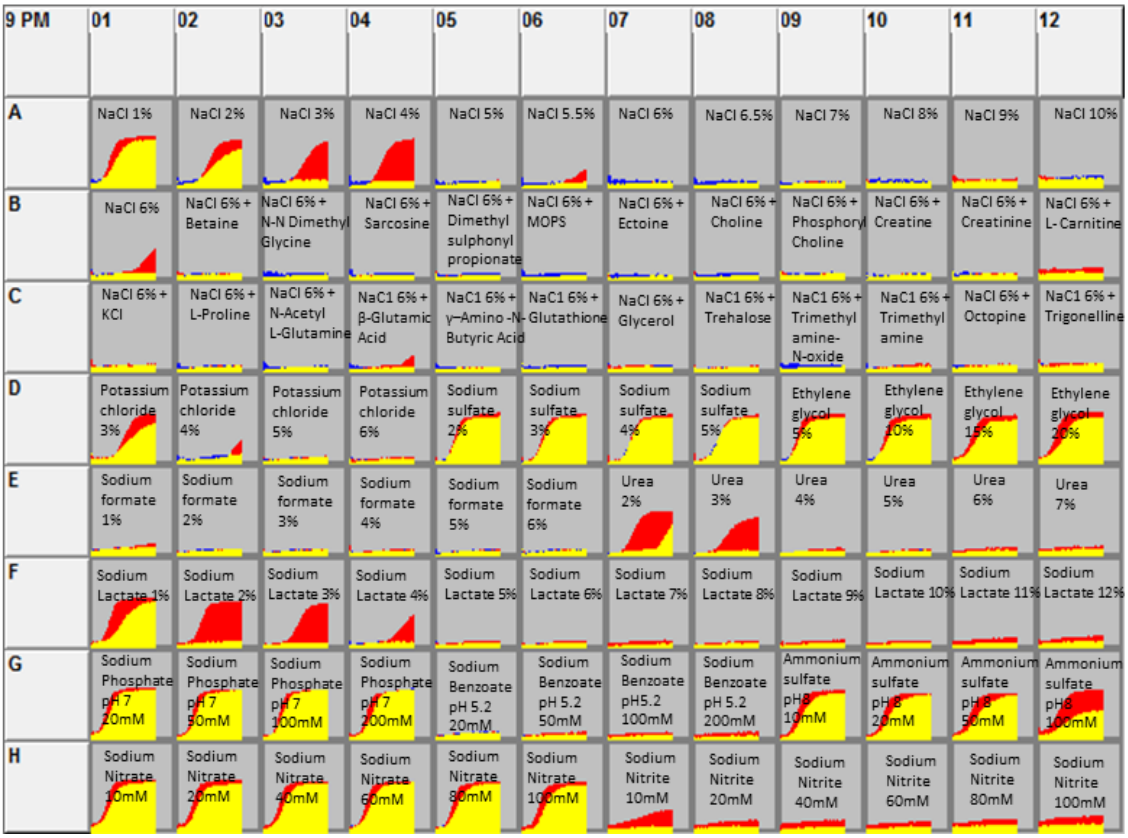


Figure 5.17: Parametric BioLog graph depicting osmolyte activities (PM9) by isolate_99 vs. *S. turgidiscabies*. Isolate_99 is recorded as a blue tracing and *S. turgidiscabies* as a red. Areas of overlap (no change) of the two kinetic curves are coloured yellow, whereas differences are highlighted as patches of blue or red. *S. turgidiscabies* showed notable increased metabolism (shown in red) at 3% NaCl (A3), 4% NaCl (A4), 2% urea (E7), 3% urea (E8), 2% sodium lactate (F2), 3% sodium lactate (F3), 4% sodium lactate and 100mM ammonium sulphate (G12).

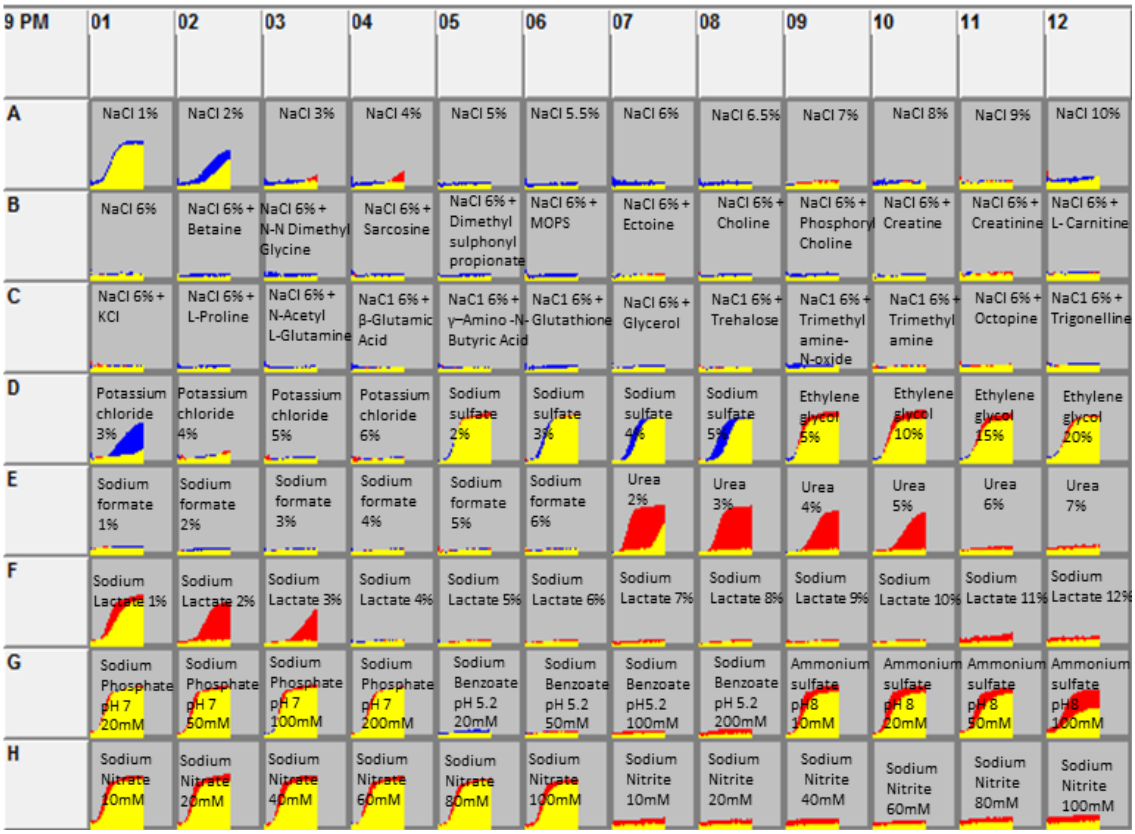


Figure 5.18: Parametric BioLog graph depicting osmolyte activities (PM9) by isolate_99 vs. *S. graminilatus*. Isolate_99 is recorded as a blue tracing and *S. graminilatus* as a red. Areas of overlap (no change) of the two kinetic curves are coloured yellow, whereas differences are highlighted as patches of blue or red. Isolate_99 showed notable increased metabolism (shown in blue) at 3% potassium chloride (D1), whereas *S. graminilatus* showed notable increased metabolism (shown in red) at 2-5% urea (E7-E10, 2% sodium lactate (F2), 3% sodium lactate (F3) and 100mM ammonium sulphate (G12).

Finally, metabolic activity of the three strains over a broad range of pH 3.5–10 was determined using PM10 Biolog plate. The strains were viable within a pH range from 5.0 to 10; however, while isolate_99 could not grow at pH 4.5, both *S. turgidiscabies* and *S. graminilatus* strains showed a general tolerance to low pH in the presence of some amino acids as noted in wells D4, 6, 7, 8 and 10 (Figures 5.19 & 5.20). Probably, the activity of decarboxylases that generate alkaline amines by the catabolism of these compounds (Viti *et al.*, 2007) gave protection against the effects of incubation at pH 4.5. However, the difference in phenotypic growth generated from PM9 and PM10 plates might be associated with the difference between these strains in terms of genes involved in various stress responses (osmotic stress, detoxification, heat and cold shock, detoxification stress). Annotation of the genomes suggested that *S. turgidiscabies* and *S. graminilatus* had resistance genes conferring environmental advantages over isolate_99, where 199, 136 and 118 stress response genes were predicted in the genomes of *S. turgidiscabies*, *S. graminilatus* and isolate_99 respectively (Figures 5.6, 5.7 and 5.8). Further, previous studies have shown that the optimum concentration of NaCl for most *Streptomyces* strains was 5-10% (Urzi and Realini, 1998; Walsh *et al.*, 2005). Another previous study carried by Tresner *et al.* (1968) has shown that salt tolerance in *Streptomyces* can be divided into 3 categories of low (<4%), intermediate (>7%) and high (>10%). However, solidified cultural media were applied for the all previous mentioned studies which contain agar that might mediate the inhibitory effect of high salt concentrations compared to liquid medium that is used in Biolog plates. Moreover, it has been reported that *Streptomyces* species grow well at neutral pH, however others are acidophiles, alkaliphiles, halophiles or thermophiles (Goodfellow, 2012). Thus, some acidophilic actinomycetes have been observed to grow in the range from 3.5 to 6.5, with optimal rates at pH 4.5 to 5.5 (Kim *et al.*, 2003).

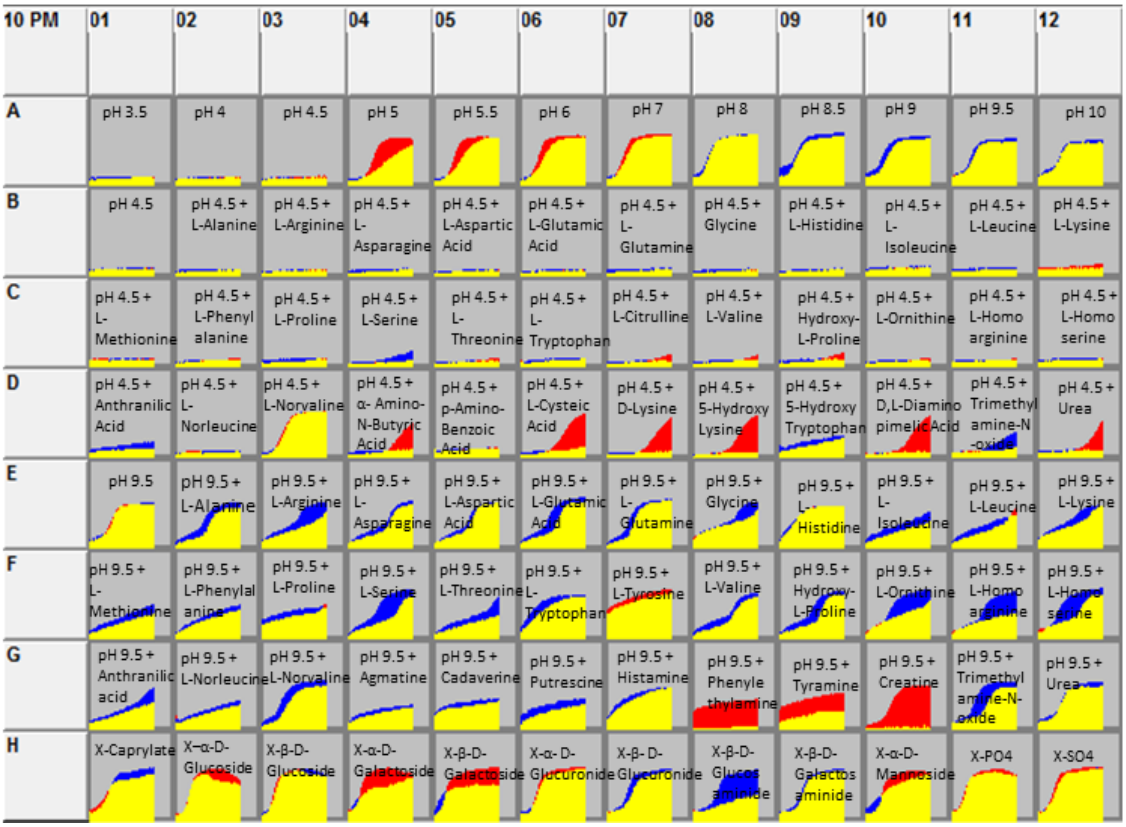


Figure 5.19: Parametric BioLog graph depicting pH condition activities (PM10) by isolate_99 vs. *S. turgidiscabies*. Isolate_99 is recorded as a blue tracing and *S. turgidiscabies* as a red. Areas of overlap (no change) of the two kinetic curves are coloured yellow, whereas differences are highlighted as patches of blue or red. Isolate_99 showed notable increased metabolism (shown in blue) at pH 9.5 + L-arginine (E3), pH 9.5 + L-serine (F4), pH 9.5 + L-ornithine (F10), pH 9.5 + L-homo arginine (F11), pH 9.5 + L-homo serine (F12), pH 9.5 + trimethyl amine-N-oxide (G11) and X-β-D-glucose aminide (H8), whereas *S. turgidiscabies* showed notable increased metabolism (shown in red) at pH 4.5 + α-amino-N-butyric acid (D4), pH 4.5 + L-cysteic acid (D6), pH 4.5 + D-lysine (D7), pH 4.5 + 5-hydroxy lysine (D8), pH 4.5 + D,L-diamino pimelic acid (D10), pH 4.5 + urea (D12), pH 9.5 + phenyle thylamine (G8), pH 9.5 + tyramine (G9) and pH 9.5 + creatine (G10).

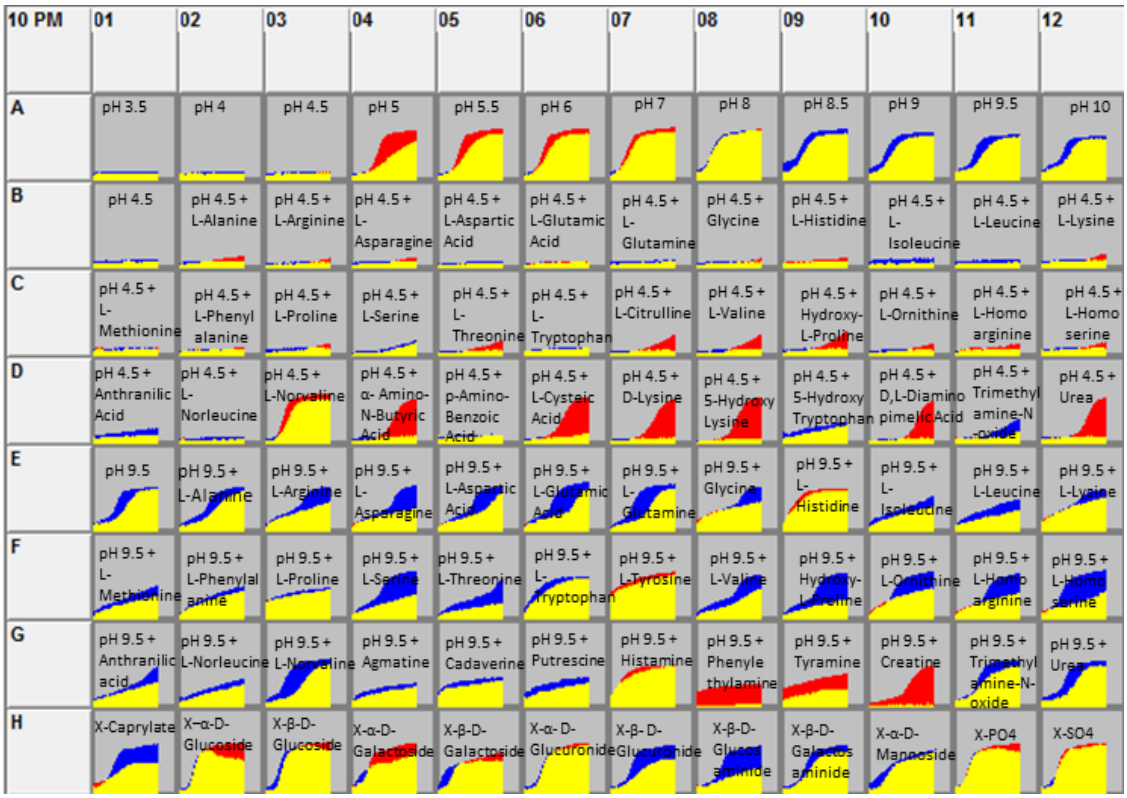


Figure 5.20: Parametric BioLog graph depicting pH condition activities (PM10) by isolate_99 vs. *S. graminilatus*. Isolate_99 is recorded as a blue tracing and *S. graminilatus* as a red. Areas of overlap (no change) of the two kinetic curves are coloured yellow, whereas differences are highlighted as patches of blue or red. Isolate_99 showed notable increased metabolism (shown in blue) at pH 9.5 + L-arginine (E3), pH 9.5 + L-serine (F4), pH 9.5 + L-ornithine (F10), pH 9.5 + L-homo arginine (F11), pH 9.5 + L-homo serine (F12), pH 9.5 + trimethyl amine-N-oxide (G11) and X-β-D-glucose aminide (H8), whereas *S. graminilatus* showed notable increased metabolism (shown in red) at pH 4.5 + α-amino-N-butyric acid (D4), pH 4.5 + L-cysteic acid (D6), pH 4.5 + D-lysine (D7), pH 4.5 + 5-hydroxy lysine (D8), pH 4.5 + D,L-diamino pimelic acid (D10), pH 4.5 + urea (D12), pH 9.5 + phenyle thylamine (G8), pH 9.5 + tyramine (G9) and pH 9.5 + creatine (G10).

Chapter 6:

Genome Mining for Biosynthetic Gene Clusters (BGCs) using antiSMASH and ClusterFinder Bioinformatics Tools of Isolate_99

5.1 Genome mining

Fundamentally, genome mining is a bioinformatics approach used for the discovery of novel natural products (NPs), which *in silico* predicts Biosynthetic Gene Clusters (BGCs). BGCs are groups of genes, usually encoded within a single locus, responsible for secondary metabolites (SMs) biosynthesis pathways, which represent the main sources of new antibiotics and pharmaceuticals (Cruz-Morales *et al.*, 2016). Although SMs have traditionally been described in plants, green algae and fungi, the vast majority of these new bioactive metabolites have obtained from bacteria (Ziemert *et al.*, 2016). Sequencing of bacterial genomes has revealed the potential of these microorganisms to harbour putative BGCs which display promising potential for discovering new bioactive natural products. Among these, the genus *Streptomyces* has become an especially attractive screening source for decades in search for new drugs (Ziemert *et al.*, 2016). Since the 1940's, large numbers of significant bioactive molecules have been identified from this genus including antibiotics, immunomodulators, anticancer agents, antiviral drugs, herbicides, and insecticides (Bibb, 2005; Hopwood, 2007). Furthermore, genome analyses of different model *Streptomyces* strains, such as *Streptomyces coelicolor* A3(2), *Streptomyces griseus* NBRC 13350 and *Streptomyces avermitilis* MA-4680 (Bentley *et al.*, 2002; Harrison and Studholme, 2014), showed that each strain harbours on average about twenty five secondary metabolic gene clusters in their genomes. These encode the enzymes necessary to synthesize twenty or more potential secondary metabolites belonging to different structural classes, even though fewer than six metabolites were known at that time (Bentley *et al.*, 2002; Gregory, 2008). More recently the complete genome of *Streptomyces albus* J1074 has been published (Olano *et al.* 2014) which demonstrated that this strain possesses twenty seven BGCs in its genome. More importantly, it has been noticed that there is no or very little overlap in the BGCs which occurred in the genomes of these species, suggesting that further bacterial genome sequencing could lead to the discovery of large numbers of novel bioactive natural products (Medema *et al.*, 2014).

5.2 antiSMASH and ClusterFinder

Nowadays, several bioinformatics tools have been designed to search for BGCs with the ability to predict a wide range of SMs classes. In 2011, antiSMASH (antibiotics & Secondary Metabolite Analysis SHell), a comprehensive pipeline for identification and analysing of gene clusters encoding biosynthetic enzymes for a wide range of SM compound classes was presented (Medema *et al.*, 2011). This tool identifies polyketides, nonribosomal peptides, terpenes, aminoglycosides, aminocoumarins, indolocarbazoles, lantibiotics, bacteriocins, nucleosides, beta-lactams, butyrolactones, and others in both bacterial and fungal nucleotide sequences. In 2013, antiSMASH 2.0 was released (Blin *et al.*, 2013). This new version was re-designed and highly improved to include more features making antiSMASH 2.0 the most comprehensive resource for identifying and analysing novel secondary metabolite biosynthetic pathways in microorganisms. These features are:

- Supports input of multiple related sequences simultaneously (multi-FASTA/GenBank/EMBL), which allows the analysis of draft genomes comprising multiple contigs.
- Ability to perform direct analysis of protein sequences.
- The capacity to detect additional classes of secondary metabolites, including oligosaccharide antibiotics, phenazines, thiopeptides, homo-serine lactones, phosphonates and furans.
- Predicting the core structure of the cluster end product of lantipeptides, in addition to polyketides and non-ribosomal peptides.
- ClusterBlast functionality was extended to identify sub-clusters involved in the biosynthesis of specific chemical building blocks.

ClusterFinder, a new bioinformatics tool, which uses a hidden Markov model to detect and extract putative biosynthetic gene clusters of unknown types in sequenced genomes was fully integrated into antiSMASH (Weber *et al.*, 2015). Clusters identified using this algorithm are further categorised into saccharide (CF-saccharide), fatty acid (CF- fatty acid) and putative (CF-putative). It searches for

enzymes involved in synthesis of secondary metabolites by using the PFAM domain definition (Cimermancic *et al.*, 2014).

5.3 Biosynthetic gene clusters (BGCs) detection

To examine the potential of isolate_99, as a secondary metabolite producer that could be stimulated to produce an unknown antimicrobial compound, in the presence of lead, which inhibited the growth of *E. coli* (Chapter 3), we searched for BGCs for secondary metabolites within the obtained draft genome sequence of this isolate. From the results (Table 6.1), a total of 26 putative BGCs for biosynthetic secondary metabolites were predicted using antiSMASH to be involved in the biosynthesis of polyketides (PKs), nonribosomal peptides (NRPs), lantipeptide, siderophore, terpene, bacteriocin, melanin, ectoine and other. In addition, 33 putative BGCs were also found using ClusterFinder bioinformatics tool (Table 6.2), which were further identified into putative, fatty acid and saccharide clusters and one additional cluster (cluster 17, Table 6.2), involving in Polyketide synthesis which were not identified by antiSMASH. Nine clusters containing modular enzyme-coding genes (polyketides (PKs) and /or nonribosomal peptides (NRPs) were identified in isolate_99. Five of them contain PKs genes belonging to type I (cluster 10, Table 6.1 and cluster 17, Table 6.2) and type II PKs (cluster 5, 6 and 11, Table 6.1).

Table 6.1: Biosynthetic gene clusters of Isolate_99 obtained from AntiSMASH 3.0 tool.

Cluster	Type	Most similar known cluster (%)	Cluster	Type	Most similar known cluster (%)
Cluster 1	Lantipeptide	Unknown	Cluster 14	Nrps	Sessilin (33%)
Cluster 2	Siderophore	Desferrioxamine_B (83%)	Cluster 15	Siderophore	Tetrocarcin_A (4%)
Cluster 3	Terpene	Albaflavenone (100%)	Cluster 16	Nrps	A54145 (10%)
Cluster 4	Terpene	Hopene (30%)	Cluster 17	Siderophore	Unknown
Cluster 5	Otherks	SCO-2138 (42%)	Cluster 18	Nrps	Teicoplanin (6%)
Cluster 6	Otherks	Acarviostatin (22%)	Cluster 19	Othe	Unknown
Cluster 7	Terpene-Butyrolactone	Gamma-butyrolactone (66%)	Cluster 20	Bacteriocin	Unknown
Cluster 8	Bacteriocin	Unknown	Cluster 21	Melanin	Melanin (60%)
Cluster 9	Bacteriocin	Unknown	Cluster 22	Ectoine	Ectoine (100%)
Cluster 10	Otherks-T1pks	Clavulanic acid (2%)	Cluster 23	Terpene	Hopene (15%)
Cluster 11	Otherks-T2pks	Cosmomycin_D (22%)	Cluster 24	Other	Unknown
Cluster 12	Butyrolactone-Otherks	Granaticin (8%)	Cluster 25	Nrps	Polyoxypeptin (13%)
Cluster 13	Terpene	Unknown	Cluster 26	Nrps	Piericidin_A1 (50%)

Clusters 10 and 17 showed similarities to clavulanic acid (2%) and herboxidiene (2%) known clusters respectively. The other PKs (type II), clusters 5, 6, and 11 showed similarities to SCO-2138 (42%), acarviostatin (22%) and cosmomycin_D (22%) respectively. Clusters 14, 16, 25 and 26, that contain NRPs genes, showed similarities to sessilin (33%), A54145 (10%), polyoxypeptin (13%) and piericidin A1 (50%) respectively. Polyketide and nonribosomal peptide compounds are well-known to be pharmaceutically important and their pathways have attracted much attention (Donadio *et al.*, 2007; Meier & Burkart 2009), and below is a brief description to the natural products that might be connected to PKs and NRPs genes identified in our sequenced strain. Clavulanic acid is a β-lactam antibiotic used to combat resistance in bacteria which produce the β-lactamase enzyme when combined with amoxicillin, cloxacillin or cefotaxime. It has been used to treat infections caused by β-lactamase-producing strains of *E. coli*, *Haemophilus influenzae, Klebsiella* spp. and *Staph. aureus* (where amoxicillin alone is not appropriate); acute otitis media and acute sinusitis; urinary tract infections in children and osteomyelitis due to *Haemophilus influenzae* or unknown pathogen in children under 5 years, together with cloxacillin and either ceftriaxone or cefotaxime (WHO, 2000). Herboxidiene is a pesticide which was isolated from *Streptomyces chromofuscus* A7847 as a novel polyketide which controls several annual weed species (Miller-Wideman *et al.*, 1992). Later, Herboxidiene was found to have promising activity in reducing cholesterol levels and has been used in clinics to treat heart disease (Shao *et al.*, 2012). SCO-2138 peptide was first found in the 6-frame translations of the *S. coelicolor* and *S. griseus* genomes (Kersten *et al.*, 2012), which represents undiscovered classes of RNPs at the time of writing this chapter. Geng *et al* (2008) discovered acarviostatin, a compound which is secreted by *S. coelicoflavus* ZG0656.

Table 6.2: Biosynthetic gene clusters of Isolate_99 obtained from ClusterFinder tool.

Cluster	Type	Most similar known cluster (%)	Cluster	Type	Most similar known cluster (%)
Cluster 1	Cf_putative	Acarbose (7%)	Cluster 18	Cf_putative	Unknown
Cluster 2	Cf_putative	Salinomycin (6%)	Cluster 19	Cf_putative	Unknown
Cluster 3	Cf_putative	Chartreusin (23%)	Cluster 20	Cf_putative	Unknown
Cluster 4	Cf_putative	Enduracidin (27%)	Cluster 21	Cf_putative	Unknown
Cluster 5	Cf_putative	Sch47554 (10%)	Cluster 22	Cf_putative	Unknown
Cluster 6	Cf_saccharide	Kiamycin (10%)	Cluster 23	Cf_putative	Unknown
Cluster 7	Cf_fatty_acid	Colabomycin (11%)	Cluster 24	Cf_putative	Unknown
Cluster 8	Cf_putative	Tirandamycin (13%)	Cluster 25	Cf_putative	Unknown
Cluster 9	Cf_putative	Kanamycin (15%)	Cluster 26	Cf_putative	Unknown
Cluster 10	Cf_putative	Pyrrolomycin (5%)	Cluster 27	Cf_putative	Unknown
Cluster 11	Cf_putative	Daptomycin (6%)	Cluster 28	Cf_putative	Unknown
Cluster 12	Cf_putative	Desotamide (9%)	Cluster 29	Cf_fatty_acid	Unknown
Cluster 13	Cf_putative	Thiolutin (12%)	Cluster 30	Cf_fatty_acid	Unknown
Cluster 14	Cf_putative	Hopene (38%)	Cluster 31	Cf_fatty_acid	Unknown
Cluster 15	Cf_putative	ECO-02301 (35%)	Cluster 32	Cf_fatty_acid	Unknown
Cluster 16	Cf_putative	Pactamycin (11%)	Cluster 33	Cf_saccharide	Unknown
Cluster 17	Otherks-T1pks-	Herboxidiene (2%)	Cluster 34	Cf_saccharide	Unknown

	Cf_fatty_acid				

This compound is among α-Amylase and α-glucodase inhibitors, which are well-known treatments and prophylactics for diabetes, obesity, or other secondary symptoms caused by these diseases (Qin *et al.*, 2011). The anthracycline cosmomycin D antibiotic is the main product isolated from *S. cosmosus* TMF5l8 (Ando *et al.*, 1985), which has antitumor activity, showing a crucial role in the successful treatment of many types of cancer, despite some side effects related to cardiotoxicity (Carvalho *et al.*, 2010). The NRPs sessilin has been known for its powerful activity against a wide range of organisms, such as fungi, bacteria, protozoa and plants (D'aes *et al.*, 2014). This metabolite was widely produced by *Pseudomonas sp. CMR12a*, displaying strong biocontrol activity against bacterial and fungal plant pathogens (Hua and Höfte, 2014). A54145 is a complex of calcium-dependent cyclic lipopeptide antibiotics used for the treatment of infections caused by Gram-positive pathogens (Miao *et al.*, 2005), which is produced by *S. fradiae* (Boeck *et al.*, 1990). Polyoxypeptin, which was isolated from a culture broth of *Streptomyces sp*. MK498-98 F14 (Umezawa *et al.*, 1998), has attracted a great deal of attention because of its ability to induce apoptosis in human pancreatic carcinoma AsPC-1 cells (Noguchi *et al.*, 2000). Piericidin A1 was first isolated from *S. mobaraensis* in the late 1950s, later from *S. piomogeues* and also from a symbiotic Streptomycete of beewolf digger wasps (Kroiss *et al.*, 2010). Piericidin A1 has been reported to have diverse biological activities, which include potent inhibitory activity toward mitochondrial NADH dehydrogenase by competing with ubiquinone reductase. It has also been reported to show antimicrobial and antifungal activity (Kitagawa *et al.*, 2010; Liu *et al.*, 2012). Moreover, a study by Hwang *et al*(2008) has shown that piericidin A1 acts as a highly selective antitumor agent in animal models.

Subsequently, we compared the abundance of PKs and NRPs gene clusters in our isolate with those of two representative genome-elucidated *Streptomyces* strains: *S. coelicolor* A3(2), which is well studied as a model actinomycetes *and S. avermitilis* MA-4680T, an industrial avermectin producer. It has been seen that the number of PKs and NRPs gene clusters in our isolate were comparable to *S. avermitilis*, which contains 5 PKs and 4 NRPs clusters. Whereas, *S. coelicolor* contains 7 clusters of PKs and 2 clusters of NPRs. However, besides PKs and NRPs, isolate_99 genome displayed

other putative secondary metabolite related genes predicted by antiSMASH (Table 6.1) (Medema *et al.*, 2011). Such genes and clusters have also been found in the representative *Streptomyces* strains, *S. coelicolor* and *S. avermitilis*. Of these putative clusters, 3 bacteriocin clusters that did not show any homology to known BGCs domains were identified. In contrast, 2 bacteriocins clusters for *S. coelicolor* and *S. avermitilis* were identified, one cluster with no known homology and the other showed homology to the informatipeptin encoding gene for each strain. Bacteriocins, are among ribosomal peptide antibiotics, which show activity to suppress the growth of similar or closely related bacterial species and antibiotic-resistant strains (Cotter *et al.*, 2013). One ribosomal lantipeptide biosynthetic gene cluster with unknown homology was successfully identified through the implementation of a lanthipeptide-specific analysis module in antiSMASH version 3 (Blin *et al.*, 2013). Version 3 of antiSMASH can predict the core peptide molecular mass and sequence after leader peptide cleavage motifs which are identified via pHMMs, when one or more open reading frames encoding putative lantipeptide prepropeptides are found (Blin *et al.*, 2014). Ectoine (1,4,5,6-tetrahydro-2-methyl-4-pyrimidinecarboxylic acid), is a cluster always identified in the members of the *Streptomyces* genus (Bursy *et al.*, 2008). It is a compatible solute which acts as protective substances, for example, by acting as an osmolyte in case of extreme osmotic stress, where it enables bacterial tolerance of environmental stress such as extreme salt and temperature levels (Pastor *et al.* 2010). In this case, one ectoine cluster was identified in our isolate, as well as one cluster for each *of S. coelicolor* and *S. avermitilis* with a 100% homology.

Furthermore, 3 siderophore clusters, one melanin cluster, which is considered an important criterion for morphological and taxonomical studies, and 3 terpene clusters were identified in isolate_99. These types of gene clusters are found to be conserved within the core genome of *Streptomyces* strains that are annotated to be involved in secondary metabolism (Kim *et al.*, 2015). Terpenes can provide stability to bacterial membranes at high temperatures and under conditions of extreme acidity. They are large and diverse class of organic compounds that are considered to be secreted by plants and fungi (Pastor *et al.* 2010), however, numerous bacteria,

especially Gram positive microorganisms such as *Streptomyces* and other actinomycetes were also found to produce terpene compounds (Cane and Ikeda, 2012).

antiSMASH supports detection of 24 secondary metabolite classes by comparing each biosynthetic gene on the uploaded DNA sequence against a manually curated collection of profile hidden Markov models (pHHMs) (Blin *et al.*, 2013). Two gene clusters related to biosynthetic metabolites were also detected in our isolate under type of ''other'' BGCs with no homology (Table 6.1). However, a key limitation to antiSMASH BGCs detection algorithm was that, despite many major classes of secondary metabolites being covered by its detection logic, it was still limited to the detection of known types of biosynthetic gene clusters (Weber *et al.*, 2015). To overcome this limitation, the ClusterFinder algorithm, which uses a hidden Markov model to probabilistically predict BGC-like regions in genomes based on the frequencies of observed PFAM domains inside and outside a comprehensive set of known BGCs was integrated within antiSMASH (Cimermancic *et al.*, 2015). In this case, further clusters were identified in our strain. From the results, 16 putative BGSc with diverse homology and 17 putative with unknown BGCs similarities were identified using ClusterFinder (Table 6.2). However, the presence of these many putative clusters with unknown homologies might reveal novel insights into isolate_99, and to identify putative triggers that might stimulate the production of new secondary metabolites.

5.4 Genetic dereplication and comparison with known pathways

The Known ClusterBlast module within antiSMASH 3.0, currently contains 1172 known biosynthetic gene clusters (Weber *et al.*, 2015). Using this enabled us to compare the obtained identified BGCs with those encoding the biosynthetic pathways for known end products (Figure 6.1). The rapid identification of known compounds is an important feature to avoid the duplication of already isolated compounds and ensures focus is on discovery of novel natural products. Further, comparative analysis, inferred from homology, of known and unknown BGCs may provide hints concerning the function of certain genes within the cluster.

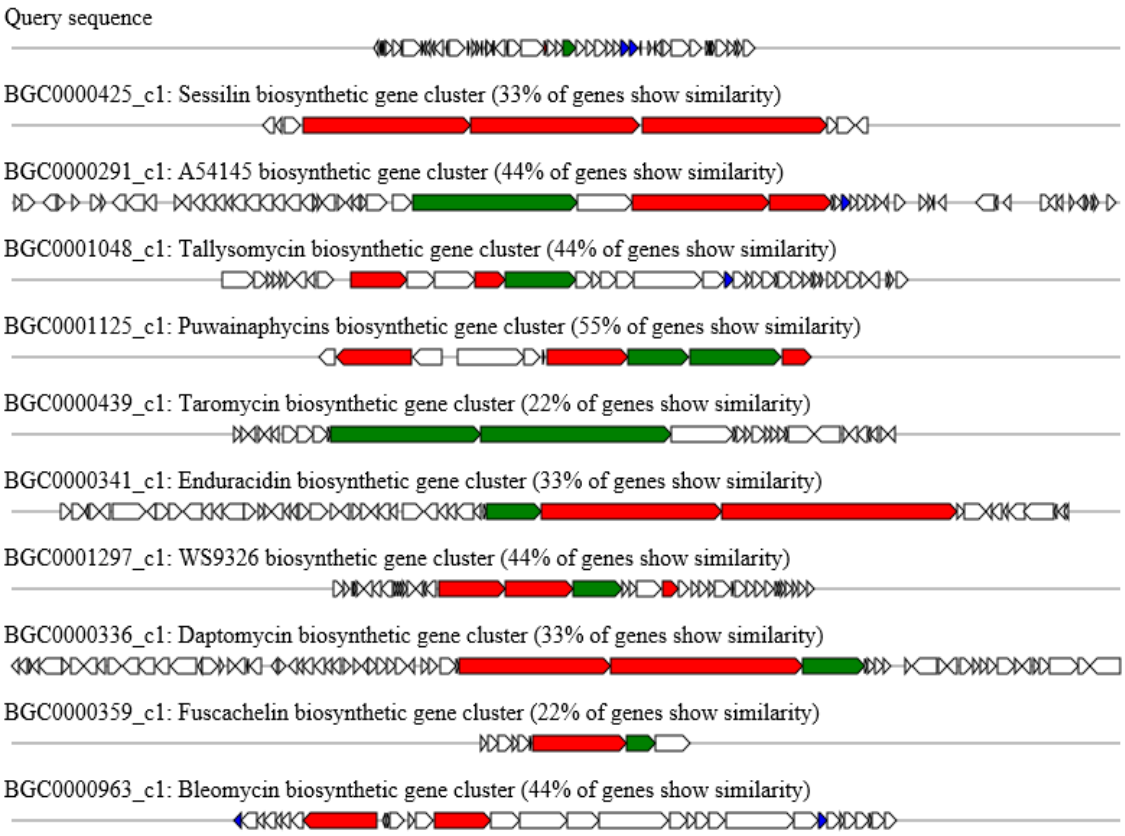


Figure 6.1: Example output of a Known ClusterBlast output (cluster 14, Table 6.1), using the sessilin biosynthetic gene cluster (GenBank JQ309920). The significance thresholds used are the same as for the ClusterBlast module (Medema *et al.*, 2011). Following the sessilin gene cluster itself, several other BGCs involved in the biosynthesis of similar NRPs are shown as next best hits. The percentage of genes in the query cluster that are present in the hit cluster is included as extra information.

5.5 Comparative analysis of gene clusters

Although most of the 26 biosynthetic gene clusters that were identified in isolate_99 were close or had a range of similarities to other gene clusters within the same genus of *Streptomyces* such as cluster 5 (Table 6.1, Figure 6.2,) which showed homology to *S. davawensis* strain JCM 4913, *S. hygroscopicus* subsp. *Jinggangensis, S. sp.* e14, *S. collinus* Tu 365, *S. pluripotens* strain MUSC 135, *S. sp*. MUSC 125 and *S*. *avermitilis* MA-4680. One gene cluster, cluster 25 (Table 6.1, Figure 6.3), did not show similarity to any strain of *Streptomyces*, but it showed similarities to distantly related actinomycete genera of *Nocardiopsis, Micromonospora, Gordonia and Rhodococcus*. Further, it displayed similarities to non-related genera of *Clostridium* and *Bacillus*. It has been reported that secondary metabolite gene clusters and their genes can be transferred horizontally among bacteria during evolution (Donadio *et al.*, 2005). A sequence and phylogenetic analysis study of ketosynthase gene from the biosynthetic pathway of aromatic polyketide of 99 *Streptomyces* isolates conducted by Metsa-Ketela *et al.* (2002) has shown that ketosynthase gene has been horizontally transferred among these *Streptomyces* species. Another interesting study carried out by Piel *et al.* (2004) has found that pederin, a group of antitumor compounds, which are found in terrestrial beetles and marine sponges and used by apparently all members of the rove beetle genera *Paederus* and *Paederidus* as a chemical defense against predators, are produced by bacterial symbionts, which are very close relatives of *Pseudomonas aeruginosa*. The authors show evidence that a *P. aeruginosa-like bacterium* has acquired the putative pederin biosynthesis (ped) gene cluster that produces pederin using a genomic island obtained through horizontal transfer. Genomic islands are large regions of DNA that can be horizontally transferred which can enhance metabolic and colonising capabilities, and allow bacterial evolution in quantum leaps (Hacker and Kaper, 2000). Fischbach *et al.* (2007) have reported another example of the role of genomic elements such as pathogenicity islands involved in horizontal transmission of gene clusters. They show that the gene cluster responsible for producing the iron scavenging agent yersiniabactin has been found not just in *Yersinia pestis*, but also found in other

bacteria including the nematode symbiont *Photorhabdus luminescens,* the plant pathogen *Pseudomonas syringae*, pathogenic strains of *Escherichia coli,* and even the Gram positive marine bacterium *Salinispora tropica.* However, biosynthetic gene clusters can also horizontally transferred by plasmids as well as pathogenicity islands during evolution (Fischbach *et al.*, 2007; Piel *et al.*, 2004). Moreover, a study by Lopez (2003) showed evidence that transposon genes can be involved in horizontal acquisition of the putative polyketide epothilone (*epo*) biosynthetic gene in a range of Gram-positive actinomycetes and cyanobacteria.

Interestingly, our 99 isolate was found to have transposon coding sequence in its genome (Figure 4.4, chapter 4). The same transposon was also found in both *Bacillus* and *Clostridium* genera under accession numbers NC_021856 and NC_029048 respectively. Thus, and from the above evidences, the presence of similar transposon sequences in isolate_99 and in distantly non related genera might explain why the biosynthetic gene cluster 25 shows similarities to non-related genera of *Clostridium* and *Bacillus*, as transposon could provide a transfer mechanism by involving as ‘‘vectors’’ (e.g. conjugative plasmids or transducing bacteriophages) (Piel *et al.*, 2004) during the evolution. These results support the idea that the biosynthetic gene cluster(s), may be prone to transposition between distantly related or non-related bacterial genera during their genomes evolutions. However, cluster 25 might be unique for this isolate, at least with the *Streptomyces* genus, and could be a source for new natural products that have not been characterised yet.

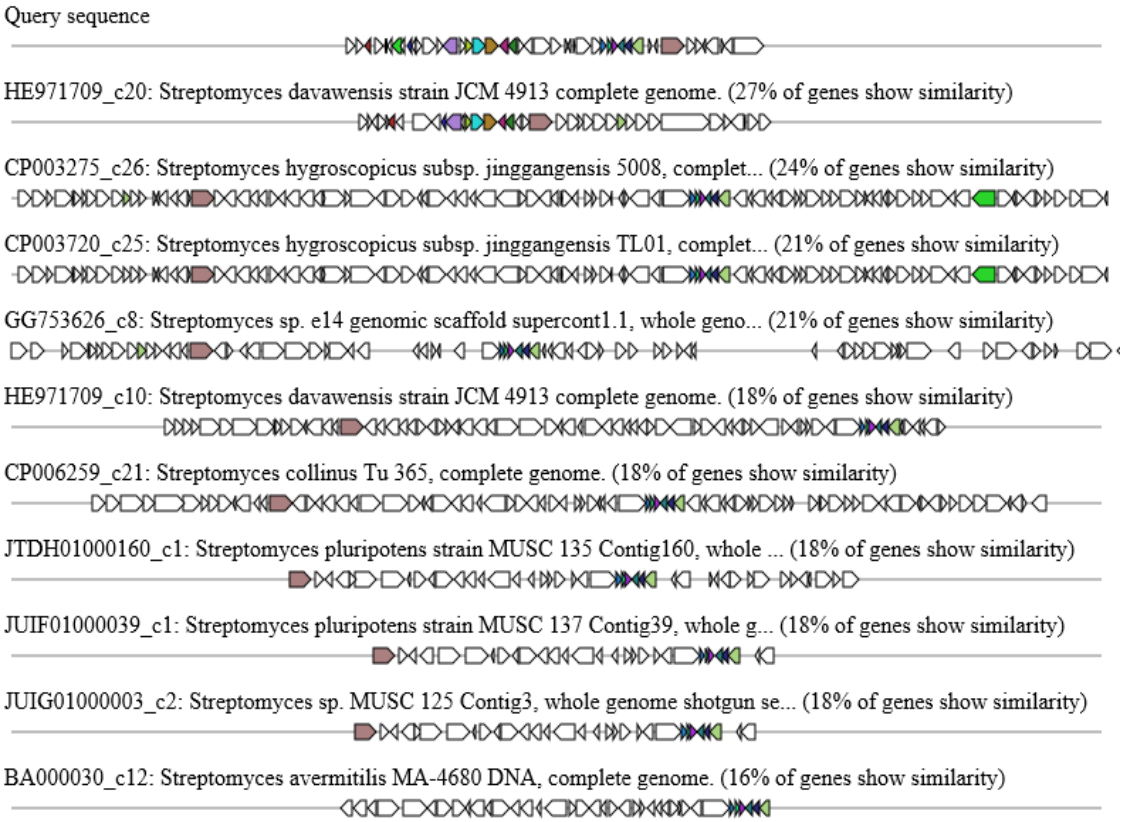


Figure 6.2: Output of ClusterBlast alignment of biosynthetic gene clusters homologous to the query gene cluster 5. All the ten best hits to the cluster 5 from isolate_99 are from *Streptomyces* genus. Homologous genes are given the same colours.

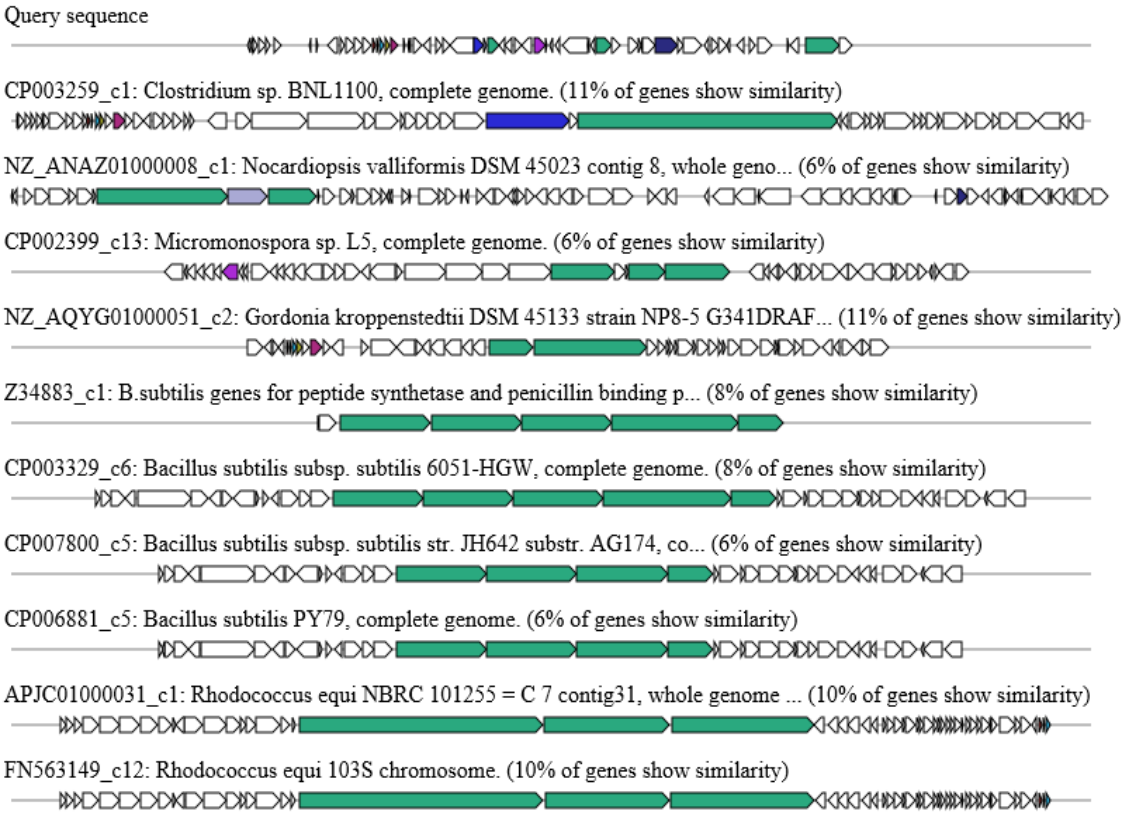


Figure 6.3: Output of ClusterBlast alignment of biosynthetic gene clusters homologous to the query gene cluster 25. All the ten best hits to the cluster 5 from isolate_99 are from distantly related genera (*Nocardiopsis, Micromonospora, Gordonia and Rhodococcus*) and no-related genera (*Clostridium and Bacillus*). Homologous genes are given the same colours.

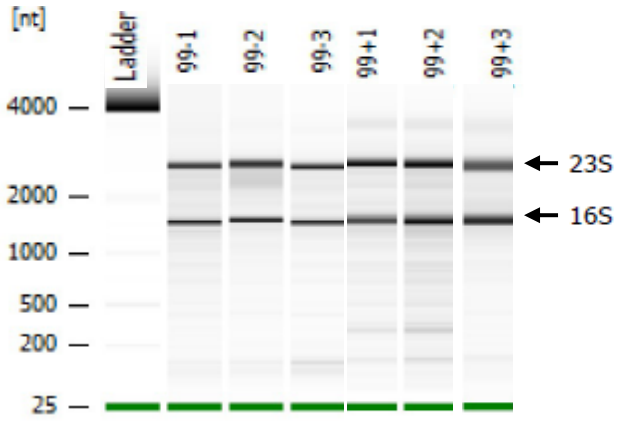
Chapter 7:

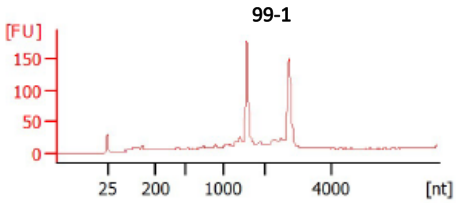
RNA Sequencing and Transcriptomic Studies of Biosynthesis Gene Clusters of Isolate_99 under Lead (Pb) Metal Stress

7.1 RNA-Sequencing

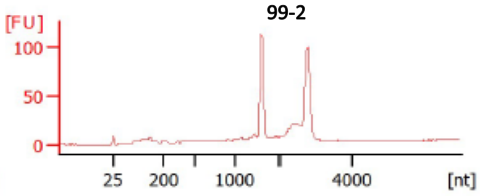
In order to investigate the transcriptomic response of biosynthetic gene clusters of isolate_99 under lead metal stress, the total RNA was isolated at the stationary phase of production of bioactive secondary metabolites (Chapter 3, section 3.5) from cultures grown in minimal medium (MM) supplemented either with 1.5 mM Pb or without Pb. The stationary phase sample was taken at day 7 (the maximum inhibition zone against *E. coli*) from both MM supplemented with or without Pb.

The initial concentration of RNA extraction from the biomass sample was greater than 10,158 pg/µl for the Pb free control replicates (99-1, 99-2 and 99-3) and 165 pg/µl for the Pb sample replicates (99+1, 99+2 and 99+3) (Figure 7.1). The RNA integrity of the samples was checked using the Bioanalyzer RNA 600 Pico assay. Both samples with Pb and without Pb had intact 16S rRNA and 23S rRNA bands with RNA integrity number (RIN) greater than 7.9 for samples without Pb and 6.8 for samples with Pb (Figure 7.1). RIN value ranges on a scale from 10 for intact to 1 for completely degraded (Imbeaud *et al.*, 2005). It can be seen that RIN scores for Pb free samples were relatively high (7.9, 8.2 and 8.6) compared to samples with Pb supplement (7.2, 6.8 and 6.9). Although, RIN score >7 is recommended for RNA sequencing from the supplier (Life Technologies catalog number 4475936), it was decided to accept all samples as it was very hard to have RIN values >7 for samples supplemented with Pb. The presence of heavy metal (Pb) might have a negative effect on the integrity of RNA samples. However, researchers have shown that RNA Integrity Number can be affected by several endogenous and exogenous factors as well as other factors including “AU-rich” sequence, transcript length, GC content, 16S/23S rRNA ratio, secondary structure and RNA protein complex (Wang *et al.,* 2016).

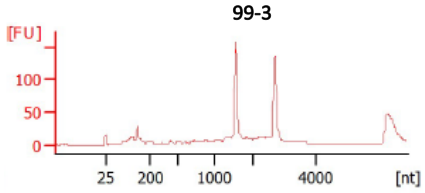




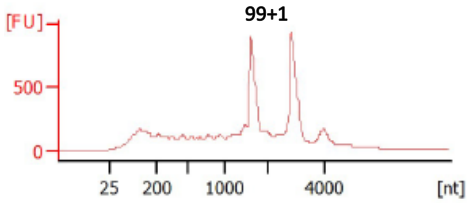
Overall results for sample 99-1		
rRNA Ratio 23s/16s	RNA Integrity Number (RIN)	RNA Concentration [pg/µl]
1.0	7.9	18,786



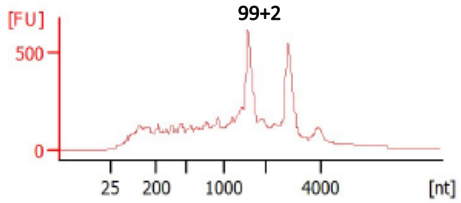
Overall results for sample 99-2		
rRNA Ratio 23s/16s	RNA Integrity Number (RIN)	RNA Concentration [pg/µl]
1.1	8.6	15,849



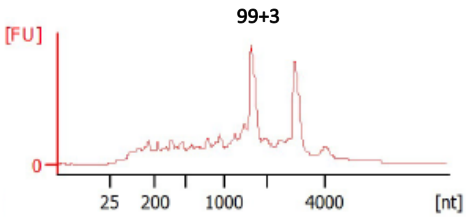
Overall results for sample 99-3		
rRNA Ratio 23s/16s	RNA Integrity Number (RIN)	RNA Concentration [pg/µl]
1.0	8.2	10,158



Overall results for sample 99+1		
rRNA Ratio 23s/16s	RNA Integrity Number (RIN)	RNA Concentration [pg/µl]
1.3	7.2	279



Overall results for sample 99+2		
rRNA Ratio 23s/16s	RNA Integrity Number (RIN)	RNA Concentration [pg/µl]
1.2	6.8	183



Overall results for sample 99+3		
rRNA Ratio 23s/16s	RNA Integrity Number (RIN)	RNA Concentration [pg/µl]
0.9	6.9	165

Figure 7.1: Agilent Bioanalyzer results to show 23S/16S rRNA ratio and the integrity of the RNA samples used for RNA-Seq experiments.

Ribosomal RNA of 16S and 23S were depleted from the samples using the Ribo-Zero Magnetic Gram-positive bacteria assay following the manufacturer's instructions. It can be seen that the two distinct bands for 16S rRNA and 23S rRNA were removed successfully from the samples after rRNA depletion step as expected (Figure 7.2).

The RNA depleted samples were fragmented and ligated to barcoded oligonucleotides and then converted into cDNA using SuperScript III Enzyme Mix (Life Technologies, part of Ion Total RNA-Seq Kit v2). The final cDNA libraries were purified using the Magnetic Bead Clean up Module and analysed by using the Bioanalizer High Sensitivity DNA assay. The size distribution of the amplified cDNA fragment libraries was found to be in the range of 50 - 160 bp fragments and the molar concentration (nM) was determined in the range of 50 -1000 pb. To determine the library dilution required for template preparation, each sample was diluted to 500 pM. The three biological replicates were further pooled at a concentration of 20 pM for each. The two samples, each from a pool of three biological replicates were emulsified, amplified using an Ion PGM OT2 200 pb assay and enriched to obtain the template for the sequencing. The percentage of template ion spheres was 24.3% for Pb supplemented samples and 21% for Pb free samples, which was in the range of 10 – 30% as recommended by the manufacturer.

Each three replicate samples were loaded on to a 316 v2 chip and sequenced using the Ion Torrent PGM (Life Technologies). The sequencing resulted in 1.97 million total reads for Pb free samples with overall 31X coverage and 1.26 million total reads for Pb supplemented samples with overall 25X coverage.

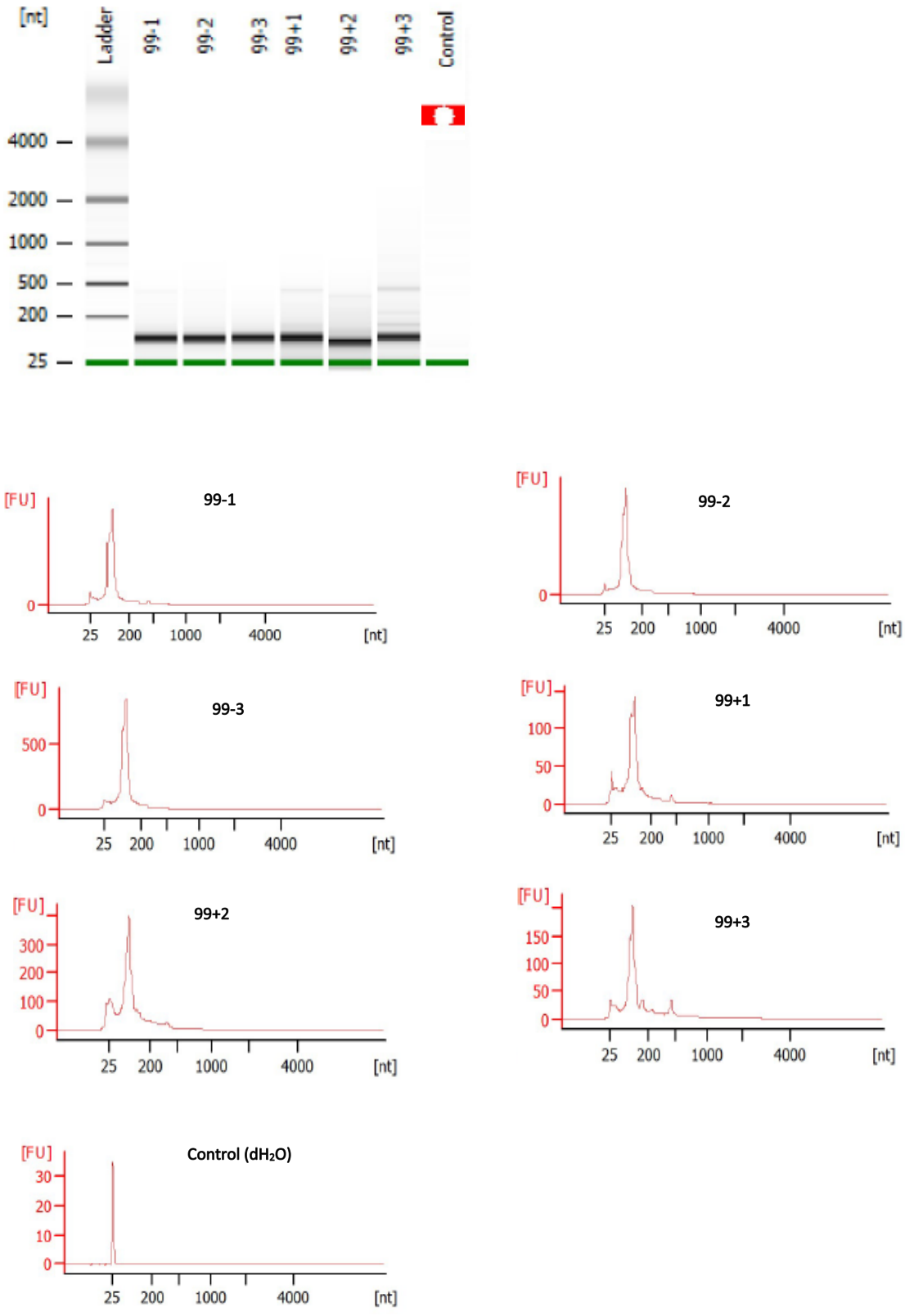
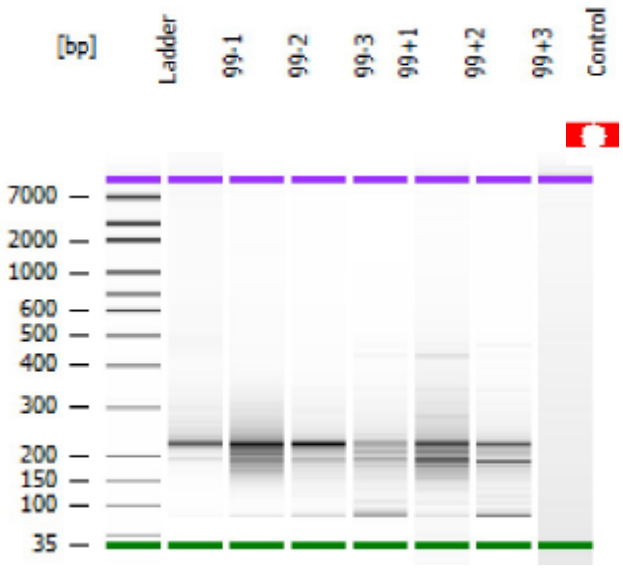
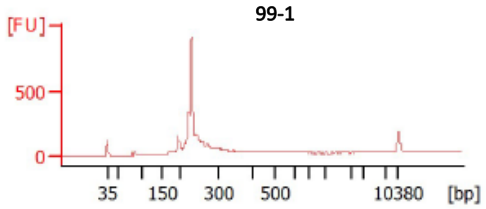
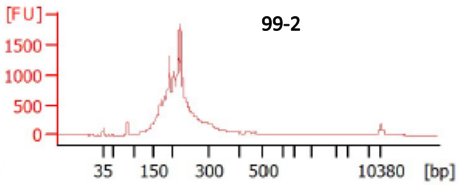


Figure 7.2: Agilent Bioanalyzer results using RNA 6000 Pico assay after depletion of 16S rRNA and 23S rRNA from the RNA samples used for RNA-Seq experiments.

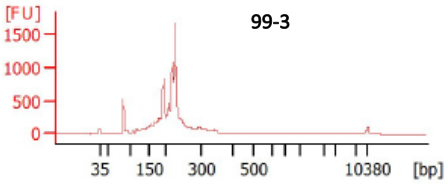




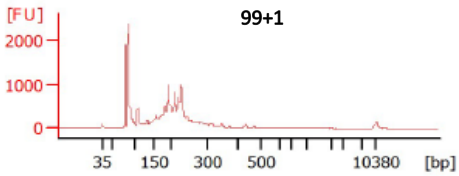
Region table for sample 99-1 (50-1000 pb)		
Average library size (pb)	Size distribution in CV [%]	Molarity [pmol/l]
284	38.7	14,666



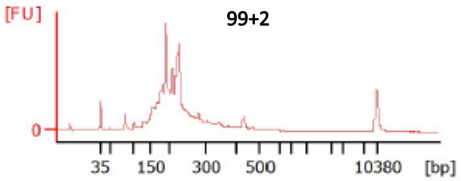
Region table for sample 99-2 (50-1000 pb)		
Average library size (pb)	Size distribution in CV [%]	Molarity [pmol/l]
268	37.4	45,814



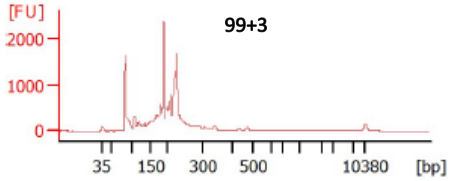
Region table for sample 99-3 (50-1000 pb)		
Average library size (pb)	Size distribution in CV [%]	Molarity [pmol/l]
260	35.4	45,897



Region table for sample 99+1 (50-1000 pb)		
Average library size (pb)	Size distribution in CV [%]	Molarity [pmol/l]
268	33.8	4,954



Region table for sample 99+2 (50-1000 pb)		
Average library size (pb)	Size distribution in CV [%]	Molarity [pmol/l]
277	35.4	9,719



Region table for sample 99+3 (50-1000 pb)		
Average library size (pb)	Size distribution in CV [%]	Molarity [pmol/l]
268	38.7	5,702

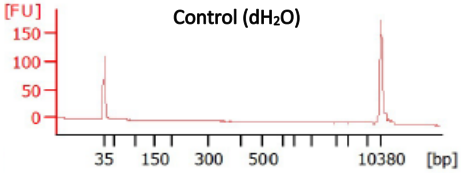


Figure 7.3: Agilent Bioanalyzer results using High Sensitivity DNA assay showing the average library size (Pb), the size distribution (CV %) and the molarity (pmol/l) the RNA samples used for RNA-Seq experiments.

7.2 Data analysis results with TopHat, Cufflinks and CummeRbund

The genome sequence of isolate_99 annotated in RAST (Aziz *et al.*, 2008) in Genbank format was used as reference. The raw RNA-Seq data were obtained from the RNA-Seq experiments, downloaded in BAM format from the Ion Torrent Server and were uploaded onto the Galaxy platform (Afgan *et al.*, 2016). The TopHat algorithm (Trapnell *et al.*, 2009) was used to align short contiguous reads. The alignment file from TopHat then was used to assemble and reconstruct the transcriptome using the Cufflinks program (Trapnell *et al.*, 2012). It assembles the overlapping of aligned reads into transcripts using a probabilistic approach, then merges multiple conditions and estimates the transcript abundances using Cuffmerge and Cuffcompare respectively. It can also use Cuffdiff to calculate the differential gene expression. Finally, the CummeRbund tool (Trapnell *et al.*, 2012) was utilised to render the Cuffdiff output into visual representations like, density (Figure 7.5), bar (Figure 7.6), scatter (Figure 7.7) and volcano (Figure 7.8) plots.

Beforehand, the samples were subjected to principal component analysis (PCA). A PCA is a statistical method, which analyses the most important influencing factors while retaining most of the variability on the dataset and plotting them accordingly (Ringner, 2008). Basically, when analysing complex data sets, PCA filters those variables out that most influence the data, the so called principal components (PC) and can be used to plot the data on PC1 against PC2. In the context of RNA-Seq analysis, PCA essentially determines whether samples can be grouped significantly (Ghosh and Chan, 2016). As expected the sample replicates of each condition were clustered (Figure 7.4).

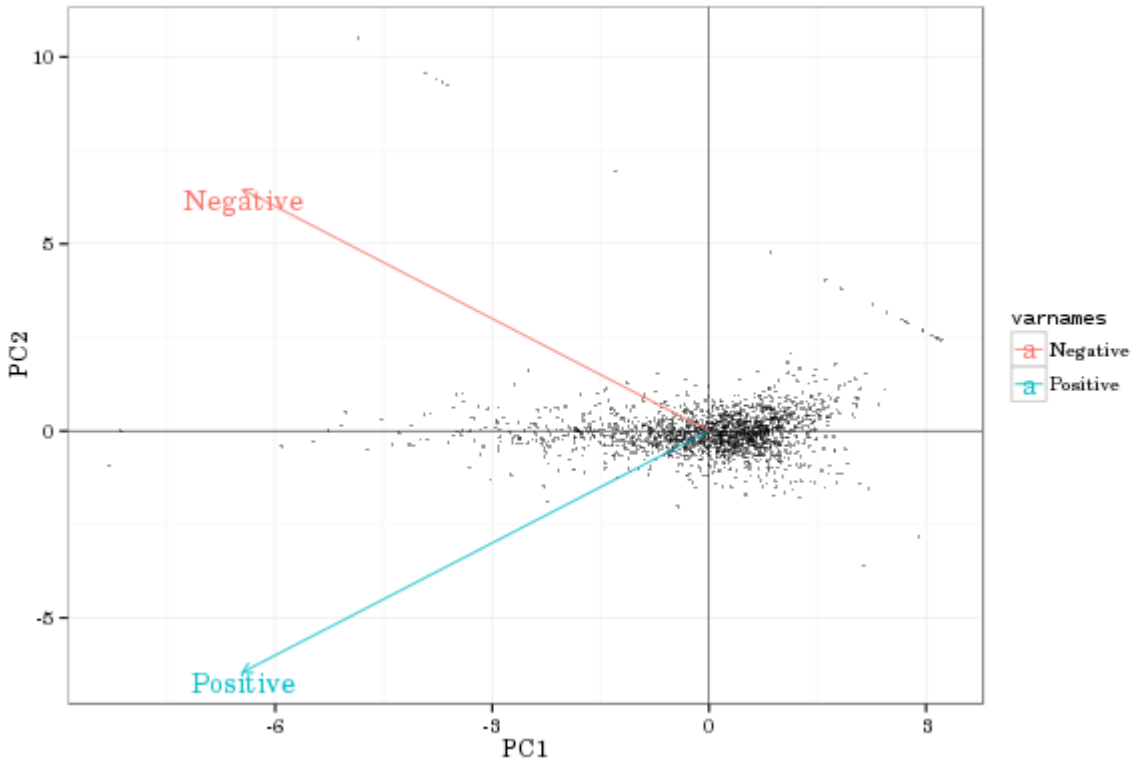


Figure 7.4: Principal component analysis of samples generated with CummeRbund tool (Trapnell *et al.*, 2012) showing the sample replicates, the blue line (Positive) is for Pb supplemented samples and the red line (Negative) for Pb free samples.

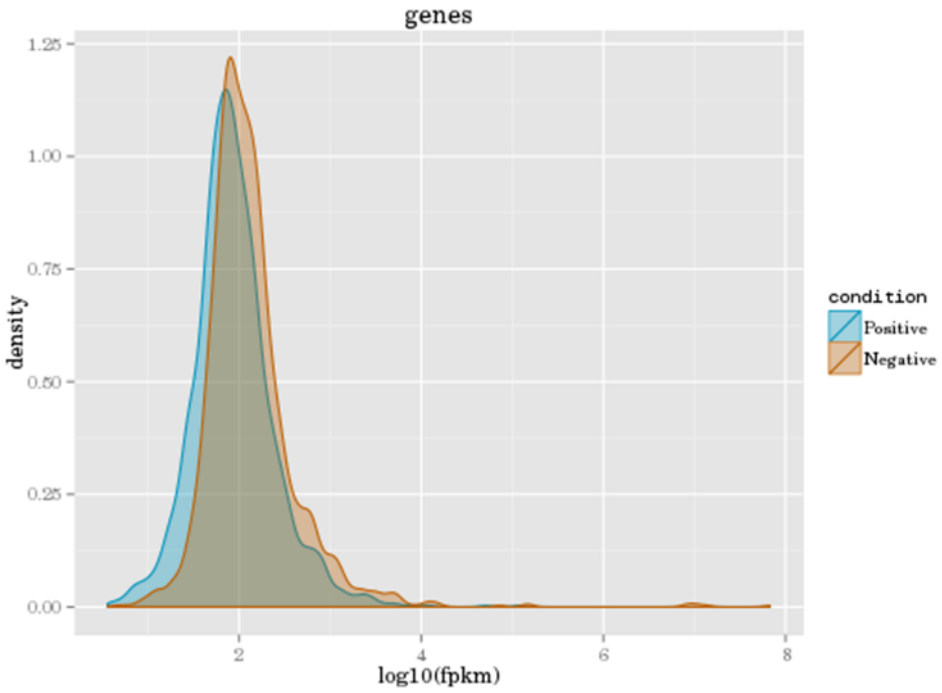


Figure 7.5: Density plots show the expression level distribution for all genes of isolate_99 in simulated experimental conditions Positive (samples with Pb supplement) and Negative (samples without Pb), measured in FPKM (fragments per kilobase of transcript per million fragments mapped). It can be seen that no outlier samples were identified and the samples displayed smoothed empirical densities for the individual samples in each condition.

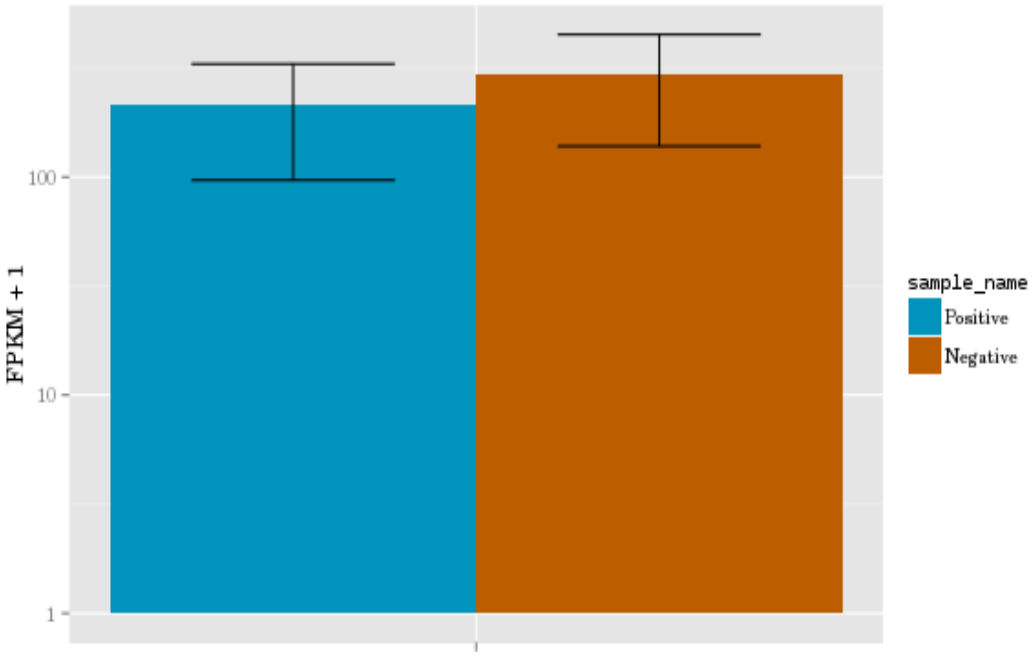


Figure 7.6: Expression plot shows differences in the expression of Isolate_99 across conditions Positive (samples with Pb supplement) and Negative (samples without Pb), measured in FPKM. It can be seen that more genes are expressed in samples without Pb than with Pb supplements.

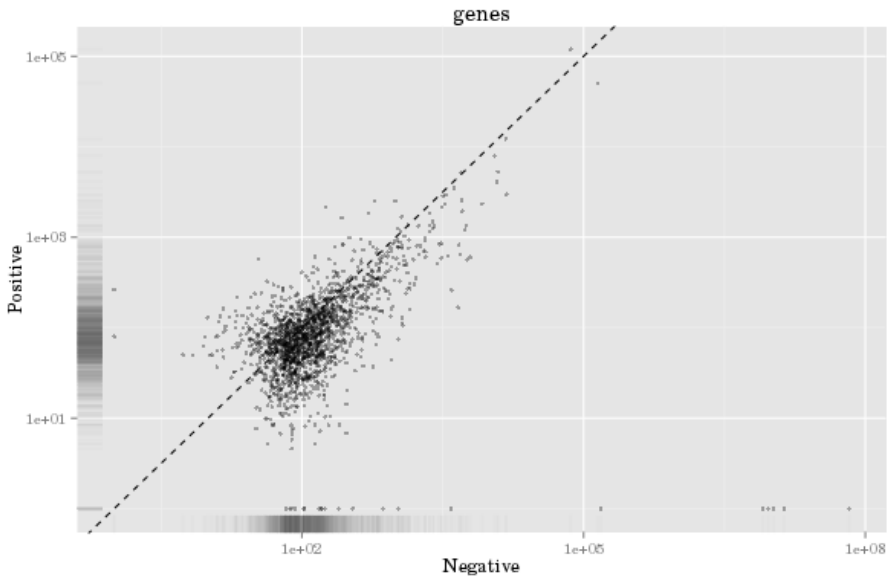


Figure 7.7: Scatter plots highlight general similarities and specific expression genes between conditions Positive (samples supplemented with Pb) and Negative (samples without Pb supplement). There is an outlier that has a lower expression on Positive samples than on Negative samples. This means that the gene is behaving differently between the treatments applied to the same bacterial isolate.

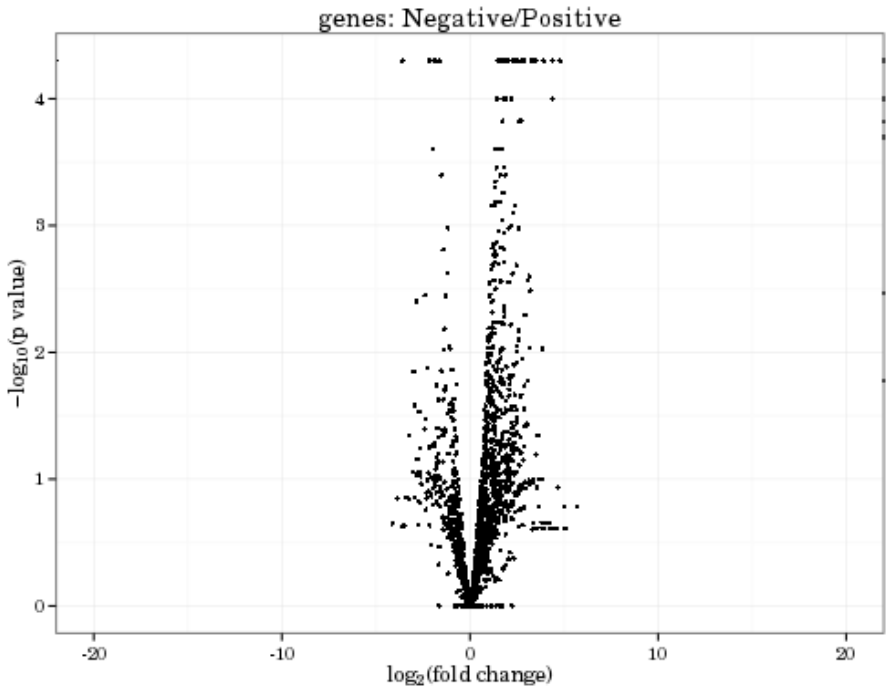


Figure 7.8: Volcano plots reveal genes whose expression differet significantly (*p*-value < 0.05) between conditions Positive (samples supplemented with Pb) and Negative (samples without Pb supplement). The plot shows that some genes (found towards the top of the plot that are far to either the left- or the right-hand side) displaying a huge change in expression when isolate_99 is supplemented with Pb.

7.3 Differential expression of genes involved in biosynthetic gene clusters of isolate_99

The draft genome sequence of isolate_99 from the whole genome sequencing using Ion Torrent PGM technology (Chapter 4) was used as a reference in a Genbank (.gbk) format. The bioinformatics tool CLC software version 8.0 (Qiagen Bioinformatics) was used to analyse the data for differential expression.

7.4 Differential expression of bacteriocin cluster of isolate_99

The transcriptional levels of biosynthesis gene clusters of isolate_99 were determined in the presence of lead. Out of 26 biosynthetic gene clusters predicted in isolate_99 by antiSMASH (Chapter 5, Table 5.1), only two clusters of bacteriocin type were up-regulated as a whole (Table 7.1).

Very little is known about bacteriocin activities from *Streptomyces* strains. The best studied case concerned the bacteriocin activities from *Streptomyces* genus was carried out by Zhang *et al.* (2003). The authors have shown that the antagonistic activity of ipomicin produced by *Streptomyces ipomoea* displays limited to closely related strains, i.e. primarily other strains of *S. ipomoea.* Also, a more recent study conducted by Farris *et al.* (2011) has shown that the bacteriocin produced by *Streptomyces scopuliridis* RB72 displayed broad spectrum antibacterial activities. The bacteriocin from this strain has demonstrated a potent activity against the Gram-positive bacteria including *Streptomyces avermitilis*, *S. coelicolor* A3 (2), *S. lividans* and *S. venezuelae*, and the Gram-negative bacteria *Escherichia coli* DH10B and *Klebsiella pneumonia* ATCC 13883 (Farris *et al.* 2011). Nevertheless, the above results are consistent with our results where the activity of isolate_99 was observed both on a related group, *Streptomyces* species in this case (Figure 3.4, Chapter 3), and unrelated group, *E. coli* in this case, (Figure 3.5, Chapter 3) of microorganisms. Moreover, the increased expression of the bacteriocin biosynthetic gene cluster in

this experiment suggests that a bacteriocin has been produced as the phenotype is bacteriocin-like.

Also, it can be seen that the transcripts of genes coding for ribosomally synthesized and post-translationally modified peptides (RiPPs), which is a major class of natural products started in the first decade of the 21st century as the result of genome sequencing efforts (Arnison *et al.*, 2013), and identified in bacteriocin clusters were significantly up-regulated ($p < 0.05$). The RiPPs genes increased by up to 13-fold (cluster 8, Table 7.2 & cluster 20, Table 7.3). RiPPs have attracted extensive interest for their diverse chemical structures and potent biological activities such as antibacterial, antifungal, allelopathic, and antiviral (Arnison *et al.*, 2013; Letzel *et al.*, 2014; Zheng *et al.*, 2015). However, the vast majority of genetically encoded natural products of RiPPs remain unknown (Skinnider *et al.*, 2016). In general, RiPPs follow a simple biosynthetic pathway in which precursor peptide consisting of an *N*-terminal leader peptide and a *C*-terminal core peptide, encoded by a single gene is translated, the leader peptide, is usually important for recognition by many of the post-translational modification enzymes and for export, is removed by a series of transporters, peptidases or a combination of both, and the remaining active peptide moiety is further processed by other enzymes, often encoded by genes within close proximity to the precursor gene (Arnison *et al.*, 2013; Letzel *et al.*, 2014).

The genetic basis of RiPPs is well understood and this can be used to guide comparison of well characterised biosynthetic genes or gene clusters against new genome sequences to identify the characterised or new putative RiPPs and in some cases, even predict the structure. In this case, the prediction informatics for secondary metabolomes (PRISM) platform (Skinnider *et al.*, 2016) can be utilised to discover and characterise of RiPPs.

This platform however, combining genome mining approach and LC–MS/MS data of crude extracts in a high-throughput manner to predict and identify the numerous families of RiPPs.

Based on the strongest BLAST protein sequences matches, the two RiPPs genes detected in isolate_99 had similarities of 95%, 94%, 91%, 81% and 75% with sequences of RiPPs found in *Streptomyces turgidiscabies* (WP_006374931), *S. sp*. WC-3618 (WP_053740852), *S. sp*. F-525 (WP_033281085), *S. sp*. CB02923 (WP_073762900) and *S. scabrisporus* (WP_078982410) respectively to the RiPP sequence of cluster 8, and similarities of 86%, 86%, 80%, 73% and 72% with sequences of RiPPs found in *S. sp*. 303MFCo15 (WP_020129143), *Actinobacter bacterium* OV320 (WP_054238789), *S. tsukubensis* (WP_077964138), *S. lushanensis* (WP_066951091) and *S. sp*. PpalLS-921 (SCD53824) respectively to the RiPP sequence of cluster 20. However, none of the RiPP types from the mentioned above strains has been studied or reported to have bioactivity. These results reveal that these two RiPP clusters are unique in isolate_99 genome and may be potential sources of bioactive RiPPs, as at least 1,833 of 2,231 (82%) of genetically encoded RiPPs remain unknown in sequenced genes (Skinnider *et al.*, 2016).

Another notable finding is that the bacteriocin clusters in isolate_99 have been found to contain genes of the radical *S*-adenosylmethionine (SAM) family (Table 7.2 & Table 7.3). Enzymes of this family are involved in biosynthetic pathways, where they function in various reactions, including methylation, radical formation, anaerobic oxidation, and protein ring formation (Sofia *et al.*, 2001; Kaminska *et al.*, 2010). The SAM genes were significantly ($p < 0.05$) upregulated by 4.8-fold and 3.4 in cluster 8 (Table 7.2) and cluster 20 (Table 7.3) respectively. However, it has been reported that radical SAM encoding genes are rare in bacteriocin-associated clusters, with the exception of subtilosin A, propionicin F and thuricin CD clusters (Zheng *et al.*, 2000; Brede *et al.*, 2004; Rea *et al.*, 2010). Subtilosin A and thuricin CD are antimicrobials belong to sactibiotics (sulphur to alpha-carbon antibiotic) class. Sactibiotics are peptides in which a sulphur bridge is posttranslationally formed between a cysteine residue and the α-carbon of another residue. This sulphur linkage is thought to be due to the associated radical SAM enzyme whose gene is co-localized in all sactipetide gene clusters (Murphy *et al.*, 2011; Letzel *et al.*, 2014).

The significant involvement of SAM enzymes in initiating the onset of bioactive metabolites in *Streptomyces* was proved by Okamoto *et al*(2003) and Kim *et al*(2003). Both research groups show that SAM enhanced actinorhodin production in *Streptomyces coelicolor* A3(2) and S. *lividans* TK23 is through activation of *act*II-ORF4 transcriptional factor, and also through the exogenous addition of SAM to the culture medium, thereby increasing the expression of genes in the actinorhodin biosynthetic cluster. Also, another study carried out by Sufrin *et al.* (2009) has shown that salinosporamide A, a bioactive product from marine *Salinispora tropica* now undergoing Phase I clinical trials for cancer treatment (Lechner *et al.*, 2011), is enzymatically derived from SAM.

In the last several years, a number of newly characterised radical SAM enzymes that carry out novel chemistry and demonstrate involvement in the antibiotics and active compounds production from *Streptomyces* have been reported. These include, YtkT - SAM protein that involved in Yatakemycin production from *Streptomyces* sp. TPA0356, which is a novel antitumor antibiotic belonging to the family of CC-1065 (Huang *et al.*, 2012); Fom3-SAM enzyme that involved in Fosfomycin production from *Streptomyces wedmorensis*, which is a broad-spectrum antibiotic used against multi-drug resistant bacteria (Allen and Wang 2014). Most recently, Nikomycins and polyoxins, which are antifungal peptidylnucleoside antibiotics active against human and plant pathogens, biosynthesis in *Streptomyces cacaoi* and *S. tendae* are catalysed by NikJ and PolH radical SAM enzymes respectively (Lilla and Yokoyama, 2016).

The two genes encoding radical SAM enzymes in isolate_99 were found adjacent to RiPP genes in both cluster 8 and 20 (Tables 7.2 and 7.3). This may indicate new classes of RiPP molecules were produced that are derived from a new type of radical SAM protein in the presence of Pb. This is supported by the fact that bacteriocin clusters predicted by antiSMASH did not show any homology to known biosynthetic gene clusters (Chapter 3, Table 6.1). However, further studies to analyse and characterise these putative encoding genes are needed.

Bacteriocins are antimicrobial peptides that have bacteriostatic or bactericidal effects on other bacteria. Bacteriocins as defined by Galves *et al.* (1989), are

ribosomally synthesised, hydrophilic or hydrophobic peptides with a narrower spectrum of antimicrobial activity than other antibiotics and found to be more effective against closely related species. However, the definition has been expanded beyond closely related species as several studies have been shown the potential effect of bacteriocins against unrelated bacteria. Despite some reports that most bacteriocins from Gram-positive bacteria (as most producers of bacteriocins) were unable to kill Gram-negative bacteria (Helander *et al.*, 1997; Chen & Hoover, 2003; Cotter *et al.*, 2005; Deegan *et al.*, 2006; Gillor *et al.*, 2008), others have reported that some bacteriocins have a broad spectrum of activities against both Gram-positive/negative bacteria. In the study carried out by Stevens *et al.* (1991), the nisin bacteriocin, produced by *Lactococcus lactis* subsp. *lactis*, displayed activity against several *Salmonella* species and other Gram-negative bacteria. In other work, Lyon *et al.* (1993) have found that Propionicin PLG-1 from *Propionibacterium thoenii* P127 is active against *Pseudomonas fluorescens*, *Vibrio parahaemolyticus*, *Yersinia enterocolitica* as well as *Listeria monocytogenes* and *Corynebacterium* sp. Similarly, a study conducted by Lopez-Lara *et al.* (1996) has shown that enterocin AS-48 which is produced by *Enterococcus faecalis* is widely effective against both related and unrelated bacteria, such as *Listeria, Bacillus*, *Salmonella choleraesuis*, and *E. coli*. Furthermore, in more recent studies in which bacteriocins were applied against Gram-negative microorganisms, such as the bacteriocin subtilosin A produced by *Bacillus subtilis* which shows activity against *Proteus mirabilis* and *Salmonella enterica* Typhi (Shelburne *et al.*, 2007), lactocyclicin Q, a novel bacteriocin produced by *Lactococcus* sp. Strain QU 12 which is active against *Salmonella enterica* Typhimurium and *E. coli* (Sawa *et al.*, 2009), the bacteriocin carnocyclin A from *Carnobacterium maltaromaticum* UAL307 which inhibits *E. coli* DH5α and *Pseudomonas aeruginosa* ATCC 14207 (Martin-Visscher *et al.*, 2010), and the bacteriocin BacTN635 produced by *Lactobacillus plantarum* sp. TN635 which displays a broad spectrum of activity against Gram-negative bacteria (*Salmonella enterica* ATCC43972, *Pseudomonas aeruginosa* ATCC 49189, *Hafnia* sp. and *Serratia* sp.) and the pathogenic fungus *Candida tropicalis* R2 CIP203 (Smaoui *et al.*, 2010).

Surprisingly, heavy metal related genes, lead uptake protein PbrT (cluster 8, table 7.10) and heavy metal translocating P-type ATPase (cluster 20, table 22), were predicted in isolate_99. These genes might be involved in metabolites productions. The transcripts of these genes were upregulated in the presence of Pb by 5.1-fold and 13.3-fold respectively. PbrT is a lead uptake permease belonging to the iron/lead transporter superfamily (Jarosławiecka *et al.*, 2014), which is involved in the influx of Pb. Also it participates (along with other encoded proteins) in a lead-resistance mechanism in bacteria (Hynninen *et al.*, 2009). Similarly, heavy metal translocating P-type ATPases have been found to play roles in metal uptake and homeostasis (Rutherford *et al.*, 1999), and also act as transporters that confer tolerance to the metal ion substrate through acting as an efflux pump (Rensing *et al.*, 1999; Okkeri & Haltia 2006).

Table 7.1: Differential expression analysis of secondary metabolism clusters of isolate_99 under lead metal stress which identified by antiSMASH 3.0. Samples without Pb supplement were used as a control. The cutoff was set at a *p*-value of 0.05.

Cluster	Type	Up/Down-regulated
Cluster 1	Lantipeptide	Down
Cluster 2	Siderophore	Down
Cluster 3	Terpene	Down
Cluster 4	Terpene	Down
Cluster 5	Otherks	Down
Cluster 6	Otherks	Down
Cluster 7	Terpene-Butyrolactone	Down
Cluster 8	Bacteriocin	Up
Cluster 9	Bacteriocin	Down
Cluster 10	Otherks-T1pks	Down
Cluster 11	Otherks-T2pks	Down
Cluster 12	Butyrolactone- Otherks	Down
Cluster 13	Terpene	Down
Cluster 14	Nrps	Down
Cluster 15	Siderophore	Down
Cluster 16	Nrps	Down
Cluster 17	Siderophore	Down
Cluster 18	Nrps	Down
Cluster 19	Othe	Down
Cluster 20	Bacteriocin	Up
Cluster 21	Melanin	Down
Cluster 22	Ectoine	Down
Cluster 23	Terpene	Down
Cluster 24	Other	Down
Cluster 25	Nrps	Down

Cluster 26	Nrps	Down

Table 7.2: Down and up-regulated genes involved in a biosynthetic gene cluster No. 8 (Bacteriocin) and their functions obtained from antiSMASH/NCBI/BacMet databases. Samples without Pb supplement were used as a control. The cutoff was set at a *p*-value of 0.05.

Gene type	Description (antiSMASH/ NCBI / BacMet)	Fold change	*p*-value
Transport-related gene	Drug resistance transporter (EmrB/QacA)/ efflux|putative transmembrane efflux protein/ multidrug efflux pump SmfY	5.2	0.3
Biosynthetic gene	Alpha/beta hydrolase fold protein/ Alpha/beta hydrolase fold protein / Efflux pump protein FarA	1.4	1
Other gene	No hits/ Hypothetical protein/ Nickel transport permease system	2.9	0.6
Other gene	No hits/ Hypothetical protein/ LysR famil	27.9	0.49
Other gene	No hits/ Hypothetical protein/ Mutidrug resistance protein MdtA	2.6	0.5
Other gene	No hits/ Hypothetical protein/ Mutidrug resistance protein PmpM	2.8	1
Other gene	No hits/ RiPP maturation radical SAM protein/ Arsenical pump-driving ATPase	11.5	1
Other gene	No hits/ RiPP maturation radical SAM protein/ Outer membrane protein OprM	13	0.00014
Biosynthetic gene	Radical SAM domain protein Radical SAM domain protein CopA family copper resistance protein	4.8	1.3E-06

Other gene	No hits/ No hits/ ABC transporter	2	0.9
Other gene	No hits/ Hypothetical protein/ Copper homeostasis protein CutC	15.1	0.7
Other gene	No hits/ Hypothetical protein/ Putative cation efflux system protein	8.4	1.95E-56
Other gene	No hits/ Hypothetical protein/ Multidrug resistance protein	7.4	2.46E-13
Other gene	No hits/ No hits/ Multidrug resistance protein	3.1	1
Biosynthetic gene	Lantibiotic dehydratase/ Lanthionine biosynthesis protein/ Lead uptake protein PbrT	13.3	0.02
Other gene	No hitis/ Lanthionine biosynthesis protein/ Na+ driven multidrug efflux pump	8	1

Table 7.3: Down and up-regulated genes involved in a biosynthetic gene cluster No. 20 (Bacteriocin) and their functions obtained from antiSMASH/NCBI/BacMet databases.

Gene type	Description (antiSMASH/ NCBI / BacMet)	Fold change	*p*-value
Transport-related gene	Major facilitator transporter/ Efflux|putative multidrug resistance protein/ Major facilitator superfamily permease	5.6	0.06
Other gene	No hits/ Peptidase S9/ Copper	1.4	1
Other gene	No hits/ S9 family peptidase/ Copper resistance B precursor	3.2	0.4
Other gene	No hits/ hypothetical protein OV320_2954/ No hits	8.3	8.28E-09
Biosynthetic gene	Radical SAM domain protein/ RiPP maturation radical SAM protein 1/ Copper-translocating P-type ATPase	3.4	0.004
Other gene	No hits/ RiPP maturation radical SAM protein/ Multidrug resistance protein MdtO	4.5	1
Other gene	No hits/ Hypothetical protein/ Protein KlaC	2	1
Biosynthetic gene	3-hydroxyisobutyrate dehydrogenase/ NADP oxidoreductase/ Putative enoyl	1.4	1
Regulatory gene	TetR family/ TetR family/ TetR family	11.5	1
Other gene	No hits/ XRE family transcriptional regulator/	7.7	1

	No hits		
Other gene	No hits/ Transcriptional regulator Heavy metal translocating P-type ATPase	5.1	0.56
Other gene	No hits/ Hypothetical protein/ CzrB protein	4.5	1
Biosynthetic gene	Short chain dehydrogenase/ Short chain dehydrogenase/ enoyl-reductase	3.2	1

7.4 Transcriptional regulator families response to heavy metal (Pb)

The following transcriptional regulator families: AraC, GntR, MarR, AsnC, MerR, IclR, LacI, TetR, SARP, PadR, LuxR and LysR families were predected in biosynthetic gene clusters of isolate_99. TetR, MarR and GntR transcriptional regulation families were more abundant than the others (Table 7.4). It can be noticed that the transcript levels of these regulator genes were different between the samples with and without Pb added to the medium. The transcriptional levels of MerR in cluster 4 (Terpene), SARP in cluster 11 (Otherks-t2pks), TetR in cluster 20 (Bacteriocin) and GntR and TetR in cluster 23 (Terpene) were notably increased in the presence of lead metal. Although the changes in the transcript levels of MerR (cluster 4), TetR (cluster 20) and GntR (cluster 23) were not significant ($p > 0.05$), MerR, TetR and GntR transcripts increased by 10.8, 11.5 and 17 fold respectively when exposed to Pb. However, there were significant changes in the levels of SARP (cluster 11) and TetR (cluster 23) transcriptional genes when exposed to Pb ($p < 0.05$). The transcripts of these two genes were increased by 33.1 and 27 fold respectively (Table 7.4).

On the other hand, transcript levels of only two transcriptional regulatory genes were notably decreased in presence of Pb. The transcript levels of GntR in cluster 18 (Nrps) and MerR in cluster 24 (Other) were decreased by 10-fold each. However, the fold changes of these two genes were not significant ($p > 0.05$) (Table 7.4).

Table 7.4: Transcriptional regulator genes involved in biosynthetic gene clusters obtained from AntiSMASH 3.0 that were differentially expressed under lead metal stress.

Cluster/Position	Type	Transcriptional regulator gene	Fold change	*p*-value
3.1	Terpene	AraC family	1.2	0.09
3.2		GntR family	1.1	1
4.1	Terpene	MarR family	-1.1	1
4.2		AsnC family	2	1
4.3		MerR family	10.8	0.54
5.1	Otherks	IclR family	2.2	0.39
5.2		MerR family	-1	1
6.1	Otherks	LacI family	-1	1
6.2		TetR family	-1.3	0.78
6.3		SARP family	-1.6	0.34
7.1	Terpene-butyrolactone	TetR family	1.2	0.84
10.3	Otherks-t1pks	SARP family	1	1
10.4		MarR family	-1.2	1
11.1	Otherks-t2pks	SARP family	33.1	0.01
11.2		PadR family	-1.4	0.49
13.2	Terpene	LacI family	1	1
14.1	Nrps	IclR family	-2	0.34
14.2		LuxR family	3.3	0.7
14.3		LysR family	-1	1
15.1	Siderophore	ArsR family	4.5	1
16.1	Nrps	AsnC family	-1.3	1
16.5		TetR family	-1.3	0.68
17.1	Siderophore	TetR family	2.8	0.35
18.1	Nrps	IclR family	1.7	0.38
18.2		MerR family	2	1
18.3		LuxR family	-1.6	0.68

18.6		MarR family	-3.3	0.4
18.7		GntR family	-10.2	0.5
19.1	Other	GntR family	2	0.35
20.1	Bacteriocin	TetR family	11.5	0.25
23.1	Terpene	GntR family	17	0.2
23.2		TetR family	27	0.03
24.1	Other	LuxR family	1.5	0.37
24.2		MerR family	-10	0.51
24.3		ArsR family	-1.2	1
24.4		TetR family	-1.4	0.41
25.1	Nrps	AraC family	-2.4	0.68
25.2		MarR family	-5.3	0.01
25.3		TetR family	1.2	0.85
26.1	Nrps	LacI family	1.2	0.45
26.2		MarR family	1.4	1

The MerR family members are dimeric proteins that display a high degree of sequence similarity in the N-terminal DNA binding domains (Newberry & Brennan 2004). The MerR transcriptional regulatory genes have been shown to be activated in response to stress signals in bacteria, such as exposure to heavy metals, oxygen radicals, or cytotoxic compounds (Jarosławiecka & Piotrowska-Seget 2014). A subset of MerR family regulators has been identified which responds to Cd(II), Zn(II), Co(II), Cu(I), Ag(I), Au(I), Hg(II) and Pb(II) (Brown *et al.*, 2003). Of these regulatory genes, ZntR and PbrR are among the well-known identified lead metal specific regulatory genes (Brown *et al.*, 2003; Taghavi *et al.*, 2009). However, these two regulatory genes were identified in isolate_99 (> 60% identity) using BacMet database (Pal *et al.*, 2014). In *Streptomyces coelicolor*, the *bldC* locus, which is required for formation of aerial hyphae encodes a small DNA-Binding protein found in members of the MerR family of transcriptional activators which are needed for development and production of the polyketide antibiotics actinorhodin and undecylprodigiosin (Hunt *et al.*, 2005).

The TetR family of transcriptional regulators are well represented and widely distributed among bacterial species, whose proteins control genes involved in the biosynthesis of antibiotics, multidrug resistance, osmotic stress, and pathogenicity of Gram-negative and Gram-positive bacteria (Ramos *et al.* 2005). The TetR family of regulators is a large and important family of one-component signal transduction systems that consist of an N-terminal DNA binding domain and a larger C-terminal domain (Cuthbertson & Nodwell 2013). However, various TetR regulators have been identified in species of *Streptomyces* and related actinobacteria which are involved in the biosynthesis clusters for antibiotics and other secondary metabolites. Of these, ActR which is involved in biosynthesis of actinorhodin that acts a typical type II polyketide synthase in *S. coelicolor* (Das& Khosla 2009); LanK has found to regulate the biosynthesis for the glycosylated angucyclic polyketide antibiotic landomycin A in *S. cyanogenus* (Ostash *et al.*, 2008) and SimR, which is involved in the biosynthesis cluster for simocyclinone D8 in *Streptomyces antibioticus* Tü 6040 which is a structurally complex inhibitor of DNA gyrase (Galm *et al.*, 2002; Edwards *et al.*, 2009).

Transcriptional regulators of the *Streptomyces* antibiotic regulatory protein (SARP) family have been shown to play a role in the activation of secondary metabolites production in *Streptomyces* species. In *S. coelicolor* they are positive regulators in the production of undecylprodigiosin and actinorhodin antibiotics (Bibb, 2005); in *S. peucetius* they regulate daunorubicin biosynthesis (Madduri and Hutchinson, 1995); and they regulate tylosin antibiotic production in *S. fradiae* (Bate *et al.*, 2002). However, SARP transcriptional regulators are a specific family of paralogous proteins that show a high specificity for regulating extracellular signals of γ-butyrolactone molecules which are produced in many if not all streptomycetes and in several other genera of actinomycetes (Bibb, 2005).

The GntR family is a large group of proteins present in diverse bacteria and regulating various biological processes (Aravind & Anantharaman, 2003). In *Streptomyces,* the GntR-family regulator DasR has been shown to regulate the sugar phosphotransferase system of N-acetylglucosamine (GlcNAc) metabolism to control the signal for the switch to antibiotic production and morphological development (Rigali *et al.*, 2008). However, the authors have shown that under famine conditions, GlcNAc addition stimulated antibiotic production in *S. clavuligerus, S. collinus, S. griseus, S. hygroscopicus* and *S. venezuelae* grown on MM agar (5 mM GlcNAc or higher) by functioning as an allosteric effector of the pleiotropic transcriptional repressor DasR regulon, which controls the GlcNAc transport and metabolism as well as antibiotic production. The AraC family of transcriptional regulators is another large group of regulatory proteins in bacteria. Members of this group are involved in the transcriptional regulation of a variety of cellular processes in bacteria, including carbon metabolism (Gallegos *et al.*, 1997); stress responses, e.g., response to alkylating agents in *E. coli*, *S. typhimurium*, *Mycobacterium tuberculosis* and *Bacillus subtilis*, response to oxidative stress in *E. coli*, *S. typhimurium* and *Mycobacterium tuberculosis*, tolerance to antibiotics and organic solvents in *Providencia stuartii , E. coli* and *Klebsiella pneumoniae* (Gallegos *et al.*, 1997; Yang *et al.*, 2011); and virulence in *Mycobacterium tuberculosis* (Gallegos *et al.*, 1997; Frota *et al.*, 2004). However, the AdpA protein which belongs to the AraC family has been found to play a major role in regulating A-factor, the representative of the γ-butyrolactones, which

is control secondary metabolism or morphological differentiation in various species of *Streptomyces* (Horinouchi and Beppu, 1994). In a study carried out by Kato *et al.* (2005), they have revealed that an A-factor-deficient mutant of *S. griseus* neither produces any of the secondary metabolites, including streptomycin, nor forms aerial mycelium or spores, and when A-factor was added to such mutants, the defects in secondary metabolism and morphological differentiation were simultaneously restored. Thus A-factor was believed to act as a chemical signalling molecule, or microbial hormone, for both secondary metabolism and morphological differentiation (Horinouchi, 2002).

The LuxR family of transcriptional regulators are well known to be involved in transcriptional activator of Quorum Sensing (QS) and coordinates the expression of a variety of genes, including those encoding virulence factors (Fuqua *et al.*, 1996; Chen & Xie 2011). However, some members of the LuxR family have been found to positively regulate biosynthesis of bioactive metabolites in Actinobacteria. Previously published studies demonstrate a major role of LuxR regulators in the antibiotic biosynthesis *Streptomyces* spp. Examples include geldanamycin biosynthesis in S. *hygroscopicus* (He *et al.*, 2008), amphotericin biosynthesis in S. *nodosus* (Carmody *et al.*, 2004), nystatin biosynthesis in S. *noursei* (Sekurova *et al.*, 2004) and pikromycin biosynthesis in S. *venezuelae* (Wilson *et al.*, 2001). Moreover, the production of salinosporamide A was enhanced through the activation of the biosynthesis of specific precursor chloroethylmalonyl-CoA in *Salinispora tropica* as a result of overexpression of salR2 which is a member of luxR family (Lechner *et al.*, 2011).

Other common regulator families that found in isolate_99 were MarR, AsnC, IclR, LacI, PadR and LysR (Table 7.4). There is no evidence that these transcriptional regulatory genes are involved in the secondary metabolites production in bacteria. However, they serve and play a key role in a drug efflux pump and antibiotic resistance, virulence, amino acids metabolism, carbon metabolism, pilli biosynthesis, DNA transactions during DNA repair and recombination, mRNA translation, degradation of aromatic compounds, biofilm formation and motility (Ellison & Miller,

2006; Calvo &Matthews, 1994; Thaw *et al.*, 2006; Yamamoto & Ishihama, 2003; Antonio *et al.*, 2006; Aguilar *et al.*, 2014).

7.5 Transport-related genes responding to heavy metal (Pb)

The transcript levels of transported-related genes involved in biosynthetic gene clusters of isolate_99 were slightly down-regulated in presence of lead (Table 7.2). The overall changes in the transcript levels of transport-related genes were not significant ($p > 0.05$). However, the drug resistance transporter gene (EmrB/QacA) in cluster 8 (Bacteriocin), major facilitator transporter gene in cluster 16 (Nrps) and major facilitator transporter gene in cluster 20 (Bacteriocin) were notably up-regulated. There were significant changes ($p < 0.05$) in the levels of drug resistance transporter gene (cluster 8) and major facilitator transporter gene (cluster 20) increasing by 5.2 and 5.5-fold respectively. Although the transcript of major facilitator transporter gene (cluster 16) was up-regulated by 15-fold, the change in the transcript level was not significant ($p > 0.05$).

It can be noticed that the ABC and major facilitator are most common transported-related genes involved in biosynthetic gene cluster of isolate_99 (Table 7.5). Both the ABC systems and the major facilitator superfamily can serve as importers that mediate the uptake of nutrients in bacteria including mono- and oligosaccharides, organic and inorganic ions, amino acids, peptides, iron-siderophores, polyamine cations, vitamins and metals also served as exports of various molecules, such as peptides, lipids, hydrophobic drugs, polysaccharides, and proteins (Davidson *et al.*, 2008; Theodoulou & Kerr, 2015; Reddy *et al.*, 2012). In the presence of high concentrations of metal ions, genes that encode for metal transporters are repressed in an effort to decrease the cytosolic uptake of that metal ion and thereby limit negative effects on bacterial cell components (Ma *et al.*, 2009). Metal transporters including the ABC transporters and the major facilitator family systems have been identified and characterised to be involved in metal homeostasis (Nies, 2003; Ma *et al.*, 2009).

Table 7.5: Transport-related genes involved in biosynthetic gene clusters obtained from AntiSMASH 3.0 that were differentially expressed under lead metal stress.

Cluster/Position	Type	Transport related gene	Fold change	*p*-value
4.1	Terpene	Drug resistance transporter, EmrB/QacA	-1.2	1
5.1 5.2	Otherks	ABC transporter permease protein ABC transporter ATP-binding protein	-1.4 1.3	0.8 0.54
6.1	Otherks	ABC transporter permease protein	1.3	0.67
8.1	Bacteriocin	Drug resistance transporter, EmrB/QacA	5.2	0.02
9.1	Bacteriocin	ABC transporter ATP-binding protein	-5.6	1
10.1	Otherks-t1pks	Drug resistance transporter, EmrB/QacA	2	0.16
11.1	Otherks-t2pks	Drug resistance transporter, EmrB/QacA	1.8	0.14
13.1 13.2 13.4	Terpene	ABC transporter, carbohydrate uptake Inner membrane translocator ABC transporter ATP-binding protein	1 -2.2 -6	1 1 1
14.1	Nrps	Major facilitator transporter	-1.2	1
15.1	Siderophore	Major facilitator transporter	2	0.62
16.1 16.2 16.3 16.4	Nrps	ABC transporter ATP-binding protein Major facilitator transporter Sodium/hydrogen exchanger Drug resistance transporter, EmrB/QacA	1 15 4.5 1	1 0.12 1 1
20.1	Bacteriocin	Major facilitator transporter	5.6	0.002

22.1	Ectoine	Polar amino acid ABC transporter	-1.7	0.03
23.1 23.2	Terpene	ABC transporter ATP-binding protein Sodium: dicarboxylate symporter	-1.1 -1.2	1 1
25.1 25.2	Nrps	Inner membrane translocator ABC transporter ATP-binding protein	3 -1.6	0.18 0.05

7.6 Heavy metal associated genes involved in biosynthetic clusters response to heavy metal (Pb).

The transcriptional levels of heavy metal associated genes involved in biosynthetic gene clusters of isolate_99 were determined in presence of lead metal. Out of 26 biosynthetic gene clusters detected in isolate_99 by antiSMASH (Chapter 5, Table 5.1), 11 biosynthetic gene clusters were displayed to have heavy metal associated genes either in their core or other (hypothetical) biosynthetic clusters (for more details see appendix C). The transcript levels of heavy metal genes involved in core biosynthetic clusters of lantipeptite (cluster 1), otherks (cluster 6), butyrolactone-otherks (cluster 12), NRPS (cluster 14), other (cluster 19), other (cluster 24) and NPRS (cluster 26) were slightly down-regulated in the presence of Pb, however, no significant change was observed ($p > 0.05$) apart from lantipeptide cluster ($p < 0.05$). On the other hand, heavy metal genes found in core biosynthetic gene clusters of otherks-tipks (cluster 10), siderophore (cluster 17), other (cluster 19) and NRPS (cluster 25) were up-regulated in the presence of Pb. Additionally, it has been noticed that the transcripts of heavy metal genes involved in other genes of bacteriocin (cluster 20, table 7.3) and other (cluster 24) biosynthetic clusters increased in presence of lead. However, although, the changes in the transcript levels of these genes were not significant ($p > 0.05$), the heavy metal genes increased by 5.1 and 8.7-fold respectively.

Chapter 8:

General Discussion

Besides traditional selection for preliminary isolation which is based on morphology characterisations for representative actinomycetes, we applied a stereoscopic microscope identification approach on selective medium which could exclude many unwanted microorganisms that did not exhibit similarities with actinomycetes. In addition, we recommend purifying actinomycetes on different culture media to elucidate their morphological and physiological characterisations.

It can be deduced from the molecular identification results that primer sets used for 16S rRNA gene amplification were highly conserved as different genera of actinomycetes were identified (Table 3.1, Chapter 3) including rare genera such as *Lentzea, Kribbella* and *Nocardioides*. In addition, the distance tree of these strains obtained using neighbour joining method (Figure 3.3, Chapter 3) exhibited that using leBIBI tool (Falndrois *et al.*, 2015) for determining and identification (Table 3.1, Chapter 3) by producing approximate maximum likelihood phylogenetic tree approach is an accurate new tool to find the most similar hit returns from environmental samples through phylogenetic reconstruction.

It seems to be that the induction of silent gene clusters by the heavy metal (Pb) resulted in isolates 85 and 99 screeting unknown compounds affecting the growth of the neighbouring cultures (Figure 3.4, chapter 3). These two promising isolates exhibited a very strong activity against *C. albicans* and *E.coli* respectively when they were grown on MM agar whereas they did not exhibit activity in broth medium. This suggests utilising solid agar media for screening actinomycetes for bioactive metabolites production from such medium state sensitive strains as in our study. To our knowledge, there are no published studies showing the activation effects of lead metal to promote the silent gene cluster in actinomycetes for secondary metabolite productions. However, a recent study conducted by Haferburg & Kothe (2013) showed that the biosynthetic pattern of secondary metabolites can be altered and new potentially therapeutic compounds might be discovered through activation of silent gene clusters in *Streptomyces* by heavy metals induction.

In order to avoid and minimise the interference by heavy metal ions with the ingredients of complex media, the agar plate method with Minimal Medium (MM) was used. Despite this, isolates grown on complex media supplemented with lead could not produce bioactive metabolites against tested microorganisms (Figure 3.6 B, Chapter 3). However, supplementation of MM with heavy metals as simulation of environmental conditions to which the isolate might have adopted can be induced to produce novel bioactive metabolites.

Based on obtained results from lead metal stress and bioassay against ESKAPE pathogens, we can conclude that microorganisms that are isolated from environments contaminated with heavy metals represent a point of interest for the research of novel undiscovered secondary bioactive metabolites.

The quality of the final draft whole genome sequences generated in this study using Ion Torrent Technology were concurrent to the data produced by other current next generation sequencing technologies, which are available from GenBank in NCBI. The data were sufficient with good coverage to perform *de novo* assemblies and the obtained draft sequences were in the average of the other genomes from *Streptomyces* and had the same ratio of GC content. It was possible with Ion Torrent PGM to generate a total data output of around 1 Gb using 316v2 chip with a total reads of 3.6 Million. However, the total bases are depended on the loading percentage, total number of reads and read length. Also, it is now possible with Ion Torrent PGM to generate 400bp read throughput of up to 2 GB using 318v2 chip type with 4-5.5 Million reads. Although low bead loading percentage can result from a low percentage of templated ISPs, which subsequently can be the result from many issues discussed above, we have developed a loading protocol (Appendix B) with high performance of >70% each time and achieved a record of 91% positive ISPs loading chip (Figure 4.2, Chapter 4).

Applying PHAST and ResFinder web services can easily identify prophage sequences and acquired antibiotic gene resistance in the genomes of *Streptomyces.* It may also help develop genetic tools, and *in silico* identifications can enable more accurate and efficient investigations in the laboratory.

The 16S rRNA gene sequence phylogenies, the most fundamental and crucial molecular marker for proposing a new species, had grouped the isolate_99 strain into the genus *Streptomyces* and distinguished it from any known *Streptomyces* spp. According to 16S rRNA phylogenetic tree (Figure 5.1, Chapter 5), the most closely related strains to isolate_99 were *S. turgidiscabies* and *S. graminilatus*. Two more phylogeny trees using 23S rRNA gene (Figure 5.4, chapter 5) and WGS reads were (Figure 5.3, Chapter 5) constructed to further investigate the relatedness between the isolate_99 and the most related strains. Interestingly, there was an excellent correlation between the phylogeny of the three trees constructed based on 16S rRNA, 23S rRNA and WGS. However, this can highlight the strength of ribosomal genes and WGS reads in identifying relationships between closely related species in genus of *Streptomyces*.

Applying the state-of-the-art bioinformatics tools using ANI and *in silico* DDH (Tables 5.1 & 5.2, Chapter 5) confirmed that the isolate_99 cannot belong to the same species or subspecies of *S. turgidiscabies* or *S. graminilatus*, and provided clear evidence that the isolate_99 is the newest member of the genus *Streptomyces*. Additionally, comparative data based on proteome analysis with PATRIC bioinformatics tool (Figure 5.5, Chapter 5) and genome function with RAST tool (Figures 5.6, 5.7 & 5.8, Chapter 5) also supports its novelty.

The phenotypic microarray (PM) experiments produced a great amount of high-quality data that gave useful information about phenotypic differences between isolate_99 and both *S. turgidiscabies* and *S. graminilatus* strains (Figures 5.9-5.20, Chapter 5). However, the correlation between the high throughput data obtained from the both next generation sequencing and phenotypic microarray showed the power of these technologies of discrimination between these three strains and assigning isolate_99 as a novel species of *Streptomyces*.

A comprehensive analysis was performed for identifying secondary metabolite biosynthesis gene clusters in isolate_99 employing both bioinformatics tools,

antiSMASH (Medema *et al.*, 2011) and ClusterFinder (Cimermancic *et al.*, 2014). A total of 26 gene clusters for putative of different classes of biosynthetic secondary metabolites were predicted using antiSMASH (Table 6.1, Chapter 6). However, ClusterFinder identified an additional cluster involving in PKs that was not identified by antiSMASH. In addition, ClusterFinder also identified another 33 putative biosynthetic gene clusters of unknown types that cannot be detected using antiSMASH Table 6.2, Chapter 6). The genome mining revealed nine PKS and NRPS gene clusters that might produce a wide range of medically important compounds are present in isolate_99. Abundance of these genes was also comparable to those of well-studied and industrially useful strains such *S. coelicolor* and *S. avermitilis*.

Implementing the Known ClusterBlast module with the new version of antiSMASH version 3.0 (Weber *et al.*, 2015) was necessary to avoid discovery of the same molecules produced by experimentally characterised biosynthetic gene clusters. Further, ClusterBlast was implemented to find similarities of BGCs identified in isolate_99 to other gene clusters within the same genus of *Streptomyces*, distantly related or other no-related genera. Surprisingly the biosynthetic gene cluster 25 did not display similarities to *Streptomyces* but instead showed similarities to non-related genera including *Clostridium* and *Bacillus* (Figure 6.3, Chapter 6). More interestingly, the presence of the same transposon in isolate_99, *Clostridium* and *Bacillus* might support the idea that horizontal gene transfer of this biosynthetic gene cluster occurred during the evolution of these genera. This cluster or the other clusters which were identified to demonstrate no similar known clusters (Table 6.1 and Table 6.2, Chapter 6) may be potential sources for novel natural products that have not been observed under standard laboratory conditions within the genus of *Streptomyces*.

Finally, the RNA sequencing and transcriptomic experiment was dedicated to the analysis of the transcriptional biosynthesis gene clusters of Isolate_99 under lead metal stress. It has been observed that the presence of lead metal had a negative effect on the integrity value of RNA compared to the RNA samples without metal (Figure 7.1, Chapter 7). Nevertheless, the ribosomal RNA of 16S and 23S were

successfully depleted from the samples (Figure 7.2, Chapter 7). Additionally, the average library sizes of the samples were between 260 to 284 bp (Figure 7.3, Chapter 7) which are highly recommended for 200 bp assay (average size 200 – 300 bp, life technologies) for the sequencing by Ion Torrent Technology.

Generally, although the number of expressed genes in samples without Pb supplement were more than in samples with Pb supplements (Figure 7.6, Chapter 7), the expression genes that differ significantly (*p*-value < 0.05) were found more in samples with Pb supplement than in samples without Pb supplement with a massive expression for some genes (Figure 7.8, Chapter 7).

Interestingly, out of 26 biosynthetic gene clusters identified in isolate_99 by antiSMASH only two clusters of bacteriocin type were up-regulated as a whole (Table 7.1, Chapter 7). Moreover, the transcripts of all genes including RiPPs and radical SAM enzymes that are involved in bacteriocin clusters of isolate_99 were increased and significantly up-regulated in presence of Pb (Table 7.2 & 7.3, Chapter 7). These results suggest that bacteriocin activities from isolate_99 might inhibit the growth of the neighbouring *Streptomyces* species as related group (Figure 3.4, Chapter 3) and the growth of *E. coli* as unrelated group (Figure 3.5, Chapter 3) of bacteria.

Several transcriptional regulatory genes have been found in biosynthetic gene clusters (BGCs) of isolate_99. Of these, TetR, MarR and GntR transcriptional families were more abundant than the others (Table 7.4, Chapter 7). However, not all of these genes involved in the identified in BGCs were found to be differentially regulated between the samples of with and without adding Pb to the medium. The following transcriptional regulator genes AraC, GntR, MerR, TetR, SARP and LuxR families were found to be involved in transcriptional activators of secondary metabolites production in bacteria including *Streptomyces* (Hunt *et al.*, 2005; Ramos *et al.* 2005; Bibb, 2005; Rigali *et al.*, 2008; Lechner *et al.*, 2011). On the other hand, MarR, AsnC, IclR, LacI, PadR and LysR transcriptional genes were found to play a vital role in bacterial life cycles rather than involving in the secondary metabolites production (Ellison & Miller, 2006; Calvo &Matthews, 1994; Thaw *et al.*, 2006;

Yamamoto & Ishihama, 2003; Antonio *et al.*, 2006; Aguilar *et al.*, 2014). Similarly, transport-related genes and heavy metal genes involved in BGCs of isolate_99 were also found to display different transcript levels under Pb stress.

Chapter 9:

Conclusions and Future Work

9.1 Conclusions

The main goal of this work was to discover new sources for novel antibiotics to combat the rise of antibiotic resistance by means of bacterial genome mining and transcriptomic analysis of silent secondary metabolite gene clusters under induction of heavy metals. The key findings of the thesis can be summarised as follows:

- Heavy metals contaminated sites can be rich sources of new *Streptomyces* strains including rare genera of actinomycetes that may produce pharmaceutically important bioactive metabolites (Chapter 3).
- The agar plate method with Minimal Medium (MM) is an excellent approach for simulation of environmental conditions for establishing induction of antibiotic silent gene clusters with heavy metals (Chapter 3).
- Promising *Streptomyces* isolates were stimulated to produce bioactive metabolites under the One Strain−Many Compounds (OSMAC) approach against ESKAPE pathogens utilising ISPs (ISP2-ISP7) media (Chapter 3).
- For Whole Genome Sequencing (WGS), we have developed a protocol for loading Ion 316v2/318v2 chip types where the loading percentage was > 70% each time and we have achieved a record of 91% positive ISPs loading chip (Chapter 4).
- Generating library sizes around 400 bp for WGS was led to high loading percentage and consequently the total amount of NGS data was increased leading to obtain sufficient coverage depths for *de novo* assemblies of the sequenced *Streptomyces* genomes (Chapter 4).
- The limitation of *de novo* genome assemblies without the use of a reference genome was overcome by insertion of long read length libraries of around 400 bp (Chapter 4).
- The limitation of *de novo* assemblies of high GC content *Streptomyces* genomes was overcome by enabling single-cell mode and choosing *k-mer* lengths of 21,33,55,77 using Spades assembler bioinformatics tool (Chapter 4).

- RestFinder and PHAST bioinformatics tools can serve as accurate *in silico* genetic tools to identify acquired antibiotic gene resistance and phages sequences in the genomes of *Streptomyces* (Chapter 4).
- The values from *in Silico* DDH and ANI have shown clear evidence that isolate_99 is a new member of the genus *Streptomyces* (Chapter 5).
- The phylogenetic relationships analysis based on 16S rRNA gene, 23S rRNA gene and WGS trees indicated that both S. *turgidiscabies* and *S. graminilatus* are the most closely related strains to the isolate_99 (Chapter 5).
- The results obtained from proteome analysis, Phenotypic Microarray (PM) and genome function based experiments have differentiated isolate_99 from both *S. turgidiscabies* and *S. graminilatus* strains (Chapter 5).
- Notably, the correlation between the high throughput data obtained from the both next generation sequencing and phenotypic microarray showed the power of these technologies to discrimination between the most closely related bacterial strains and assigning isolate_99 as a novel species of *Streptomyces* (Chapter 5).
- The results of genome mining for Biosynthetic Gene Clusters (BGCs) revealed that isolate_99 harbours 26 putative BGCs of differential classes of biosynthetic secondary metabolites. Some of these may be potential sources for novel natural products that have not been discovered under standard laboratory conditions within the genus of *Streptomyces* (Chapter 6).
- The RNA integrity number (RIN) values for samples without lead metal (Pb) were higher compared to samples with Pb supplement. This indicates that lead metal has a negative effect on the integrity of RNA samples isolated from bacteria. (Chapter 7).
- The presence of Pb did not affect the depletion of 16S rRNA and 23S rRNA genes from the RNA samples, also it showed no effect on the average library sizes of the RNA samples, which were within the recommended sizes, for the sequencing by Ion Torrent Technology using 200 bp assay (Chapter 7).
- The transcriptomic RNA analysis of 26 biosynthetic gene clusters identified in isolate_99 by antiSMASH revealed that the Bacteriocin gene clusters were the

only clusters that were significantly up-regulated in the presence of lead metal. These results suggest that bacteriocin activities from isolate_99 might inhibit the growth of the neighbouring *Streptomyces* species as a related group and also inhibit the growth of *E. coli* as an unrelated group of bacteria (Chapter 7).

- Finally, the transcriptomic and computational analyses suggest that there is also the potential to identify novel RiPP and radical SAM genetic loci in isolate_99. This may, in turn, lead to the discovery of novel antimicrobial compounds to treat multidrug resistant infections (chapter 7).

9.2 Future Work

The work warrants further investigations. From the theoretical point of view, it is still not entirely clear why the presence of Pb-metal in MM agar induced the silent gene clusters in these *Streptomyces* isolates to secrete bioactive compounds affecting the growth of tested microorganisms. Future work would be to identify and characterise of the bioactive compounds produced in cultures under Pb-metal stress by HPLC-MS, LC-MS, MS-MS and NMR spectroscopy. After patenting of the novel compounds, further development could be possible, including medicinal and industrial applications. It also remains to be investigated, whether the compounds produced in these cultures are synthesised only in presence of Pb-metal, or if their production is induced under metal-rich conditions to amounts exerting biological activity.

Also, for future work would be to analyse and characterise the two radical SAM protein encoding genes which may be responsible for the modification of RiPPs precursors and carry out novel chemistry that is involved in production of new bioactive metabolites. Additionally, investigation of genes surrounding the encoded radical SAM proteins like transporter, regulatory and putative genes that might cause antibiotic production in the presence of lead.

Chapter 10:

References

Abou-Shanab, R.A.I., van Berkum, P. and Angle, J.S. (2007). Heavy metal resistance and genotypic analysis of metal resistance genes in gram-positive and gram-negative bacteria present in Ni-rich serpentine soil and in the rhizosphere of *Alyssum murale*. *Chemosphere, 68*: 360–367

Afgan, E., Baker, D., van den Beek, M. *et al.*, The Galaxy platform for accessible, reproducible andcollaborative biomedical analyses: 2016 update. *Nucleic Acids Research, 44*: W3-W10

Aguilar, C., Schmid, N., Lardi, M., Pessi, G. and Eberl, L. (2014). The IclR-Family Regulator BapR Controls Biofilm Formation in B. cenocepacia H111. *PLoS ONE, 9*: 1-7

Aigle, B. and Corre, C. (2012). Waking up *Streptomyces* Secondary Metabolism by Constitutive Expression of Activators or Genetic Disruption of Repressors. *Methods in Enzymology, 517*: 343-366

Akins, R.L. and Rybak, M.J. (2001). Bactericidal Activities of Two Daptomycin Regimens against Clinical Strains of Glycopeptide Intermediate-Resistant *Staphylococcus aureus*, Vancomycin-Resistant *Enterococcus faecium*, and Methicillin-Resistant *Staphylococcus aureus* Isolates in an In Vitro Pharmacodynamic Model with Simulated Endocardial Vegetations. *Antimicrobial Agent and Chemotherapy, 45*: 454-459

Alam, M., Merlo, M., Takano, E. and Breitling, R. (2010). Genome-based phylogenetic analysis of Streptomyces and its relatives. *Molecular Phylogenetics and Evolution, 54*: 763–772

Allen, K. and Wang, S. (2014). Initial Characterization of Fom3 from *Streptomyces wedmorensis*: The Methyltransferase in Fosfomycin Biosynthesis. *Arch Biochem Biophys, 543*: 67–73

Altschul, P., Gish, W., Miller, W., Myers, E. and Lipman, D. (1990). Basic Local Alignment Search Tool. *J. Mol. Biol., 215*: 403-410

Amoroso, M., Benimeli, C. and Cuozzo, S. (2013). *Actinobacteria, Application in Bioremidation and Production of Indertrial Enzymes*. New York: CRC Press.

Amoroso, M., Castro, G., Carlino, F., Romero, N., Hill, R. and Oliver, G. (1998). Screening of heavy metal-tolerant actinomycetesisolated from the Salí River. *The Journal of General and Applied Microbiology, 44*: 129–132

Ando, T., Hirayama, K., Takahashi, R., Horino, I., Etoh, Y., Morioka, H., Shibai, H. and Murai, A. (1985). Cosmomycin D, a New Anthracycline Antibiotic. *Agricultural and Biological Chemistry, 49: 259-262*

Andrews, S. (2011). FastQC A Quality Control tool for High Throughput Sequence Data. Babraham Bioinformatics.

Antonio, J., Krell, T., Guazzaroni, M. *et al.* (2005). Membersof theIclRfamilyof bacterial transcriptional regulators functionas activatorsand/or repressors. *FEMS Microbiol Rev, 30*: 157–186

Aouiche, A., Bijani, C., Zitouni A, Mathieu F. and, Sabaou, N. (2014). Antimicrobial activity of saquayamycins produced by *Streptomyces* spp. PAL114 isolated from a Saharan soil. Journal de Mycologie Médicale, 2:17-23

Aravind, L.and Anantharaman, V. (2003). HutC/FarR-like bacterial transcription factors of the GntR family contain a small molecule-binding domain of the chorismate lyase fold. FEMS Microbiol Lett., 222: 17–23

Arnison, P., Bibb, M., Bierbaum, G. *et al.* (2013). Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep., 30*: 108–160

Asmub, M., Mullenders, L. H. F., Hartwig, A. (2000). Interference by toxic metalcompounds with isolated zinc finger DNA repair proteins. *Toxicology Letters, 112*: 227–231.

Auch, A., von Jan, M., Klenk, HP., and Göker, M. (2010). Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Standards in Genomic Sciences, 2*:117-134

Aziz, A., Bartels, D., Best, A. *et al.*, (2008). The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics, 9*:1-15

Balcazar, J.L. *(2014). Bacteriophages as Vehicles for Antibiotic Resistance Genes in the Environment. PLoS Pathog 10 (7).*

Bali, V., Panesar, P., Bera, M. and Kennedy, J. (2016). Bacteriocins: Recent Trends and Potential Applications. *Critical Reviews in Food Science and Nutrition, 56*:817–834

Bansal, A. K., and Meyer, T. E. (2002). Evolutionary analysis by wholegenome comparisons. *Journal of Bacteriology, 184*:2260–2272

Bate, N., Stratigopoulos, G. and Cundliffe, E. (2002). Differential roles of two SARP-encoding regulatory genes during tylosin biosynthesis. *Molecular Microbiology , 43*: 449 – 458

Bentley, S.D., Chater, K.F, Cerdeño-Tárraga A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabbinowitsch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G., Parkhill, J. and Hopwood, D. A. (2002). Complete genome sequence of the model actinomycetes *Streptomyces coelicolor*. *Nature, 417*: 141-147.

Bérdy, J. (2005). Bioactive Microbial Metabolites. *J. Antibiot. 58*(1):1–26.

Bertels, F., Silander, O., Pachkov, M., Rainey, P. and van Nimwegen, E. (2014). Automated reconstruction of whole genome phylogenies from short sequence reads. *Mol Biol Evol, 3*: 1-21

Bibb, M. J. (2005). Regulation of secondary metabolism in streptomycetes. *Current Opinion in Microbiology, 8*: 208–215.

Blin, K., Medema, M., Kazempour, D., *et al.*, (2013). antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. *Nucl Acids Res, 41*: W204-W212

Blin, K., Kazempour, D., Wohlleben, W. and Weber, T. (2014). Improved Lanthipeptide Detection and Prediction for antiSMASH. *PLoS ONE, 9(2)*: 1-7

Bochner, B. (2009). Global phenotypic characterization of bacteria. Fems Microbiology

Reviews, 33: 191-205.

Bochner, B., Gadzinski, P., Panomitros, E. (2001). Phenotype microarrays for highthroughput phenotypic testing and assay of gene function. *Genome Res 11*:1246–1255

Bode, H. B., Bethe, B., Hoefs, R. and Zeeck, A. (2002). Big effects from small changes: possible ways to explore nature's chemical diversity. *Chembiochem, 3*: 619–627

Boeck, L., Papiska, H., Wetzel, R., Mynderse, J., Fukuda, D., Mertz, F., Berry, D. (1990) A54145, a new lipopeptide antibiotic complex: discovery, taxonomy, fermentation and HPLC. *J Antibiot (Tokyo) 43*:587–593

Brown, L., Stoyanov, V., Kidd, P., and Hobman, L. (2003). The MerR family of transcriptional regulators. *FEMS, 27*: 145-163

Bouras, N, Meklat, A., Toumatia, O., Mokrane, S., Holtz, M.D., Strelkov, S.E. and Sabaou, N. (2013). Bioactive potential of a new strain of *Streptomyces* sp.PP14 isolated from Canadian soil. *African Journal of Microbiology Research, 7*: 3199-3208

Brede, D., Faye, T., Johnsborg, O., Odegard, I., Nes, I., *et al.* (2004). Molecular and genetic characterization of propionicin F, a bacteriocin from Propionibacterium freudenreichii. *Appl Environ Microbiol, 70*: 7303–7310

Brink, A.J., Coetzee, j., Clay, C.G., Sithole, S., Richards, G.A., Poirel, L. and Nordmanne, P. (2012). Emergence of New Delhi Metallo-Beta-Lactamase (NDM-1) and Klebsiella pneumoniae Carbapenemase (KPC-2) in South Africa. *Journal of Clinical Microbiology, 95*(11): 525-527

Bushnell, G., Mitrani-Gold, F. and Mundy, L.M. (2013). Emergence of New Delhi metallo-β-lactamase type 1-producing *Enterobacteriacea*e and non - *Enterobacteriaceae*: global case detection and bacterial surveillance. *International Journal of Infectious Diseases, 17*: 325-333

Bystrykh, L., Fernández-Moreno, M., Herrema, J., Malpartida, F., Hopwood, D and Dijkhuizen, L.(1996). Production of Actinorhodin-Related ''Blue Pigments'' by *Streptomyces* coelicolor A3(2). *Journal of Bacteriology, 178*: 2238–2244

Bursy, J., Kuhlmann, A., Pittelkow, M. *et al.* (2008). Synthesis and Uptake of the Compatible Solutes Ectoine and 5-Hydroxyectoine by Streptomyces coelicolor A3(2) in Response to Salt and Heat Stresses. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY, 47*: 7286–7296

Calvo, .J, and Matthews, R. (1994). The leucine-responsive regulatory protein, a global regulator of metabolism in Escherichia coli. *Am Soc Microbiol., 58*: 466–490.

Cane, D.E. and Ikeda, H. (2012). Exploration and mining of the bacterial terpenome. A*cc. Chem. Res., 45*: 463–472.

Carmody, M., Byrne, B., Murphy, B. *et al.* (2004). Analysis and manipulation of amphotericin biosynthetic genes by means of modified phage KC515 transduction techniques. *Gene, 343*: 107–115

Carvalho, H., Garrido, L.M., Furlan, R.L., Padilla, G., Agnoletto, M., Guecheva, T., Henriques, J.A., Saffi, J.,and Menck, C.F. (2010). DNA damage induced by the anthracycline cosmomycin D in DNA repair-deficient cells. *Cancer Chemother. Pharmacol. 65*: 989 –994.

Chai, Y.-J., Cui, C.-B., Li, C.-W., Wu, C.-J., Tian, C-K, and Hua, W. (2012). Activation of the Dormant Secondary Metabolite Production by Introducing Gentamicin-Resistance in a Marine-Derived Penicillium purpurogenum G59. *Marine Drugs, 10*: 559-582.

Chakravorty, S., Helb, D, Burday, M., Connell, A. and Alland, D. (2007). A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *Journal of Microbiological Methods, 69*: 330–339

Chander, J., Singla, N. and Handa, U. (2013). Human cervicofacial mycetoma caused by

Streptomyces griseus: First case report. Journal of Microbiology, *Immunology and Infection, 12*: 244-247

Chen, Y. Liu, T., Yu, C. Chiang, T. and Hwang, C. (2013). Effects of GC Bias in Next-Generation-Sequencing Data. PLoS ONE, 8(4).

Chen, H. & Hoover, D. (2003). Bacteriocins and their food applications*. Compr Rev Food Sci F 2*: 82–100.

Chen, J. and Xie, J. (2011). Role and Regulation of Bacterial LuxR-Like Regulators. *Journal of Cellular Biochemistry, 112*: 2694–2702

Cheon, Y., Kim, J., Park, J. *et al.* (2014). A biosynthetic pathway for hexanoic acid production inKluyveromyces marxianus. *Journal of Biotechnology, 182*:30–36

Chiang, Y-M., Chang, S-L., Oakle, B. R. and Wang, C. C. (2011). Recent advances in awakening silent biosynthetic gene clusters and linking orphan clusters to natural products in microorganisms. *Current Opinion in Chemical Biology, 15*:137–143

Christian, O.E., Compton, J., Christian, K.R., Mooberry, S.L., Valeriote, F.A. and Crews, P. (2005). Using Jasplakinolide to Turn on Pathways that Enable the Isolation of New Chaetoglobosins from Phomospis asparagi. *Journal of Natural Products, 68*: 1592–1597

Chun, J. and Goodfellow, M. (1995). A Phylogenetic Analysis of the Genus *Nocardia* with 16s rRNA Gene Sequences. *International Journal of Systematic Bacteriology, 45*: 240-245

Cimermancic, P., Medema, M., Claesen, J., *et al.* (2014). Insights into Secondary Metabolism from a Global Analysis of Prokaryotic Biosynthetic Gene Clusters. *Cell 158,* 412–421,

Clarridge, J.E. (2004). Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases. *Clinical Microbiology Reviews*, 17: 840–862

Cold Spring Harbor Laboratory Press. (2007). DNA loading buffer (6X). CSH Protocols. doi:10.1101/pdb.rec11045.

Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., Klam-Syed-Mohideen, A.S., McCarell, D.M., Marsh, T., Garrity, G.M., and Tiedje, J.M. (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Research, 37*: D141–D145.

Cotter, P. D., Ross, R. P. and Hill, C. (2013). Bacteriocins—a viable alternative to antibiotics?.*Nat Rev Microbiol, 11*:95–105

Cotter P.D., Hill, C. & Ross, R. P. (2005b). Bacteriocins: developing innate immunity for food. *Nat Rev Microbiol 3*: 777–788

Cruz-Morales, P., Kopp, J., Martínez-Guerrero, C. *et al.*, (2016): Phylogenomic analysis of natural products biosynthetic gene clusters allows discovery of arseno-organic metabolites in model streptomycetes. *Genome Biology and Evolution Advance Access, 11*: 1-27

Cuthbertson, L. and Nodwell, J. (2013). The TetR Family of Regulators. Microbiology and *Molecular Biology Reviews, 77*: 440–475

D'aes, J., Kieu, N., Léclère, V. *et al.*, (2014). To settle or to move? The interplay between two classes of cyclic lipopeptides in the biocontrol strain Pseudomonas CMR12a. *Environmental Microbiology,* 1-91

Deegan, L., Cotter, P., Hill, C. and Ross, P. (2006). Bacteriocins: biological tools for bio-preservation and shelf-life extension. *Int Dairy J 16*: 1058–1071

Denef, V., Park, J., Tsoi, T., *et al.* .(2004). Biphenyl and Benzoate Metabolism in a Genomic Context: Outlining Genome-Wide Metabolic Networks in Burkholderia xenovorans LB400. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY, 70*: 4961–4970

Desai, A., Marwah, V., Yadav, A., Jha, V., Dhaygude, K., Bangar, U. and Jere a. (2013). Identification of Optimum Sequencing Depth Especially for De Novo Genome Assembly of Small Genomes Using Next Generation Sequencing Data. PLoS ONE 8(4)

Dharmik, P.G. and Gomashe, A.V. (2013). Isolation, Identification and Antioxidant Activity of Melanin Pigment from Actinomycete (*Streptomyces* Species) Isolated from Garden Soil, Nagpur District, India. *International Journal of Pure and Applied Sciences and Technology, 18*: 69-72

Donadio, S., Sosio, M., Stegmann, E., Weber, T., Wohlleben, W. (2005). Comparative analysis and insights into the evolution of gene clusters for glycopeptide antibiotic biosynthesis. *Mol Gen Genomics, 274*: 40–50

Donadio, S., Monciardini, P. and Sosio, M. (2007). Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics. *Nat. Prod. Rep., 24*: 1073–1109

Darmon, E. and Leach, R. (2014). Bacterial Genome Instability. *Microbiology and Molecular Biology Reviews, 78*: 1-39

Das, A. and Khosla, C. (2009). Biosynthesis of aromatic polyketides in bacteria. Acc. Chem. Res., 42: 631–639.

Das, S., Ward, L.R. and Burke, C. (2010). Screening of marine *Streptomyces* spp. for potential use as probiotics in aquaculture. *Aquacultur, 7*: 32-42

Davidson, A., Dassa, E., Orelle, C. and Chen, J. (2008). Structure, Function, and Evolution of Bacterial ATP-Binding Cassette Systems. *Microbiology and Molecular Biology Reviews, 72*: 317–364

Dunne, E.F., Burman W.J. and Wilson, M.L. (1998). *Streptomyces pneumonia* in a patient with human immunodeficiency virus infection: case report and review of the literature on invasive *Streptomyces* infections. *Clinical Infection Diseases, 27*: 93-96

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research, 32*:1792–1797.

Edwards, M., Flatman, R. and Mitchenall, L . *et al.* (2009). A crystal structure of the bifunctional antibiotic simocyclinone D8, bound to DNA gyrase. *Science, 326*: 1415–1418.

Ekblom, P and Wolf, J. (2014): *Afield guide to whole-genome sequencing, assembly and annotation. Evolutionary Applications: 1026-1042*

Ellison, D., and Miller, L. (2006). Regulation of virulence by members of the MarR/SlyA family. *Curr. Opin. Microbiol., 9*: 153–159

Espinoza, L.E., Davelos Baines, A.L. and Lowe, K.L. (2013). Biochemical, Nutrient and Inhibitory Characteristics of *Streptomyces* Cultured from a hypersaline Estuary, the Laguna Madre (Texas). *Journal of Biological Sciences 13*: 18-27

Fahal, A.H. and Sabaa AHA. (2010). Mycetoma in children in Sudan. *Transactions of the Royal Society of Tropical Medicine and Hygiene, 104*:117-121

Flandrois, J-P., Perrier, G. and Gouy, M. (2015). leBIBI^{QBPP} : a set of databases and a webtool for automatic phylogenetic analysis of prokaryotic sequences. *BMC Bioinformatics, 16*:251

Fischbach, M., Walsh, C. and Clardy, J. (2007). The evolution of gene collectives: How natural selection drives chemical innovation. *PNAS, 105*: 4601–4608

Foster, J., Beckstrom-Sternberg, S., Pearson, T., *et al.*, (2009). Whole-Genome-Based Phylogeny and Divergence of the Genus Brucella. *JOURNAL OF BACTERIOLOGY, 8*: 2864–2870

Fuqua, C., Winans, S., Greenberg, E. (1996) Census and consensus in bacterial ecosystems: the LuxR-LuxI family of quorum-sensing transcriptional regulators. *Annu Rev Microbiol, 50*: 727–751.

Gallegos, M, Schleif, R., Bairoch, A., Hofmann, K. and Ramos, J. (1997). AraC/XylS Family of Transcriptional Regulators. Microbiolog and Molecular Biology Reviews, 61: 393–410

Galm, U., Schimana, J., Fiedler, H., Schmidt, J., Li, S. and Heide, L. (2002). Cloning and analysis of the simocyclinone biosynthetic gene cluster of Streptomyces antibioticus Tü 6040. *Arch. Microbiol., 178*: 102–114

Galvez, A., Maqueda, M., Martinez-Bueno, M. and Valdivia, E. (1989). BACTERICIDAL AND BACTERIOLYTIC ACTION OF PEPTIDE ANTIBIOTIC AS-48 AGAINST GRAM-POSITIVE AND GRAM-NEGATIVE BACTERIA AND OTHER ORGANISMS. *Res. Microbiol., 140*, 57-68

Geng, P., Meng, X., Bai, G. and Luo, G. (2008). Profiling of Acarviostatin Family Secondary Metabolites Secreted by Streptomyces coelicoflavus ZG0656 Using Ultraperformance Liquid Chromatography Coupled with Electrospray Ionization Mass Spectrometry. *Anal. Chem., 80*: 7554–7561

Ghosh, S. and Chan, C. (2016). Analysis of RNA-Seq Data Using TopHat and Cufflinks. Methods in Molecular Biology, 1374: 339-361

Gillor, O., Etzion, A. and Riley, M. (2008). The dual role of bacteriocins as anti- and probiotics. *Appl Microbiol Biot 81*: 591–606

Goodfellow, M. (2012). Phylum Actinobacteria. In Bergey's. Manual of Systematic Bacteriology, 2nd edn, vol. 5. Springer, New York, 56–57.

Goodfellow, M., Kumar, Y., Labeda, D.P. and Sembiring, L. (2007). The *Streptomyces* violaceusniger clade: a home for Streptomycetes with rugose ornamented spores. *Antonie van Leeuwenhoek, 92*: 173-199

Goodfellow, M., Mordarski, M.,Williams, S.T. and Mordarski, M. (1983). *The biology of the actinomycetes*. London: Academic Press Inc., 1-88

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P. and Tiedje, J. M. (2007). DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol 57*: 81–91.

Gregory, L.C. (2008). Mining microbial genomes for new natural products and biosynthetic pathways. *Microbiology, 154,* 1555–1569

Gunatilaka, A.A.L .(2006). Natural products from plant-associated microorganisms: distribution, structural diversity, bioactivity, and implications of their occurrence. *Journal of Natural Products, 69*: 509–526

Guo, J., Lin, Y., Zhao, M., Sun, R., Wang, T., Tang, M. and Wei, G. (2009). *Streptomyces* plumbiresistens sp. nov., a leadresistant actinomycete isolated from lead-polluted soil in north-west China. *International Journal of Systematic and Evolutionary Microbiology, 59*: 1326–1330

Guo, Y., Zheng, W., Rong, X and Huang, Y. (2008). A multilocus phylogeny of the *Streptomyces* griseus 16S rRNA gene clade: use of multilocus sequence analysis for streptomycete systematics*. International Journal of Systematic and Evolutionary Microbiology, 58*: 149–159

Gupta, R.S. (2011). Origin of diderm (Gram-negative) bacteria: antibiotic selection pressure rather than endosymbiosis likely led to the evolution of bacterial cells with two membranes. *Antonie van Leeuwenhoek, 100*: 171–182

Gupta, A. and Gupta U. (2014). Next Generation Sequencing and Its Applications. Elsevier Inc., ISBN: 978-0-12-386882-4

Gurevich, A., Saveliev, V., Vyahhi, N. and Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics, 29*: 1072-1075.

Hacker, J., and Kaper, J. (2000). Pathogenicity islands and the evolution of microbes. *Annu. Rev. Microbiol, 54*: 641–679

Haferburg, G. and Kothe, U. (2013). Activation of Silent Genes in Actinobacteria by Exploiting Metal Stress. In: Amoroso, M., Benimeli, C. and Cuozzo, S. (Eds.), *Actinobacteria, Application in Bioremidation and Production of Indertrial Enzymes*. New York: CRC Press, pp: 55- 72

Haferburg, G., Groth, I., Moellmann, U, Kothe, E.and Sattler, I. (2009). Arousing sleeping genes: shifts in secondary metabolism of metal tolerant actinobacteria under conditions of heavy metal stress. *Biometals, 22*: 225–234

Hall, B.G. (2011). *Phylogenetic trees made easy: a how-to manual*. 4th ed. Sunderland (MA), Sinauer Associates, ISBN: 978-0-87893-606-9

Han, J-H., Cho, M.-H. and Kim, S.B. (2012). Ribosomal and protein coding gene based multigene phylogeny on the family Streptomycetacea. *Systematic and Applied Microbiolog, 35*: 1-6

Hayakawa, M. and Monomura, H. (1987). Humic Acid-Vitamin Agar, a New Medium for the Selective Isolation of Soil Actinomycetes. *Journal of fermentation technology, 66*:501-509

Harisha, S. (2007). *Biotechnology Procedures and Experiments Handbook*. Infinity Science Press LLC, Hingham, ISBN: 978-1-934015-11-7.

Harrison, J. and Studholme, D. (2014). Recently published Streptomyces genome sequences. *Microbial Biotechnology, 7*(5), 373–380

Hartwig, A., Asmuss, M., Ehleben, I., Herzer, U., Kostelac, D., Pelzer, A., Schwerdtle, T., and Burkle,A. (2002). Interference bytoxic metalions withDNA repair processes and cellcycle control: molecula rmechanisms. *Environmental Health Perspectives, 110:*

797–799

Hazra, B., Bandyopadhyay, S., Saha, S.K., Banerjee, D.P. and Dutta, G. A. (1998). Study of mycetoma in eastern India. The Journal of Communicable Diseases, 30: 7-11

He, W., Lei, J., Liu, Y. and Wang, Y. (2008). The LuxR family members GdmRI and GdmRII are positive regulators of geldanamycin biosynthesis in *Streptomyces hygroscopicus* 17997. *Arch Microbiol, 189*: 501–510

Health Protection Agency (HPA). (2012). Identification of Aerobic Actinomycetes. *UK Standards for Microbiology Investigations, 11*:1-15

Helander, I., vonWright, A. & Mattila-Sandholm, T. (1997). Potential of lactic acid bacteria and novel antimicrobials against gram-negative bacteria. *Trends Food Sci Tech, 8*: 146–150**.**

Hernandez, D., François, P., Farinelli, L., Østerås and M., Schrenzel, J. (2008): De novo bacterial genome sequencing: Millions of very short reads assembled on a desktop Computer. Genome Research, 18: 802-809

Herzyk, P. (2014). *Next-Generation Sequencing*. Elsevier Inc., ISBN: 978-0-12-386882-4

Higginbotham, S.J. and Murphy, C.D. (2010). Identification and characterisation of a

Streptomyces sp. Isolate exhibiting activity against methicillin-resistant *Staphylococcu saureus***.** *Microbiological Research, 165*: 82- 86

Höfs, R., Walker, M. and Zeeck, A. (2000). Hexacyclinic acid, a polyketide from *Streptomyces* with a novel carbon skeleton. Angewandte Chemie, 39: 3258–3261.

Hopwood, D. (2007). *Streptomyces in Nature and Medicine*. Oxford University Press, Inc., New York, ISBN:13 978–0–19–515066–7

Horinouchi, S. (2002). A microbial hormone, A-factor, as a master switch for morphological differentiation and secondary metabolism in Streptomyces griseus. *Front. Biosci.,7*: d2045–d2057.

Horinouchi, S. and Beppu, T. (1994). A-factor as a microbial hormone that controls cellular differentiation and secondary metabolism in *Streptomyces griseus. Mol. Microbiol. 12*, 859 -864

Hua, G. and Höfte, M. (2014). The involvement of phenazines and cyclic lipopeptide sessilin in biocontrol of Rhizoctonia root rot on bean (Phaseolus vulgaris) by Pseudomonas sp. CMR12a is influenced by substrate composition. *Plant Soil, 13*: 1-11

Huang, W., Xu, H., Li, Y. *et al.* (2012). Characterization of Yatakemycin Gene Cluster Revealing a Radical SAdenosylmethionine Dependent Methyltransferase and Highlighting Spirocyclopropane Biosynthesis. J. Am. Chem. Soc., 134: 8831−8840

Hunt, A., Servin-Gonzalez, L., Kelemen, G. and Buttner, M. (2005). The bldC developmental locus of Streptomyces coelicolor encodes a member of a family of small DNA-binding proteins related to the DNA-binding domains of the MerR family. *J. Bacteriol, 187*:716 –728

Hwang, J.H., Kim, J.Y., Cha, M.R., Ryoo, I.J., Choo, S.J., Cho, S.M., Tsukumo, Y., Tomida, A., Shin-Ya, K., Hwang, Y.I., *et al.* (2008). Etoposide-resistant HT-29 human colon carcinoma cells during glucose deprivation are sensitive to piericidin A, a GRP78 down-regulator. *J. Cell. Physiol., 215*: 243–250.

Hynninen, A., Touzé, T., Pitkänen, L. (2009). An efflux transporter PbrA and a phosphatase PbrB cooperate in a lead-resistance mechanism in bacteria. *Molecular Microbiology, 74(2)*: 384–394

Ibrahim, A., El-Tayeb, M, Elbadawi, Y. and Al-Salamah, A. (2011). Isolation and characterization of novel potent Cr(VI) reducing alkaliphilic Amphibacillus sp. KSUCr3 from hypersaline soda lakes. *Electronic Journal of Biotechnology, 14*: 1-14

Imbeaud, S., Graudens, E., Boulanger, V., *et al.* (2005). Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces. *Nucleic Acids Research, 33*: 1-12

Jarosławiecka, A. and Piotrowska-Seget, S. (2014). Lead resistance in micro-organisms. *Microbiology, 160*: 12–25

Joshi, A. and Jaiswal, P. (2013). Microorganisms living in zinc contaminated soil - a review. *Journal of Pharmacy and Biological Sciences, 6*: 67-72

Joshi, M.V., Bingell, D.R.D., Johnson, E.G., Sparks, J.P., Gibson, D.M. and Loria, R. (2007). The AraC/XylS regulator TxtR modulates thaxtomin biosynthesis and virulence in *Streptomyces* scabies. *Molecular Microbiology, 3*: 633–642

Faccio, G., Kruus, K., Saloheimo, M. and Thöny-Meyer, L. (2012). Bacterial tyrosinases and their applications. *Process Biochemistry, 47*: 1749–1760

Fraise, A.P. (2002). Biocide abuse and antimicrobial resistance – a cause for concern? *Journal of Antimicrobial Chemotherapy, 49*: 11-12

Frota, C., Papavinasasundaram, K., Davis, E. and Colston, M. (2004). The AraC Family Transcriptional Regulator Rv1931c Plays a Role in the Virulence of Mycobacterium tuberculosis. *INFECTION AND IMMUNITY, 72*: 5483–5486

Farris, M.H., Duffy, C., Findlay, R.H. and Olson, J.B. (2011). *Streptomyces scopuliridis* sp. nov., a bacteriocin-producing soil streptomycete. *Int J Syst Evol Microbio, 61*: 2112–2116

Kämpfer, P. (2006). The Family Streptomycetaceae, Part I: Taxonomy. Prokaryotes, 3:538-604

Kämpfer, P., Buczolits, S., Albrecht, A., Busse and H., Stackebrandt, E. (2003). Towards a standardized format for the description of a novel species (of an established genus): *Ochrobactrum gallinifaecis* sp. nov. *International Journal of Systematic and Evolutionary Microbiology, 53*: 893–896

Kämpfer, P. (2012). Phylum Actinobacteria. In Bergey's Manual of Systematic Bacteriology, 2nd edn, vol. 5. Springer, New York, 1446–1806.

Kaminska, K., Purta, E., Hansen, L., Bujnicki, J., Vester, B., *et al.* (2010) Insights into the structure, function and evolution of the radical-SAM 23S rRNA methyltransferase

Cfr that confers antibiotic resistance in bacteria. *Nucleic Acids Research, 38*: 1652–1663

Katahara, T., Koyama, N., Mastuda, J. *et al.* (2004). Antimicrobial Activity of Saturated Fatty Acids and Fatty Amines against Methicillin-Resistant Staphylococcus aureus. *Biol. Pharm. Bull, 27*:1321-1326

Kato, J., Ohnishi, Y. and Horinouchi, S. (2005). Autorepression of AdpA of the AraC/XylS Family, a Key Transcriptional Activator in the A-factor Regulatory Cascade in *Streptomyces griseus*. *J. Mol. Biol , 350*: 12–26

Kauko, D., Haukka, K., AbuOun, M. *et al.* (2010). Phenotype MicroArray™ in the metabolic characterisation of Salmonella serotypes Agona, Enteritidis, Give, Hvittingfoss, Infantis, Newport and Typhimurium. *Eur J Clin Microbiol Infect Dis, 29*:311–317

Keller, PM, Rampini, SK, Buchler, AC, Eich, G, Wanner, PM and Bloemberg, GV (2010). Recognition of Potentially Novel Human Disease-Associated Pathogens by Implementation of Systematic 16S Rrna Gene Sequencing in the Diagnostic Laboratory. *J Clin Microbiol, 48*: 3397–3402

Kersten, R., Yang, Y., Xu, Y. *et al.*, (2012). A mass spectrometry-guided genome mining approach for natural product peptidogenomics. *Nat Chem Biol., 11*: 794–802

Khatri, B., Fielder, M., Jones, J. *et al.* (2013). High Throughput Phenotypic Analysis of Mycobacterium tuberculosis and Mycobacterium bovis Strains' Metabolism Using Biolog Phenotype Microarrays. *PLoS ONE, 8*:1-15

Kieser, T., Bibb, M. J., Buttner, M. J, Chater, K. F and Hopwood, D. A. (2000). *Practical Streptomyces Genetics*. The John Innes Foundation, Norwich, England

Kim, D., Huh, J., Yang, Y. *et al.* (2003). Accumulation of S-Adenosyl-L-Methionine Enhances Production of Actinorhodin but Inhibits Sporulation in Streptomyces lividans TK23. *JOURNAL OF BACTERIOLOGY, 185*: 592–600

Kim, E.J., Chung, H.J., Suh, H.J., Hah, Y.C., Roe, J.H. (1998). Transcriptional and post-transcriptional regulation by nickel of sodN gene encoding nickel-containing superoxide dismutase from *Streptomyces* coelicolor Muller. *Molecular Microbiology, 27*: 187–195.

Kim, J., Kim, Y., Jeong, Y., Roe, J., Kim, B. and Cho, B. (2015). Comparative Genomics Reveals the Core and Accessory Genomes of Streptomyces Species. *J. Microbiol. Biotechnol., 25*: 1599–1605

Kim, M., Oh, H., Park, S. and Chun, J. (2014). Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *International Journal of Systematic and Evolutionary Microbiology, 64*: 346–351

Kim, S.B., Lonsdale, J., Seong, C.N. and Goodfellow, M. (2003). Streptacidiphilus gen. nov., acidophilic actinomycetes with wall chemotype I and emendation of the family Streptomycetaceae (Waksman and Henrici 1943^{AL}) emend. Rainey *et al.* 1997. *Antonie van Leeuwenhoek 83*: 107–116.

Kirby, R., Sangal, V., Tucker, N.P., Zakrzewska-Czerwinska, J., Wierzbicka, K., Herron, P.R., Chu, C.-H., Chandra, G., Fahal, A.H., Goodfellow, M. and Hoskisson, P.A. (2012). Draft Genome Sequence of the Human Pathogen *Streptomyces somaliensis*, a Significant Cause of Actinomycetoma. *Journal of Bacteriology,13*: 3544-3545

Kitagawa, M., Ikeda, S., Tashiro, E., Soga, T., and Imoto, M. (2010). Metabolomic identification of the target of the filopodia protrusion inhibitor glucopiericidin *A. Chem. Biol., 17:* 989–998.

Konstantinidis K., and Tiedje, M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America, 102*: 2567–2572.

Korf, I. (2003). Serial BLAST searching. *BIOINFORMATICS, 19*: 1492–1496

Kroiss, J., Kaltenpoth, M., Schneider, B., Schwinger, M.G., Hertweck, C., Maddula, R.K., Strohm, E., and Svatos, A. (2010). Symbiotic Streptomycetes provide antibiotic combination prophylaxis for wasp offspring. *Nat. Chem. Biol., 6*: 261–263.

Kumar, A., Bisht, B.S. and Joshi, V.D. (2011). Bioremediation potential of three acclimated bacteria with reference to heavy metal removal from waste. *International Journal of Environmental Science, 2*: 896-908

Kumar, Y., Aiemsum-ang, P., Ward, A. and Goodfellow, M. (2007). Diversity and geographical distribution o fmembers of the *Streptomyces violaceusniger* 16S rRNAgene clade detected by clade-specific PCR primers. *Federation of European Microbiological Societies, 62*: 54-63

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L. (2004). Versatile and open software for comparing large genomes. *Genome Biology, 5*:R12

Labeda, D., Goodfellow, M., Brown, R., Ward, A., Lanoot, B., Vanncanneyt, M., Swings, J., Kim, S.-B., Liu, Z., Chun, J., Tamura, T., Oguchi, A., Kikuchi, T., Kikuchi, H., Nishii, T., Tsuji, K., Yamaguchi, Y., Tase, A., Takahashi, M., Sakane, T., Suzuki, K. and Hatano, K. (2012). Phylogenetic study of the species within the family Streptomycetaceae. *Antonie van Leeuwenhoek, 10*: 73-104

Labeda, D. and Kroppensted, M. (2000). Phylogenetic analysis of *Saccharothrix* and

related taxa: proposal for *Actinosynnemataceae* fam. nov. *International Journal of Systematic and Evolutionary Microbiology, 50*: 331–336

Labeda, D., Goodfellow, M., Brown, R., Ward, A., Lanoot, B., Vanncanneyt, M., Swings, J., Kim, S., Liu, Z., Chun, J., Tamura, T., Oguchi, A., Kikuchi, T., Kikuchi, H., Nishii, T., Tsuji, K., Yamaguchi, Y., Tase, A., Takahashi, M., Sakane, T., Suzuki, K.I., and Hatano, K. (2012). Phylogenetic study of the species within the family Streptomycetaceae. *Antonie van Leeuwenhoek, 101*:73–104

Lagesen, K., Hallin, P., Rødland, E., Staerfeldt, H., Rognes, T. and Ussery D. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res., 9*: 3100–3108.

Land, M., Hauser, L., Jun, S., *et al.* (2015). Insights from 20 years of bacterial genome sequencing. *Funct Integr Genomics, 15*:141–161

Lau, Y., Yin, W. and Chan, K. (2014). Enterobacter asburiae Strain L1: Complete Genome and Whole Genome Optical Mapping Analysis of a Quorum Sensing Bacterium. *Sensors, 14*: 13913-13924

Lechner, A., Alessandra, S., Gulder, T., Hafner, M. and , Bradley, M. .(2011). Selective Overproduction of the Proteasome Inhibitor Salinosporamide A via Precursor Pathway Regulation. *Chem Biol 18*: 1527–1536.

Lechevalier, H., Lechevalier, M. and Becker, B. (1966). Comparision of the chemical composition of cell-walls of Nocardiae with that of other aerobic actinomycetes. *International Nournal of Systematic Bacteriology, 16*: 151-160

Lee, H-J. and Whang, K-S. (2014). *Streptomyces graminifolii* sp. nov., isolated from bamboo (Sasa borealis) litter. International Journal of Systematic and Evolutionary Microbiology, (Under Press)

Letzel, A., Pidot, S. and Hertweck, C. (2014). Genome mining for ribosomally synthesized and post-translationally modified peptides (RiPPs) in anaerobic bacteria. *BMC Genomics, 15*: 1-21

Leung, E, Weil, D., Raviglione, M. and Nakatani, H. (2011). The WHO policy package to combat antimicrobial resistance. Bull World Health Organ, 89: 390–392

Lilla, E. and Yokoyama, K. (2016). Carbon extension in peptidylnucleoside biosynthesis by radical SASAM enzymes. Nature CHEMICAL BIOLOGY, 1-5.

Lin, Y.B., Wang, X.Y., Li, H.F., Wang, N.N., Wang, H.S., Tang, M. and Wei, G.-H. (2011). *Streptomyces* zinciresistens sp. nov., a zinc-resistant actinomycete isolated from soil

from a copper and zinc mine. International Journal of Systematic and Evolutionary Microbiology, 61: 616–620

Liu, G., Chater, K.F., Chandra, C., Niu, G. and Tan, H. (2013). Molecular Regulation of Antibiotic Biosynthesis in *Streptomyces*. *Microbiology and Molecular Biology Review, 77*: 112–143

Liu, Q., Yao, F., Chooi, Y. *et al.*, (2012). Elucidation of Piericidin A1 Biosynthetic Locus Revealed a Thioesterase-Dependent Mechanism of a-Pyridone Ring Formation. *Chemistry & Biology 19*, 243–253

Liu, Y., Wang, Y., Walsh, T. *et al.* (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

Lopez-Maury, L., Garcia-Dominguez, M., Florencio, F.J., Reyes, J.C. (2002). A two component signal transduction system involved in nickel sensing in the cyanobacterium Synechocystis sp. PCC 6803. *Molecular Microbiology, 43*: 247–256.

Lopez-Lara, I., Galvez, M. Martinez-Bueno, M., and Valdivia, E. (1991). Purification, characterization, and biological effects of a second bacteriocin from *Enterococcus faecalis ssp*. liquefaciens S-48 and its mutant strain B-48-28. *Can. J. Microbiol., 37*: 769 –774

Locci, R. (1994). Actinomycetes as plant pathogens. *European Journal of Plant Pathology, 100*: 179-200

Loytynoja, A. and Goldman, N. (2008). Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science, 320*: 1632-1635

Lucas, A. and Manna, A. (2013). Phenotypic characterization of sarR mutant in Staphylococcus aureus. *Microbial Pathogenesis, 57*: 52-61

Ma, Z., Jacobsen, F. and Giedroc, D. (2009). Coordination Chemistry of Bacterial Metal Transport and Sensing. *Chem. Rev., 109*, 4644 – 4681

Madduri, K., and Hutchinson, C. (1995). Functional characterization and transcriptional analysis of the dnrR1 locus, which controls daunorubicin biosynthesis in Streptomyces peucetius. *J Bacteriol, 177*: 1208–1215

Magarvey, N.A., Keller J.M., Bernan, V., Dworkin, M. and Sherman, D.H. (2004). Isolation and Characterization of Novel Marine-Derived Actinomycete Taxa Rich in Bioactive Metabolites. *Applied and Environmental Microbiology, 70*: 7520–7529

Mäkinen, V., Salmela, L., and Ylinen, J. (2012). Normalized N50 assembly metric using gap-restricted co-linear chaining. BMC Bioinformatics 13, 255.

Mara, K., Decorosi, F., Viti, C., *et al.* (2012). Molecular and phenotypic characterization of Acinetobacter strains able to degrade diesel fuel. *Research in Microbiology 163*, 161-172

Mardis, E. R. (2011). A decade's perspective on DNA sequencing technology.

Nature, 470: 198–203.

Margulies, M. *et al.* (2005). Genome sequencing in microfabricated highdensity picolitre reactors. Nature, 437: 376–380

Martin-Visscher, L., Yoganathan, S., Sit, C., Lohans, C. and Vederas, J. (2010). The activity of bacteriocins from *Carnobacterium maltaromaticum* UAL307 against Gram-negative bacteria in combination with EDTA treatment. *FEMS Microbiol Lett, 317*: 152–159

Mattioni, S., Develoux ,M., Brun, S., Martin, A., Jaureguy, F.,Naggara, N. and Bouchaud, O. (2013). Management of mycetomas in France. *Médecine et maladies infectieuses, 43*: 286–294

Mazur, A., Stasiak, G., Wielbo, A. *et al.* (2013). Phenotype profiling of Rhizobium leguminosarum bv. trifolii clover nodule isolates reveal their both versatile and specialized metabolic capabilities. *Arch Microbiol, 195*:255–267

Medema, M., Blin, K., Cimermancic, P. *et al.* (2011). antiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters. *Nucleic Acids Research, 39*: W339-W346

Medema, M, Claesen, J. Kurita, K. *et al.*, (2014). Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell, 158*: 412–421

Meier, J. and Burkart, M. (2009). The chemical biology of modular biosynthetic enzymes. *Chem. Soc. Rev. 38*: 2012–2045

Meier-Kolthoff, J., Auch, A., Klenk, P. and Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics 14*: 1-14

Metsa-Ketela, M., Halo, L,, Munukka, E., Hakala, J., Mantsala, P., Ylihonko, K. (2002). Molecular evolution of aromatic polyketides and comparative sequence analysis of polyketide ketosynthase and 16S ribosomal DNA genes from various streptomyces species. *Appl Environ Microbiol, 68*:4472–4479

Miao, M., Brost, R., Chapple, J., She, K., and Baltz, R. (2006). The lipopeptide antibiotic A54145 biosynthetic gene cluster from Streptomyces fradiae. *J Ind Microbiol Biotechnol, 33*: 129–140

Miller-Wideman, M., Makker, N., Tran, N., Isaak, B. and Stonard, R. (1992). HERBOXIDIENE, A NEW HERBICIDAL SUBSTANCE FROM Streptomyces chromofuscus A7847. *THE JOURNAL OF ANTIBIOTICS, 6*:914-921

Mizrahi-Man, O., Davenport, E.R., and Gilad, Y. (2013). Taxonomic Classification of Bacterial 16S rRNA Genes Using Short Sequencing Reads: Evaluation of Effective Study Designs. *PLoS ONE, 8*: 1-14

Molinari, G. (2007). Natural products in drug discovery: present status and perspectives. *In "Pharmaceutical Biotechnology"*. Landes Bioscience, Austin, USA.

Moore, J.M., Bradshaw, E., Seipke, R.F., Hutchings, M.I. and McArthur, M. (2012). Use and Discovery of Chemical Elicitors That Stimulate Biosynthetic Gene Clusters in Streptomyces Bacteria. Methods in Enzymology, 517: 368 - 383

Moore, J.M., Bradshaw, E., Seipke, R.F., Hutchings, M.I. and McArthur, M. (2012). Use and Discovery of Chemical Elicitors That Stimulate Biosynthetic Gene Clusters in *Streptomyces* Bacteria. *Methods in Enzymology, 517*: 368 - 383

Moraga, N.B, Poma, H.R., Amoroso, M.J.and Rajal1, V.B. (2013). Isolation and characterization of indigenous *Streptomyces* and Lentzea strains from soils containing boron compounds in Argentina. *Journal of Basic Microbiology, 00*: 1-10

Mukherjee, A., Mammel, M., LeClerc, J. and Cebula, T. (2008). Altered Utilization of N-Acetyl-D-Galactosamine by Escherichia coli O157:H7 from the 2006 Spinach Outbreak. *JOURNAL OF BACTERIOLOGY, 5*: 1710–1717

Munoz-Price, L. S., Poirel, L., Bonomo, R. A. *et al.* (2013). Clinical epidemiology of the global expansion of Klebsiella pneumoniae carbapenemases. Lancet Infect, 13: 785–96

Munshi, A., Shafi, G., Aliya, N., & Jyothy, A. (2009). Histone modifications dictate specific biological readouts. *Journal of Genetics and Genomics, 36*: 75–88.

Murphy, K., O'Sullivan, O., Rea, M., Cotter, P., Ross, R., *et al.* (2011). Genome Mining for Radical SAM Protein Determinants Reveals Multiple Sactibiotic-LikeGene Clusters. *PLoS ONE 6(7)*: 1-12

Naik, M.M. and Dubey, S.K. (2013). Lead resistant bacteria: Lead resistance mechanisms, their applications in lead bioremediation and biomonitoring. *Ecotoxicology and Environmental Safety, 9*: 1–7

Nakashima, T., Miyano, R.**,** Iwatsuki, M. *et al.* (2016). Iminimycin A, the new iminium metabolite produced by Streptomyces griseus OS-3601. The Journal of Antibiotics 69, 611-615

Nakatsuji, T., Kao, M., Fang, J. *et al.* (2009). Antimicrobial Property of Lauric Acid Against Propionibacterium acnes: Its Therapeutic Potential for Inflammatory Acne Vulgaris. *J Invest Dermatol, 129*: 2480–2488

Newberry, K. and Brennan, R. (2004). The Structural Mechanism for Transcription Activation by MerR Family Member Multidrug Transporter Activation, N Terminus. THE JOURNAL OF BIOLOGICAL CHEMISTRY, 279: 20356–20362

Nies, D.H. (1999). Microbial heavy-metal resistance. *Applied Microbiology and Biotechnology, 51*: 730–750.

Nies, D. H. (2003). E¥ux-mediated heavy metal resistance in prokaryotes. *FEMS Microbiology Reviews, 27*: 313-339

Nett, M., Ikeda, H., and Moore, B. S. (2009). Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Natural Product Reports, 26*: 1362–1384

Newman, D.J., Cragg, G.M., and Snader, K.M. (2003). Natural products as sources of new drugs over the period 1981–2002. *J. Nat. Prod. 66*: 1022-1037.

Noguchi, Y., Yamada, T., Uchiro, H. and Kobayashi, S. (2000). Synthetic study of polyoxypeptin: stereoselective synthesis of the acyl side-chain segment. *Tetrahedron Letters, 41*: 7499–7502

Noomnual, S., Thasana, N., Sungkeeree, P. *et al.* (2016). Streptanoate, a new anticancer butanoate from Streptomyces sp. DC3. *The Journal of Antibiotics, 69*: 124-127

Nurk, P., Bankevich, A., Gurevich, A., *et al.* (2013): Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads. *Research in Computational Molecular Biology, 7821*: 158-170

Okamoto, S., Lezhava, A., Hosaka, T., Okamoto-Hosoya, Y. and Ochi, A. (2003). Enhanced Expression of S-Adenosylmethionine Synthetase Causes Overproduction of Actinorhodin in Streptomyces coelicolor A3(2). *JOURNAL OF BACTERIOLOGY, 185*: 601–609

Okamoto, S., Taguchi, T., Ochi, K., and Ichinose, K. (2009). Biosynthesis of Actinorhodin and Related Antibiotics: Discovery of Alternative Routes for Quinone Formation Encoded in the act Gene Cluster. *Chemistry and Biology, 16*: 226–236

Okkeri, J. and Haltia, T. (2006). The metal-binding sites of the zinc-transporting P-type ATPase of Escherichia coli. Lys693 and Asp714 in the seventh and eighth transmembrane segments of ZntA contribute to the coupling of metal binding and ATPase activity. *Biochimica et Biophysica Acta, 1757*: 1485–1495

Olano, C., García, I., González, A. *et al.*, (2014). Activation and identification of five clusters for secondary metabolites in Streptomyces albus J1074. *Microbial Biotechnology, 3*: 242–256

Onaka, H., Mori, Y., Igarashi, Y. and Furumai, T. (2011). Mycolic Acid-Containing Bacteria Induce Natural-Product Biosynthesis in *Streptomyces* Speciesnd. *Applied and Environmental Microbilogy, 2*: 400 - 406

Orro, A., Cappelletti, M., D'Ursi, P. et at. (2015). Genome and Phenotype Microarray Analyses of Rhodococcus sp. BCP1 and Rhodococcus opacus R7: Genetic Determinants and Metabolic Abilities with Environmental Relevance. *PLoS ONE, 10*: 1-41

Ostash, I., Ostash, B., Luzhetskyy, A., Bechthold, A., Walker, S. and Fedorenko V. (2008). Coordination of export and glycosylation of landomycins in Streptomyces cyanogenus S136. *FEMS Microbiol. Lett., 285*:195–202.

Overbeek, R., Olson, R., Pusch, G. *et al.*, (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research,42*: D206–D214

Overhage, J., Bains, M., Brazas, M.D. and Hancock, R.E.W. (2008). Swarming of Pseudomonas aeruginosa Is a Complex Adaptation Leading to Increased Production of Virulence Factors and Antibiotic Resistance. *Journal of Bacteriology, 8*: 2671-2679

Pal, C., Bengtsson-Palme, J., Rensing, C., Kristiansson, E. and Larsson, D. (2014). BacMet: antibacterial biocide and metal resistance genes database. *Nucleic Acids Research, 42*: D737–D743

Park, M. and Rafii, F. (2014). Global Phenotypic Characterization of Effects of Fluoroquinolone Resistance Selection on the Metabolic Activities and Drug Susceptibilities of Clostridium perfringens Strains. *International Journal of Microbiology,* 1-9

Parte, A. (2013). LPSN—list of prokaryotic names with standing in nomenclature. *Nucleic Acids Research, 42*: D613–D616

Passari, A., Chandra, P., Leo, V., *et al.* (2017). Production of Potent Antimicrobial Compounds from *Streptomyces cyaneofuscatus* Associated with Fresh Water Sediment. Frontiers in Microbiology, 1: 1-13

Pastor, M., Salvador, M., Argandona, M., Bernal, V., Reina-Bueno, M., Csonka, L.N., Iborra, J.L., Vargas, C., Nieto, J.J. and Canovas, M .(2010). Ectoines in cell stress protection: uses and biotechnological production. *Biotechnol Adv, 28*:782–801

Patel, J. B. 2001. 16S rRNA gene sequencing for bacterial pathogen identification in the clinical laboratory. Journal of Molecular Diagnosis, 6:313–321**.**

Pendleton, J., Gorman, S. and Gilmore, B. (2013): Clinical relevance of the ESKAPE pathogens. *Expert Rev. Anti I 308 nfect. Ther,* (11): 297-308

Penn, O., Privman, E., Ashkenazy, H., Landan, G., Grauer, D. and Pupko, T. (2010). GUIDANCE: A web server for assessing alignment confidence score. Nucleic Acids Research, 38: W23-W28

Phongsopitanun, W., Thawai, C., Suwanborirux, K., Kudo, T., Ohkuma, M. and Tanasupawat, S. (2014). *Streptomyces* chumphonensis sp. nov., isolated from marine sediments in Thailand. International Journal of Systematic and Evolutionary Microbiology (Under Press).

Pei, A., Nossa, C., Chokshi, P., Blaser, M., Yang, L., Rosmarin, D. and Pei, Z. (2009). Diversity of 23S rRNA Genes within Individual Prokaryotic Genomes. *PLoS ONE 4(5)*: 1-9

Piette, A., Derouaux, A., Gerkens, P. *et al.*, (2005). From dormant to germinating spores of Streptomyces coelicolor A3(2): New perspectives from the crp null mutant. *Journal of Proteome Research*, 4: 1699–1708.

Piel, J., Hofer, I., and Hui, D. (2004). Evidence for a Symbiosis Island Involved in Horizontal Acquisition of Pederin Biosynthetic Capabilities by the Bacterial Symbiont of Paederus fuscipes Beetles. *JOURNAL OF BACTERIOLOGY, 186* :1280–1286

Pleg, A.Y., Seifert, H. and Paterson, D.L. (2008). *Acinetobacter baumannii*: Emergence of a Successful Pathogen. *Clinical Microbiology Reviews, 21*: 538-582

Plonka, P. M. and Grabacka, M. (2006). Melanin synthesis in microorganisms-biotechnological and medical aspects. *Acta Biochimica Polanica, 53*: 429 - 443

Pollak, F. C. and Berger, R. G.(1996). Geosmin and related volatiles in bioreactor-cultured Streptomyces citreus CBS 109.60. *Appl Environ Microbiol, 62*: 1295–1299

Quast, C., Pruesse, E., Yilmaz, P. *et al.*, (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research, 41*: D590–D596

Queener, S.W and Day, L.E. (1986). *Antibiotic-Producing Streptomyces*. Orlando Academic Press Inc., 1-25.

Quintana, E.T., Wierzbicka, K., Mackiewicz, P., Osman, A., Fahal, A.H., Hamid, M.E., Zakrzewska-Czerwinska, J., Maldonado, L.A. and Goodfellow M. (2008). *Streptomyces* sudanensis sp. nov., a new pathogen isolated from patients with actinomycetoma. *Antonie Van Leeuwenhoek, 93*: 305–313

Qin, Z., Munnoch J., Devine, R., Holmes, N., Seipke, R., Wilkonson, K., Wilkonson, B. and Hutchings, S. Formicamycins, antibacterial polyketides produced by

Streptomyces formicae isolated from African Tetraponera plant-ants. *Chem. Sci., 8*: 3218–3227

Qin, X., Ren, L., Yang, X., Bai, F., Wang, L., Geng, P., Bai, G. and Shen, Y. (2011). Structures of human pancreatic α-amylase in complex with acarviostatins: Implications for drug design against type II diabetes. *Journal of Structural Biology, 174*:196–202

Ramos, J., Martinez-Bueno, M., Molina-Henares, A. *et al.* (2005). The TetR Family of Transcriptional Repressors. *MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS, 69*: 326–356

Rateb, M.E., Houssen, W.Y., Harrison,W., Deng,H., Okoro,C.K., Asenjo, J.A., Andrews,B.A., Bull, A.T., Goodfellow,M., Ebel, R. and Jaspars, M. (2011). Diverse Metabolic Profiles of a *Streptomyces* Strain Isolated from a Hyper-arid Environment. Journal of Natural Products, 74:1965–1971

Ravel, J., Schrempf, H., Hill,R.T. (1998) .Mercury resistance is encoded by transferable giant linear plasmids in two Chesapeake Bay *Streptomyces* strains. *Applied and Environmental Microbiology 64*:3383–3388.

Rea, M., Sit, C., Clayton, E. *et al.* (2010). Thuricin CD, a posttranslationally modified bacteriocin with a narrow spectrum of activity against Clostridium difficile. *PNAS, 107*: 1-6

Reddy, V., Shlykov, M., Castillo, R. *et al.* (2012). The Major Facilitator Superfamily (MFS) Revisited. *FEBS, 279*: 2022–2035

Rensing, C., Ghosh, M., and Rosen, B.P. (1999). Families of softmetal- ion-transporting ATPases. *J. Bacteriol, 181* :5891–5897

Rice, L. B. (2008). Federal funding for the study of antimicrobial resistance

in nosocomial pathogens: no ESKAPE. J Infect Dis 197, 1079–1081.

Richter, M. and Rossello-Mora. (2009). Shifting the genomic gold standard for theprokaryotic species definition. *PNAS, 45*: 19126–19131

Rigali, S., Titgemeyer, F., Barends, S., Mulder, S., Andreas W. Thomae, Hopwood, D.A. and van Wezel, G.P. (2008). Feast or famine: the global regulator DasR links nutrient stress to antibiotic production by *Streptomyces*. European *Molecular Biology Organization Reports, 9*: 670-675

Ringnér, M. (2008). What is principal component analysis?. *Nat Biotechnol, 26*: 303–304

Robinson, D., McCarthy, J., and Smyth,K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics 26*: 139–140.

Rocha, E. and Danchin, A. (2002). Base composition bias might result from competition for metabolic resources. *TRENDS in Genetics 6*: 291-294

Rodriguez, L., Grajales, A., Arrieta-Ortiz, M., Salazar, C., Restrepo, S. and Berna, A. (2012). Genomes-based phylogeny of the genus Xanthomonas. *BMC Microbiology, 12*: 1-14

Rong, X., Gue, Y. and Huang, Y. (2010). Proposal toreclassifythe *Streptomyces albidoflavus* clade on the basisof multilocus sequenceanalysis and DNA DNA hybridization, and taxonomic elucidationof *Streptomyces griseus* subsp. solvifaciens. *Systematic and Applied Microbiology*, 32: 314–322

Rong, X. and Huang, Y. (2012). Taxonomic evaluation of the *Streptomyces* hygroscopicus clade using multilocus sequence analysis and DNA–DNA hybridization, validating the MLSA scheme for systematics of the whole genus. *International Journal of Systematic and Evolutionary Microbiology, 35*: 7–18

Rong, X. and Huang, Y. (2010). Taxonomic evaluation of the *Streptomyces* griseus clade using multilocus sequence analysis and DNA–DNA hybridization, with proposal to combine 29 species and three subspecies as 11 genomic species. *International Journal of Systematic and Evolutionary Microbiology, 60*: 696–703

Rothberg, J.M., Hinz, W., Rearick, T.M., Schultz. (2011). "An integrated semiconductor device enabling non-optical genome sequencing", *Nature, 475*: 348-352.

Rouse, M., Rotger, M., Piper, K. *et al.* (2005). In Vitro and In Vivo Evaluations of the Activities of Lauric Acid Monoester Formulations against Staphylococcus aureus. *ANTIMICROBIAL AGENTS AND CHEMOTHERAPY, 49*: 3187-3191

Rusk, N. (2011). Torrents of sequence. Nature Method 8: 44-44

Rutherford, J., Cavet, J., and Robinson, N. (1999). Cobaltdependent transcriptional switching by a dual-effector MerR-like protein regulates a cobalt-exporting variant CPx-type ATPase. *J. Biol. Chem., 274*: 25827–25832.

Ryu, Y-G., Butler, M. J., Chater, K. F. and Lee, K. J. (2006). Engineering of Primary Carbohydrate Metabolism for Increased Production of Actinorhodin in *Streptomyces* coelicolor. *Applied and Environmental Microbiology, 72*: 7132–7139

Sacchi, C., Whitney, A., Mayer, L., Morey, R., Steigerwalt, A., Boras, A., Weyant, R. and Popovic, T. (2002). Sequencing of 16S rRNA gene: a rapid tool for identification of Bacillus anthracis. *Emerging Infectious Diseases, 8*:1117–1123

Saitou, N. and Nie, M. (1987). The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Journal of Molecular Biology and Evolution, 4*: 406-425.

Sanchini, A., Dematheis, F., Semmler, T. and Lewin, A. (2017). Metabolic phenotype of clinical and environmental Mycobacterium avium subsp. hominissuis isolates. *Peerj,* 1-16

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America, 74*: 5463–5467.

Sanglier, J.J., Haag, H., Huck, T.A., and Fehr, T. (1993). "Novel bioactive compounds from actinomycetes: a short review (1988–1992)" . *Research in Microbiology, 144*: 633–642

Saribas, S. and Bagdatli, Y. (2004). Vancomycin Tolerance in Enterococci. *Chemotherapy, 50*(11): 250-254

Sawa, N., Zendo, T., Kiyofuji, J. *et al.* (2009). Identification and Characterization of Lactocyclicin Q, a Novel Cyclic Bacteriocin Produced by Lactococcus sp. Strain QU 12. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY, 75*: 1552–1558

Scaria, J., Suzuki, H., Ptak, C. (2015). Comparative genomic and phenomic analysis of Clostridium difficile and Clostridium sordellii, two related pathogens with differing host tissue preference. *BMC Genomics, 16*:1-16

Schmidt, A., Haferburg,G., Sineriz,M., Merten, D., Buchel, G., Kothe, E. (2005). Heavy metal resistance mechanisms in actinobacteria or survival in AMD contaminated soils. *Geochemistry, 65*: 131–144.

Shelburne, C., An, F., Dholpe, V., Ramamoorthy, A., Lopatin, D. and Lantz, M. (2007). The spectrum of antimicrobial activity of the bacteriocin subtilosin A. *Journal of Antimicrobial Chemotherapy, 59*: 297–300

Shephered, M.D, Kharel, M.K, Bosserman, M.A. and Rohr, J. (2010). Laboratory Maintenance of *Streptomyces* species. *Current Protocols in Microbiology, 10*: 1-10

Shirling, E.B. and Gottlieb, D. (1966). Methods for characterization of *Streptomyces species*. *International Journal of Systematic Bacteriology, 16*: 313-340.

Shirling, E.B. and Gottlieb, D. (1972). Cooperative Description of Type Strains of *Streptomyces*. *International Journal of Systematic Bacteriology, 22*: 265-394.

Schütze, E., Kothe, E. (2012) Bio-Geo Interactions in Metal-Contaminated Soils. In: Kothe, E., Varma, A. (Eds.), *Soil Biology 31*. Springer-Verlag, Berlin Heidelberg, pp: 163–182

Seipke, R. (2015). Strain-Level Diversity of Secondary Metabolism in *Streptomyces albus*. *PLOS ONE,* 1-14

Sekurova, O., Brautaset, T., Sletta, H. *et al.* (2004). In vivo analysis of the regulatory genes in the nystatin biosynthetic gene cluster of Streptomyces noursei ATCC 11455 reveals their differential control over antibiotic biosynthesis. *J. Bacteriol., 186*: 1345–1354

Seyedsayamdost, M.R. (2014). High-throughput platform for the discovery of elicitors of silent bacterial gene clusters. *PNAS, 1*: 1-6

Seyedsayamdost, M.R., Traxler, M.F., Clardy, J. and Kolter, R. (2012). Old Meets New: Using Interspecies Interactions to Detect Secondary Metabolite Production in Actinomycetes. *Methods in Enzymology, 517*: 89-107

Shao, L., Zi, J., Zeng, J. and Zhan, J. (2012). Identification of the Herboxidiene Biosynthetic Gene Cluster in Streptomyces chromofuscus ATCC 49982. *Applied and Environmental Microbiology, 1*: 2034–2038

Shea, A., Wolcott, M, Daefler, S. and Rozak, D. (2012). Biolog Phenotype Microarrays. *Methods in Molecular Biology, 881*: 331-373

Sims, D. Sudbery, I, Ilott, N., Heger, A. and Ponting, C. (2014). Sequencing depth and coverage: key considerations in genomic analyses. NATURE REVIEWS, 15: 121-132

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, 1*: 1-18

Skinnider, M., Johnston, C., Edgar, R. *et al.* (2016). Genomic charting of ribosomally synthesized natural product chemical space facilitates targeted mining. *PNAS, 3*: E6343–E6351

Skrivanova, E., Marounek, M., Dlouha., G. and Kanak, J. (2005). Susceptibility of Clostridium perfringens to C2–C18 fatty acids. *Letters in Applied Microbiology, 41*: 77–81

Slijkhuis, H., John, J. and Kylstra, D. (1984). Microthrix purvicella, a Filamentous Bacterium from Activated Sludge: Growth on Tween 80 as Carbon and Energy Source. *Journal of General Microbiology, 130*: 2035-2042.

Smaoui, S., Elleuch, L., Bejar, W. *et al.* (2010). Inhibition of Fungi and Gram-Negative Bacteria by Bacteriocin BacTN635 Produced by Lactobacillus plantarum sp. TN635. *Appl Biochem Biotechnol, 162* :1132–1146

Sofia, H., Chen, G., Hetzler, B., Reyes-Spindola, J., and Miller, N (2001). Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods. *Nucleic Acids Res., 29*:1097–1106

Soliev, A.B., Hosokawa, K. and Enomoto, K. (2011). Bioactive Pigments fromMarine Bacteria: Applications and Physiological Roles. *Journal of Hindawi Publishing Corporation, 2011: 1-17*

Stackebrandt, E. (2011). Molecular taxonomic parameters. *In focus*, 5: 59-61

Stackebrandt, E. and Ebers, J. (2006). Taxonomic parameters revisited:tarnished gold standards. *Microbiol Today 33*, 152–155.

Stackebrandt, E. and Goebel, B. M. (1994). Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int J Syst Bacteriol 44*, 846-849.

Stackebrandt, E., Rainey, F.A., Ward-Raine, N.L. (1997). Proposal for a new hierarchic classification system, Actinobacteria classis nov. *International Journal of Systematic Bacteriology, 47*: 479–491.

Stevens, K., Sheldon, B., Klapes, N. and Lleanhammer, T. (1991). Nisin Treatment for Inactivation of Salmonella Species and Other Gram-Negative Bacteria. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY, 57*: 3613-3615

Streshinskaya, G., Shashkov, A., Usov, A., Evtushenko, L. and Naumova, I. (2002). Cell Wall Teichoic Acids of Actinomycetes of Three Genera of the Order Actinomycetales. *Biochemistry (Moscow), 67*: 778-785

Świątek, M.A., Urem, M., Tenconi, E., Rigali, S. and van Wezel1, G.P. (2012). Engineering of N-acetylglucosamine metabolism for improved antibiotic production in *Streptomyces coelicolor* A3(2) and an unsuspected role of NagA in glucosamine metabolism. *Bioengineered, 3*: 280–285

Takizawa, M., Colwell R.R. and Hill R.T. (1993). Isolation and diversity of actinomycetes in the Chesapeake Bay. *Applied and Environmental Microbiology 59*: 997-1002

Taghavi, S., Lesaulnier, C., Monchy, S., Wattiez, R., Mergeay, M. & van der Lelie, D. (2009). Lead(II) resistance in Cupriavidus metallidurans CH34: interplay between plasmid and chromosomally- located functions. *Antonie van Leeuwenhoek, 96*: 171–182

Tamburini, E. and Mastromei, G. (2000). Do bacterial cryptic genes really exist?. *Research in Microbiology, 151*: 179–182

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution, 30*: 2725-2729.

Tanaka, Y., Kasahara, K., Hirose, Y., Murakami, K., Kugimiya, R. and Ochi, K. (2013). Activation and Products of the Cryptic Secondary Metabolite Biosynthetic Gene Clusters by Rifampin Resistance (rpoB) Mutations in Actinomycetes. *Journal of Bacteriology, 13*: 2959 –2970

Tang, Z., Xiao, C., Zhuang, Y., Chu, J., Zhang, S., Herron, P.R., Hunter, I.S and Guo, M. (2011). Improved oxytetracycline production in *Streptomyces rimosus* M4018 by metabolic engineering of the G6PDH gene in the pentose phosphate pathway. *Enzyme and Microbial Technology, 49*: 17–24

Tettelin, H.,Riley,D.,Cattuto,C.,andMedini,D.(2008).Comparative genomics:thebacterialpan-genome. *Curr. Opin.Microbiol. 11,* 472–477.

Thaw, P., Sedelnikova, S., Muranova, T. *et al.* (2006). Structural insight into gene transcriptional regulation and effector binding by the Lrp/AsnC family. *Nucleic Acids Research, 34*: 1439–1449

Theodoulou, F. and Kerr, I. (2015). ABC transporter research: going strong 40 years on. *Biochem. Soc. Trans., 43*: 1033–1040

Thibodeaux, C.J., Melancon, C.E. and Liu, H.W. (2008). Natural-product sugar biosynthesis and enzymatic glycodiversification. *Angew. Chem. Int. Ed., 47*: 9814–9859.

Thompson, C., Silva, G., Vieira, N., *et al.* (2013). Genomic Taxonomy of the Genus Prochlorococcus. *Microbial Ecology, 66*: 1-11

Thompson, J., D., Plewniak, F. and Poch, O. (1999). Acomprehensive comparison of multiple sequence alignment programs. *Nucleic Acids Research, 27*:2682-2690

Tindall, B. J., Rossello´ -Mo´ ra, R., Busse, H. J., Ludwig, W. & Ka¨ mpfer, P. (2010). Notes on the characterization of prokaryote strains for taxonomic purposes. *Int J Syst Evol Microbiol, 60:* 249–266.

Trapnell, C., Pachter, L., Salzberg, S. (2009). TopHat: discovering splice junctions with

RNA-Seq. Bioinformatics, 9: 1105–1111

Trapnell, C., Roberts, A., Goff, L. *et al.* (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols, 3*: 562-578

Turcatti, G., Romieu, A., Fedurco, M. & Tairi, A. (2008). "A new class of cleavable fluorescent nucleotides: Synthesis and optimization as reversible terminators for DNA sequencing by synthesis", Nucleic acids research, : 36, no. 4.

Ueberschaar, N., Ndejouong, B.S., Ding, L., Maier, A., Fiebig, H.H. and Hertweck, C. (2011). Hydrazidomycins, cytotoxic alkylhydrazides from *Streptomyces* atratus. *Bioorganic & Medicinal Chemistry Letters . 19*: 5839–5841.

Umezawa, K., Nakazawa, K., Uemura, T. and *et al.* (1998). Polyoxypeptin Isolated from Streptomyces: A Bioactive Cyclic DepsipeptideContaining the Novel Amino Acid 3-Hydroxy-3-methylproline. *Tetrahedron Letters, 39*: 1389-1392

Valouev, A., Ichikawa, J., Tonthat, T., Stuart, J., Ranade, S., Peckham, H., Zeng, K., Malek, J.A., Costa, G., McKernan, K., Sidow, A., Fire, A. & Johnson, S.M. (2008). A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence dictated positioning. *Genome Research, 18*:1051-1063.

van Wezel, G.P., Krabben, P., Traag, B.A., Keijser, B.J.F, Kerste, R., Vijgenboom, E., Heijnen, J.J. and Kraal, B. (2006). Unlocking *Streptomyces* spp. for Use as Sustainable Industrial Production Platforms by Morphological Engineering. *Applied and Environmental Microbiology, 72*: 5283–5288

Varghese, N., Mukherjee, S., *et al.* (2015). Microbial species delineation using whole genome sequences. *Nucleic Acids Research, 43*:6761–6771

Waksman, S., Henrici, A. (1943). The nomenclature and classification of the actinomycetes. *Journal of Bacteriology*, *46*: 337–341.

Wang, L., Nie, J., Sicotte, H. *et al.* (2016). Measure transcript integrity using RNA-seq data. *BMC Bioinformatics, 17*:1-16

Wattam, A., Abraham, A., Dalay, O. *et al.*, (2014). "PATRIC, the bacterial bioinformatics database and analysis resource*." Nucl Acids Res, 42 (D1)*: D581-D591.

Wayne, G., Brenner, J., Colwell, R., Grimont, P. A. D., Kandler, O., Krichevsky, M. I., Moore, H., Moore, C., Murray, E. *et al.*, .(1987). Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int J Syst Bacteriol 37*, 463–464.

Weber, T., Blin, K., Duddela, S., *et al.*, (2015). antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research, 5*: 1-7

Welsh, O., Vera-Cabrera, L., Welsh., E. and Salinas, M.C. (2012). Actinomycetoma and advances in its treatment. *Clinics in Dermatology, 30*: 372–381

WHO. (2017). GLOBAL PRIORITY LIST OF ANTIBIOTIC-RESISTANT BACTERIA TO GUIDE RESEARCH, DISCOVERY, AND DEVELOPMENT OF NEW ANTIBIOTICS

WHO. (2001). Global Strategy for Containment of Antimicrobial Resistance

WHO.(2000). Model Prescribing Information: Drugs used in Bacterial Infections.

Wilson, D., Xue, Y., Reynolds, K. and Sherman, D. (2001). Characterization and analysis of the PikD regulatory factor in the pikromycin biosynthetic pathway of Streptomyces venezuelae. *J. Bacteriol. 183*: 3468 –3475

Viti, C., Decorosi, F., Tatti, E., Giovannetti, L. (2007). Characterization of chromate-resistant and-reducing bacteria by traditional means and by a high-throughput phenomic technique for bioremediation purposes. *Biotechnol, 23*: 553-559

Woese, C., Stackebrandt, E., Macke, T. and Fox, G. (1985). A phylogenetic definition of the major eubacterial taxa. *Systemic and Applied Microbiology, 6*:143–151

Woodford, N. and Livermore, D.M. (2009). Infections caused by Gram‐positive bacteria: a review of the global challenge. *The Journal of Infection, 59*: 4‐16.

Yadav, A.K., Kumar, R., Saikia, R., Bora, T.C. and Arora, D.K. (2009). Novel copper resistant and antimicrobial *Streptomyces* isolated from Bay of Bengal, India. *Journal de Mycologie Médicale, 19*: 234-240

Yamamoto, K. and Ishihama, A. (2003). Two different modes of transcription repression of the Escherichia coli acetate operon by IclR. *Mol. Microbiol., 47*: 183–194

Yan, Q., Power, K., Cooney, S. *et al.* (2013). Complete genome sequence and phenotype microarray analysis of *Cronobacter sakazakii* SP291: a persistent isolate cultured from apowdered infant formula production facility. *Frontiers in microbiology, 4*:1-20.

Yang, J., Tauschek, M. and Robins-Browne, R. (2011). Control of bacterial virulence by AraC-like regulators that respond to chemical signals. *Trends in Microbiology, 19*: 128-135

Zabala,D., Braña, A. F., Flórez, A.B., Salas, J.A. and Méndez, C. (2013). Engineering precursor rmetabolite pools for increasing production of antitumor mithramycinsin *Streptomyces argillaceus*. *Metabolic Engineering*, *20*: 187–197

Zankari, E., Hasman, H., Cosentino, S. *et al.* (2012). Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother, 67*: 2640-2644

Zhang, X., Clark, C. and Pettis, G. (2003). Interstrain inhibition in the sweet potato pathogen Streptomyces ipomoeae: purification and characterization of a highly specific bacteriocin and cloning of its structural gene. *Appl Environ Microbiol, 69*: 2201–2208

Zheng, G., Hehn, R., Zuber, P. (2000). Mutational analysis of the sbo-alb locus of Bacillus subtilis: Identification of genes required for subtilosin production and immunity. *Journal of Bacteriology, 182*: 3266–3273

Zheng, Q., Wang, Q., Wang, S., Wu, J., Gao, Q., and Liu, W. (2015). Thiopeptide antibiotics exhibit a dual mode of action against intracellular pathogens by affecting both host and microbe. *Chem. Biol., 22*: 1002–1007**.**

Zhou, Y., Liang, Y., Lynch, K. *et al.* (2011). PHAST: A Fast Phage Search Tool. *Nucleic Acids Research, 39*: W347–W352

Zhu, H.H., Guo, J., Yao, Q., Yang, S.Z., Deng, M.R., Phuong le, T.B., Hanh, V.T. and Ryan M.J. (2007). *Streptomyces* vietnamensis sp. nov., a streptomycete with violet–

blue diffusible pigment isolated from soil in Vietnam. *International Journal of Systematic and Evolutionary Microbiology, 57*: 1770–1774

Ziemert, N., Alanjary, M and Weber, T. (2016). The evolution of genome mining in microbes – a review. *Nat. Prod. Rep., 33*, 988

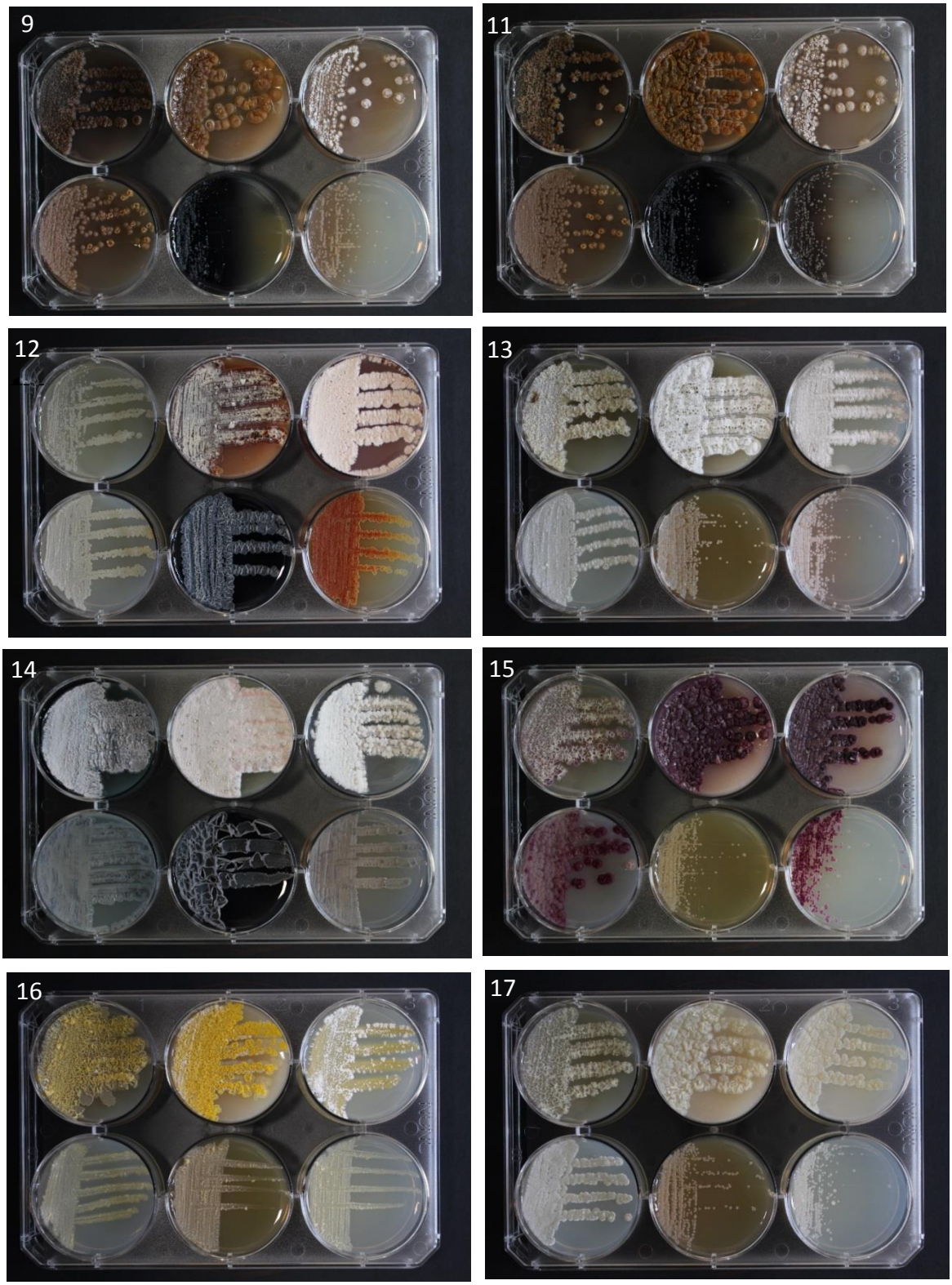
Ziemert, N. and Jensen P. (2012). Phylogenetic Approaches to Natural Product Structure Prediction. *Methods Enzymol, 517*: 161–182.

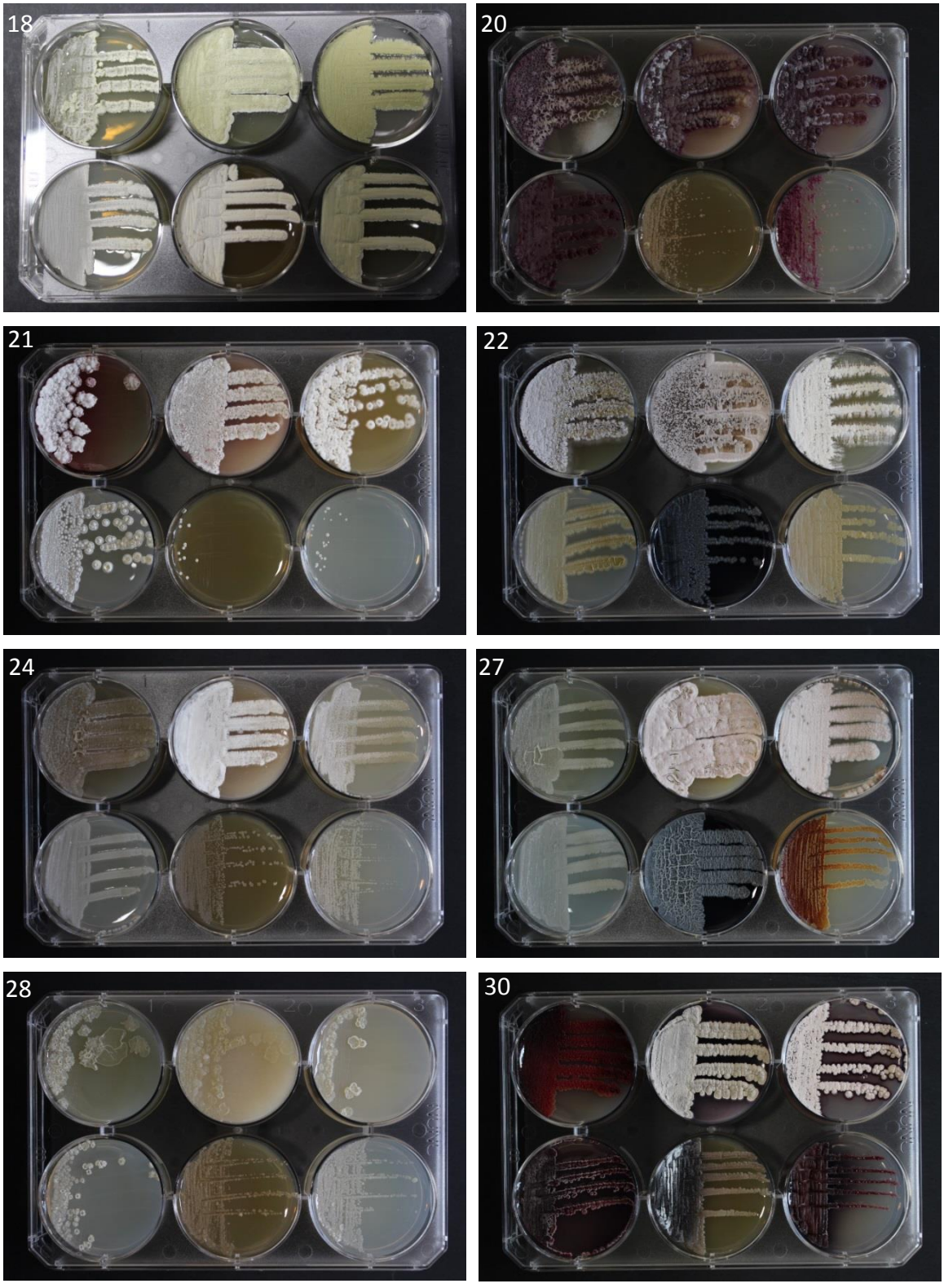
Ziemert, N., Podel, S., Penn, K., Badger, J., Allen, E. and Jensen, P. (2012). The Natural Product Domain Seeker NaPDoS: A Phylogeny Based Bioinformatic Tool to Classify Secondary Metabolite Gene Diversity. *PLoS ONE 7(3*): 1-9

Zingue, D., Bouam, A., Militello, M. and Drancourt, M. (2017). High-Throughput Carbon Substrate Profiling of *Mycobacterium ulcerans* Suggests Potential Environmental Reservoirs. *PLOS Neglected Tropical Diseases, 1*, 1-12

Appendix A

Morphology and growth characteristics of the strains isolated in this study on the different ISPs (ISP2-ISP7) media.




18
20
21
22
24
27
28
30

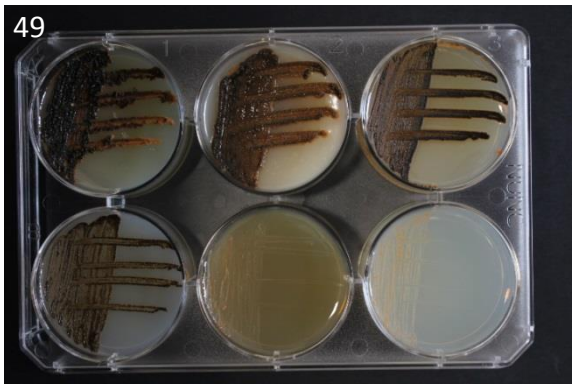

31
32
33
36
37
38
39
40


41


42


43


47


49


50


51


52


53


54


55


56


57


58


59


60


63


64


69

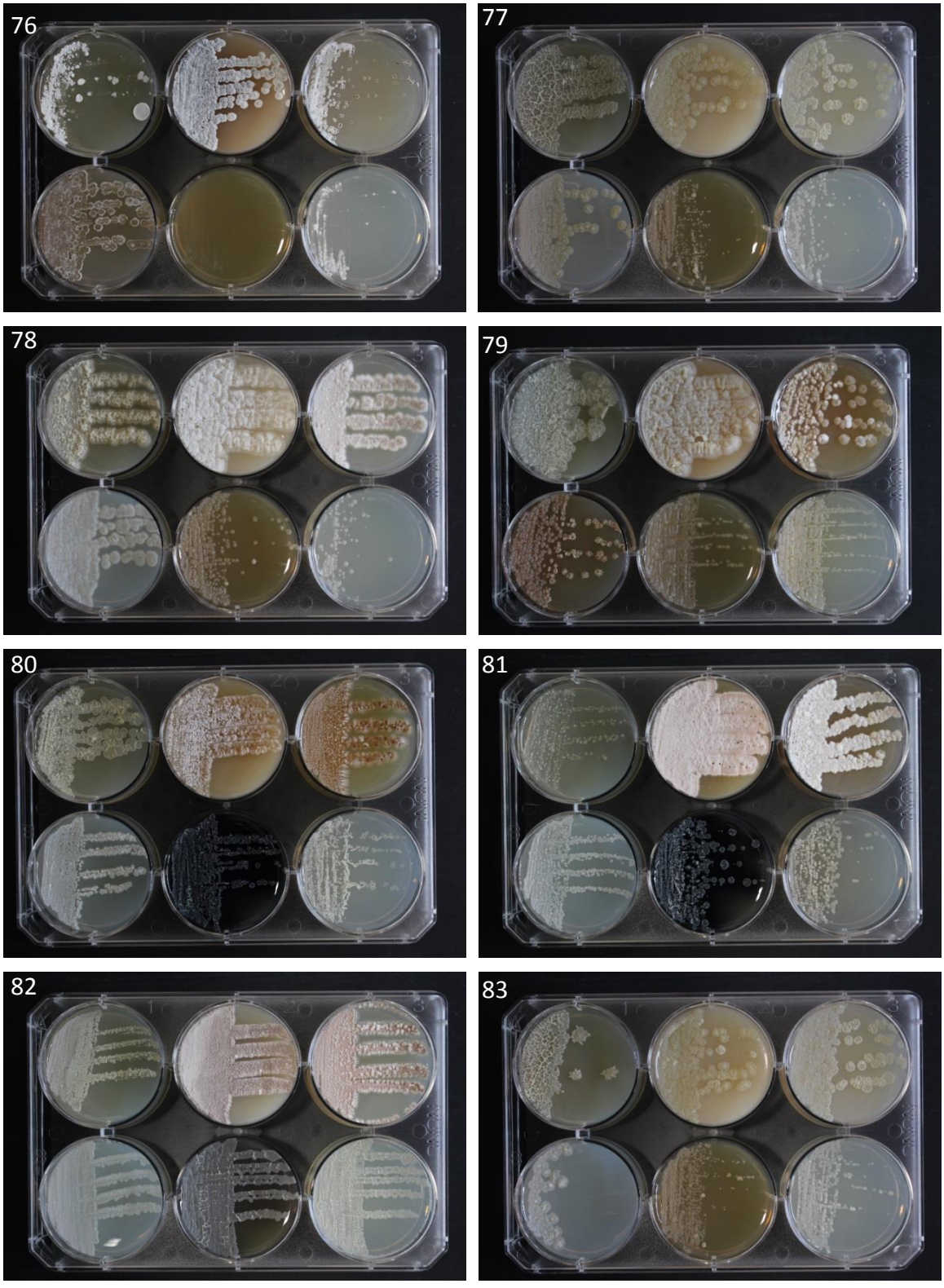

70


71


73


74


75


76
77
78
79
80
81
82
83


85


87


88


89


90


92


93


94


96


99


100


101


102


103


104


105


106


108


109


110


111


112


113


114


115


116


117

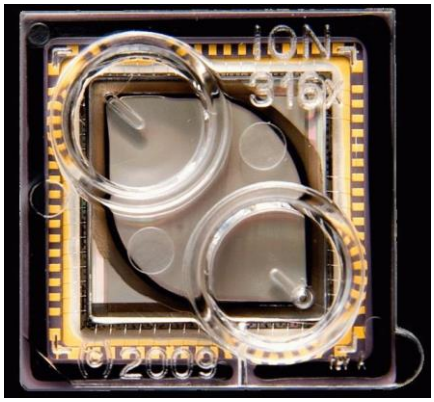

118

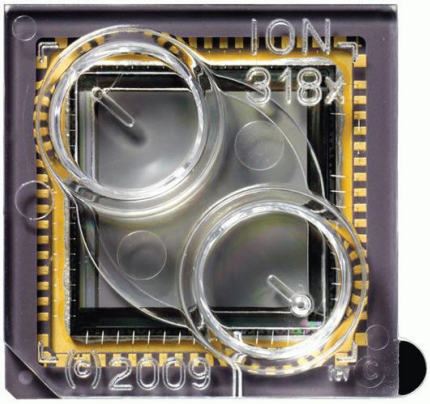

119


120

Appendix B

Chip Loading Protocol 316 v2/ 318 v2





Notes: Use unweighted buckets.

Use a used chip (same type) to balance the MiniFuge containing 30ul of annealing buffer (from Sequencing Kit in fridge).

Remove liquid from chip after chip check

1. Tilt chip 45 degree angle with the loading port down (where the little nudge is) and insert pipette tip firmly and remove as much liquid as possible.
2. Place the chip upside down in the bucket and orient the chip tab pointing in (towards centre of the MiniFuge).
3. Centrifuge 10 sec and remove any liquid from the bucket with a Kimwipe tissue.

Loading templated ISPs onto the chip

IMPORTANT: Keep the chip in the bucket and tilt it at a 45 degree angle. The pipette tip needs to be at 90 degrees to the chip. Loading port should face down (close to chip tab).

1. Following the polymerase incubation, collect the entire sample of approximately 30 ul into Rainin SR-L200F pipette tip and insert the tip firmly into the loading port of the chip.
2. Dial down the pipette to gently and slowly deposit the ISPs at a rate of 1ul per second. Avoid introducing any bubbles and thus keep a small amount of sample in the pipette tip.
3. Transfer the chip to MiniFuge with the bucket to the centrifuge with the chip tab poining to the centre. Centrifuge for 30 sec (Cycle 1).
4. Mix the sample in the chip by:
    a) Setting the volume to 25 ul.
    b) Tilt the chip to a 45 degree angle so that the loading port is the lower port, and insert the tip firmly into loading port.
    c) Slowly pipette the sample in and out of the chip 10 times and avoid introducing any bubbles.
5. Centrifuge the chip for 30 sec with the spin tab.

	Cycle 1	**Cycle 2**	**Cycle 3**	**Cycle 4**	**Cycle 5**
Spin Tab	In	Out	In	Out	In

At cycle 5, the last time after mixing dial the sample back into the chip as for cycle 1

DO NOT SPIN THE CHIP UPSIDE DOWN AT ANY TIME

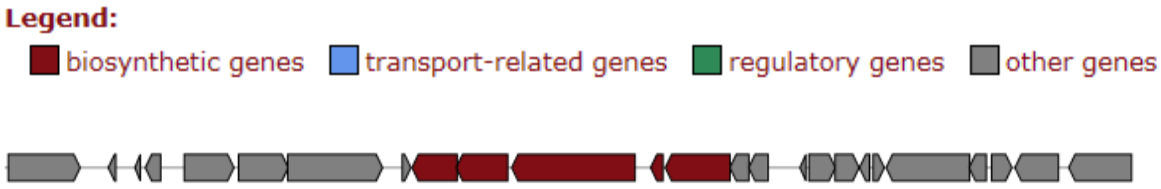
6. Tilt the chip at 45 degree angle and slowly remove as much liquid as possible, if any liquid remains spin the chip for 5 sec with the chip tab

pointing out and facing upwards and remove any additional liquid. Do not flush the chip.

7. Proceed to select Planned Run and perform the run.

Appendix C

Tables show fold change and p- value for all identified biosynthesis gene clusters in isolate_99 employing antiSMASH that their functions obtained from antiSMASH/ NCBI/ BacMet databases. The colour codes correspond to the antiSMASH prediction for each cluster. Red highlighted genes and rows show biosynthetic, blue highlighted show transport-related, green highlighted show regulatory and grey or white show other genes and proteins.



Cluster No. 1 (Lantipeptide)

Gene type	Description (antiSMASH/NCBI/BacMet)	Fold change	*p*-value
Biosynthetic gene	Lantibiotic dehydratase domain protein/ Methyltranferase/ heavy metal translocating P-type ATPase	-3.1	0.012
Biosynthetic gene	Lanthionine synthetase C family protein/ Lanthionine synthetase C-like protein/ Putative multidrug resistance protein B	-2.5	0.78
Biosynthetic gene	Lantibiotic dehydratase domain protein/ Lantibiotic dehydratase/ multicopper oxidase, type 2	-2.9	0.2
Biosynthetic gene	None/ FxLD family lantipeptide/ Copper transporter	-1.0	1
Biosynthetic gene	Methyltranferase/ protein-L-isoaspartate(D-aspartate) methyltransferase/ protein-L-isoaspartate O-methyltransferase	-9.2	1.61E-08

Cluster No. 2 (Siderophore)

Gene type	Description (antiSMASH/NCBI/BacMet)	Fold change	*p*-value
Biosynthetic gene	Decarboxylase, pyridoxal-dependent/ Amino acid decarboxylase pyridoxal-dependent protein/ Glutamate decarboxylase	-2.6	1
Biosynthetic gene	Lysine/ornithine N-monooxygenase/ Monooxygenase family protein/ Mercuric reductase MerA	-2.1	1
Biosynthetic gene	Putative siderophore biosynthesis protein/ Acetyltransferase/ Sensor protein IrlS	1.31	3.41E-06
Biosynthetic gene	None/ Siderophore biosynthesis protein, IucA/IucC family/ Ferrous-iron efflux pump FieF	-24.1	0.55
Biosynthetic gene	Asparagine synthase/ Glutamine--fructose-6-phosphate aminotransferase/ Putative enoyl-(acyl-carrier-protein) reductase II	-1.2	1

Cluster No. 3 (Terpene)

Gene type	Description (antiSMASH/NCBI/BacMet)	Fold change	*p*-value
Biosynthetic	Aminotransferase class-III/	-5.6	1

gene	4-aminobutyrate aminotransferase/ Type 11 methyltransferase		
Biosynthetic gene	Cytochrome P450/ Cytochrome P450 hydroxylase/ Tellurite resistance protein	1	1
Biosynthetic gene	Peptide deformylase/ Peptide deformylase/ No hits	-1.3	1
Biosynthetic gene	Terpene Synthase/ Multidrug MFS transporter/ modC gene product	-1.2	1
Biosynthetic gene	Cytochrome P450/ Cytochrome P450/ cadmium-transporting ATPase	-1.2	1
Regulatory gene	AraC family transcriptional regulator/ AraC family transcriptional regulator/ Multiple antibiotic resistance protein MarA	1.8	0.7
Regulatory gene	GntR family transcriptional regulator/ GntR family transcriptional regulator/ Transcriptional repressor CopY	1.1	1

Cluster No. 4 (Terpene)

Gene type	Description (antiSMASH/NCBI/BacMet)	Fold change	*p*-value
Transport-related gene	Drug resistance transporter, EmrB/QacA/ Efflux|putative transport protein/ Antiseptic resistance protein	-1.2	1
Regulatory gene	MarR family/ MarR family/ MarR family	-1.1	1
Regulatory gene	AsnC family/ AsnC family /unnamed protein product	2	1

Regulatory gene	MerR family/ MerR family/ MerR family	10.8	0.54
Biosynthetic gene	Cytochrome p450/ Cytochrome p450/ Fe(3+)-pyochelin receptor	-3.3	1
Biosynthetic gene	No hits/ squalene-hopene cyclase/ glycerol uptake facilitator protein	2.5	0.0009
Biosynthetic gene	Polyprenyle synthetase/ polyprenyle synthetase / apolipoprotein N-acyltransferase	4.3	0.005
Biosynthetic gene	Acetyl-CoA carboxylase/ methylmalonyl-CoA carboxyltransferase/ magnesium and cobalt efflux protein CorC	1.1	1
Biosynthetic gene	Alkaline serine protease/ peptidase/ Copper-transporting P-type ATPase	1.1	1

Cluster No. 5 (Otherks)

Gene type	Description (antiSMASH/NCBI/BacMet)	Fold change	*p*-value
Biosynthetic gene	Cytochrome p450/ Cytochrome p450/ Fe(3+)-pyochelin receptor	-3.3	1
Regulatory gene	MarR family/ MarR family/ MarR family	-1.1	1
Regulatory gene	AsnC family/ AsnC family /unnamed protein product	2	1
Regulatory gene	MerR family/ MerR family/ MerR family	10.8	0.54
Biosynthetic gene	Cytochrome p450/ Cytochrome p450/ Fe(3+)-pyochelin receptor	-3.3	1

Biosynthetic gene	No hits/ squalene-hopene cyclase/ glycerol uptake facilitator protein	2.5	0.0009
Biosynthetic gene	Polyprenyle synthetase/ polyprenyle synthetase /apolipoprotein N-acyltransferase	4.3	0.005
Transport-related gene	Drug resistance transporter, EmrB/QacA/ Efflux|putative transport protein/ Antiseptic resistance protein	-1.2	1
Biosynthetic gene	Alkaline serine protease/ peptidase/ Copper-transporting P-type ATPase	1.1	1

Cluster No. 6 (Otherks)

Gene type	Description (antiSMASH/NCBI/BacMet)	Fold change	*p*-value
Biosynthetic gene	Alpha-glucosidase / Alpha-glucosidase/ Multicopper oxidase family	-1.2	1
Transport-related gene	ABC transporter permease protein/ ABC transporter permease/ molybdate ABC transporter	1.3	0.67
Biosynthetic gene	Sugar-binding lipoprotein/ Sugar ABC transporter substrate-binding protein/ Multidrug resistance protein	2.5	0.2
Regulatory gene	LacI family/ LacI family/ Copper transporter MctB	-1	1
Biosynthetic gene	Beta-ketoacyl synthase type I polyketide synthase Nickel-cobalt-cadmium resistance protein	-3.1	0.45
Biosynthetic gene	Reductase SDR type I polyketide synthase heavy metal translocating P-type ATPase	-1.24	1

Biosynthetic gene	Malonyl CoA-acyl carrier protein transacylase type I polyketide synthase unnamed protein produc	4.5	1
Biosynthetic gene	Acyl-CoA dehydrogenase Acyl-CoA dehydrogenase SoxR family transcriptional regulator	-5.9	1
Biosynthetic gene	Alkaline serine protease serine protease iron dependent repressor	1.27	1
Biosynthetic gene	Nitrilase NAD+ synthase apolipoprotein N-acyltransferase	1.3	1
Biosynthetic gene	Oxidoreductase Oxidoreductase cation transport ATPase	-1	1
Regulatory gene	TetR family TetR family TetR family	-1.2	1
Biosynthetic gene	Aldo/keto reductase Aldo/keto reductase multicopper oxidase type 2	-1.5	
Regulatory gene	SARP family Tetratricopeptide repeat protein Putative cation efflux system protein	-1.6	1

Cluster No. 7 (Terpene)

Gene type	Description (antiSMASH/ NCBI / BacMet)	Fold change	*p*-value
Biosynthetic gene	Pullulanase type I/ malto-oligosyltrehalose trehalohydrolase/ NADPH-dependent ferric-chelate reductase	2.1	1
Biosynthetic	Methionine aminopeptidase/	1.7	1

gene	peptidase M24/ copper resistance protein A		
Regulatory gene	Terpene synthase/ Geosmin synthase/ Aconitate hydratase	-2.8	0.18
Regulatory gene	TetR family/ TetR family/ TetR family	-1.3	1
Biosynthetic gene	No hits/ A-factor biosynthesis protein/ Phosphate transport system permease protein	-5.9	1
Regulatory gene	TetR family/ TetR family/ TetR family	1.2	1
Biosynthetic gene	Alkaline serine protease/ Protease/ Multicopper oxidase family protein	1.2	1

Cluster No. 8 (Bacteriocin)

Gene type	Description (antiSMASH/ NCBI / BacMet)	Fold change	*p*-value
Transport-related gene	Drug resistance transporter (EmrB/QacA)/ efflux|putative transmembrane efflux protein/ multidrug efflux pump SmfY	5.2	0.3
Biosynthetic gene	Alpha/beta hydrolase fold protein/ Alpha/beta hydrolase fold protein / Efflux pump protein FarA	1.4	1
Other gene	No hits/ Hypothetical protein/ Nickel transport permease system	2.9	0.6
Other gene	No hits/	27.9	0.49

	Hypothetical protein/ LysR famil		
Other gene	No hits/ Hypothetical protein/ Mutidrug resistance protein MdtA	2.6	0.5
Other gene	No hits/ Hypothetical protein/ Mutidrug resistance protein PmpM	2.8	1
Other gene	No hits/ RiPP maturation radical SAM protein/ Arsenical pump-driving ATPase	11.5	1
Other gene	No hits/ RiPP maturation radical SAM protein/ Outer membrane protein OprM	13	0.00014
Biosynthetic gene	Radical SAM domain protein Radical SAM domain protein CopA family copper resistance protein	4.8	1.3E-06
Other gene	No hits/ No hits/ ABC transporter	2	0.9
Other gene	No hits/ Hypothetical protein/ Copper homeostasis protein CutC	15.1	0.7
Other gene	No hits/ Hypothetical protein/ Putative cation efflux system protein	8.4	1.95E-56
Other gene	No hits/ Hypothetical protein/ Multidrug resistance protein	7.4	2.46E-13
Other gene	No hits/ No hits/ Multidrug resistance protein	3.1	1
Biosynthetic	Lantibiotic dehydratase/	13.3	0.02

gene	Lanthionine biosynthesis protein/ Lead uptake protein PbrT		
Other gene	No hitis/ Lanthionine biosynthesis protein/ Na+ driven multidrug efflux pump	8	1

Cluster No. 9 (Bacteriocin)

Gene type	Description (antiSMASH/ NCBI / BacMet)	Fold change	*p*-value
Transport-related gene	ABC transporter/ ABC transporter/ Zinc ABC transporter ATP binding protein	-5.6	1
Biosynthetic gene	No hits/ Endonuclease/ copper resistance protein CopC	-9.3	0.3
Biosynthetic gene	No hits/ Protease/ Copper oxidase	-1.9	1

Cluster No. 10 (Otherks-t1pks)

Gene type	Description (antiSMASH/ NCBI / BacMet)	Fold change	*p*-value
Regulatory gene	Sensor histidine kinase/ Sensor histidine kinase/ heavy metal sensor signal transduction histidine kinase	-15.3	0.8
Regulatory gene	Response regulator/ DNA-binding response regulator/ Heavy metal response regulator	8	1

Biosynthetic gene	Alkaline serine protease/ peptidase S8/ MerA	-1.6	1
Regulatory gene	SARP family/ LuxR family Putative multidrug resistance protein	1	1
Biosynthetic gene	O-methyltranferase/ hypothetical protein/ methyltransferase type 11	1.2	1
Biosynthetic gene	2-isopropylmalate synthase/ 4-hydroxy-2-oxovalerate aldolase/ arsenical resistance protein ArsH	4.5	1
Biosynthetic gene	Aminotransferase/ Aminotransferase/ Methionine gamma-lyase	-5.9	1
Biosynthetic gene	Malonyl CoA-acyl carrier protein transacylase/ type I polyketide synthase/ Quaternary ammonium compound-resistanc	4.7	0.2
Biosynthetic gene	Beta-keto synthase/ type I polyketide synthase/ putative multidrug resistance efflux protein	3.2	0.6
Biosynthetic gene	Short chain dehydrogenase/ acyl transferase domain protein/ zntA gene product	3.8	0.7
Biosynthetic gene	Beta-ketoacyl synthase/ type I polyketide synthase/ Multidrug efflux pump SmfY	2.3	0.5
Transport-related gene	Drug resistance transporter (EmrB/QacA)/ multidrug efflux|putative transporter/ multidrug resistance membrane translocase	2	0.9
Regulatory gene	MarR family/ MarR family/ MarR family	-1.2	1

Cluster No. 11 (Otherks-t2pks)

Gene type	Description (antiSMASH/ NCBI / BacMet)	Fold change	*p*-value
Biosynthetic gene	1-deoxy-D-xylose-5-phosphate synthase/ alpha-ketoglutarate decarboxylase/ copper-translocating P-type ATPase	1.1	1
Biosynthetic gene	Monooxygenase FAD binding/ Oxygenase/ unnamed protein product	-10.1	1
Transport-related gene	Drug resistance transporter (EmrB/QacA)/ multidrug efflux (EmrB/QacA subfamily)/ multidrug resistance protein	1.7	0.8
Biosynthetic gene	Cyclase/ dehydrase Cyclase/ chaperone protein DnaK	-5.6	1
Biosynthetic gene	AMP-dependent synthatase and ligase/ Hypothetical protein/ Extracellular tungstate binding protein	-1.2	1
Biosynthetic gene	Malonyl CoA-acyl carrier protein transacylase/ polyketide synthase/ Molybdenum ABC transporter, ATP-binding protein	4.5	1
Biosynthetic gene	3-oxoacyl/ 3-oxoacyl-ACP synthase/ mercuric reductase	1	1
Regulatory gene	SARP family/ hypothetical protein/ Response regulator arlR	33.1	0.18
Biosynthetic gene	Beta-ketoacyl synthase/ Beta-ketoacyl synthase/ Multicopper oxidase type 2	-5.6	1
Regulatory	PadR family/	-1.4	1

gene	PadrR family/ Methyltransferase		
Biosynthetic gene	Crotonyl-CoA reductase alcohol dehydrogenase/ type 11 methyltransferase	-1.4	1

Cluster No. 12 (Butyrolactone-otherks)

Gene type	Description (antiSMASH/ NCBI / BacMet)	Fold change	*p*-value
Biosynthetic gene	carnitine dehydratase/ carnitine dehydratase/ Zinc transporter	8.7	1
Biosynthetic gene	No hits/ Hypothetical protein/ RND multidrug efflux transporter	1	1
Regulatory gene	Isoprenylcysteine carboxyl methyltransferase/ Hypothetical protein Heavy metal translocating P-type ATPase	-10	1
Biosynthetic gene	Carbamoyl phosphate synthase/ aminodeoxychorismate/anthranilate synthase component II/ Heavy metal translocating P-type ATPase	-5.3	1
Regulatory gene	serine/threonine protein kinase/ serine/threonine protein kinase/ copB-like protein	-1.4	1

Cluster No. 13 (Terpene)

Gene type	Description (antiSMASH/ NCBI / BacMet)	Fold change	*p*-value
Biosynthetic	No hits/	6.1	0.57

gene	Hypothetical protein/ Putative membrane fusion protein		
Biosynthetic gene	Cytochrome P450/ Steroid C27-monooxygenase/ copper-exporting ATPase	8	1
Biosynthetic gene	Acetyl-CoA acetyltransferase/ Acetyl-CoA acetyltransferase/ Copper-exporting ATPase	1	1
Regulatory gene	Serine/threonine protein kinase/ Serine/threonine protein kinase/ Apolipoprotein N-acyltransferase	2.2	1
Transport-related gene	ABC(binding protein) transporter/ ABC(binding protein) transporter/ Multidrug resistance protein MdtB	1	1

Cluster No. 14 (Nrps)

Gene type	Description (antiSMASH/ NCBI / BacMet)	Fold change	*p*-value
Regulatory gene	IclR family/ Hypothetical protein/ Integral membrane protei	-1.9	1
Biosynthetic gene	Oxidoreductase/ LLM class F420-dependent oxidoreductase/ Putative Ferritin protein	-1.1	1
Biosynthetic gene	Methyltransferase/ 3' terminal RNA ribose 2'-O-methyltransferase Hen1 Heavy metal translocating P-type ATPase	-2.4	0.5
Biosynthetic gene	Condensation domain containing protein/ Hypothetical protein/ chromate transporter	-1.6	1
Biosynthetic	AMP-dependent synthetase and ligase/	-1.8	1

gene	non-ribosomal peptide synthetase/ protein-l-isoaspartate o-methyltransferase		
Biosynthetic gene	Beta-lactamase/ Hypothetical protein/ Copper homeostasis protein CutC	-1.2	1
Biosynthetic gene	ornithine carbamoyltransferase/ ornithine carbamoyltransferase/ cation-transporting ATPase, p-type	-5.6	1
Transport-related genes	Major facilitator transporter/ Efflux|putative multidrug resistance protein Major facilitator superfamily protein	-1.2	1
Biosynthetic gene	Dioxygenase TauD/TfdA/ Hypothetical protein/ Putative sensor histidine kinase	2	1
Regulatory gene	LuxR family/ Hypothetical protein/ Sensor histidine kinase PmrB	3.3	1
Regulatory gene	LysR family/ LysR family/ Activator of hydrogen peroxide-inducible genes	-1	1
Biosynthetic gene	Aminotransferase class v/ cysteine desulfurase/ Methionine gamma-lyase	-1	1

Cluster No. 15 (Siderophore)

Gene type	Description (antiSMASH/ NCBI / BacMet)	Fold change	*p*-value
Transport-related gene	Major facilitator transporter gene/ Multidrug efflux|putative transporter/ Multidrug resistance protein	2.1	1
Regulatory gene	ArsR family/ Transcriptional regulator/	4.5	1

	ArsR family		
Biosynthetic gene	No hits/ Iron transporter/ Iron transport er	1	1
Biosynthetic gene	No hits/ Siderophore biosynthesis protein, IucA/IucC family/ MexX	2.1	1
Regulatory gene	Serine/threonine protein kinase/ Serine/threonine protein kinase/ Arsenical pump membrane protein	3.8	0.22

Cluster No. 16 (Nrps)

Gene type	Description (antiSMASH/ NCBI / BacMet)	Fold change	*p*-value
Transport-related gene	ABC transporter ATP-binding protein/ ABC transporter ATP-binding protein/ ABC transporter ATP-binding protein	1	1
Biosynthetic gene	Aminotransferase class-III/ Aspartate aminotransferase family protein/ Membrane fusion protein MtrC	-1.2	1
Biosynthetic gene	AsnC family/ AsnC family/ Multidrug resistance protein MdtA	-1.2	1
Biosynthetic gene	Aldehyde dehydrogenase/ gamma-aminobutyraldehyde dehydrogenase/ Probable multidrug resistance protein	1.3	1
Biosynthetic gene	Sugar-binding lipoprotein/ ABC transporter substrate-binding protein/ Multidrug resistance protein A	-14.8	1
Biosynthetic gene	Beta-lactamase/ Beta-lactamase/	-4.8	0.75

	Putative alkylmercury lyase		
Regulatory gene	RNA polymerase sigma factor/ RNA polymerase sigma factor/ RNA polymerase sigma factor cnrH	-10.2	1
Biosynthetic gene	AMP-dependent synthetase & ligase/ Non-ribosomal peptide synthetase/ Copper/silver efflux system membrane fusion protein CusB	-5.9	1
Biosynthetic gene	Aminotransferase/ Hypothetical protein/ Methionine gamma-lyase	-2.2	1
Biosynthetic gene	Tryptophan halogenase/ Tryptophan halogenase/ Inner membrane protein VmeB	-1.5	1
Biosynthetic gene	Flavine reductase/ Oxidase/ High-affinity nickel transport protein	1	1
Biosynthetic gene	Metallo-beta lactamase/ MBL fold metallo-hydrolase/ Membrane fusion protein MexC	-13.8	1

Cluster No. 17 (Siderophore)

Gene type	Description (antiSMASH/ NCBI / BacMet)	Fold change	*p*-value
Biosynthetic gene	AMP-dependent synthetase & ligase/ Long-chain-fatty-acid--CoA ligase/ Multidrug resistance protein MdtK	1.7	1
Regulatory gene	TetR family/ TetR family/ TetR family	-2.8	1
Biosynthetic gene	Putative siderophore biosynthesis protein/ N-acetyltransferase/	-6.1	1

	Efflux pump membrane transporter BepE		
Biosynthetic gene	No hits/ Iron transporter/ ATPase P	4.5	1
Biosynthetic gene	Aminotransferase class-III/ 2,4-diaminobutyrate 4-aminotransferase/ Heavy metal translocating P-type ATPase	1.12	1

Cluster No. 18 (Nrps)

Gene type	Description (antiSMASH/ NCBI / BacMet)	Fold change	*p*-value
Regulatory gene	IclR family/ IclR family/ IclR family	1.7	0.38
Regulatory gene	MerR family/ MerR family/ MerR family	2	1
Biosynthetic gene	Hydrolase/ Hydrolase/ Probable cobalt/nickel-exporting P-typ	-1.3	1
Biosynthetic gene	Condensation domain-containing protein/ Non-ribosomal peptide synthetase/ Mercury operon coregulator protein	1.9	0.11
Biosynthetic gene	AMP-dependent synthetase & ligase/ Non-ribosomal peptide synthetase/ zinc/manganese/iron ABC transporter	1.4	1
Biosynthetic gene	Acyl-CoA dehydrogenase/ Hypothetical protein/ Nickel-cobalt-cadmium resistance protein	-2.2	1
Regulatory gene	LuxR family/ LuxR family/ LuxR family	-1.6	0.68

Regulatory gene	MarR family/ MarR family/ MarR family	-3.3	0.4
Regulatory gene	GntR family/ GntR family/ GntR family	-10.2	0.5

Cluster No. 19 (Other)

Gene type	Description (antiSMASH/NCBI/BacMet)	Fold change	*p*-value
Biosynthetic gene	Alcohol dehydrogenase/ Alcohol dehydrogenase/ putative methyltransferase	-5.6	1
Biosynthetic gene	Pyridine nucleotide-disulfide oxidoreductase/ FAD-dependent oxidoreductase/ Mercuric reductase	-10	1
Biosynthetic gene	Aldo/keto reductase / Aldo/keto reductase/ Heavy metal sensor signal transduction	-1.6	1
Biosynthetic gene	Acetylornithine deacetylase/ Hypothetical protein/ Heavy metal-translocating P-type ATPase	1.2	1
Regulatory gene	GntR family/ GntR family/ GntR family	2	0.35
Biosynthetic gene	Phenylalanine- specific permeas/ Amino acid permease/ Amino acid permease-associated region	-17.7	0.97
Biosynthetic gene	Crotonyl CoA reductase/ Alcohol dehydrogenase Putative inner membrane protein	-5.6	1

Cluster No. 20 (Bacteriocin)

Gene type	Description (antiSMASH/ NCBI / BacMet)	Fold change	*p*-value
Transport-related gene	Major facilitator transporter/ Efflux|putative multidrug resistance protein/ Major facilitator superfamily permease	5.6	0.06
Other gene	No hits/ Peptidase S9/ Copper	1.4	1
Other gene	No hits/ S9 family peptidase/ Copper resistance B precursor	3.2	0.4
Other gene	No hits/ hypothetical protein OV320_2954/ No hits	8.3	8.28E-09
Biosynthetic gene	Radical SAM domain protein/ RiPP maturation radical SAM protein 1/ Copper-translocating P-type ATPase	3.4	0.004
Other gene	No hits/ RiPP maturation radical SAM protein/ Multidrug resistance protein MdtO	4.5	1
Other gene	No hits/ Hypothetical protein/ Protein KlaC	2	1
Biosynthetic gene	3-hydroxyisobutyrate dehydrogenase/ NADP oxidoreductase/ Putative enoyl	1.4	1
Regulatory gene	TetR family/ TetR family/ TetR family	11.5	1
Other gene	No hits/	7.7	1

	XRE family transcriptional regulator/ No hits		
Other gene	No hits/ Transcriptional regulator Heavy metal translocating P-type ATPase	5.1	0.56
Other gene	No hits/ Hypothetical protein/ CzrB protein	4.5	1
Biosynthetic gene	Short chain dehydrogenase/ Short chain dehydrogenase/ enoyl-reductase	3.2	1

Cluster No. 21 (Melanin)

Gene type	Description (antiSMASH/ NCBI / BacMet)	Fold change	*p*-value
Biosynthetic gene	GTP-binding protein LepA/ ATP-binding protein/ Nickel and cobalt resistance protein CnrA	-1.7	1
Biosynthetic gene	No hits/ Tyrosinase co-factor/ NcrC OS=Serratia marcescens	1	1
Biosynthetic gene	Glycosyl transferase group / Glycosyl transferase/ Multicopper oxidase	4.5	1

Cluster No. 22 (Ectoine)

Gene type	Description (antiSMASH/ NCBI / BacMet)	Fold change	*p*-value
Transport-related gene	Polar amino acid ABC transporter/ ABC transporter permease protein / ABC transporter	-1.7	0.43

Biosynthetic gene	Glutamine-binding lipoprotein/ ABC transporter substrate-binding protein/ Multidrug efflux pump subunit AcrA	-3.3	0.005
Biosynthetic gene	Aminotransferase class v/ Aspartate aminotransferase/ Methyl viologen resistance protein	1	1
Biosynthetic gene	N-acetyl transferase/ Diaminobutyrate acetyltransferase/ Diamine N-acetyltransferase	-1.2	1
Biosynthetic gene	Aminotransferase class v/ Aminotransferase class v/ Methionine gamma-lyase	-1.8	1

Cluster No. 23 (Terpene)

Gene type	Description (antiSMASH/ NCBI / BacMet)	Fold change	*p*-value
Biosynthetic gene	Dehydrogenase/ Dehydrogenase/ zinc/cadmium/mercury/lead-transporting ATPase	1.1	1
Transport-related gene	ABC transporter ATP binding protein/ ABC transporter ATP-binding component/ ABC transporter related protein	-1.1	1
Biosynthetic gene	No hits/ Squalene synthase HpnC Multidrug resistance protein MdtO	-1.1	1
Biosynthetic gene	No hits/ Squalene synthase HpnC/ Iron-dependent repressor DtxR	2.4	0.32
Biosynthetic gene	Dehydrogenase Dehydrogenase	-2.2	1

	Enoyl-[acyl-carrier-protein]		
Regulatory gene	GntR family Transcriptional regulator MntR GntR family	16.9	1
Transport-related gene	Sodium:dicarboxylate symporter DAACS|2.A.23 proton/sodium Putative multidrug resistance protein	-1.2	1
Regulatory gene	TetR family TetR family Lead, cadmium, zinc and mercury-transporter	27.1	0.38

Cluster No. 24 (Other)

Gene type	Description (NCBI / BacMet)	Fold change	*p*-value
Biosynthetic gene	PfkB domain protein/ 1-phosphofructokinase / Heavy metal translocating P ATPase	-1.4	0.8
Biosynthetic gene	Sugar-binding lipoprotein ACB transporter substrate-binding protein / Heavy metal translocating P ATPase	-2.3	1.12
Regulatory gene	LuxR family/ LuxR family / EvgA positive	1.5	0.37
Biosynthetic gene	Alkyl hydroperoxide reductase/ Alkyl hydroperoxide reductase / Molybdenum ABC Transporter periplasmic binding-protein	4.5	1
Biosynthetic gene	Methyltransferase/ Type 11 methyltransferase / Methyltransferase	-1.2	1
Other gene	Heavy metal translocating P ATPase/ Heavy metal translocating P ATPase /	8.7	0.5

	Heavy metal translocating P ATPase		
Other gene	Heavy metal translocating P ATPase/ Heavy metal translocating P ATPase / Heavy metal translocating P ATPase	1	1
Biosynthetic gene	No hits/ Hypothetical protein / Hypothetical protein	-1.2	1
Regulatory gene	MerR family/ MerR family / MerR family	-10	0.51
Regulatory gene	ArsR famil/ ArsR family / ArsR family	-1.2	1
Biosynthetic gene	No hits/ Trans-aconitate methyletransferase / Arsenite S-adenosylmethyl transferase	1.6	0.42
Regulatory gene	TetR family TetR family / TetR family	-1.4	0.41
Biosynthetic gene	Glucose-1-phosphate adenylyl/ Glucosamine-phosphate acetyletransferase / Heavy metal translocating P ATPase/	-1.7	0.07
Regulatory gene	Histidine kinase/ Histidine kinase/ Cation transporting ATAPase	-1.3	0.5

Cluster No. 25 (Nrps)

Gene type	Description (antiSMASH/ NCBI / BacMet)	Fold change	*p*-value
Regulatory gene	AraC family/ AraC family/	-2.4	1

	AraC family		
Regulatory gene	MarR family/ MarR family/ MarR family	-5.3	1
Regulatory gene	TetR family/ TetR family/ TetR family	-1.3	1
Biosynthetic gene	No hits/ non-ribosomal peptide synthetase 4/ multidrug resistance protein D	4.5	1
Biosynthetic gene	Beta-lactamase/ Serine hydrolase/ Superoxide dismutase [Fe]	-4.2	0.8
Regulatory gene	TetR family/ TetR family/ TetR family	1.2	1
Biosynthetic gene	Alcohol dehydrogenase/ Mycothiol dehydrogenase/ Arsenite S-adenosylmethyltransferase	1.5	1
Biosynthetic gene	Metallo-beta-lactamase/ Metallo-beta-lactamase/ Unnamed protein product	1	1
Biosynthetic gene	Short-chain dehydrogenase/ Type I polyketide synthase/ Heavy metal translocating P-type ATPase	1.4	1
Biosynthetic gene	Phosphpantetheine-binding domain-containing protein/ Non-ribosomal peptide synthetase/ Heavy metal translocating P-type ATPase	2	1
Transport-related gene	Inner-membrane translocator/ Sugar ABC transporter permease protein/ Methyl viologen resistance protein	2.9	0.96
Transport-	ABC transporter/	-1.68	0.54

related gene	Sugar ABC transporter permease protein/ Phosphate transport ATP-binding protein		
Biosynthetic gene	3-hydroxyisobutyrate dehydrogenase/ 6-phosphogluconate dehydrogenase/ Heavy metal translocating P-type ATPase	1.77	0.23
Biosynthetic gene	Condensation domain-containing protein/ Non-ribosomal peptide synthetase/ multidrug resistance efflux pump	1	1

Cluster No. 26 (Nrps)

Gene type	Description (antiSMASH/ NCBI / BacMet)	Fold change	*p*-value
Biosynthetic gene	Aldehyde dehydrogenase/ Malonate-semialdehyde dehydrogenase/ Cadmium-translocating P-type ATPase	-1.2	1
Biosynthetic gene	Acetyl-CoA carboxylase/ Methylmalonyl-CoA carboxyltransferase/ Putative multicopper oxidase	1.76	1.05E-05
Regulatory gene	LacI family/ LacI family/ LacI family	1.2	1
Regulatory gene	MarR family/ MarR family/ MarR family	1.4	1
Biosynthetic gene	Methyltransferase/ Methyltransferase/ Arsenite S-adenosylmethyltransferase	11.6	1
Biosynthetic gene	Condensation domain-containing protein/ Non-ribosomal peptide synthetase/ ABC transporter	1	1
Biosynthetic gene	Beta-ketoacyl synthase/ Hypothetical protein IQ63_08835/	-1.9	0.55

	Mercuric reductase		
Biosynthetic gene	Methltransferase/ SAM-dependent methyltransferase/ Methyltransferase type 11	-1.69	1
Biosynthetic gene	AMP-dependent synthetase and ligase/ Non-ribosomal peptide synthetase/ heavy metal translocating p-type ATPase	-5.6	1
Biosynthetic gene	No hits/ Type I polyketide synthase/ Efflux transporter	-1	1

Appendix D

Conferences and workshops attendance

April 2014, Society for General Microbiology (SGM) annual meeting conference, **Liverpool, UK**

September 2014, Society for General Microbiology (SGM) focus meeting-Emerging challenges & opportunities in soil microbiology, **Loughborough University, UK**

November 2014, Grand Challenges & Microbiology policy workshop, **Glasgow, UK**

March 2015, Society for General Microbiology (SGM) annual meeting conference, **Birmingham, UK**

April 2015, Microbial genetic analysis workshop-life technologies, **Birmingham, UK**

August 2015, 9th European Conference on Marine Natural Products, **Glasgow, UK**

March 2016, Society for General Microbiology (SGM) annual meeting conference, **Liverpool, UK**

August 2017, Glasgow Microbiology Collective Event, **Glasgow, UK**

Training courses and Summer Schools

Bioinformatics Training course on Molecular Phylogenetics (10-12 December 2014), School of the Biological Sciences & Bioinformatics, **University of Cambridge, UK**

Bioinformatics Training course on Jalview program for multiple sequence alignment editing, visualisation and analysis (15th December 2014), **University College London (UCL), UK**

Bioinformatics Training course on Next Generation Sequencing (NGS) Data Analysis (17-19 December 2014), School of the Biological Sciences & Bioinformatics, **University of Cambridge, UK**

Ensembl Genome Browser (14th April 2015), **University of Glasgow, UK**

Sampling, Isolation and Cultivation of Marine Microorganisms (12-24 July 2015), MaCuMBA Summer School, **Royal Netherlands Institute for Sea Research, Texel, The Netherlands**

Molecular Phylogenetics (25th January 2016), Centre for Genome-Enabled Biology and Medicine, **University of Aberdeen, UK**

RNA Sequencing and Differential Expression (1st February, 2016), Centre for Genome-Enabled Biology and Medicine, **University of Aberdeen, UK**

Multiple Sequence Alignment and Analysis with Jalview (14th October, 2016), **University of Strathclyde, UK**