Metagenomic guided isolation of sporoactinomycetes from soil and screening for antimycobacterial activity

by

Samuel Hewat

University of Cape Town

Thesis presented for a Master of Science in the department of Molecular and Cell Biology, Faculty of Science, University of Cape Town, South Africa

February 2024

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

Table of Contents

Acknowledgements 3

List of Abbreviations 4

Abstract 6

Chapter 1 8

Introduction

Chapter 2 31

Metagenomic analysis: the actinobacterial diversity of two soil samples from the University of Cape Town using selective culture-independent methods

Chapter 3 50

Isolation and identification of the actinobacterial biodiversity of the Kramer and Tennis Court soil samples using selective culture-dependent methods

Chapter 4 89

Screening of actinobacterial isolates for antimycobacterial activity

Chapter 5 108

General discussion

Appendices 115

Acknowledgements

My thanks go to my supervisor Dr Paul Meyers for all the support and guidance that he has given me throughout my time as a UCT post-graduate student. I am incredibly lucky to have worked under a supervisor that is approachable, extremely knowledgeable, supportive, kind, caring, and light-hearted. Working under Dr Meyers allowed me to enjoy the research process and his help and guidance allowed produce a thesis that I am proud of. Many thanks goes out to Dr Daniel Watson for all his help and work that went into completing my antibiotic screening results (Chapter 4). His help and guidance allowed me to go the extra mile and produce antibiotic results that I am excited for people to read.

I acknowledge my lab mates, past and present, who were responsible for making lab 202/203 a safe and enjoyable work space where I was able to feel at home. I thank Ashleigh Dreyer for her amazing friendship and support throughout my post graduate studies. My thanks also goes to Adeebah Rakiep for her support and for showing me the path to completing a masters degree successfully.

A special thanks goes out to my family and friends for their love and support during my years of research and study. To my partner, Nicky Butters, without your unconditional love I would not have been able to get to where I am today.

List of Abbreviations

7H9	-	Middlebrook 7H9 (agar/broth)
amp	-	ampicillin
ASM1	-	*Amycolatopsis* selective medium 1
ASM2	-	*Amycolatopsis* selective medium 2
ASM3	-	*Amycolatopsis* selective medium 3
ANI	-	average nucleotide identity
bp	-	base pair (DNA)
BLAST	-	Basic Local Alignment Search Tool
blastn	-	nucleotide-nucleotide BLAST search
BSA	-	bovine serum albumin
CHX	-	cycloheximide
CZ	-	Czapek solution (agar/broth)
Da	-	daltons
dNTP	-	deoxyribonucleotide triphosphate
dDDH	-	digital DNA-DNA hybridization
DDH	-	DNA-DNA hybridization
eDNA	-	environmental DNA
F	-	forward
GGDC	-	Genome-To-Genome Distance Calculator
gDNA	-	genomic DNA
GNPS	-	Global Natural Products Social
gyrB	-	gyrase subunit B gene
HPLC	-	high performance liquid chromatography
ISP4	-	International *Streptomyces* Project medium No. 4
ISP2	-	International *Streptomyces* Project medium No. 2
K	-	Kramer
KSM	-	*Kribbella* selective medium
LC-MS	-	liquid chromatography-mass spectrometry

LPSN	-	list of prokaryotic names withstanding in nomenclature
LB	-	Luria-Bertani
M	-	molar
MEGA	-	Molecular Evolutionary Genetic Analysis
MDR	-	multidrug-resistant
MP	-	maximum parsimony (phylogenetic tree)
MUSCLE	-	Multiple Sequence Comparison by Log-Expectation
NA	-	nalidixic acid
NJ	-	neighbour-joining (phylogenetic tree)
NRPS	-	non-ribosomal peptide synthetase
OD	-	optical density
OTUs	-	Operational Taxonomic Units
pH	-	potential Hydrogen
PKS	-	polyketide synthase
PCR	-	polymerase chain reaction
R	-	reverse
recA	-	recombinase A gene
rpm	-	revolutions per minute
sp. nov.	-	species nova, “new species”
TC	-	Tennis Court
TDR	-	totally drug-resistant
TB	-	tuberculosis
U	-	units
UCT	-	University of Cape Town
v/v	-	volume for volume
XDR	-	extensively drug-resistant
ZOI	-	zone of inhibition
λ	-	phage Lambda
x *g*	-	times gravity (g-force – relative centrifugal force)

Abstract

Metagenomic analysis was used to analyse the actinobacterial diversity in the Kramer (K) and Tennis Court (TC) soil samples collected at the University of Cape Town (UCT). Fifty (50) and 52 16S-rRNA gene sequences were amplified for the K and TC samples, respectively, from environmental DNA (eDNA) using actinobacterium-specific primers. EzBioCloud analyses of the 16S-rRNA gene sequences showed a total of 10 orders present in each of the K and TC samples with the orders *Acidimicrobiales*, *Pseudonocardiales* and *Propionibacteriales* being the most abundant in both samples. Rarefaction curves showed that the full actinobacterial diversity of the samples was not captured.

Using selective media chosen based on the results of the metagenomic analyses, a total of 63 presumptive actinobacterial strains was isolated from the two soil samples, but 21 of these strains were discarded. Of the 42 actinobacteria studied, partial 16S-rRNA gene sequences were analysed for 40 of the 42 presumptive actinobacterial strains: 23 strains from the K sample and 17 strains from the TC sample. The 40 strains were identified to the genus level revealing 38 *Streptomyces* strains and two non-*Streptomyces* strains: *Actinomadura* strain 1KK1 and *Nonomuraea* strain 7TK1. The phylogenetic position of each identified strain was investigated by generating phylogenetic trees. Thirty (30) of the 38 *Streptomyces* strains identified were potential-clones, which were initially divided into three potential-clone groups. Clone analysis carried out on 23 of the 30 *Streptomyces* potential-clone strains revealed that the three potential-clone groups divided into five clone groups and eight distinctive *Streptomyces* strains based on their partial *gyrB* gene sequences.

Forty (40) of the 42 isolates studied were screened for antimycobacterial activity against *Mycobacterium aurum* strain A+ on four different media. A total of 35 strain-medium combinations (from 25 different strains) showed strong or very strong antimycobacterial activity. These 35 strain-medium combinations were tested for activity against *Escherichia coli* strain ATCC 25922 and *Staphylococcus aureus* strain ATCC 25923 to investigate the antibiotic spectrum of each active actinobacterial strain. Crude extracts of the most active actinobacterial strains were subjected to Global Natural Products Social (GNPS) molecular networking analyses to assess if the compounds produced are known compounds. None of the active actinobacterial strains showed activity against *E. coli* ATCC 25922, with just two strains (*Actinomadura* strain 1KK1 on R2A medium and *Streptomyces* strain 49TA3 on 7H9 and R2A media) showing strong or very strong activity against *S. aureus* ATCC 25923. Crude extracts of 26 of the 35 strain-medium combinations tested returned no matches on the GNPS database, with nine combinations having matches to know compounds. These no-match combinations could potentially be novel antibiotics with activity against *Mycobacterium tuberculosis.*

Chapter 1

Introduction

Table of Contents

1.1 Actinobacteria 10
1.2 Actinobacterial habitats 11
  1.2.1 Terrestrial actinobacteria.. 11
  1.2.2 Aquatic actinobacteria.. 12
    1.2.2.1 Fresh water actinobacteria.. 12
    1.2.2.2 Marine actinobacteria 12
1.3 Actinobacterial taxonomy 13
  1.3.1 Molecular classification 13
    1.3.1.1 16S-rRNA gene analysis.. 14
    1.3.1.2 DNA-DNA hybridization (DDH) 14
    1.3.1.3 Average Nucleotide Identity (ANI).. 15
1.4 Selective isolation of actinobacteria 16
  1.4.1 Selective media.. 17
1.5 Metagenomics... 17
1.6 Bioactive compounds produced by actinobacteria 18
  1.6.1 Antibiotics.. 19
  1.6.2 Antibiotic resistance 21
    1.6.2.1 Tuberculosis (TB) antibiotic resistance 21
  1.6.3 New methods for discovery of bioactive compounds 22
1.7 Aims 23
1.8 References.. 24

1.1 Actinobacteria

Actinobacteria, one of the largest phyla in the domain bacteria, comprises of Gram-positive bacteria that have a high guanine and cytosine (G+C) content (65-75%) in their DNA. They have large genomes compared to other bacteria, with genome sizes ranging from 1 megabases (Mb) to 12 Mb in size. Actinobacteria comprise of unicellular cocci or rods or multicellular filamentous bacteria (Miao and Davies, 2010; Salwan and Sharma, 2020; van Bergeijk *et al.*, 2020).

Filamentous actinobacteria, often referred to as actinomycetes, usually form a substrate mycelium consisting of branched hyphae and may also produce an aerial mycelium (Barka *et al.*, 2016). The appearance of actinomycete colonies is compact, leathery and dry looking on culture media. Actinobacteria inhabit a wide range of habitats both terrestrial and aquatic. They can be found most commonly in soil, fresh water and marine environments where they play an important role in the decomposition of organic materials such as cellulose and chitin. This means that they play a significant role in the carbon cycle, replenishing nutrients back into their environments (Anandan *et al.*, 2016).

As well as their ability to live in diverse habitats, actinobacteria produce a diverse array of bioactive natural compounds. Actinobacteria produce a variety of secondary metabolites, many with high pharmacological and commercial interest. Of all the microbial taxa, actinobacteria are the most prolific producers of bioactive metabolites. Notably, actinobacteria are prolific producers of antibiotics, almost 80% of the world's antibiotics are derived from actinobacteria (Anandan et al., 2016). In addition to antibiotics, actinobacteria are also known to produce anticancer compounds, immunosuppressants, anthelmintics, biofertilizers, herbicides and antiviral compounds (van Bergeijk *et al.*, 2020). This chapter will discuss actinobacterial habitats, actinobacterial taxonomy, culture-dependent and culture-independent methods to study them and the natural products they produce.

1.2 Actinobacterial habitats

Actinobacteria have adapted to occupy a wide range of ecological environments and conditions. They are present in soil, fresh water and marine environments (Goodfellow and Williams, 1983; Anandan *et al.*, 2016). Actinobacteria are abundant in the soil and can be found in surface soils to depths of over two metres below ground (Goodfellow and Williams, 1983). Actinobacteria are also capable of being part of the microbiomes of higher eukaryotes, such as plants, as endophytes (Conti *et al.*, 2016; van Bergeijk *et al.*, 2020). Actinobacteria are also present in the human body, where they make up more than a third of the healthy human microbiota (Miao and Davies, 2010).

1.2.1 Terrestrial actinobacteria

One gram of soil can contain up to 1×10^{10} microorganisms belonging to thousands of different species (Rosselló-Mora and Amann, 2001). Soil is heterogenous and provides many different microbial niches for microorganisms, including micropores and different soil aggregates. This allows for a high diversity of soil microorganisms to live in soil habitats (Torsvik and Øvreås, 2002; Daniel, 2004). There actinobacteria play their important role in organic matter turnover and the carbon cycle, replenishing the supply of nutrients to the soil (Anandan *et al.*, 2016).

Actinobacteria have been found in most soil types. Anandan *et al.* (2016) documented that actinobacterial species have been isolated from mangrove, sandy, black alkaline, sandy loam, alkaline desert and sub-tropical desert soils. Soil remains the most important habitat for actinobacteria (Anandan *et al.*, 2016). In the soil surrounding plant roots, the rhizosphere, actinobacteria play a major part in the microbial community where they break down hardy plant material. This makes the rhizosphere one of the best habitats for the isolation of actinobacteria (Anandan *et al.*, 2016).

There is still much untapped diversity in soil, with only 1% of the total microbial diversity of soil culturable on traditional lab media (Amann, 1995; Arjun and Harikrishnan, 2011). This means soil remains a potential source of novel actinobacteria and bioactive active compounds. However, different approaches may be required to capture and exploit this untapped potential.

1.2.2 Aquatic actinobacteria

Actinobacteria are found in both fresh water and marine environments. Members of the phylum present in aquatic environments may sometimes be there because they have been washed in from surrounding terrestrial environments (Anandan *et al.*, 2016).

1.2.2.1 Fresh water actinobacteria

Cross (1981) provided evidence that actinobacteria can be readily isolated from freshwater sites. Genera such as *Actinoplanes*, *Micromonospora*, *Rhodococcus* and *Streptomyces* were shown to be strongly represented in fresh water (Cross, 1981). *Micromonospora* species are considered to be common in fresh water habitats with their presence confirmed in streams, rivers and river sediments (Anandan *et al*., 2016). Evidence suggests that, while freshwater actinobacteria may be active in freshwater ecosystems, they are not specifically adapted to living in those environments (Anandan *et al.*, 2016).

1.2.2.2 Marine actinobacteria

The marine environment contains an even greater biodiversity of actinobacteria when compared to the terrestrial environment. Many actinobacterial species have been found to be indigenous to the marine environment. This ability is illustrated by some actinomycetes, *Streptomyces* and *Rhodococcus* species included, being shown to possess putative genes for gas vesicle production. These genes are associated with the ability to survive in aqueous environments (van Keulen *et al.*, 2005). Members of the genera *Rhodococcus*, *Streptomyces*, *Salinispora*, '*Solwaraspora*', *Salinibacterium* and others have been designated as indigenous marine actinobacteria (Bull *et al.*, 2005).

Currently, novel secondary metabolites are being discovered at a faster rate from marine actinobacteria than terrestrial actinobacteria (Anandan *et al.*, 2016). Grossart *et al.* (2004) studied marine organic aggregates and found 10% of the aggregates to be made up of actinobacteria. This showed that the seas and oceans are an understudied and an untapped source of novel actinobacterial biodiversity and new metabolites (Goodfellow, 2010; Anandan *et al.*, 2016).

1.3 Actinobacterial taxonomy

Taxonomy is an important organising principle in microbiology and is based on classification, identification and nomenclature. Classification is the arrangement of organisms into specific taxa (groups) based on matching characteristics; nomenclature is the assignment of names to taxa based on published rules and identification determines if an isolate belongs to a taxon with a validly published name (Kämpfer and Glaeser, 2012).

The basis of modern bacterial taxonomy is a polyphasic approach (Colwell, 1970), which utilizes phenotypic, chemotaxonomic, genotypic and phylogenetic information in the classification of a bacterium. This approach is used for the delineation of taxa at all levels (Chun and Rainey, 2014) and has led to marked improvements in the classification, identification and nomenclature of prokaryotes (Zhi *et al.*, 2012). The polyphasic approach is effective because the interplay between genotypic and phenotypic datasets provides an effective basis for describing the diversity of prokaryotes (Tindall et al. 2010).

Bacterial taxonomy began in the 19th century. Initially the classification of bacteria was purely on the basis of phenotypic markers and cell morphology (presence of flagella, shape, colony morphology etc.)(Lehmann and Neumann, 1896). The next addition to taxonomy was chemotaxonomy: the analyses of the chemical constituents of cells as taxonomically useful characteristics. The use of chemotaxonomy helped to classify bacteria and confirm previous taxonomic assignments (Minnikin *et al.*, 1975). In the genomics era, there have been calls to place less emphasis on chemotaxonomy in species descriptions. These calls have included the suggestion that chemotaxonomy is no longer necessary for species delineation, because genomic data is sufficient for this purpose (Vandamme and Sutcliffe, 2021).

1.3.1 Molecular classification

Molecular classification involves grouping organisms according to similarities in their genetic material. Tremendous achievements have been made using DNA-based analyses to classify microorganisms and modern prokaryote taxonomy has been strongly influenced by developments in genetic methods (Tindall *et al.*, 2010).

1.3.1.1 16S-rRNA gene analysis

Fox *et al.* (1977) suggested the characterization of ribosomal nucleic acids (RNAs) as the basis for microbial taxonomy. The 16S rRNA gene was chosen as a way to "comparatively catalogue" with the potential for phylogenetic and general taxonomic classification. The 16S rRNA gene has the following features: functional constancy, ubiquitous distribution in bacteria, a large size (1.6 kb) and the presence of both conserved and variable regions. These characteristics mean it fulfils all the requirements for a phylogenetic marker molecule (Ludwig and Klenk, 2001). 16S rRNA gene sequences are used extensively in phylogenetic analyses, which have helped with the classification of all bacteria (Woese *et al.*, 1990). Phylogenetic trees take evolutionary relationships and varying rates of evolution into consideration and assign species to natural classifications. Classifications based on phylogenetic analyses are theoretically sound (Doolittle, 1999), with 16S rRNA gene sequence variations in archaea and bacteria being seen to provide the backbone for their classification (Tindall *et al.*, 2010).

16S rRNA gene sequences are still invaluable in the description of prokaryotic diversity and are essential in the allocation of unknown microorganisms to a taxon (Kämpfer and Glaeser, 2012). Evidence suggests that a 16S rRNA gene sequence similarity of less than 98.6% (for a 0.5% maximum probability of error) means the compared strains do not belong to the same species (Meier-Kolthoff *et al.*, 2013). Despite being one of the most widely used datasets, 16S rRNA gene sequences have their limitations and there are factors that require consideration when evaluating 16S rRNA gene sequences (Tindall *et al.*, 2010). The main limitation to the 16S rRNA gene is its low resolving power at high levels of relatedness (Stackebrandt and Goebel, 1994). Therefore, it is suggested that, at a threshold sequence similarity level of 98.2% to 99% (depending on the investigated taxonomic group), any species delineation conclusions drawn are supported by DNA-DNA hybridization (DDH) data (Meier-Kolthoff *et al.*, 2013).

1.3.1.2 DNA-DNA hybridization (DDH)

DDH was developed in the 1980s and became the basis for a bacterial species definition (Wayne *et al.*, 1987; Stackebrandt and Goebel, 1994). DDH is still recognised as being the genotypic 'gold standard' for the delineation of prokaryotic

species (Tindall *et al.*, 2010). The underlying principle of DDH is quantifying the reassociation of complementary regions of single stranded DNA of two organisms (Ludwig, 2007). DDH is the required protocol to delineate species when 16S rRNA gene relatedness is greater than 98.6% (for a 0.5% maximum probability of error). The definition of a species when using DDH is a DNA relatedness value by DNA-DNA hybridization of 70% or greater between the bacterial strains in question. A value of less than 70% DNA relatedness by DDH indicates different species (Wayne *et al.*, 1987).

Traditional "wet-lab" DDH can be tedious, complex, time-consuming, error-prone and cannot be used to establish a comparative database. In the era of genomics, this has made traditional DDH obsolete (Richter and Rosselló-Móra, 2009). To overcome the problems of traditional DDH, digital DDH (dDDH) is used. Digital DDH, which is based on whole-genome sequences, shows a better correlation with 16S rRNA gene sequence analyses and can outperform average nucleotide identity (ANI) in some cases (Auch *et al.*, 2010a). The Genome-To-Genome Distance Calculator (GGDC) is an example of putting dDDH into practice. The GGDC (https://www.dsmz.de/services/online-tools/genome-to-genome-distance-calculator-ggdc) is a web server using an *in silico* method, which performs genome-to-genome comparisons and accurately mimics traditional DDH without the challenges of the wet-lab method (Auch *et al.*, 2010b).

1.3.1.3 Average Nucleotide Identity (ANI)

ANI is another method that is effective for the delineation of species (Richter and Rosselló-Móra, 2009). ANI is the average nucleotide identity of the total genomic sequence shared between two strains. It has been shown to be a robust and sensitive method for measuring evolutionary relatedness between closely related bacterial strains. ANI shows a strong linear correlation to DDH values, with the 70% DDH standard corresponding to approximately 95% ANI (Konstantinidis and Tiedje, 2005). If the genome sequence of a prokaryotic strain shows greater than 95-96% ANI to the genome of the type strain of a species with a validly-published name, then it belongs to that genomic species. The ANI threshold for species delineation can, however, differ depending on the specific taxon being studied (Ciufo *et al.*, 2018; Palmer *et al.*, 2020).

1.4 Selective isolation of actinobacteria

The diversity of terrestrial and marine microorganisms has been exploited for many years through the isolation of environmental microbial strains. The isolation of these microorganisms, particularly actinobacteria, has led to the discovery of many natural products of economic value. However, the rate at which novel natural products are being discovered has decreased (Daniel, 2004). This is largely as a result of the rediscovery of known bioactive metabolites, especially antibiotics. There is therefore the need to isolate and characterize poorly studied actinobacterial taxa, allowing us access to novel organisms and the novel bioactive compounds they produce (Connon and Giovannoni, 2002; Goodfellow, 2010). This can be done through the selective isolation of rare actinobacteria (i.e. the actinobacteria that belong to all genera other than *Streptomyces*).

Selective isolation can be applied through all the steps of the isolation process. Pre-treatment of the sample, how the sample is prepared, the choice of selective media, the incubation conditions and colony selection once the microorganisms have grown are all steps during which selection can be applied. Pre-treatments can range from air drying to chemical pre-treatments. They are used to target the isolation of actinobacteria by inhibiting the growth of unwanted organisms, such as Gram-negative bacteria. Sample preparation is done to extract bacteria effectively from particulate matter in the soil or marine sediment. This is important for bacteria that show mycelial growth, because they can bind to soil aggregates.

Colony selection of desired actinobacteria requires experience, as inaccurate selection of actinobacterial colonies can result in a laborious research process (Goodfellow, 2010). On agar culture media, actinobacteria appear compact and leathery. The colonies are also often sunken into the surface of the agar and are frequently covered with aerial mycelium (Anandan *et al.*, 2016). Other parameters, such as the temperature and duration of incubation are also conditions during which selection can be introduced. Even selection conditions, such as the use of gaseous compounds like carbon monoxide have been used to isolate anaerobic actinobacteria (Goodfellow, 2010).

1.4.1 Selective media

Arguably the choice of selective media is where the greatest selection for actinobacteria can be done. For a selective medium, the carbon source, nitrogen source, pH and choice of antibiotics can all be altered to create conditions where a desired taxon can grow and the majority of others cannot (Goodfellow, 2010). This prevents the growth of fast growing microbial species, allowing unique populations of microorganisms to be cultured (Connon and Giovannoni, 2002). With the wide array of actinobacterial diversity, it is impossible to recommend a single selective medium. There are many recommended approaches and, as a result, innumerable media have been suggested for the isolation of actinobacteria (Goodfellow, 2010).

There are many examples of the use of selective media to isolate interesting and novel actinobacteria. Vickers *et al.* (1984) used combinations of amino acids, carbohydrates and antibiotics to favour the isolation of uncommon members of the genus *Streptomyces*. Their raffinose-histidine agar was successful in the selective isolation of many *Streptomyces* species and resulted in the isolation and description of *Streptomyces aureus* and *Streptomyces sanglieri* (Vickers *et al.*, 1984). A similar approach was followed by Tan *et al.* (2006) with the selective isolation of members of the genus *Amycolatopsis* from soil. Three (SM1, SM2 and SM3) of the twenty selective agar media tested supported the growth of all of the representatives of the genus *Amycolatopsis* tested (Tan *et al.*, 2006).

1.5 Metagenomics

Culture-based isolations of bacteria are not representative of the microbial diversity of a sample (Holben and Harris, 1995). It has been estimated that less than 1% of the total microbial diversity in the terrestrial environment (and even less in the marine environment) can be isolated in pure culture (Amann, 1995; Arjun and Harikrishnan, 2011). This leads to a bias in culture-based isolations whereby only 1% or less of microbes that are able to grow on lab media can be cultured, allowing access to only a small proportion of the microbial community (Brady *et al.*, 2009).

To overcome these limitations, culture-independent molecular approaches have been developed. Metagenomics, the study of genetic material recovered directly from the environment, allows us to better describe a microbial community in their native habitat (Arjun and Harikrishnan, 2011). Metagenomic methods do not rely on the isolation and cultivation of microorganisms and therefore reduce the bias associated with culture-based methods (Handelsman *et al.*, 1998).

Metagenomics can be studied through the construction of genetic libraries. The method involves the isolation of DNA directly from the environment. The environmental DNA is then inserted into an appropriate vector and then transformed into a bacterial strain, usually competent *Escherichia coli.* This allows the soil metagenome to be screened (Handelsman *et al.*, 1998). The extraction of nucleic acids followed by PCR amplification of a target gene and the direct cloning of environmental DNA from soil are the two main methods used to isolate environmental DNA for the construction of genetic libraries (Daniel, 2004). Next-generation sequencing (NGS) methods have allowed for metagenomics to have a higher throughput, a simpler library preparation and they require no cloning. This has made it more appealing to do metagenomic studies, which has caused a substantial increase in the number of metagenomic studies performed (Pagani *et al.*, 2012).

Metagenomics has opened the door to new opportunities. The genetic diversity of the rich and underexplored soil metagenome may allow for the discovery of new bioactive compounds and industrial enzymes (Daniel, 2004). Metagenomics also offers the opportunity to exploit biochemical pathways of uncultured microorganisms, such as actinomycetes, through genome mining of uncultured actinobacteria (Niu, 2018).

1.6 Bioactive compounds produced by actinobacteria

Bioactive compounds can be produced by eukaryotic and prokaryotic organisms. These compounds are used and exploited by humans for a variety of medicinal purposes. Plants produce a great variety of compounds and have been used for millennia as medicines (Miao and Davies, 2010). Soil microorganisms have been exploited for many years for their production of bioactive compounds. Most natural products of economic value, such as antibiotics or other pharmaceuticals, are derived from cultured soil microorganisms (Anandan *et al.*, 2016).

Actinomycetes have an extensive secondary metabolism, meaning they produce a wide array of compounds that are not required for growth. They have the ability to synthesize many different biologically active metabolites, such as antibiotics, antiviral compounds, immunosuppressants, herbicides, pesticides, anti-parasitic compounds, siderophores, redox-active agents, anticancer compounds and useful extracellular enzymes (Oskay *et al.*, 2004; Miao and Davies, 2010; van Bergeijk *et al.*, 2020). Actinomycetes are the most prolific producers of microbial bioactive metabolites. Forty five (45) percent of all the bioactive compounds obtained from microbes are produced by actinomycetes (as of 2009) (Selvameenal *et al.*, 2009). Passari *et al.* (2015) showed non-ribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) type I pathways to be the dominant pathways in the production of antimicrobial secondary metabolites in actinomycetes.

Actinomycetes, particularly members of the genus *Streptomyces*, have become renowned as the principal source of therapeutic pharmaceuticals (Miao and Davies, 2010). They have a unique ability to produce many antibiotics, which are used for treating a variety of conditions (Mahajan and Balachandran, 2012).

1.6.1 Antibiotics

Antibiotics started being used in the 1940s. These are antimicrobial compounds that are active against microorganisms of all types and antibiotics are used in medicine to treat microbial infections. The discovery of these antimicrobial agents made the treatment of most types of infectious diseases possible (Martínková *et al.*, 2009).The first actinomycete antibiotic discovered was actinomycin in 1940 (Cassell and Mekalanos, 2001). In 1943, streptomycin was discovered from *Streptomyces griseus* by the Waksman group and this sparked huge interest in the screening of actinomycetes for antibiotics (Comroe, 1978; Mahajan and Balachandran, 2012). Many more antibiotics have since been discovered from actinobacteria. Almost 80% of the world's antibiotics are known to be derived from actinobacteria, mostly from the genera *Streptomyces* and *Micromonospora* (Anandan *et al.*, 2016). Figure 1.1 shows the variety of structural classes of antibiotics produced by actinobacteria, which include all the important antimicrobial drug classes (Genilloud, 2017). There are thousands of antibiotics produced by actinobacteria. *Streptomyces* species alone produce around 7600 described compounds, many of which are potent antibiotics (Anandan *et al.*, 2016).

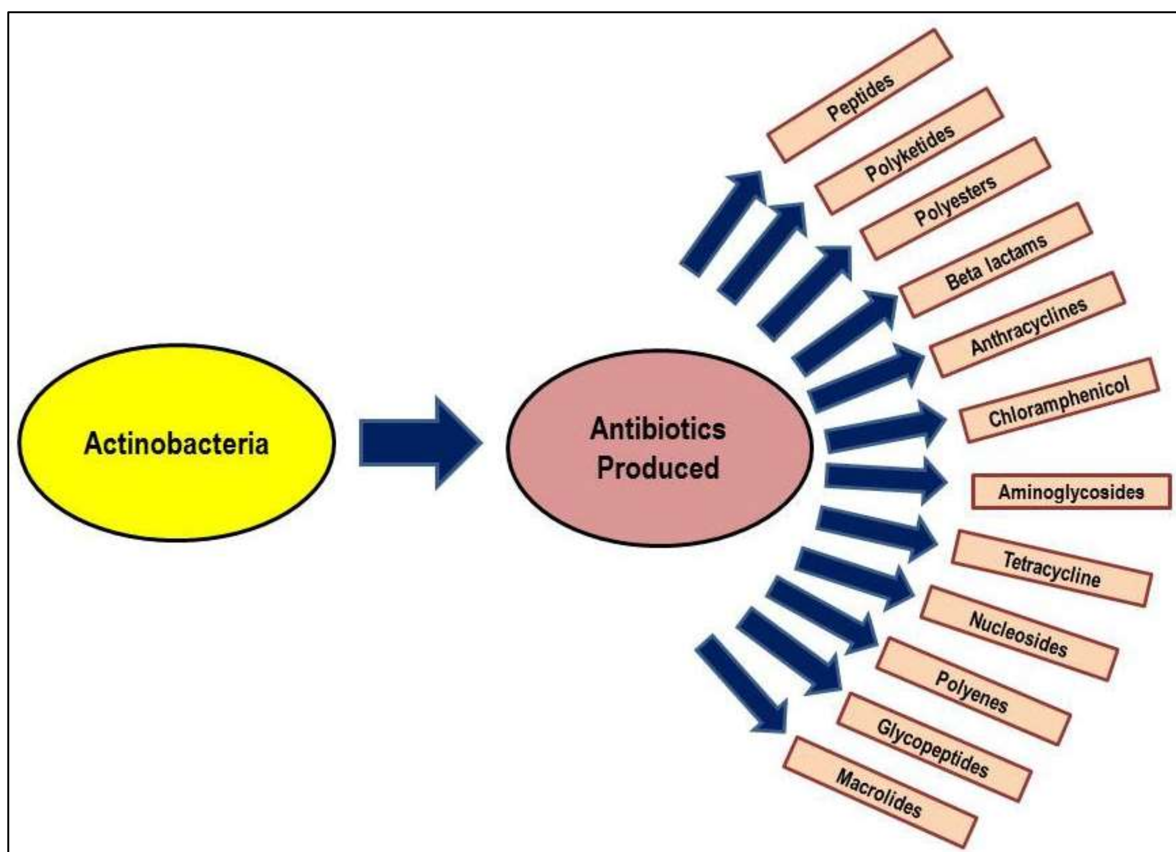

Figure 1.1: Structural classes of antibiotics produced by actinobacteria (Anandan *et al.*, 2016).

It is thought actinobacteria produce antibiotic compounds as biological weaponry to kill competing organisms and gain a competitive advantage (Hibbing *et al.*, 2010; Laskaris *et al.*, 2010). However, when present at sub-inhibitory concentrations antibiotics perform a different function. The result of sensitive promoter-reporter libraries or RNA microarrays shows these compounds to modulate transcription patterns in bacterial and eukaryotic cells at sub-inhibitory concentrations (Yim *et al.*, 2006). It is proposed that these cellular responses occur through the binding of the bioactive compounds to receptors in cells, making them effectively cell-cell signalling agents (Fajardo and Martínez, 2008).

Recently, there has been a dramatic decrease in the success of novel drug discovery through the isolation and screening of actinobacteria. The chance of finding new antibiotics through the screening of randomly chosen actinobacteria is estimated at less than one per million (van Bergeijk *et al.*, 2020). This decrease has coincided with the growing incidence of antimicrobial-resistant pathogens (Niu, 2018).

1.6.2 Antibiotic resistance

Antibiotic resistance is a major healthcare problem. It is the process by which microbial pathogens gain resistance to the antibiotic classes specifically recommended for their treatment (Davies and Davies, 2010). To combat the problem, two things are required: the discovery of new antibiotic compounds with new mechanisms of action and a better understanding of the mechanisms by which antibiotic resistance is developed (Niu, 2018).

Antibiotic resistance is not a new phenomenon. Antibiotic-resistant strains of bacteria are common in the environment irrespective of the use of antibiotics in medicine. In the presence of plasmids, antibiotic resistance can be transferred between any two bacteria through horizontal gene transfer (Passari *et al.*, 2015). Horizontal gene transfer is, however, just one of a wide range of biochemical and physiological mechanisms that may be responsible for resistance. The process by which resistance emerges against antimicrobial agents, like antibiotics, and is spread is extremely complex. This has meant that little has been achieved in the effective prevention and control of the development of drug resistance (Davies and Davies, 2010). Antibiotic resistance can lead to the development of "superbugs" that have enhanced morbidity and mortality due to their resistance to antibiotics (Davies and Davies, 2010). A perfect example of this is tuberculosis (TB).

1.6.2.1 Tuberculosis (TB) antibiotic resistance

TB is a disease caused by the bacillus *Mycobacterium tuberculosis* that has infected humans for nine thousand years and remains the most widely disseminated pathogen worldwide (Hershkovitz *et al.*, 2008). About a quarter of the world's population has been infected with *M. tuberculosis.* It is spread when people sick with TB expel the bacteria into the air by coughing. The disease typically affects the lungs, but can affect other sites in the body. There were approximately 10 million cases and 1.5 million deaths due to TB in 2020, showing that it remains a major healthcare problem worldwide (World Health Organization, 2021).

Today drug-susceptible TB is treated with a 6-month program of antibiotics. The treatment consists of a cocktail of the four first-line anti-TB drugs (isoniazid, rifampicin, ethambutol and pyrazinamide) and tends to have considerable success (World Health Organization, 2021). However, antibiotic resistance has compromised (and continues

to compromise) TB treatment (Sotgiu *et al.*, 2009). Multidrug-resistant (MDR) *M. tuberculosis*, strains resistant to isoniazid and rifampicin, are currently the most prevalent drug resistant strains (World Health Organization, 2021). More recently, extensively drug-resistant (XDR) *M. tuberculosis* strains (Sotgiu *et al.*, 2009) and totally drug resistant (TDR) *M. tuberculosis* strains (Velayati *et al.*, 2009) have emerged. MDR-TB strains affect 3-4% of people diagnosed with TB for the first time and an estimated 18-21% of people previously treated for TB (World Health Organization, 2021). Therefore, in order to effectively treat people in future, there is a need to discover new antibiotics that will be effective against drug resistant *M. tuberculosis* strains.

1.6.3 New methods for discovery of bioactive compounds

It was thought that the diversity of molecules with antibiotic activity was close to being exhausted. The reality is that the majority of easily accessible bioactive compounds, the “low hanging fruit”, have been exploited (Baltz, 2006). However, now we are in the “genomic era” and information from bacterial genome sequencing has revealed a vast reservoir of biosynthetic gene clusters (BGCs) for natural products. Even extensively studied actinobacteria have the capacity to produce many more structurally different bioactive compounds than determined from experiments with laboratory-grown cultures of the organisms (Miao and Davies, 2010; van Bergeijk *et al.*, 2020).

There are two appealing sources for novel natural product discovery: new biosynthetic pathways from uncultured actinobacteria and uncharacterised biosynthetic pathways within the genomes of cultured actinobacteria (Niu, 2018). The failure to uncover the full potential of actinobacterial genomes from cultured actinobacteria is due to our inability to activate the expression of all antibiotic biosynthetic pathways in the laboratory (van Bergeijk *et al.*, 2020). Genomics-driven strategies have emerged that connect unexpressed biosynthetic pathways to chemical entities for novel product discovery (Niu, 2018). Different genome mining methodologies, in combination with powerful computational tools, have reinvigorated the discovery of actinobacteria-derived natural products (Niu, 2018; van Bergeijk *et al.*, 2020).

1.7 Aims

This study had three primary aims: (1) a metagenomic analysis of the actinobacterial diversity in two soil samples, (2) the investigation of the culturable actinobacterial diversity from the same two soil samples and (3) the screening of the isolated actinobacterial strains for antimycobacterial activity. The soil samples were taken from two sites at UCT. The K sample was top soil taken from a wooded area on UCT Middle Campus and the TC sample was top soil taken from a pine forest above Upper Campus. Both top soils were covered in decaying organic matter, making them a good environment for actinobacteria to thrive.

The project started with an assessment of the actinobacterial diversity of the soil from the two sampling sites using a culture-independent metagenomic approach. The results of the metagenomic analyses guided the choice of selective isolation media. For the metagenomic work, an actinobacterial 16S rRNA gene library was created and the actinobacterial diversity was assessed through the sequencing of the gene library. Based on the metagenomic results, four media were chosen for use in isolating actinobacteria from the original soil samples. Isolations were done based on standard plate cultivation techniques.

Isolated actinobacteria were screened for antimycobacterial activity using agar overlays with *Mycobacterium aurum* strain A+ as the target organism. Those strains with strong or very strong activity against *M. aurum* A+ were also screened for activity against *Staphylococcus aureus* ATCC 25923 and *Escherichia coli* ATCC 25922 to assess the antimicrobial spectrum of these isolates.

1.8 References

Amann, R. I. (1995) 'Fluorescently labelled, rRNA-targeted oligonucleotide probes in the study of microbial ecology', Molecular Ecology. John Wiley & Sons, Ltd, pp. 543–554. doi: 10.1111/j.1365-294X.1995.tb00255.x.

Anandan, R. *et al.* (2016) An Introduction to Actinobacteria, Actinobacteria - Basics and Biotechnological Applications. Intech. Available at: https://www.intechopen.com/books/5056.

Arjun, J. K. and Harikrishnan, K. (2011) 'Metagenomic analysis of bacterial diversity in the rice rhizosphere soil microbiome', Biotechnology Bioinf. Bioengineering, 1(3), pp. 361–367.

Auch, A. F. *et al.* (2010a) 'Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison', Standards in Genomic Sciences, 2(1), pp. 117–134. doi: 10.4056/sigs.531120.

Auch, A. F. *et al.* (2010b) 'Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs', Standards in Genomic Sciences, 2(1), pp. 142–148. doi: 10.4056/sigs.541628.

Baltz, R. H. (2006) 'Marcel Faber Roundtable: Is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration?', Journal of Industrial Microbiology and Biotechnology, 33(7), pp. 507–513. doi: 10.1007/S10295-005-0077-9.

Barka, E. A. *et al.* (2016) 'Correction for Barka et al., Taxonomy, Physiology, and Natural Products of Actinobacteria', Microbiology and Molecular Biology Reviews, 80(4). doi: 10.1128/mmbr.00044-16.

Brady, S. F. *et al.* (2009) 'Metagenomic approaches to natural products from free-living and symbiotic organisms', Natural Product Reports. Royal Society of Chemistry, pp. 1488–1503. doi: 10.1039/b817078a.

Bull, A. T. *et al.* (2005) 'Marine actinobacteria: Perspectives, challenges, future directions', Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology, 87(1), pp. 65–79. doi: 10.1007/s10482-004-6562-8.

Cassell, G. H. and Mekalanos, J. (2001) ‘Development of antimicrobial agents in the era of new and reemerging infectious diseases and increasing antibiotic resistance’, Journal of the American Medical Association, 285(5), pp. 601–605. doi: 10.1001/jama.285.5.601.

Chun, J. and Rainey, F. A. (2014) ‘Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea’, International Journal of Systematic and Evolutionary Microbiology, 64(PART 2), pp. 316–324. doi: 10.1099/ijs.0.054171-0.

Ciufo, S. *et al.* (2018) ‘Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI’, International Journal of Systematic and Evolutionary Microbiology, 68(7), pp. 2386–2392. doi: 10.1099/ijsem.0.002809.

Colwell, R. R. (1970) ‘Polyphasic taxonomy of the genus vibrio: numerical taxonomy of *Vibrio cholerae*, *Vibrio parahaemolyticus*, and related *Vibrio* species.’, Journal of Bacteriology, 104(1), pp. 410–433. doi: 10.1128/JB.104.1.410-433.1970.

Comroe, J. H. (1978) ‘Pay dirt - The story of Streptomycin. Part 1. From Waksman to Waksman’, American Review of Respiratory Disease, 117(4 I), pp. 773–781. Available at: https://www.atsjournals.org/doi/pdf/10.1164/arrd.1978.117.4.773 (Accessed: 14 June 2022).

Connon, S. A. and Giovannoni, S. J. (2002) ‘High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates’, Applied and Environmental Microbiology, 68(8), pp. 3878–3885. doi: 10.1128/AEM.68.8.3878-3885.2002.

Conti, R. *et al.* (2016) ‘Endophytic Actinobacteria from the Brazilian Medicinal Plant *Lychnophora ericoides* Mart. and the Biological Potential of Their Secondary Metabolites’, Chemistry and Biodiversity, pp. 727–736. doi: 10.1002/cbdv.201500225.

Cross, T. (1981) ‘Aquatic Actinomycetes: A Critical Survey of the Occurrence, Growth and Role of Actinomycetes in Aquatic Habitats’, Journal of Applied Bacteriology, 50(3), pp. 397–423. doi: 10.1111/j.1365-2672.1981.tb04245.x.

Daniel, R. (2004) ‘The soil metagenome - A rich resource for the discovery of novel natural products’, Current Opinion in Biotechnology. Elsevier Current Trends, pp. 199–204. doi: 10.1016/j.copbio.2004.04.005.

Davies, J. and Davies, D. (2010) ‘Origins and evolution of antibiotic resistance.’, Microbiology and Molecular Biology Reviews, 73(3), pp. 417–433. doi: 10.1128/mmbr.00016-10.

Doolittle, W. F. (1999) ‘Phylogenetic classification and the universal tree’, Science. Science, pp. 2124–2128. doi: 10.1126/science.284.5423.2124.

Fajardo, A. and Martínez, J. L. (2008) ‘Antibiotics as signals that trigger specific bacterial responses’, Current Opinion in Microbiology, 11(2), pp. 161–167. doi: 10.1016/j.mib.2008.02.006.

Fox, G. E. *et al.* (1977) ‘Comparative cataloging of 16S ribosomal ribonucleic acid: molecular approach to procaryotic systematics’, International Journal of Systematic Bacteriology, 27(1), pp. 44–57. doi: 10.1099/00207713-27-1-44.

Genilloud, O. (2017) ‘Actinomycetes: Still a source of novel antibiotics’, Natural Product Reports, 34(10), pp. 1203–1232. doi: 10.1039/c7np00026j.

Goodfellow, M. (2010) ‘Selective Isolation of Actinobacteria’, in Manual of Industrial Microbiology and Biotechnology. John Wiley & Sons, Ltd, pp. 13–27. doi: 10.1128/9781555816827.ch2.

Goodfellow, M. and Williams, S. T. (1983) ‘Ecology of actinomycetes.’, Annual review of microbiology, pp. 189–216. doi: 10.1146/annurev.mi.37.100183.001201.

Grossart, H. P. *et al.* (2004) ‘Antagonistic activity of bacteria isolated from organic aggregates of the German Wadden Sea’, FEMS Microbiology Ecology, 47(3), pp. 387–396. doi: 10.1016/S0168-6496(03)00305-2.

Handelsman, J. *et al.* (1998) ‘Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products’, Chemistry and Biology, 5(10). doi: 10.1016/S1074-5521(98)90108-9.

Hershkovitz, I. *et al.* (2008) ‘Detection and molecular characterization of 9000-year-old *Mycobacterium tuberculosis* from a neolithic settlement in the Eastern mediterranean’, PLoS ONE, 3(10), p. e3426. doi: 10.1371/journal.pone.0003426.

Hibbing, M. E. *et al.* (2010) ‘Bacterial competition: Surviving and thriving in the microbial jungle’, Nature Reviews Microbiology. NIH Public Access, pp. 15–25. doi: 10.1038/nrmicro2259.

Holben, W. E. and Harris, D. (1995) ‘DNA-based monitoring of total bacterial community structure in environmental samples’, Molecular Ecology, pp. 627–632. doi: 10.1111/j.1365-294X.1995.tb00263.x.

Kämpfer, P. and Glaeser, S. P. (2012) ‘Prokaryotic taxonomy in the sequencing era - the polyphasic approach revisited’, Environmental Microbiology, 14(2), pp. 291–317. doi: 10.1111/j.1462-2920.2011.02615.x.

Konstantinidis, K. T. and Tiedje, J. 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(7), pp. 2567–2572. doi: 10.1073/pnas.0409727102.

Laskaris, P. *et al.* (2010) ‘Coevolution of antibiotic production and counter-resistance in soil bacteria’, Environmental Microbiology, 12(3), pp. 783–796. doi: 10.1111/j.1462-2920.2009.02125.x.

Lehmann, K. B. and Neumann, R. O. (1896) ‘Atlas und Grundriss der Bakteriologie und Lehrbuch der Speciellen Bakteriologischen Diagnostik’.

Ludwig, W. (2007) ‘Nucleic acid techniques in bacterial systematics and identification’, International Journal of Food Microbiology, 120(3), pp. 225–236. doi: 10.1016/j.ijfoodmicro.2007.06.023.

Ludwig, W. and Klenk, H.-P. (2001) ‘Overview: A Phylogenetic Backbone and Taxonomic Framework for Procaryotic Systematics’, in Bergey’s Manual® of Systematic Bacteriology, pp. 49–65. doi: 10.1007/978-0-387-21609-6_8.

Mahajan, G. B. and Balachandran, L. (2012) ‘Antibacterial agents from actinomycetes’, Frontiers in Bioscience, (4), pp. 240–253. doi: 10.1109/epqu.2011.6128966.

Martínková, L. *et al.* (2009) ‘Biodegradation potential of the genus *Rhodococcus*, Environment International, 35(1), pp. 162–177. doi: 10.1016/j.envint.2008.07.018.

Meier-Kolthoff, J. P., Göker, M., Spröer, C., & Klenk, H. P. (2013). When should a DDH experiment be mandatory in microbial taxonomy? Archives of Microbiology, 195(6), 413–418. https://doi.org/10.1007/s00203-013-0888-4

Miao, V. and Davies, J. (2010) 'Actinobacteria: The good, the bad, and the ugly', in Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology, pp. 143–150. doi: 10.1007/s10482-010-9440-6.

Minnikin, D. E. *et al.* (1975) 'Differentiation of *Mycobacterium*, *Nocardia* and related taxa by thin layer chromatographic analysis of whole organism methanolysates', Journal of General Microbiology, 88(1), pp. 200–204. doi: 10.1099/00221287-88-1-200.

Niu, G. (2018) 'Genomics-Driven Natural Product Discovery in Actinomycetes', Trends in Biotechnology, 36(3), pp. 238–241. doi: 10.1016/j.tibtech.2017.10.009.

Oskay, M. *et al.* (2004) 'Antibacterial activity of some actinomycetes isolated from farming soils of Turkey', African Journal of Biotechnology, 3(9), pp. 441–446. doi: 10.5897/ajb2004.000-2087.

Pagani, I. *et al.* (2012) 'The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata', Nucleic Acids Research, 40(D1), pp. D571–D579. doi: 10.1093/NAR/GKR1100.

Palmer, M. *et al.* (2020) 'All anis are not created equal: Implications for prokaryotic species boundaries and integration of anis into polyphasic taxonomy', International Journal of Systematic and Evolutionary Microbiology, 70(4), pp. 2937–2948. doi: 10.1099/ijsem.0.004124.

Passari, A. K. *et al.* (2015) 'Isolation, abundance and phylogenetic affiliation of endophytic actinomycetes associated with medicinal plants and screening for their in vitro antimicrobial biosynthetic potential', Frontiers in Microbiology, 6(APR). doi: 10.3389/fmicb.2015.00273.

Richter, M. and Rosselló-Móra, R. (2009) 'Shifting the genomic gold standard for the prokaryotic species definition', Proceedings of the National Academy of Sciences of the United States of America, 106(45), pp. 19126–19131. doi: 10.1073/pnas.0906412106.

Rosselló-Mora, R. and Amann, R. (2001) 'The species concept for prokaryotes', FEMS Microbiology Reviews, 25(1), pp. 39–67. doi: 10.1111/j.1574-6976.2001.tb00571.x.

Salwan, R. and Sharma, V. (2020) 'Molecular and biotechnological aspects of secondary metabolites in actinobacteria', Microbiological Research. Elsevier GmbH. doi: 10.1016/j.micres.2019.126374.

Selvameenal, L. *et al.* (2009) 'Antibiotic pigment from desert soil actinomycetes; Biological activity, purification and chemical screening', Indian Journal of Pharmaceutical Sciences, 71(5), pp. 499–504. doi: 10.4103/0250-474X.58174.

Sotgiu, G. *et al.* (2009) 'Epidemiology and clinical management of XDR-TB: A systematic review by TBNET', European Respiratory Journal, pp. 871–881. doi: 10.1183/09031936.00168008.

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', International Journal of Systematic Bacteriology. Microbiology Society, pp. 846–849. doi: 10.1099/00207713-44-4-846.

Tan, G. Y. A. *et al.* (2006) 'Exploration of *Amycolatopsis* diversity in soil using genus-specific primers and novel selective media', Systematic and Applied Microbiology, 29(7), pp. 557–569. doi: 10.1016/j.syapm.2006.01.007.

Tindall, B. J. *et al.* (2010) 'Notes on the characterization of prokaryote strains for taxonomic purposes', International Journal of Systematic and Evolutionary Microbiology, 60(1), pp. 249–266. doi: 10.1099/ijs.0.016949-0.

Torsvik, V. and Øvreås, L. (2002) 'Microbial diversity and function in soil: From genes to ecosystems', Current Opinion in Microbiology. Elsevier Current Trends, pp. 240–245. doi: 10.1016/S1369-5274(02)00324-7.

van Bergeijk, D. A. *et al.* (2020) 'Ecology and genomics of Actinobacteria: new concepts for natural product discovery', Nature Reviews Microbiology, 18(10), pp. 546–558. doi: 10.1038/s41579-020-0379-y.

van Keulen, G. *et al.* (2005) ‘Gas vesicles in actinomycetes: old buoys in novel habitats?’, Trends in Microbiology, 13(8), pp. 347–350. doi: 10.1016/j.tim.2005.06.002.

Vandamme, P. and Sutcliffe, I. (2021) ‘Out with the old and in with the new: Time to rethink twentieth century chemotaxonomic practices in bacterial taxonomy’, International Journal of Systematic and Evolutionary Microbiology. Microbiology Society, p. 5127. doi: 10.1099/ijsem.0.005127.

Velayati, A. A. *et al.* (2009) ‘Emergence of new forms of totally drug-resistant tuberculosis bacilli: Super extensively drug-resistant tuberculosis or totally drug-resistant strains in Iran’, Chest, 136(2), pp. 420–425. doi: 10.1378/chest.08-2427.

Vickers, J. C. *et al.* (1984) ‘A taxonomic approach to selective isolation of streptomycetes from soil’, in Biological, Biochemical, and Biomedical Aspects of Actinomycetes, pp. 553–561. doi: 10.1016/b978-0-12-528620-6.50045-8.

Wayne, L. G. *et al.* (1987) ‘Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics’, Systematic and Applied Microbiology, 37(4), pp. 463–464. doi: 10.1016/S0723-2020(88)80020-1.

Woese, C. R. *et al.* (1990) ‘Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya’, Proceedings of the National Academy of Sciences of the United States of America, 87(12), pp. 4576–4579. doi: 10.1073/pnas.87.12.4576.

World Health Organization (2021) Global Tuberculosis Report 2021, Licence: CC BY-NC-SA 3.0 IGO.

Yim, G. *et al.* (2006) ‘The truth about antibiotics’, International Journal of Medical Microbiology, 296(2–3), pp. 163–170. doi: 10.1016/j.ijmm.2006.01.039.

Zhi, X. Y. *et al.* (2012) ‘Prokaryotic systematics in the genomics era’, in Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology, pp. 21–34. doi: 10.1007/s10482-011-9667-x.

Chapter 2:

Metagenomic analysis: the actinobacterial diversity of two soil samples from the University of Cape Town using culture-independent methods

Summary

Metagenomic analysis was used to analyse the actinobacterial diversity in the K and TC soil samples from the UCT. Fifty (50) and 52 16S-rRNA gene sequences were obtained for the K and TC samples, respectively. The 16S-rRNA gene sequences were amplified from eDNA using the actinobacterium-specific 16S-rRNA gene primers Com 2xf and Ac118r. The 16S-rRNA gene sequences were analysed using the EzBioCloud database and the data obtained were used for subsequent analyses. All of the clone sequences were identified as being actinobacterial, with 40% of both the K and TC sample clone sequences having top hits that were uncultured actinobacteria. Rarefaction analyses based on the gene library sequences of both the K and TC samples indicated that the total actinobacterial diversity was not represented by either of the clone libraries. Pie charts were generated to analyse the diversity of the clone libraries based on sequence top hits to the order level. The K and TC samples both had 10 orders present in the diversity analysis. Eight orders (*Acidimicrobiales, Micrococcales, Micromonosporales, Mycobacteriales, Propionibacteriales, Pseudonocardiales, Streptomycetales* and *Streptosporangiales*) were present in both the TC and K clone libraries. The K clone library was most strongly represented by *Acidimicrobiales* (26%), *Pseudonocardiales* (20%) and *Propionibacteriales* (16%). The TC clone library was most strongly represented by *Acidimicrobiales* (36.5%) and *Pseudonocardiales* (21.2%).

Table of Contents

2.1 Introduction 32
2.2. Materials and methods 34
2.2.1 Sample Collection 34
2.2.2 Environmental DNA extraction 34
2.2.3 Two-step 16S-rRNA gene PCR amplification.. 34
2.2.4 DNA ligation and transformation 35
2.2.5 Plasmid DNA amplification and sequencing.. 35
2.2.6 Metagenomic sequence identification and analysis 36
2.2.7 Analysis of Biodiversity 36
2.3 Results and discussion... 37
2.3.1 Clone sequencing and rarefaction analysis... 37
2.3.2 Actinobacterial composition of the soil samples 41
2.3.2.1 Order *Pseudonocardiales* 43
2.3.2.2 Order *Propionibacteriales* 44
2.3.2.3 Order *Acidimicrobiales*.. 45
2.3.2.4 Genus Streptomyces 46
2.4 References.. 47

2.1 Introduction

Less than 1% of the total microbial diversity in the terrestrial environment can be isolated in pure culture (Amann, 1995; Arjun and Harikrishnan, 2011). There is therefore a bias in culture-based studies whereby only 1% or less of microbes that are able to grow on laboratory media can be cultured. This bias, therefore, allows access to only a small proportion of the microbial community of a sample, thereby underestimating the microbial diversity of the samples being studied (Brady *et al.*, 2009; Vartoukian *et al*., 2010).

Culture independent analysis, the study of genetic material recovered directly from the environment, allows us to better describe a microbial community in its native habitat without the limitations of culture-dependent methods (Arjun and Harikrishnan, 2011). Culture-independent molecular approaches do not rely on the isolation and cultivation of microorganisms and therefore negate the bias associated with culture-based methods. Metagenomic analysis allows access to all the DNA sequences in a sample, including the previously inaccessible sequences of the 99% of microbes that cannot be cultured under laboratory conditions (Handelsman *et al.*, 1998). It can therefore be used to identify new and rare bacterial species (Brady *et al.*, 2009).

Culture-independent analysis of a microbial community can be performed through the construction of genetic libraries or through the use of next generation sequencing (NGS). The construction of genetic libraries involves the isolation of DNA directly from the environment. This eDNA can be directly inserted into an appropriate vector. Alternatively, PCR amplification of target gene sequences (such as the amplification of the 16S-rRNA gene) can be performed on the eDNA followed by the cloning of PCR products into appropriate vectors. If the intention is to create small insert libraries, plasmids are the vector of choice. Alternatively, if large insert libraries are being created, bacterial artificial chromosomes (BACs) are the appropriate choice of vector (Daniel, 2004). The vectors are then transformed into a bacterial strain, usually competent *Escherichia coli.* This allows the metagenome to be screened.

Like culture-based isolation techniques, culture-independent techniques are subject to limitations. One major limitation is the PCR amplification process where preferential amplification can occur for certain species in a sample (Von Wintzingerode *et al.*, 1997) or the primers used in PCR can have inadequate specificity for the chosen taxa (Sipos *et al.*, 2007). Despite these limitations the PCR based approach for creating genetic libraries is one of the best tools for accessing the diversity of a microbial community (Xin *et al.*, 2008).

The aim of this part of the study was to analyse the actinobacterial diversity in two soil samples using culture-independent techniques. The metagenomic analyses of the two samples served to aid in the choice of selective media for the culture-based isolations discussed in Chapter 3.

2.2. Materials and methods

2.2.1 Sample Collection

Soil samples were collected on the 27th January 2020 at two different locations on UCT property: the TC soil sample was top soil collected in the pine forest below the UCT tennis courts on Upper Campus (33°57'24.2"S 18°27'32.7"E) and the K soil sample was top soil collected in a wooded area near the Kramer Building on UCT Middle Campus (33°57'30.0"S 18°27'54.0"E). The soil samples were collected in sterile Falcon® conical centrifuge tubes. They were processed for eDNA and the pH of the sediments measured on the day of sample collection. The remaining soil from the K and TC soil samples was placed into sterile Petri dishes and allowed to dry slowly at room temperature for later culture-based isolations.

2.2.2 Environmental DNA extraction

0.5g of soil from each soil sample was used for eDNA extraction using the ZR Soil Microbe DNA miniprep Kit™ (Zymo Research, Irvine, California). The protocol was followed as per the manufacturer's instructions, with 2 X 0.25g of soil used for DNA extraction from each soil sample.

2.2.3 Two-step 16S-rRNA gene PCR amplification

Actinobacteria-specific 16S-rRNA gene sequences were amplified from the eDNA using a two-step, or nested, PCR program. The first step being a 50 µl PCR reaction using the F1-R5 universal 16S-rRNA gene primer combination described by Cook and Meyers (2003). For the template DNA in the reaction, 2 µl of eDNA was added to the 50 µl reaction. The following cycle parameters were used: initial denaturation (96°C for 2 min), 30 cycles of denaturation (96°C for 30s), annealing (56°C for 20s), and extension (72°C for 30s), and a final extension (72°C for 5 min). The reaction yielded a ≈1500 bp 16S-rRNA gene product.

The second step was a 25 µl PCR reaction using the Com 2xf-Ac118r actinobacterium-specific 16S-rRNA gene primer set (Schäfer *et al*., 2010). For template DNA: 2 µl of the F1-R5 reaction was added to the second, 25 µl reaction. The reaction components and cycle parameters for the second reaction were as described in Schäfer *et al*. (2010) with the only change being a denaturation temperature of 96°C instead of 94°C. The following cycle parameters were used: initial denaturation (96°C for 3 min), 30 cycles of denaturation (96°C for 30s), annealing (60°C for 30s), and extension (72°C

for 30s), and a final extension (72°C for 15 min). The reaction yielded a ≈273 bp 16S-rRNA gene product. All PCR reactions were carried out using a Bio-Rad T100™ thermocycler PCR machine.

For each reaction: a 2 mM $MgCl_2$ concentration, 150μM of each dNTP and 1U Super-Therm *Taq* polymerase (JMR Holdings, USA) was used. PCR products were electrophoresed on 0.8% (w/v) agarose gels containing 0.8μg/ml ethidium bromide (EtBr) at 90V in 1 X Tris-acetate-EDTA (TAE) buffer. *Pst*I-digested bacteriophage lambda (λ) DNA was used as a molecular-weight marker. All electrophoresis gels were visualized using a ChemiDoc™ XRS+ Molecular Imager® illuminometer (Bio-Rad) at 260nm. The PCR amplified DNA was purified using the ISOLATE II PCR and Gel Kit (Bioline, www.bioline.com) following the manufacturer's instructions.

2.2.4 DNA ligation and transformation

The purified 16S-rRNA gene sequences were ligated into the pGEM®-T Easy Vector System (Promega, USA). The ligation mixture was then transformed into *E. coli* DH5α competent cells (Bioline, London, United Kingdom) and plated onto Luria-Bertani (LB) agar plates containing 100 μg/ml ampicillin (LB amp). The recommended transformation protocol was followed with the following changes: in step 2, 5 μl of the ligation reaction was added instead of 2 μl; in step 4: 15 μl of competent cells were added instead of 50 μl and in step 9, the solution was incubated for 1 hour at 37°C without shaking instead of 1.5 hours with shaking. All white colonies (suspected positive recombinants) from blue-white selection were numbered and sub-cultured onto clean LB amp plates.

2.2.5 Plasmid DNA amplification and sequencing

A 50 μl colony PCR reaction was performed on suspected positive recombinant colonies using the M13F-M13R primer set. Colonies from LB amp plates where picked up with sterile tooth picks and dipped into the reaction mixture as the template DNA. For the reaction: a 2 mM $MgCl_2$ concentration, 150μM of each dNTP and 1U Super-Therm *Taq* polymerase (JMR Holdings, USA) was used. The reaction had the following cycle parameters: initial denaturation (96°C for 2 min), 30 cycles of denaturation (96°C for 30s), annealing (56°C for 20s), and extension (72°C for 20s), and a final extension (72°C for 5 min). Primers M13F and M13R bind in the pGEM®-T Easy vector. The expected product size of the PCR product was ≈500 bp.

The presence of the 16S-rRNA gene insert was determined by agarose gel electrophoresis, as described in section 2.2.3. Positive recombinants, containing a cloned 16S-rRNA gene insert, were inoculated into 5 ml of LB amp liquid medium in sterile universal containers, incubated at 37°C with shaking for 16-18 h and then the cultures were subjected to plasmid DNA extraction. Plasmid DNA extraction was performed using the Zyppy™ Plasmid Miniprep Kit (Zymo Research, Irvine, USA) following the manufacturer's instructions. The concentrations of the purified plasmid DNAs were measured using a Nanodrop™ ND-1000 spectrophotometer and then diluted to 100 ng/µl in Zyppy™ Plasmid Miniprep Kit elution buffer. Purified plasmids were sent to Macrogen Europe Inc., (Amsterdam-Zuidoost, Netherlands) in a 96 well plate for sequencing.

2.2.6 Metagenomic sequence identification and analysis

Returned clone sequences were edited using Chromas version 2.6.4 (Technelysium Pty Ltd, Australia) and analysed using the EzBioCloud database (www.ezbiocloud.net) (Yoon *et al.*, 2017) to determine the taxonomic order to which each cloned 16S-rRNA gene sequence belongs. The results of the EzBioCloud search (sequence length and information associated with the top hit) were recorded (Tables A1 and A2). Sequences suspected of being duplicates were analysed by pairwise sequence alignments in DNAMAN version 4.13 program (Lynnon BioSoft). All duplicate sequences (100% sequence similarity to another sequence) were removed such that only one copy was included in the analysis.

2.2.7 Analysis of Biodiversity

Actinobacterial rarefaction curves (Gotelli and Colwell, 2001) were plotted for both the K and the TC sample clone libraries in Microsoft Excel. The number of unique phylotypes (<99% 16S-rRNA gene sequence similarity to their top hit on EzBioCloud) was plotted against the number of sequenced clones. Pie charts to represent the sample biodiversity were generated using Microsoft Excel based on the classification of the top hit of each clone sequence to the taxonomic level of order.

2.3 Results and discussion

2.3.1 Clone sequencing and rarefaction analysis

Both the K sample and the TC sample originally each had 55 clones sequenced. After identical clone sequences were removed, the K and TC samples had a total of 50 and 52 clone sequences, respectively. Tables 2.1 and 2.2 show the summarised EzBioCloud results for both samples after clone dereplication. The complete EzBioCloud results for the metagenomic clones can be seen in Tables A1 and A2.

The length of the sequenced 16S-rRNA gene inserts ranged from 260 bp to 280 bp in length (Tables 2.1 and 2.2). Therefore, only a small section of the 16S rRNA gene length (1.6 kb) is covered in the insert sequences. This sequence length of the 16S-rRNA gene is sufficient for reliable genus identification of the clone inserts, but not species identification. The naming of species based on short-stretch environmental Operational Taxonomic Units (OTUs), or sequences, constitutes “shaky scientific grounds” (Stackebrandt *et al.*, 2022). Therefore, in analysing the data, the sequences were analysed at the higher taxonomic rank of genus.

Table 2.1: Sequence similarity of the Kramer soil sample clone sequences to their top hit in the EzBioCloud database. The sequence length and the top hit order are also shown.

Clone no.	Sequence length	Closest relative (GenBank Accession Number)	% Similarity	Order
1	272	Uncultured Bacterial Clone (AY792234)	96.7	*Acidimicrobiales*
13	272	Uncultured Bacterial Clone (FJ592837)	97.1	*Acidimicrobiales*
21	271	Uncultured Bacterial Clone (AJ241005)	96.3	*Acidimicrobiales*
26	272	Uncultured Bacterial Clone (FJ478790)	97.4	*Acidimicrobiales*
32	272	Uncultured Bacterial Clone (FJ592837)	97.8	*Acidimicrobiales*
36	272	Uncultured Bacterial Clone (JX098497)	98.9	*Acidimicrobiales*
7	272	Uncultured Bacterial Clone (FJ478778)	98.2	*Acidimicrobiales*
8	271	Uncultured Bacterial Clone (JF179655)	99.3	*Acidimicrobiales*
11	272	Uncultured Bacterial Clone (EU753662)	96.3	*Acidimicrobiales*
33	272	Uncultured Bacterial Clone (DQ129383)	97.4	*Acidimicrobiales*
34	272	Uncultured Bacterial Clone (EF516495)	99.6	*Acidimicrobiales*
43	273	Uncultured Bacterial Clone (PAC000665)	98.2	*Acidimicrobiales*
48	273	Uncultured Bacterial Clone (PAC002681)	97.1	*Acidimicrobiales*
22	272	Uncultured Bacterial Clone (HQ910322)	98.2	*Frankiales*
41	271	Uncultured Bacterial Clone (EU132518)	99.6	*Frankiales*
9	272	*Kineosporia rhamnosa* (AB003935)	100.0	*Kineosporiales*
6	271	*Naasia aerilata* (JQ639051)	98.5	*Microbacteriales*
12	273	*Pseudolysinimonas kribbensis* (EF466129)	100.0	*Microbacteriales*
18	272	*Diaminobutyricimonas aerilata* (JQ639052)	100.0	*Microbacteriales*
40	277	*Cryobacterium psychrophilum* (AJ544063)	95.2	*Microbacteriales*
44	273	*Naasia aerilata* (JQ639051)	95.2	*Microbacteriales*
49	275	*Leifsonia lichenia* (AB278552)	97.1	*Microbacteriales*
20	276	*Arthrobacter humicola* (AB279890)	99.6	*Micrococcales*
24	272	*Hamadaea tsunoensis* (AUAX01000066)	97.8	*Micromonosporales*
37	272	*Micromonospora rifamycinica* (LRMV01000349)	99.3	*Micromonosporales*
15	279	Millisia brevis (BCRN01000038)	95.6	*Mycobacteriales*
35	273	*Mycolicibacterium confluentis* (LQOQ01000043)	98.2	*Mycobacteriales*
38	273	*Mycobacterium cookii* (AF480598)	100.0	*Mycobacteriales*
5	270	*Nocardioides dilutus* (EF466121)	98.9	*Propionibacteriales*
16	272	*Microlunatus phosphovorus* (AP012204)	100.0	*Propionibacteriales*
17	271	*Aeromicrobium ginsengisoli* (AB245394)	100.0	*Propionibacteriales*
27	272	*Nocardioides iriomotensis* (AB544079)	97.4	*Propionibacteriales*
28	270	*Nocardioides daejeonensis* (JF937066)	98.2	*Propionibacteriales*
30	298	Uncultured Bacterial Clone (JF202912)	99.6	*Propionibacteriales*
46	280	*Kribbella jiaozuonensis* (MK817655)	99.6	*Propionibacteriales*
50	273	Uncultured Bacterial Clone (JX079124)	96.7	*Propionibacteriales*
2	277	*Amycolatopsis azurea* (ANMG01000133)	97.8	*Pseudonocardiales*
3	273	*Actinomycetospora chiangmaiensis* (ARBI01000033)	100.0	*Pseudonocardiales*
19	273	*Actinophytocola corallina* (AB511316)	99.6	*Pseudonocardiales*
23	273	*Pseudonocardia spinosispora* (AJ249206)	98.5	*Pseudonocardiales*
25	273	*Actinomycetospora lutea* (AB514515)	98.5	*Pseudonocardiales*
31	273	*Amycolatopsis mediterranei* (AJ293754)	99.3	*Pseudonocardiales*
39	273	*Pseudonocardia antarctica* (AJ576010)	98.5	*Pseudonocardiales*
42	273	*Amycolatopsis mediterranei* (AJ293754)	99.3	*Pseudonocardiales*
45	274	*Actinophytocola sediminis* (KJ013500)	97.1	*Pseudonocardiales*
47	274	*Pseudonocardia alaniniphila* (EU722519)	100.0	*Pseudonocardiales*
14	278	*Streptomyces fragilis* (AY999917)	99.3	*Streptomycetales*
4	278	Uncultured Bacterial Clone (PAC000666)	95.2	*Streptosporangiales*
10	271	Uncultured Bacterial Clone (PAC000340)	99.3	*Streptosporangiales*
29	272	Uncultured Bacterial Clone (JF135243)	99.3	*Streptosporangiales*

Table 2.2: Sequence similarity of the Tennis Court soil sample clone sequences to their top hit in the EzBioCloud database. The sequence length and top hit order are also shown.

Clone no.	Sequence length	Closest relative (GenBank Accession Number)	% Similarity	Order
28	272	Uncultured Bacterial Clone (EF516583)	98.2	*Acidimicrobiales*
52	272	Uncultured Bacterial Clone (JQ427076)	98.5	*Acidimicrobiales*
4	272	Uncultured Bacterial Clone (GQ397040)	96.7	*Acidimicrobiales*
6	272	Uncultured Bacterial Clone (PAC003044)	99.6	*Acidimicrobiales*
7	272	Uncultured Bacterial Clone (PAC000665)	98.5	*Acidimicrobiales*
8	273	Uncultured Bacterial Clone (JF216358)	98.2	*Acidimicrobiales*
10	273	Uncultured Bacterial Clone (JQ407953)	97.1	*Acidimicrobiales*
20	272	AY234624_s (AY234624)	100.0	*Acidimicrobiales*
25	272	Uncultured Bacterial Clone (JF179655)	98.8	*Acidimicrobiales*
29	272	Uncultured Bacterial Clone (PAC000665)	96.7	*Acidimicrobiales*
30	272	Uncultured Bacterial Clone (PAC000665)	98.5	*Acidimicrobiales*
34	272	Uncultured Bacterial Clone (JF216358)	96.3	*Acidimicrobiales*
35	272	Uncultured Bacterial Clone (EF516495)	97.8	*Acidimicrobiales*
38	277	Uncultured Bacterial Clone (JQ978888)	96.0	*Acidimicrobiales*
39	272	Uncultured Bacterial Clone (EF516495)	98.5	*Acidimicrobiales*
41	272	Uncultured Bacterial Clone (PAC000665)	98.5	*Acidimicrobiales*
43	272	Uncultured Bacterial Clone (GQ397040)	97.1	*Acidimicrobiales*
47	258	*Desertimonas flava* (QKYK01000032)	100.0	*Acidimicrobiales*
50	272	Uncultured Bacterial Clone (PAC003044)	95.6	*Acidimicrobiales*
21	272	*Catenulispora acidiphila* (CP001700)	100.0	*Catenulisporales*
23	272	*Jatrophihabitans endophyticus* (jgi.1107904)	97.1	*Jatrophihabitantales*
24	271	*Jatrophihabitans endophyticus* (jgi.1107904)	100.0	*Jatrophihabitantales*
40	271	*Jatrophihabitans endophyticus* (jgi.1107904)	95.9	*Jatrophihabitantales*
3	274	*Lysinimonas yzui* (MG934573)	98.2	*Microbacteriales*
26	273	*Curtobacterium citreum* (X77436)	95.2	*Microbacteriales*
49	260	*Demequina lutea* (AB639016)	98.9	*Micrococcales*
42	259	AP019371_s (AP019371)	96.1	*Micromonosporales*
12	273	Uncultured Bacterial Clone (EF018761)	99.3	*Mycobacteriales*
18	273	*Mycolicibacterium mucogenicum* (AY457074)	97.4	*Mycobacteriales*
31	273	*Mycolicibacterium tokaiense* (AF480590)	97.1	*Mycobacteriales*
5	271	*Nocardioides halotolerans (*KE383922)	98.5	*Propionibacteriales*
37	271	*Nocardioides daejeonensis* (JF937066)	97.1	*Propionibacteriales*
46	282	*Nocardioides albus* (AF004988)	96.7	*Propionibacteriales*
48	272	*Nocardioides allogilvus* (MG800321)	98.2	*Propionibacteriales*
9	273	*Pseudonocardia bannensis* (FJ817375)	96.7	*Pseudonocardiales*
13	273	*Actinokineospora enzanensis* (ARFV01000009)	98.2	*Pseudonocardiales*
14	273	*Pseudonocardia tetrahydrofuranoxydans* (AJ249200)	99.6	*Pseudonocardiales*
15	273	*Amycolatopsis helveola* (AB327254)	99.6	*Pseudonocardiales*
16	274	Uncultured Bacterial Clone (JF173550)	98.5	*Pseudonocardiales*
19	273	*Amycolatopsis azurea* (ANMG01000133)	100.0	*Pseudonocardiales*
22	273	*Amycolatopsis pigmentata* (AB327253)	97.8	*Pseudonocardiales*
27	273	Uncultured Bacterial Clone (DQ129564)	99.6	*Pseudonocardiales*
33	273	*Actinophytocola corallina* (AB511316)	100.0	*Pseudonocardiales*
45	273	*Actinophytocola corallina* (AB511316)	99.6	*Pseudonocardiales*
51	274	*Amycolatopsis mediterranei* (AJ293754)	99.6	*Pseudonocardiales*
1	278	*Streptomyces haliclonae* (AB473554)	99.6	*Streptomycetales*
2	272	*Streptomyces regensis* (LFVR01000689)	100.0	*Streptomycetales*
32	262	*Streptomyces cinnabarinus* (JNXQ01000585)	99.6	*Streptomycetales*
36	280	*Streptomyces fragilis* (AY999917)	100.0	*Streptomycetales*
44	269	*Streptomyces niveiscabiei* (AF361786)	98.5	*Streptomycetales*
11	279	Uncultured Bacterial Clone (PAC000666)	93.4	*Streptosporangiales*
17	272	*Actinoallomurus amamiensis* (AB364583)	97.1	*Streptosporangiales*

All of the bacterial clones sequenced had 16S-rRNA gene inserts belonging to the phylum *Actinomycetota,* to which all actinobacteria belong. This shows the Com 2xf-Ac118r primer set to have a high specificity for sequence amplification from actinobacteria (Schäfer *et al*., 2010). Forty percent (40%) of the clone sequences from both the K sample and the TC sample were found to have sequences with top hits to uncultured bacteria (Tables 2.1 and 2.2). These sequences represent actinobacteria in the soil samples that potentially cannot be cultured on laboratory media.

Top hits on EzBioCloud with less than 99% sequence similarity represented unique phylotypes in the samples: 29 of 50 clones from the K sample and 35 of 52 clones from the TC sample were determined to be unique phylotypes. The threshold value of <99% was chosen based on the recommended threshold value for short-stetch OTUs (Stackebrandt *et al.*, 2022).

Figure 2.1 shows the rarefaction curves generated for the clone sequences of the K and TC samples. Rarefaction curves compare the sequencing data volume to the species richness of different samples to indicate whether the quantity of sequences was sufficient to capture the biodiversity in the sample analysed (Kuang *et al.*, 2018). For a rarefaction curve to indicate that the full biodiversity of a sample has been captured, the gradient of the curve must decrease significantly, eventually reaching an asymptote, which means that no new phylotypes would be found if additional sequences were analysed (Yang *et al.*, 2005). In Figure 2.1, the orange curve depicts the rarefaction curve for the TC sample clone sequences and the blue curve depicts the rarefaction curve for the K sample clone sequences.

Neither the orange nor the blue curve reaches an asymptote (Figure 2.1). This indicates that, if more clone sequences had been analysed, additional new phylotypes, and thus more actinobacterial diversity, would have been found. The clone library sequences for both samples, therefore, do not capture the full biodiversity of the soil samples. More clone sequences would need to be generated and analysed to reveal more of the actinobacterial diversity in the K and TC soils. This was not done in this study due to time lost as a result of the lockdown during the Covid-19 pandemic.

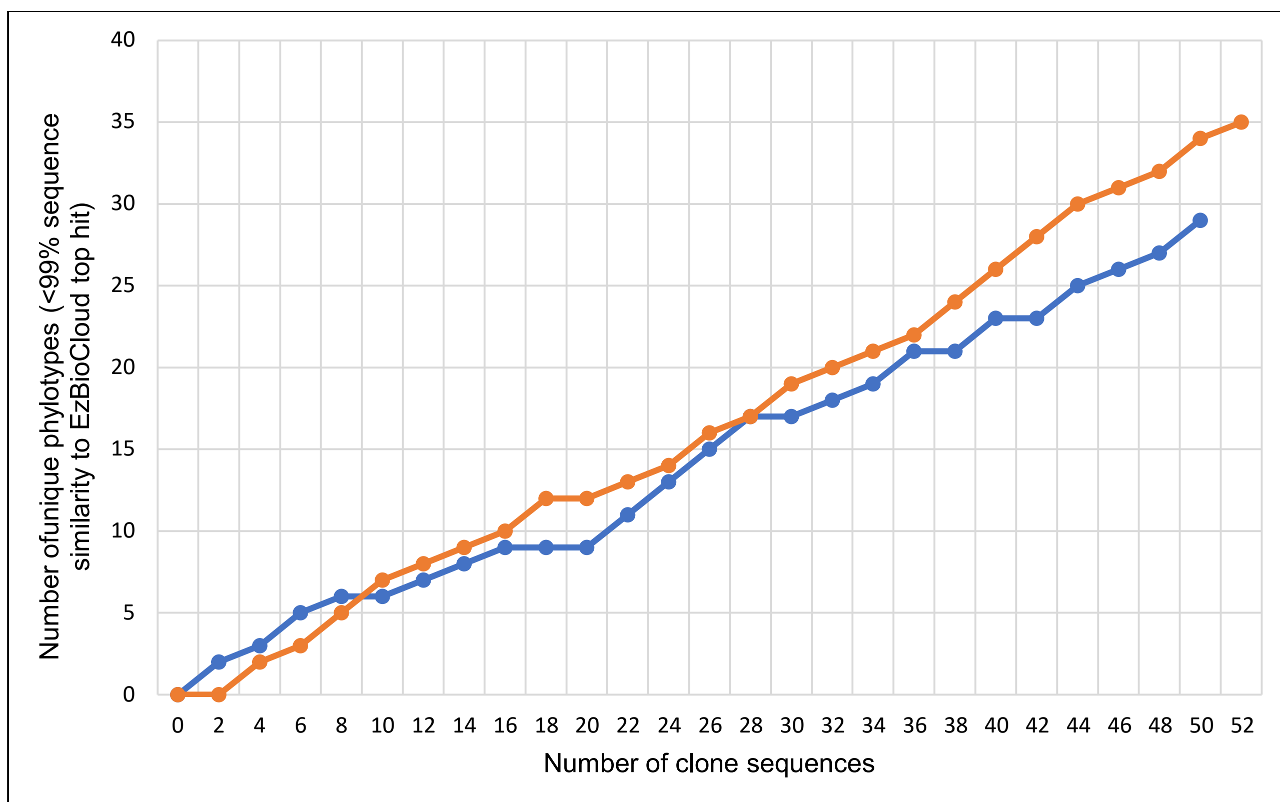

Figure 2.1: Rarefaction curve for the Kramer soil sample and Tennis Court soil sample comparing the number of unique phylotypes against the total number of clone sequences. Clone sequences were designated as unique phylotypes if the EzBioCloud top hit had a sequence similarity of <99%. The orange curve represents the Tennis Court soil sample clone library sequences and the blue curve represents the Kramer soil sample clone library sequences.

2.3.2 Actinobacterial composition of the soil samples

Pie charts were generated in Microsoft Excel to illustrate the actinobacterial diversity of the clone libraries for both the K and TC samples. The charts were constructed using the data from the EzBioCloud top hits of the clone sequences at the taxonomic level of order (Tables 2.1 and 2.2). The analyses of the clone library diversity were done at the order level, rather than the family or genus level, because some clone sequences had top hits without validly published family or genus names (Tables A1 and A2). Figure 2.2 shows the actinobacterial diversity of the K sample clone library and Figure 2.3 shows the actinobacterial diversity of the TC sample clone library.

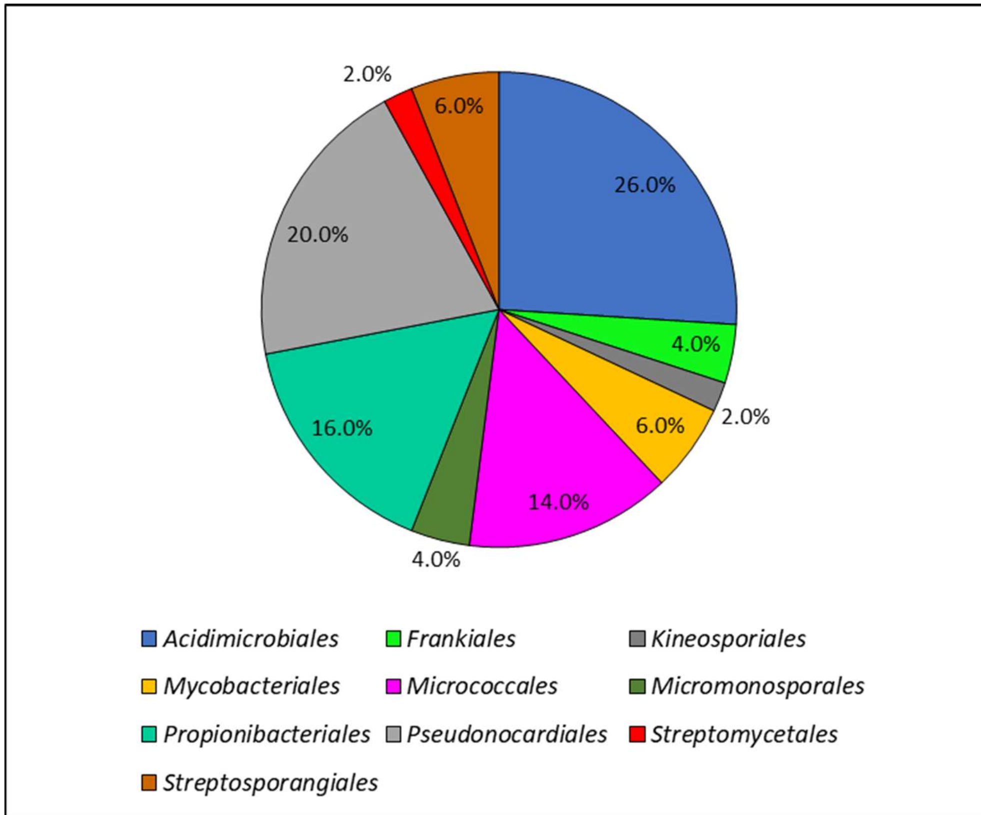

Figure 2.2: The distribution of clones from the Kramer soil sample at the order level, showing the actinobacterial diversity in the sample.

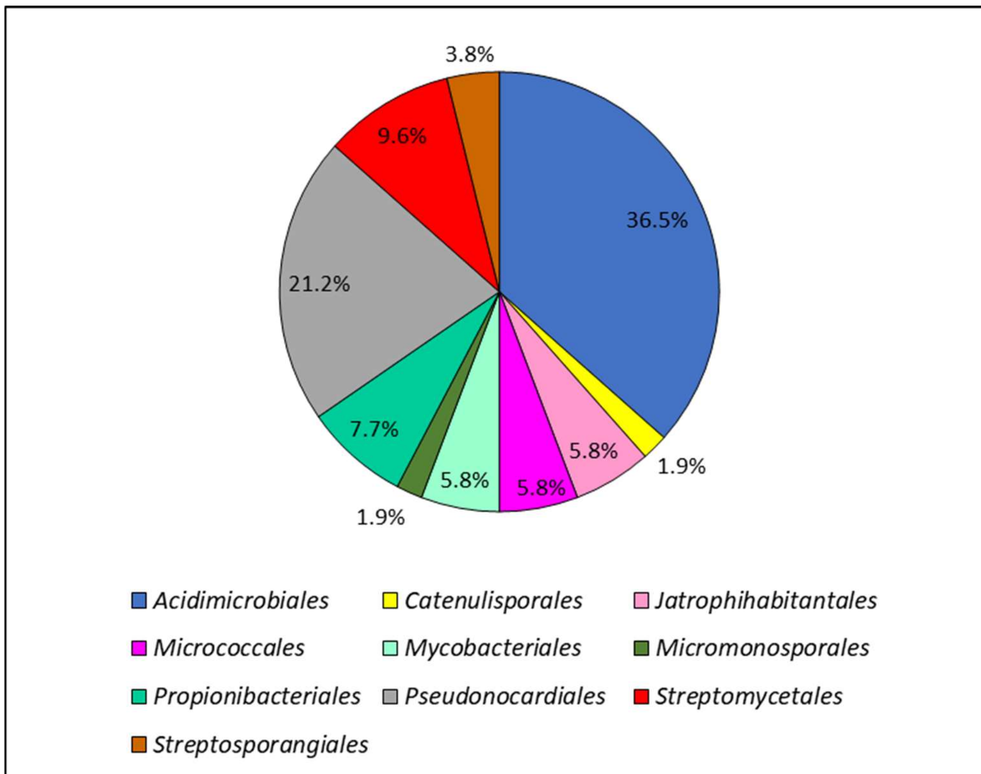

Figure 2.3: The distribution of clones from the Tennis Court soil sample at the order level, showing the actinobacterial diversity in the sample.

Both the K and TC libraries were represented by clone sequences from a wide diversity of actinobacterial orders. The K and TC samples both had clone sequences that represented 10 different orders. Eight orders (*Acidimicrobiales, Micrococcales, Micromonosporales, Mycobacteriales, Propionibacteriales Pseudonocardiales, Streptomycetales* and *Streptosporangiales*) were represented in both the K and TC clone libraries (Figures 2.2 and 2.3). The orders *Frankiales and Kineosporiales* were represented in the K clone library, but not the TC clone library. Similarly, the orders *Catenulisporales* and *Jatrophihabitantales* were represented in the TC clone library, but not the K clone library. The K sample clone library had orders range from 2% of the total clone sequences (*Streptomycetales* and *Kineosporiales*) to 26% of the total clone sequences (*Acidimicrobiales*) (Figure 2.2). The TC sample clone library orders range from 1.9% of the total clone sequences (*Catenulisporales* and *Micromonosporales*) to 36.5% of the total clone sequences (*Acidimicrobiales*) (Figure 2.3).

In this chapter, the orders *Pseudonocardiales, Propionibacteriales, Acidimicrobiales* and the genus *Streptomyces* are discussed. The orders *Pseudonocardiales, Propionibacteriales* and *Acidimicrobiales* are discussed because they, on average, made up a large proportion (>10%) of clone sequences in both clone libraries. The genus *Streptomyces* is discussed, because in culture-based isolations they are almost always the dominant species.

2.3.2.1 Order *Pseudonocardiales*

The order *Pseudonocardiales* was strongly represented in the clone libraries of both the K and TC samples. Twenty percent (20%) (10 out of 50 clone sequences) of the K sample clone sequences belong to the order *Pseudonocardiales* (Figure 2.2). Of the K sample clone sequences that are *Pseudonocardiales* representatives, the genera *Amycolatopsis* and *Pseudonocardia* were both represented by three clone sequences (Table 2.1). Twenty-one point two percent (21.2%) (11 out of 52 clone sequences) of the TC sample clone sequences are part of the order *Pseudonocardiales* (Figure 2.3). Of the TC sample clone sequences that represent the order *Pseudonocardiales*, the genera *Amycolatopsis* and *Pseudonocardia* both have four clone sequences (Table 2.2). All the clone sequences belonging the order *Pseudonocardiales* had top hits that are cultured bacteria (Table 2.1 and 2.2), meaning that they could potentially be cultured on laboratory media.

The order *Pseudonocardiales* has one family: *Pseudonocardiaceae* (https://lpsn.dsmz.de/order/pseudonocardiales). Within the family *Pseudonocardiaceae* there are 39 genera with validly published names (https://lpsn.dsmz.de/family/pseudonocardiaceae). The genera *Pseudonocardia* and *Amycolatopsis* are two of the largest with 67 (https://lpsn.dsmz.de/genus/pseudonocardia) and 85 (https://lpsn.dsmz.de/genus/amycolatopsis) validly published and correctly named species, respectively, according to the List of Prokaryotic names with Standing in Nomenclature (LPSN) (Parte, 2018). The order *Pseudonocardiales* is a taxonomic group that is particularly gifted in terms of the natural product diversity produced by its members (Lazzarini *et al.*, 2001; Doroghazi *et al.*, 2014). The genus *Amycolatopsis*, in particular, is one of the major producers of natural products within the order *Pseudonocardiales.* Evidence of this is the production of rifampicins (ansamycin-type antibiotics) by *Amycolatopsis mediterranei* and the production of vancomycin (a glycopeptide-type antibiotic) by *Amycolatopsis orientalis* (Lazzarini *et al.*, 2001).

2.3.2.2 Order *Propionibacteriales*

The order *Propionibacteriales* was represented in both the K and TC sample clone libraries. Sixteen percent (16%) (8 out of 50 clone sequences) of the K sample clone sequences belong to members of the order *Propionibacteriales* (Figure 2.2). Of the K sample clone sequences that are *Propionibacteriales* representatives, the genus *Nocardioides* makes up 5 of the 8 clone sequences, with one clone sequence representing the genus *Kribbella* (Table 2.1). Seven point seven percent (7.7%) (four out of 52 clone sequences) of the TC sample clone sequences are made up of the order *Propionibacteriales* (Figure 2.3). All 4 clone sequences representing the order *Propionibacteriales* belong to the genus *Nocardioides.* As with the order *Pseudonocardiales,* all clone sequences belonging to the order *Propionibacteriales* had top hits that are cultured bacteria (Table 2.1 and 2.2), meaning that they could potentially be cultured on laboratory media.

The order *Propionibacteriales* contains four families (https://lpsn.dsmz.de/order/propionibacteriales). The genus *Nocardioides* falls within the family *Nocardioidaceae* and has 139 species with validly published and correctly named species (https://lpsn.dsmz.de/genus/nocardioides). The genus *Kribbella* falls within the family *Kribbellaceae* and has 33 species with validly published and correctly

named species (https://lpsn.dsmz.de/genus/kribbella). Work done by Curtis *et al*. (2020) saw the isolation and description of two novel *Kribbella* species (*Kribbella capetownensis* and *Kribbella speibonae*) from samples taken at UCT. These isolations were done using a newly-developed *Kribbella*-selective medium (Curtis *et al.*, 2020).

2.3.2.3 Order *Acidimicrobiales*

The order *Acidimicrobiales* has the largest representation of clone sequences within the K and TC clone libraries. The order *Acidimicrobiales* made up 26% (13 out of 50 clone sequences) of the K sample clone sequences (Figure 2.2). Only one of the K sample clone sequences representing the order *Acidimicrobiales* had a top hit with a validly published genus name, that clone sequence being assigned to the genus *Desertimonas* (Table A1). The order *Acidimicrobiales* made up 36.5% (19 out of 52 clone sequences) of the TC sample clone sequences (Figure 2.3). Only two of the TC sample clone sequences representing the order *Acidimicrobiales* had a top hit with a validly published genus name, those clone sequence being assigned to the genera *Desertimonas* and *Aciditerrimonas,* respectively (Table A2). Between all the clone sequence representatives of the order *Acidimicrobiales* from both soil samples, two of 32 clone sequences had top hits on EzBioCloud that were cultured bacteria (Table 2.1 and 2.2). This indicates that the *Acidimicrobiales* species present in the soil samples may be unculturable. In addition, very little information about the clone sequences of the order *Acidimicrobiales* was provided at the genus level by the EzBioCloud analyses.

Acidimicrobiales, also known as *Iamiales*, is an order that contains 3 different families (Parte, 2018). The families *Iamiaceae* and *Ilumatobacteraceae* are represented in both the K and the TC sample libraries (Table A1 and A2). The pH of the Tennis Court sample was found to be pH 6.13-6.14 and the pH of the K soil sample was found to be pH 6.10-6.35. The type species of the family *Ilumatobacteraceae* is *Ilumatobacter fluminis*, which grows over the pH range of 7-11 (Matsumoto *et al.*, 2009). The type species of the family *Iamiaceae* is *Iamia majanohamensis*, which grows in the pH range of 6 to 9 (Asem *et al.*, 2018). The presence of the members of the family *Iamiaceae* in the soil samples is ,therefore, not surprising as the pH growth range of its type species fits with the pH's of the soil samples. However, the presence of the members of the families *Ilumatobacteraceae* is surprising as the soil samples are too acidic for the pH growth range (pH 7–11) of its type species, *I. fluminis*.

2.3.2.4 Genus Streptomyces

All the top hits on EzBioCloud for clone sequences belonging to the order *Streptomycetales* in the K and TC clone libraries belong to the genus *Streptomyces* (Table 2.1 and 2.2). Just 2% (one out of 50 clone sequences) of the K sample clone sequences belong to the genus *Streptomyces* (Figure 2.2). Nine point six percent (9.6%) (five out of 52 clone sequences) of the TC sample clone sequences belong to the genus *Streptomyces* (Figure 2.3). The order *Streptomycetales* has one family, *Streptomycetaceae* (https://lpsn.dsmz.de/order/streptomycetales), which contains six genera with validly published and correct names (*Allostreptomyces*, *Embleya*, *Kitasatospora*, *Streptacidiphilus*, *Streptomyces* and *Yinghuangia*; https://lpsn.dsmz.de/family/streptomycetaceae). *Streptomyces* is the largest of these genera and contains 697 species with validly published and correct names (https://lpsn.dsmz.de/genus/streptomyces). *Streptomycetales* is an order that is gifted in the abundance and diversity of the natural products that its members can produce (Lazzarini *et al.*, 2001; Doroghazi *et al.*, 2014). The genus *Streptomyces*, in particular, is an exceptional producer of natural products. *Streptomyces* species produce about 75% of commercially and medically useful antibiotics (Anandan *et al*., 2016).

2.4 References

Amann, R. I. (1995) ‘Fluorescently labelled, rRNA-targeted oligonucleotide probes in the study of microbial ecology’, Molecular Ecology. John Wiley & Sons, Ltd, pp. 543–554. doi: 10.1111/j.1365-294X.1995.tb00255.x.

Anandan, R. *et al.* (2016) An Introduction to Actinobacteria, Actinobacteria - Basics and Biotechnological Applications. Intech. Available at: https://www.intechopen.com/books/5056.

Arjun, J. K. and Harikrishnan, K. (2011) ‘Metagenomic analysis of bacterial diversity in the rice rhizosphere soil microbiome’, Biotechnology Bioinf. Bioengineering, 1(3), pp. 361–367.

Asem, M. D. *et al.* (2018) ‘*Desertimonas flava* gen. nov., sp. nov. isolated from a desert soil, and proposal of *Ilumatobacteraceae* fam. nov.’, International Journal of Systematic and Evolutionary Microbiology, 68(11), pp. 3593–3599. doi: 10.1099/ijsem.0.003038.

Brady, S. F. *et al.* (2009) ‘Metagenomic approaches to natural products from free-living and symbiotic organisms’, Natural Product Reports. Royal Society of Chemistry, pp. 1488–1503. doi: 10.1039/b817078a.

Cook, A. E. and Meyers, P. R. (2003) ‘Rapid identification of filamentous actinomycetes to the genus level using genus-specific 16S rRNA gene restriction fragment patterns’, International Journal of Systematic and Evolutionary Microbiology, 53(6), pp. 1907–1915. doi: 10.1099/ijs.0.02680-0.

Curtis, S. M. *et al.* (2020) ‘Development of a *Kribbella*-specific isolation medium and description of *Kribbella capetownensis* sp. nov. and *Kribbella speibonae* sp. nov., isolated from soil’, Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology, 113(5), pp. 617–628. doi: 10.1007/s10482-019-01365-6.

Daniel, R. (2004) ‘The soil metagenome - A rich resource for the discovery of novel natural products’, Current Opinion in Biotechnology. Elsevier Current Trends, pp. 199–204. doi: 10.1016/j.copbio.2004.04.005.

Doroghazi, J. R. *et al.* (2014) ‘Aroadmap for natural product discovery based on large-scale genomics and metabolomics’, Nature Chemical Biology, 10(11), pp. 963–968. doi: 10.1038/nCHeMBIO.1659.

Gotelli, N. J. and Colwell, R. K. (2001) ‘Quantifying biodiversity: Procedures and pitfalls in the measurement and comparison of species richness’, Ecology Letters, pp. 379–391. doi: 10.1046/j.1461-0248.2001.00230.x.

Handelsman, J. *et al.* (1998) ‘Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products’, Chemistry and Biology, 5(10). doi: 10.1016/S1074-5521(98)90108-9.

Kuang, S. *et al.* (2018) ‘Soil microbial community structure and diversity around the aging oil sludge in yellow river delta as determined by high-throughput sequencing’, Archaea, 2018. doi: 10.1155/2018/7861805.

Lazzarini, A. *et al.* (2001) ‘Rare genera of actinomycetes as potential producers of new antibiotics’, Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology, 78(3–4), pp. 399–405. doi: 10.1023/A:1010287600557.

Matsumoto, A. *et al.* (2009) ‘*Ilumatobacter fluminis* gen. nov., sp. nov., a novel actinobacterium isolated from the sediment of an estuary’, Journal of General and Applied Microbiology, 55(3), pp. 201–205. doi: 10.2323/jgam.55.201.

Parte, A. C. (2018) ‘LPSN - List of prokaryotic names with standing in nomenclature (Bacterio.net), 20 years on’, International Journal of Systematic and Evolutionary Microbiology, 68(6), pp. 1825–1829. doi: 10.1099/ijsem.0.002786.

Schäfer, J. *et al.* (2010) ‘Development of a new PCR primer system for selective amplification of Actinobacteria’, FEMS Microbiology Letters, 311(2), pp. 103–112. doi: 10.1111/j.1574-6968.2010.02069.x.

Sipos, R. *et al.* (2007) ‘Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targeting bacterial community analysis’, FEMS Microbiology Ecology, 60(2), pp. 341–350. doi: 10.1111/j.1574-6941.2007.00283.x.

Stackebrandt, E. *et al.* (2022) ‘Authors need to be prudent when assigning names to microbial isolates’, Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology, 115(1), pp. 1–5. doi: 10.1007/s10482-021-01675-8.

Vartoukian, S. R. *et al.* (2010) ‘Strategies for culture of “unculturable” bacteria’, FEMS Microbiology Letters. Blackwell Publishing Ltd, pp. 1–7. doi: 10.1111/j.1574-6968.2010.02000.x.

Von Wintzingerode, F. *et al.* (1997) ‘Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis’, FEMS Microbiology Reviews, 21(3), pp. 213–229. doi: 10.1111/j.1574-6976.1997.tb00351.x.

Xin, Y. *et al.* (2008) ‘Culture-independent nested PCR method reveals high diversity of actinobacteria associated with the marine sponges *Hymeniacidon perleve* and Sponge sp’, Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology, 94(4), pp. 533–542. doi: 10.1007/s10482-008-9270-y.

Yang, H. *et al.* (2005) ‘Niche heterogeneity determines bacterial community structure in the termite gut (*Reticulitermes santonensis*)’, Environmental Microbiology, 7(7), pp. 916–932. doi: 10.1111/j.1462-2920.2005.00760.x.

Yoon, S. H. *et al.* (2017) ‘Introducing EzBioCloud: A taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies’, International Journal of Systematic and Evolutionary Microbiology, 67(5), pp. 1613–1617. doi: 10.1099/ijsem.0.001755.

Chapter 3:

Isolation and identification of the actinobacterial biodiversity of the Kramer and Tennis Court soil samples using selective culture-dependent methods

Summary

A total of 63 presumptive actinobacterial strains was isolated from two different soil samples taken from the UCT, namely, the K soil sample and the TC soil sample. The investigation of 21 of the presumptive actinobacteria was discontinued prior to morphological and molecular analysis due to either the isolated strains being considered to be non-actinobacterial or the inability to achieve pure cultures of the isolates. Of the 42 actinobacteria studied, the partial 16S-rRNA gene sequences were amplified, sequenced and analysed for 40 of the 42 presumptive actinobacterial strains: 23 strains from the K sample and 17 strains from the TC sample. The 40 isolates were identified to the genus level based on the analysis of partial 16S-rRNA gene sequences revealing a total of 38 *Streptomyces* strains and two non-*Streptomyces* strains: *Actinomadura* strain 1KK1 and *Nonomuraea* strain 7TK1. The phylogenetic position of each identified strain was investigated by generating phylogenetic trees based on their 16S-rRNA gene sequences. Further analysis of the gyrase subunit B gene (*gyrB*) sequence and *gyrB* gene phylogenetic analysis were undertaken for *Nonomuraea* strain 7TK1. Thirty (30) of the 38 *Streptomyces* strains identified were potential-clones, which were divided into three potential-clone groups coloured blue, orange and green. Clone analysis was carried out for 23 of the 30 *Streptomyces* potential-clone strains through homology matrix construction, phenotypic analysis on International *Streptomyces* Project medium No. 4 (ISP4) plates and phylogenetic analyses using partial *gyrB* gene sequences. These analyses revealed that the three potential-clone groups divided into five clone groups and eight distinctive *Streptomyces* species based on their partial *gyrB* gene sequences.

Table of Contents

3.1 Introduction 52
3.2 Materials and methods 55
3.2.1 Sample Collection 55
3.2.2 Bacterial sample preparation 55
3.2.3 Bacterial isolation and purification 55
3.2.4 Genomic DNA extraction 56
3.2.5 16S-rRNA and *gyrB* gene amplification 57
3.2.6 DNA sequencing and analysis 57
3.2.7 Phylogenetic analyses 58
3.2.7.1 16S-rRNA gene sequences 58
3.2.7.2 gyrB gene sequences 59
3.3 Results and discussion 59
3.3.1 Bacterial isolations 59
3.3.2 16S-rRNA and *gyrB* identification 59
3.3.3 Phylogenetic analyses 63
3.3.3.1 Genus *Actinomadura* 63
3.3.3.2 Genus *Nonomuraea* 65
3.3.3.3 Genus *Streptomyces* 68
3.3.3.3.1 *Streptomyces* strains assigned to a potential-clone group.............. 68
3.3.3.3.2 *Streptomyces* strains not assigned to a potential-clone group...... 72
3.3.4 Clone analysis 77
3.3.4.1 The blue potential-clone group 80
3.3.4.2 The orange potential-clone group 81
3.3.4.3 The green potential-clone group 83
3.4 References 85

3.1 Introduction

The isolation of diverse soil microorganisms has been exploited since the 1940s. The isolation of actinobacteria, in particular, has led to the discovery of many natural products of economic and medical value. The rate at which useful natural products, for example antibiotics, are being discovered has decreased (Daniel, 2004). There has been a 30% drop in natural product based drugs used clinically between 2001 and 2008 (Li and Vederas, 2009). This decrease is largely as a result of the rediscovery of known bioactive metabolites, especially antibiotics. There is therefore the need to isolate and characterize poorly studied actinobacterial taxa. This would allow us access to novel actinobacterial organisms and the novel bioactive compounds they produce (Connon and Giovannoni, 2002; Goodfellow, 2010). Selective isolation can be used to access rare and understudied actinobacterial taxa (i.e. the actinobacteria that belong to all genera other than *Streptomyces*).

Selective isolation can be applied through all the steps of the isolation process: sample pre-treatment, sample preparation, the choice of selective media, incubation conditions and colony selection once the microorganisms have grown. Arguably the choice of selective media is where the greatest selection for actinobacteria can be achieved. For a selective medium, the carbon source, nitrogen source, pH and choice of antibiotics can all be altered to create conditions where a desired taxon can grow and the majority of others cannot (Goodfellow, 2010). This prevents the growth of fast growing microbial species, such as *Streptomyces,* allowing unique populations of microorganisms to grow and be cultured (Connon and Giovannoni, 2002). Innumerable selective media have been suggested for the isolation of actinobacteria so it is impossible to recommend a selective medium to isolate all rare actinobacterial genera (Goodfellow, 2010).

This study shows the use of selective media to bias the isolation of rare actinobacterial genera from soil samples. The genera *Amycolatopsis* and *Kribbella* were targeted for isolation based on the results of the metagenomic analyses performed (Chapter 2). Tan *et al.* (2006) showed three different media to be selective for the genus *Amycolatopsis*. The authors tested 20 selective agar media and found three (SM1, SM2 and SM3) to support the growth of all representatives of the genus *Amycolatopsis* (Tan *et al.*, 2006). Curtis (2015) formulated a selective minimal medium named

Kribbella selective medium (KSM). The medium utilizes adonitol as the carbon source and 4-L-hydroxyproline as the nitrogen source. These are compounds which are not commonly utilized by actinobacteria, but can be utilized by all *Kribbella* species (Curtis, 2015). The use of KSM resulted in the isolation of a strain belonging to a novel *Kribbella* species (Curtis *et al*., 2020).

Isolation media in this M.Sc. study were supplemented with nalidixic acid (NA) and cycloheximide (CHX) to inhibit the growth of Gram-negative bacteria and fungi, respectively. These choices were made for culture-based isolations to increase the chance of isolating rare actinobacteria, particularly *Amycolatopsis* and *Kribbella*, from the soil samples.

Molecular classification groups organisms according to the similarities in their genetic material. DNA-based analyses are predominantly used to classify microorganisms and strongly influence modern prokaryotic taxonomy (Tindall *et al.*, 2010). Variations in the 16S-rRNA gene provide the backbone to the classification of archaea and bacteria (Tindall *et al.*, 2010). The 16S-rRNA gene has functional constancy, a ubiquitous distribution in bacteria, a large size (1.6 kb) and the presence of both conserved and variable regions. These characteristics mean it fulfils all the requirements for a phylogenetic marker molecule (Ludwig and Klenk, 2001), making the 16S-rRNA gene the ideal candidate for the determination of evolutionary relationships among microorganisms. 16S-rRNA gene sequences are invaluable in the description of prokaryotic diversity and are essential in the allocation of an unknown microorganism to a taxon (Kämpfer and Glaeser, 2012).

16S-rRNA gene sequences are used extensively in phylogenetic analyses (Woese *et al.*, 1990). Phylogenetic trees take evolutionary relationships and varying rates of evolution into consideration and assign species to natural classifications (Doolittle, 1999). Various species of actinobacteria have been delineated based on their branching position in 16S-rRNA gene phylogenetic trees (Clarridge, 2004). 16S-RNA gene sequences do, however, have their limitations. The major limitation being that the 16S-rRNA gene has low resolving power at high levels of relatedness between species, which causes ambiguity in the delineation of species (Stackebrandt and Goebel, 1994; Barka *et al.*, 2016). Besides the use of 16S-rRNA gene, other molecular marker genes including *rpoB*, *atpD*, *gyrB*, *recA*, *trpB* and *ssgB* are also used to

distinguish closely related genera (Barka *et al.*, 2016). The gyrase subunit B (*gyrB*) and recombinase A (*recA*) genes have previously been used to delineate genera in the actinobacterial family *Streptosporangiaceae* (Meyers, 2014, 2015). This can be done because the *gyrB* and *recA* genes have a higher mutation rate than the 16S-rRNA gene and therefore show greater variability between closely related species.

Evidence suggests that a 16S-rRNA gene sequence similarity for actinobacteria of less than 98.8% (for a 0.5% maximum probability of error) means that the compared strains do not belong to the same species (Meier-Kolthoff *et al.*, 2013). It is suggested that at a threshold level of >98.8% 16S-rRNA gene similarity (for actinobacteria with a 0.5% maximum probability of error) that any species delineation conclusions drawn should be supported by DNA-DNA hybridization (DDH) data (Meier-Kolthoff *et al.*, 2013). Digital DDH (dDDH), which has replaced laboratory-based DDH experiments, is based on whole-genome sequences and provides a robust taxonomic classification.

In the work presented in this chapter, the aim was to isolate rare actinobacterial strains, particularly from the genera *Amycolatopsis* and *Kribbella*, and assess their phylogeny. Actinobacterial strains were isolated based on colony morphology and identified to the genus level by partial 16S-rRNA gene sequence analysis. The taxonomic affiliations of the isolates were determined through the construction of phylogenetic trees for the three genera isolated. All the actinobacterial isolates described in this chapter were then screened for antimycobacterial activity (Chapter 4).

3.2 Materials and methods

3.2.1 Sample Collection

The K and TC soil samples were collected as described in Chapter 2. After removing 0.5 g of soil from each sample for metagenomic-DNA isolation (Chapter 2), the remaining soil from each of K and TC soil samples was spread out in empty sterile Petri dishes. The soil samples were then left to dry slowly at room temperature to encourage the actinobacteria to sporulate. After completion of the metagenomic studies portion of the project (10 months), the dry soil samples were used for culture-based isolations.

3.2.2 Bacterial sample preparation

0.1 g of each soil sample was placed in 1ml of sterile dH_2O. The soil-water mixtures were then vortexed vigorously for 60s to dislodge the bacteria from the soil particles. The vortexed mixtures were left to settle for 5 minutes at room temperature. A 10-fold dilution series (to a final dilution of 10^{-5}) was then prepared for each sample with sterile dH_2O as the diluent. Using the spread plate technique, 100 µl volumes of each soil suspension dilution were spread onto agar plates. The aliquots were spread in the following distribution: 10^{-2} dilution, 2 agar plates; 10^{-3} dilution, 3 agar plates; 10^{-4} dilution, 3 agar plates; 10^{-5} dilution, 1 agar plate.

Four different media were used for the bacterial isolations: SM1 (Tan *et al*., 2006), referred to as *Amycolatopsis* selective medium 1 (ASM1) in this study; SM2 (Tan *et al*., 2006), referred to as *Amycolatopsis* selective medium 2 (ASM2) in this study; SM3 (Tan *et al*., 2006), referred to as *Amycolatopsis* selective medium 3 (ASM3) in this study and *Kribbella* Selective Medium (KSM) (Curtis, 2015), referred to as *Kribbella* selective medium no.1 (KSM1) in this study. NA (10µg/ml) and CHX (50µg/ml) were added to all actinobacterial isolation media to inhibit the growth of Gram-negative bacteria and fungi, respectively. All actinobacterial isolation plates were incubated at 30°C for a period of up to 12 weeks.

3.2.3 Bacterial isolation and purification

Bacterial plates were checked weekly for actinobacterial growth with potential actinobacterial colonies having one or more of the following characteristics: leathery or wrinkly surfaces, colonies with “volcano-like” or pointy shapes, colonies sunken into the agar medium or colonies displaying aerial mycelium that had differentiated into

spores. Colonies fitting the actinobacterial description were sub-cultured onto fresh agar plates (lacking antibiotics) of the same medium as they were isolated on. In the case of colonies on the KSM1 isolation plates, they were sub-cultured onto both KSM1 and International *Streptomyces* Project medium No. 2 (ISP2) (Shirling and Gottlieb, 1966) agar plates.

Colonies were normally sub-cultured using a sterile inoculating loop, but sterile toothpicks were used for colonies where the sterile loop was unable to remove the colony from the agar surface. Sub-cultured bacteria were generally incubated at 30°C for 7-14 days, but times varied depending on the speed at which the strains grew. Any sub-culture plates showing contamination by bacteria or fungi were sub-cultured repeatedly until pure cultures were achieved. Each new sub-cultured bacterium was given a strain number. Strain numbers were based on the sample the strain was isolated from and the isolation medium on which it was isolated. For example, strain number 42KA2 was the 42nd bacterium sub-cultured from the K sample and was isolated on ASM2 (“A2”) . Strain 7TK1 was the 7th bacterium sub-cultured from the TC sample (“T”) and was isolated on KSM1 (“K1”).

Some sub-cultured strains were clearly not actinobacteria. These strains were discarded and not studied further. Once pure cultures were obtained, duplicate stock cultures (11.5% (v/v) glycerol) were prepared from a broth culture for each strain and stored at -80°C or -20°C. Stock cultures were prepared as follows: (1) broth cultures were prepared by growing the chosen strain in 20 ml liquid medium (the same medium as the medium of the sub-culture plate) in a 250 ml Erlenmeyer flask. The flask was incubated at 30°C with shaking until sufficient cell mass had grown. (2) 1000 µl of broth culture was added to 300 µl of 50% (v/v) sterile glycerol in a sterile 1.5 ml microcentrifuge tube.

3.2.4 Genomic DNA extraction

Bacterial strains were incubated in 20 ml liquid medium under agitation at 30°C in a sterile 250ml Erlenmeyer flask for 7-14 days (or until sufficient cell mass had grown). Each bacterial strain was grown in the medium in which it showed optimal growth. A benchtop centrifuge (Eppendorf Centrifuge 5418) was used to harvest bacterial cells in sterile 1.5 ml microcentrifuge tubes (14000 rpm [16873 X *g*] for 2 min per run), until approximately 200 µl of cell mass had been collected.

Genomic DNA (gDNA) was extracted based on the method of Wang *et al*. (1996) with the following modifications: (1) the cell mass was resuspended in lysozyme buffer containing 25 mg/ml lysozyme instead of 5 mg/ml, (2) the cell mass was incubated at 37°C in lysozyme buffer overnight (14 to 18 h) instead of for 30 min and (3) the resultant gDNA pellets were redissolved in Tris-EDTA (TE) buffer (pH 8.0) of differing volumes (12µl to 60µl) depending on the size of the pellet (not 100µl as proposed by Wang *et al*., 1996). The extracted gDNA was stored at 4°C overnight and then the DNA concentration was measured using a Nanodrop™ ND-1000 spectrophotometer.

3.2.5 16S-rRNA and *gyrB* gene amplification

Amplification of both the 16S-rRNA gene and *gyrB* gene was done by polymerase chain reaction (PCR). Amplification of 16S-rRNA genes was done using the F1-R5 primer set (Cook and Meyers, 2003). The following cycle parameters were used: initial denaturation (96°C for 2 min), 30 cycles of denaturation (96°C for 30s), annealing (56°C for 20s), and extension (72°C for 30s), and a final extension (72°C for 5 min). For each reaction: 2 mM $MgCl_2$, 150µM of each dNTP, 1U Super-Therm *Taq* polymerase (JMR Holdings, USA) and 200ng/µl bovine serum albumin (BSA), to prevent inhibition of amplification, were used. The reaction yielded a ≈1500 bp 16S-rRNA gene product.

Amplification of the *gyrB* gene was done using the GgyrB-F2 (le Roes *et al*., 2008) - KgyrB-R1892 (Kirby *et al*., 2010) primer set. The following cycle parameters were used: initial denaturation (96°C for 2 min), 30 cycles of denaturation (96°C for 30s), annealing (60°C for 20s), and extension (72°C for 30s) and a final extension (72°C for 5 min). BSA (200ng/µl) was added to the reaction to prevent inhibition of amplification. The reaction yielded a ≈660 bp *gyrB* gene product. The reaction components were as for 16S-RNA gene amplification, but a higher magnesium concentration was used (4 mM $MgCl_2$).

Electrophoresis and DNA purification were caried out as described in section 2.2.3.

3.2.6 DNA sequencing and analysis

DNA concentrations were measured using a Nanodrop™ ND-1000 spectrophotometer and diluted to 50 ng/µl. The purified PCR products were sequenced using the 518F primer (5'-CCAGCAGCCGCGGTAATACG-3') for the 16S-rRNA gene and the GgyrB-F2 primer (le Roes *et al*., 2008) for the *gyrB* gene. Sanger sequencing was performed

by Macrogen Europe Inc. (Amsterdam, The Netherlands). The resulting chromatograms were analysed and edited using Chromas version 2.6.4 (Technelysium Pty. Ltd., Australia).

The edited 16S-rRNA gene partial sequences were submitted to EzBioCloud (https://www.ezbiocloud.net/) (a curated database), to accurately identify the isolated strains to the genus level (Kim *et al.*, 2012). The edited *gyrB* partial gene sequences were subjected to Basic Local Alignment Search Tool (BLAST) analysis (https://blast.ncbi.nlm.nih.gov/Blast.cgi) for comparison against the GenBank database. The standard nucleotide-nucleotide BLAST search (blastn) (Altschul *et al.*, 1997)(https://blast.ncbi.nlm.nih.gov/Blast.cgi) tool was used for the BLAST analyses. The sequences of the top hits of each actinobacterial strain were downloaded from the databases for phylogenetic analyses. If required, strain sequences were aligned and assembled in DNAMAN version 4.13 (Lynnon Biosoft).

3.2.7 Phylogenetic analyses

Phylogenetic trees were generated for each actinobacterial strain along with their top hits by sequence similarity. Molecular Evolutionary Genetic Analysis version 7 (MEGA7) (Kumar *et al.*, 2016) was used to create a multiple sequence alignment of the gene sequences using the MUSCLE algorithm (Edgar, 2004) and the default parameters. All alignments were edited to remove any columns containing ambiguous bases, insertions or deletions. Edited alignments were used to generate neighbour-joining (NJ) (Saitou and Nei, 1984), maximum parsimony (MP) (Felsenstein, 1985) and maximum likelihood (Kimura, 1980) trees using the bootstrap method (1000 replications). The three generated trees were compared to each other and matching topologies marked on the maximum likelihood tree. Only maximum likelihood trees, with matching topologies indicated, are shown in the results. The top hit gene sequences were collected as described below.

3.2.7.1 16S-rRNA gene sequences

For strains belonging to the genus *Streptomyces*: the 16S-rRNA gene sequences of the top five hits were downloaded from the EzBioCloud database (Yoon *et al.*, 2017). For non-*Streptomyces* strains: the 16S-rRNA gene sequences of the top 30 hits were downloaded from the EzBioCloud database.

3.2.7.2 *gyrB* gene sequences

For strains belonging to the genus *Streptomyces*: the *gyrB* gene sequences of the top five hits with validly published names were downloaded from the GenBank database. For non-*Streptomyces* strains: the *gyrB* gene sequences of the top 30 hits, with validly published names, were downloaded from the GenBank database.

3.3 Results and discussion

3.3.1 Bacterial isolations

Initially, 63 presumptive actinobacteria were isolated. The study of 21 of the presumptive actinobacterial strains was discontinued for one of two reasons: (1) the isolated strains did not show phenotypic traits consistent with actinobacteria when in pure culture or (2) presumptive actinobacteria in pure culture were lost due to dust mite or fungal related contamination of pure sub-culture plates. There were therefore 42 presumptive actinobacteria studied: 25 strains from the K sample and 17 strains from the TC sample.

3.3.2 16S-rRNA and *gyrB* identification

Of the isolated strains, partial 16S-rRNA gene sequences were amplified for 40 of the isolates studied. The 16S-rRNA gene could not be amplified from strain 53KA3 and strain 57KA3. The partial 16S-rRNA gene sequences were used to identify the isolates to the genus level and were also used for phylogenetic analysis of the isolated strains. The vast majority of the identified strains (38 of 40 isolates) were found to belong to the genus *Streptomyces* (Table 3.1). Two rare actinobacteria were identified in the study: *Actinomadura* strain 1KK1, highlighted in purple and *Nonomuraea* strain 7TK1, highlighted in red (Table 3.1). Thirty (30) of the 38 *Streptomyces* strains were suspected to be potential-clones based on their 16S-rRNA gene sequences. These 30 *Streptomyces* strains were divided into three different potential-clone groups: all strains within potential-clone groups are highlighted in their distinct colour (blue, green or orange) in all figures and tables in this chapter.

In the metagenomic analysis part of the project (Chapter 2), *Streptomyces* species made up 2% and 9.6% of the K and TC sample clone sequences, respectively (Figure 2.2 and Figure 2.3). The proportion of *Streptomyces* strains in the metagenomic results is therefore significantly lower than in the culture-based part of the study. This result is, however, not surprising as *Streptomyces* species historically have dominated culture-based studies, because they grow very well on laboratory media.

KSM1 was chosen as isolation medium to selectively isolate *Kribbella* strains (Curtis, 2015). Similarly, the media ASM1, ASM2 and ASM3 were chosen as isolation media to selectively isolate *Amycolatopsis* strains (Tan *et al*., 2006). No *Amycolatopsis* or *Kribbella* strains were isolated in the culture-based portion of the study (Table 3.1). This result occurred despite metagenomic results showing *Amycolatopsis* strains be to be present in both samples (Table 2.1 and 2.2) and a *Kribbella* strain being shown to be present in the K sample (Table 2.1). The aim of the choice of selective media was to isolate strains of rare and understudied actinobacterial genera from the soil samples. The medium KSM1 was then, arguably, the most effective medium used as both rare actinobacterial strains (*Nonomuraea* strain 7TK1 and *Actinomadura* strain 1KK1) were isolated on KSM1 agar plates (Table 3.1).

Along with the amplification of partial 16-rRNA gene sequences for 40 of the isolates, partial *gyrB* gene sequences were amplified for 24 of the 40 identified strains. These partial *gyrB* gene sequences were analysed using the blastn search tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and the top hit for each sequence recorded (Table 3.1). The top hit(s) of each *gyrB* gene sequence belongs to the same genus as that identified from the partial 16S-rRNA gene sequence analysis (Table 3.1). The blastn results based on the partial *gyrB* gene sequences, therefore, confirm the genus assignment of all relevant strains made based on their partial 16S-rRNA gene sequences. Analysis using the 16S-rRNA or *gyrB* gene sequence is sufficient for accurate genus identification of isolated bacteria, but not for species identification. Therefore, the author has drawn no conclusions on the species to which the isolated strains belong in this study.

Table 3.1: Genus identification of 40 actinobacterial isolates from the K and TC soil samples from UCT. The identification of the 16S-rRNA gene sequences was based on EzBioCloud analyses of partial 16S-rRNA gene sequences. The identification of *gyrB* gene sequences was based on blastn analyses using the GenBank database. The percentage sequence similarity is the similarity to the query sequence. * indicates strains with multiple top hits. Blue indicates strains assigned to the blue potential-clone group. Orange indicates strains assigned to the orange potential-clone group. Green indicates strains assigned to the green potential-clone group.

	16S-rRNA gene sequence			*gyrB* gene sequence		
Strain	**Top hit**	**Similarity (%)**	**Query sequence length (bp)**	**Top hit**	**Similarity (%)**	**Query sequence length (bp)**
Kramer sample						
1KK1	*Actinomadura montaniterrae*	99.16	832			
2KK1	*Streptomyces platensis**	100	620	*Streptomyces hygroscopicus**	98.49	595
8KK1	*Streptomyces abietis*	100	828			
10KK1	*Streptomyces platensis**	100	899	*Streptomyces hygroscopicus**	98.41	628
22KA1	*Streptomyces lydicus**	100	828			
23KA1	*Streptomyces griseochromogenes**	99.62	799			
25KA1	*Streptomyces lydicus**	100	784	*Streptomyces lydicus**	98.85	660
26KA1	*Streptomyces lydicus**	100	844	*Streptomyces lydicus**	97.43	661
28KA1	*Streptomyces platensis**	100	619	*Streptomyces hygroscopicus**	98.42	633
31KA1	*Streptomyces platensis**	100	869	*Streptomyces hygroscopicus**	98.3	589
32KA1	*Streptomyces platensis**	100	596	*Streptomyces hygroscopicus**	98.38	618
39KA2	*Streptomyces lydicus**	99.88	840	*Streptomyces lydicus**	100	651
40KA2	*Streptomyces lydicus**	100	879	*Streptomyces lydicus**	99.7	666
41KA2	*Streptomyces lydicus**	100	912	*Streptomyces lydicus**	100	630
42KA2	*Streptomyces lydicus**	99.89	871	*Streptomyces lydicus**	100	645
52KA3	*Streptomyces platensis**	100	829			
54KA3	*Streptomyces platensis**	100	622	*Streptomyces hygroscopicus**	98.39	620
55KA3	*Streptomyces lydicus**	98.88	859			
58KA3	*Streptomyces lydicus**	100	868	*Streptomyces lydicus**	100	644

Table 3.1: (continued)

	16S-rRNA gene sequence			*gyrB* gene sequence		
Strain	**Top hit**	**Similarity (%)**	**Query sequence length (bp)**	**Top hit**	**Similarity (%)**	**Query sequence length (bp)**
			Kramer sample			
59KA3	*Streptomyces platensis**	100	836	*Streptomyces hygroscopicus**	98.49	597
60KA3	*Streptomyces adustus*	99.66	870			
61KA3	*Streptomyces lydicus**	100	712	*Streptomyces lydicus**	100	618
63KA3	*Streptomyces aureus*	100	808			
			Tennis Court sample			
5TK1	*Streptomyces platensis**	100	679	*Streptomyces hygroscopicus**	99.16	592
6TK1	*Streptomyces platensis**	100	605	*Streptomyces hygroscopicus**	99.01	603
7TK1	*Nonomuraea cavernae*	99.43	701	*Nonomuraea guangzhouensis*	96.99	632
09TK1	*Streptomyces platensis**	100	864	*Streptomyces hygroscopicus**	99.18	608
12TA1	*Streptomyces chartreusis*	99.40	839			
13TA1	*Streptomyces lannensis*	100	842			
15TA1	*Streptomyces platensis**	100	849	*Streptomyces hygroscopicus**	99.15	591
16TA1	*Streptomyces platensis**	100	684	*Streptomyces hygroscopicus*	99.1	664
18TA1	*Streptomyces platensis**	100	805			
19TA1	*Streptomyces platensis**	100	839	*Streptomyces hygroscopicus*	99.25	669
20TA1	*Streptomyces platensis**	100	779			
33TA2	*Streptomyces platensis**	100	829			
36TA2	*Streptomyces platensis**	100	849	*Streptomyces hygroscopicus*	99.09	657
45TA3	*Streptomyces platensis**	100	704			
47TA3	*Streptomyces platensis**	99.88	809	*Streptomyces hygroscopicus*	99.51	616
48TA3	*Streptomyces phaeoluteichromatogenes**	99.88	864			
49TA3	*Streptomyces geldanamycininus*	99.85	669			

3.3.3 Phylogenetic analyses

Phylogenetic trees were constructed for all identified actinobacterial strains based on their partial 16S-rRNA gene sequences. In the case of *Nonomuraea* strain 7TK1, a second phylogenetic tree was constructed based on its partial *gyrB* gene sequence. These analyses were conducted to determine the taxonomic position of each strain within its genus.

3.3.3.1 Genus *Actinomadura*

Actinomadura is a genus within the family *Thermomonosporaceae* **(**https://lpsn.dsmz.de/family/thermomonosporaceae). The genus is comprised of 76 species with validly published and correct names (https://lpsn.dsmz.de/genus/actinomadura). *Actinomadura* strain 1KK1 was isolated on the KSM1 medium from the K soil sample. The strain formed small, flaky and tan coloured substrate mycelium with a white spore mass when grown on ISP2 medium and formed translucent white substrate mycelium with no visible spore mass when grown on ISP4 medium (Table A3). A phylogenetic tree was generated for *Actinomadura* strain 1KK1 using the 16S-rRNA gene sequences of the top 31 *Actinomadura* type strain hits on EzBioCloud in order to determine its phylogenetic position within the genus *Actinomadura* (Figure 3.1).

An EzBioCloud analysis showed that *Actinomadura* strain 1KK1 had highest 16S-rRNA gene sequence similarity to *Actinomadura montaniterrae* CYP1-1B^{T} (99.16% sequence similarity over 832 bp; Table 3.1). In the maximum-likelihood phylogenetic tree based on 16S-rRNA gene sequences (Figure 3.1), strain 1KK1 formed part of a seven-strain clade, which included *A. montaniterrae* CYP1-1B^{T}. However, there was weak support for this association (bootstrap value <70%; Figure 3.1). Furthermore, this association was not seen in the NJ and maximum parsimony MP trees (data not shown). The author attempted to amplify a section of the *gyrB* gene of *Actinomadura* strain 1KK1, but all attempts with multiple primer sets and DNA extraction methods proved unsuccessful. Whole-genome sequencing followed by ANI and dDDH analyses would be required to provide a definitive answer as to the species to which *Actinomadura* strain 1KK1 belongs.

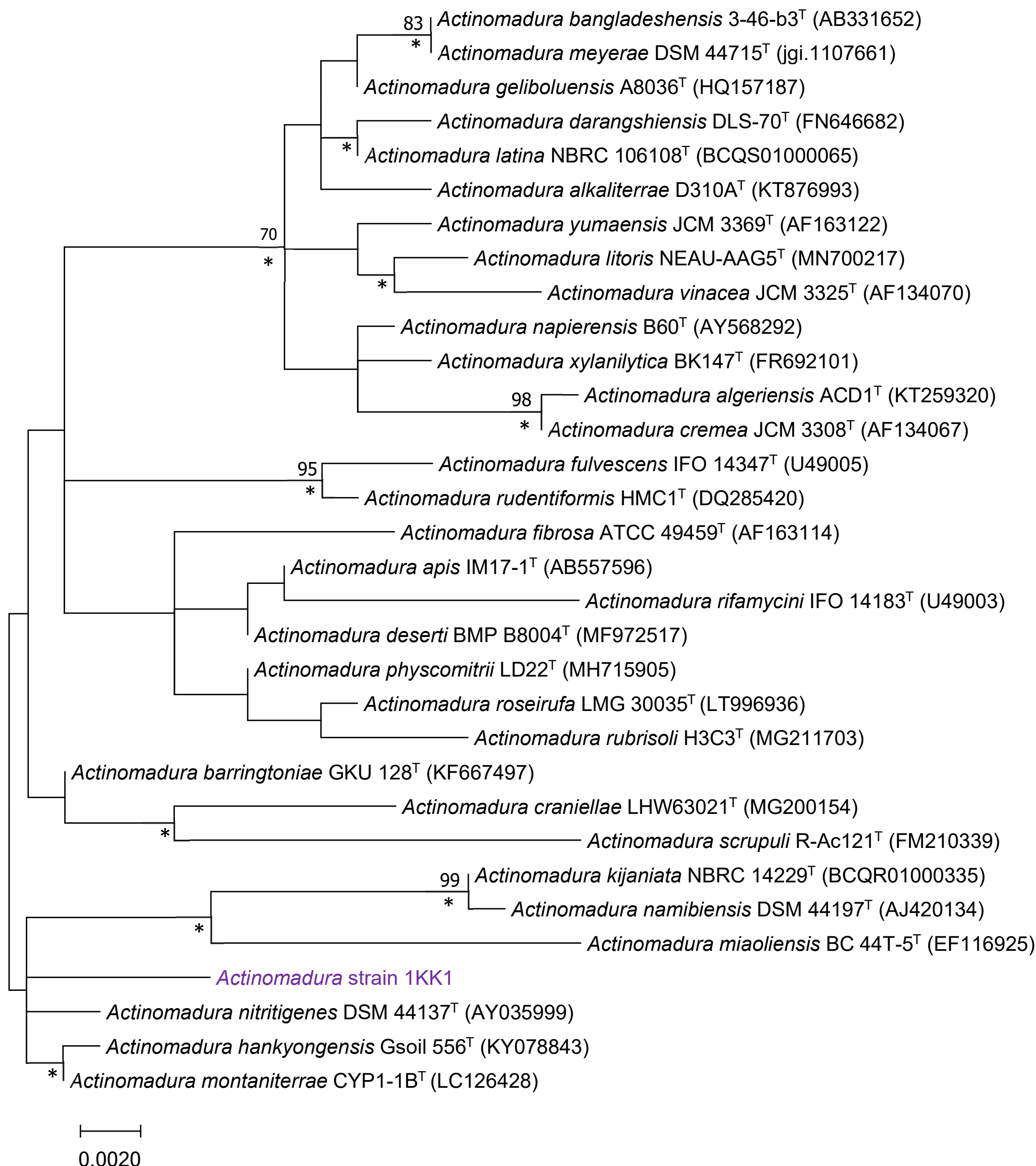

Figure 3.1: Maximum likelihood phylogenetic tree of *Actinomadura* strain 1KK1 compared to the top 31 *Actinomadura* hits by EzBioCloud analysis based on a partial 16S-rRNA gene sequence alignment of 832 bp. The value at each node represents the percentage bootstrap support calculated from 1000 resampled datasets. Only bootstrap values of ≥ 70% are displayed. The scale bar indicates two nucleotide changes per 1000 base pairs. "*" Indicates matching topology with the neighbour joining and maximum parsimony trees.

3.3.3.2 Genus *Nonomuraea*

Nonomuraea is a genus within the family *Streptosporangiaceae* (https://lpsn.dsmz.de/family/streptosporangiaceae). The genus is comprised of 66 species with validly published and correct names (https://lpsn.dsmz.de/genus/nonomuraea). *Nonomuraea* strain 7TK1 was isolated on KSM1 medium from the TC soil sample. *Nonomuraea* strain 7TK1 showed dark orange, smooth and raised colonies that were sunken into the agar when grown on ISP2 medium and was unable to grow on ISP4 medium (Table A3). A phylogenetic tree was generated for *Nonomuraea* strain 7TK1 using the 16S-rRNA gene sequences of the top 30 *Nonomuraea* type strain hits on EzBioCloud in order to determine its phylogenetic position within the genus *Nonomuraea* (Figure 3.2).

An EzBioCloud analysis showed that *Nonomuraea* strain 7TK1 had highest 16S-rRNA gene sequence similarity to *Nonomuraea cavernae* SYSU K10005^{T} (99.43% sequence similarity over 701 bp; Table 3.1). In the maximum-likelihood phylogenetic tree based on 16S-rRNA gene sequences (Figure 3.2), strain 7TK1 formed part of a ten-strain clade, which included *N. cavernae* SYSU K10005^{T}. However, there was weak support for this association (bootstrap value <70%) and this clustering was not seen in the NJ and MP trees (data not shown).

For further investigation of the phylogeny of *Nonomuraea* strain 7TK1, a *gyrB* gene phylogenetic tree was generated using the *gyrB* gene sequences of the top 30 hits from a blastn analysis (https://blast.ncbi.nlm.nih.gov/Blast.cgi; Figure 3.3). *Nonomuraea* 7TK1 grouped closely with *Nonomuraea guangzhouensis* DSM 45889^{T} with strong bootstrap support (82%; Figure 3.3). This association was also seen in the NJ and MP trees (data not shown). This result is in agreement with the *gyrB* gene blastn result where *Nonomuraea* strain 7TK1 had *N. guangzhouensis* DSM 45889^{T} as its top hit with 96.99% sequence similarity over 632 bp (Table 3.1). *Nonomuraea* strain 7TK1 and *N. guangzhouensis* DSM 45889^{T} were also part of the same ten-strain cluster in the 16S-rRNA gene phylogenetic tree (Figure 3.2). Whole-genome sequencing followed by ANI and dDDH analyses would be required to determine the species to which *Nonomuraea* strain 7TK1 belongs.

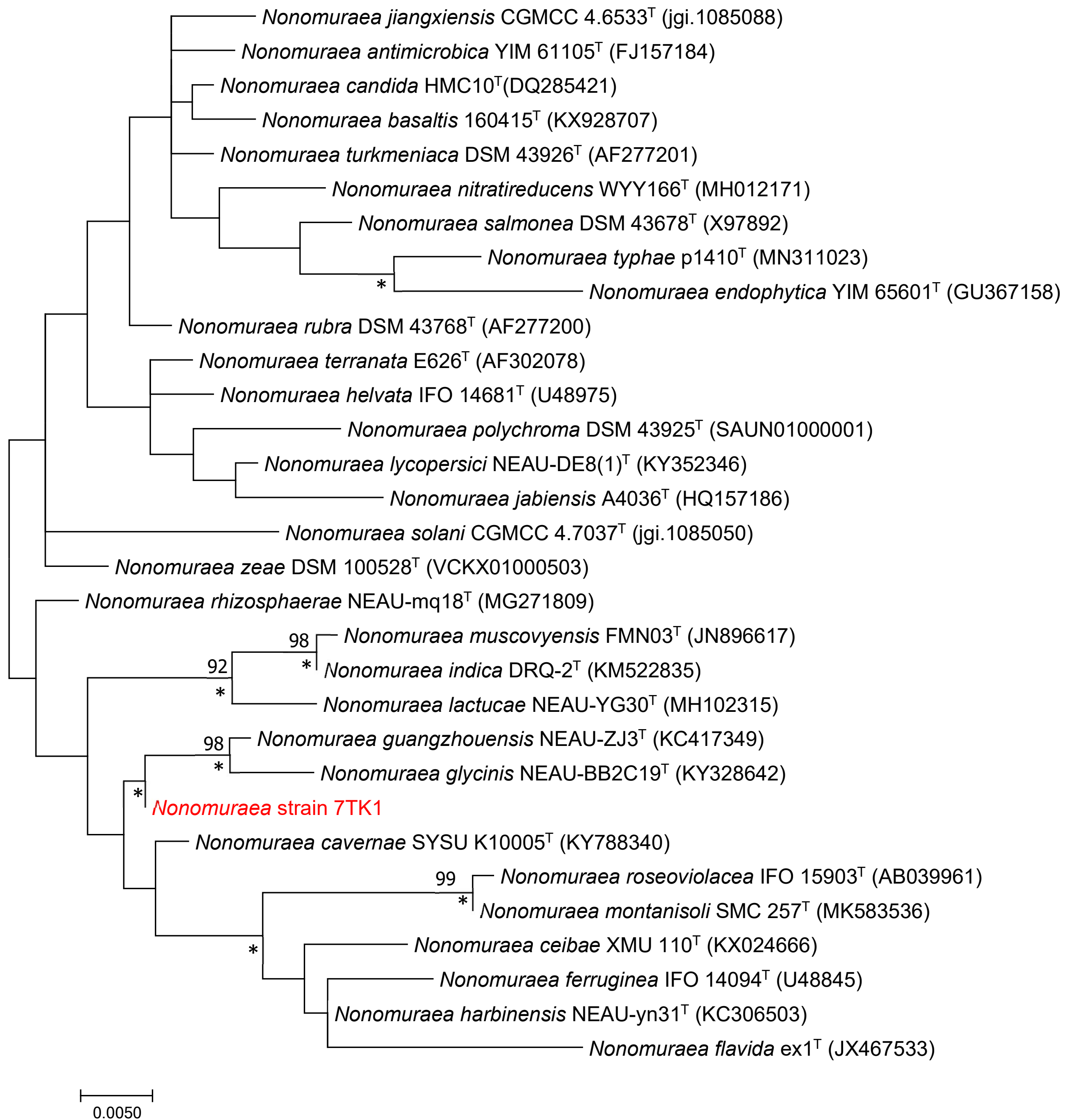

Figure 3.2: Maximum likelihood phylogenetic tree of *Nonomuraea* strain 7TK1 compared to the top 30 *Nonomuraea* hits by EzBioCloud analysis based on a partial 16S-rRNA gene sequence alignment of 701 bp. The value at each node represents the percentage bootstrap support calculated from 1000 resampled datasets. Only bootstrap values of ≥ 70% are displayed. The scale bar indicates 5 nucleotide changes per 1000 base pairs. "*" Indicates matching topology with the neighbour joining and maximum parsimony trees.

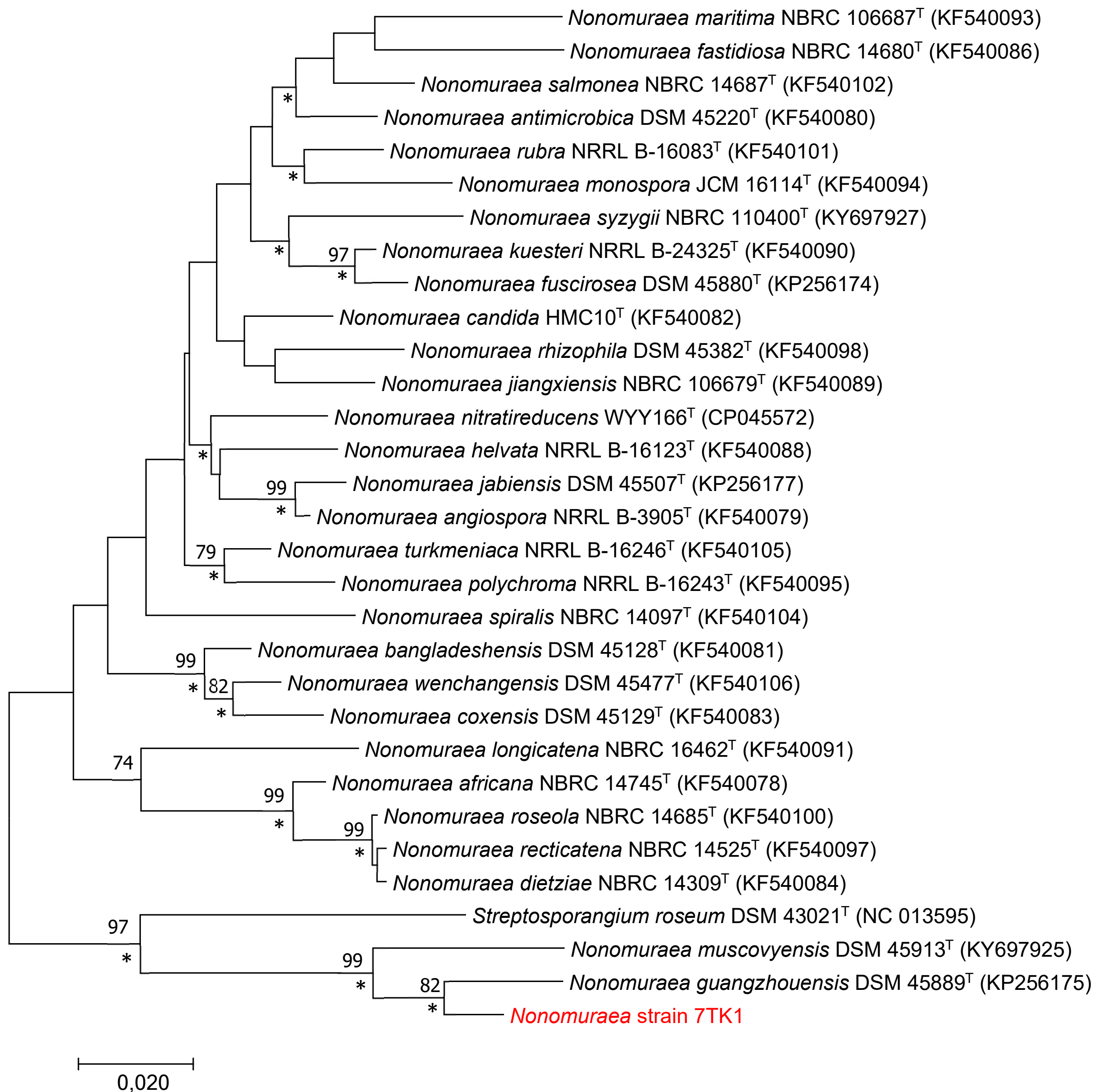

Figure 3.3: Maximum likelihood phylogenetic tree of *Nonomuraea* strain 7TK1 compared to the 30 top hits (with validly published names) by blastn analysis based on a partial *gyrB* gene sequence alignment of 632 bp. The value at each node represents the percentage bootstrap support calculated from 1000 resampled datasets. Only bootstrap values of ≥ 70% are displayed. The scale bar indicates two nucleotide changes per 100 base pairs. "*" Indicates matching topology with the neighbour joining and maximum parsimony trees.

3.3.3.3 Genus *Streptomyces*

The genus *Streptomyces* forms part of the family *Streptomycetaceae* and is one of six genera with validly published and correct names (*Allostreptomyces*, *Embleya*, *Kitasatospora*, *Streptacidiphilus*, *Streptomyces* and *Yinghuangia*; https://lpsn.dsmz.de/family/streptomycetaceae). *Streptomyces* is the largest genus within the family *Streptomycetaceae* and comprises of 697 species with validly published and correct names (https://lpsn.dsmz.de/genus/streptomyces).

In this study, two separate phylogenies of *Streptomyces* isolates were constructed: (1) A tree was constructed with all the *Streptomyces* strains assigned to potential-clone groups and (2) a second tree was constructed with all the *Streptomyces* strains not assigned to potential-clone groups. Due to the large size of the trees constructed, parts of the phylogenetic trees were compressed to make them more compact and the figures are displayed as the top and bottom sections of the trees. To determine the phylogenetic position of isolated *Streptomyces* strains within the genus *Streptomyces*, the type strains of the top five hits from EzBioCloud for each isolate were included in the phylogenetic analysis.

3.3.3.3.1 *Streptomyces* strains assigned to a potential-clone group

Figure 3.4 displays the phylogenetic analysis of 30 *Streptomyces* isolates: strains 5TK1, 6TK1, 9TK1, 15TA1, 16TA1, 18TA1, 19TA1, 20TA1, 33TA2, 36TA2, 45TA3 and 47TA3 (isolated from the TC soil sample and strains of the green potential-clone group), strains 2KK1, 10KK1, 28KA1, 31KA1, 32KA1, 52KA3, 54KA3 and 59KA3 (isolated from the K soil sample and strains of the blue potential-clone group) and strains 22KA1, 25KA1, 26KA1, 39KA2, 40KA2, 41KA2, 42KA2, 55KA3, 58KA3 and 61KA3 (isolated from the K soil sample and strains of the orange potential-clone group). The phylogenetic analysis was based on a partial 16S-rRNA gene sequence alignment of 545 bp.

An EzBioCloud analysis showed that *Streptomyces* strains assigned to the blue and green potential-clone groups all had the same three top hits (*Streptomyces platensis* JCM 4662^T, *Streptomyces decoyicus* NRRL 2666^T and *Streptomyces caniferus* NBRC 15389^T; Table 3.1). All the strains in the blue and green potential-clone groups, except for *Streptomyces* strain 47TA3, had 100% similarity to these top hits, with *Streptomyces* strain 47TA3 having 99.9% similarity to the top hits (Table 3.1). In the

maximum-likelihood phylogenetic tree based on 16S-rRNA gene sequences (Figure 3.4a), all the *Streptomyces* strains assigned to the blue and green potential-clone groups were located within the same 28-strain clade, which included *S. caniferus* NBRC 15389^{T}, *S. decoyicus* NRRL 2666^{T}, *Streptomyces kronopolitis* NEAU-ML8^{T}, *Streptomyces monomycini* NRRL B-24309^{T}, *Streptomyces olivaceiscleroticus* DSM 40595^{T}, *S. platensis* JCM 4662^{T}, *Streptomyces ramulosus* NRRL B-2714^{T} and *Streptomyces sclerotialus* NRRL ISP-5269^{T}. This association was also seen in the NJ and MP trees (data not shown). However, there was weak support for this association (bootstrap value <70%; Figure 3.4a).

An EzBioCloud analysis showed that *Streptomyces* strains assigned to the orange potential-clone group all had the same five top hits (including *Streptomyces lydicus* ATCC 25470^{T} and *Streptomyces chattanoogensis* NRRL ISP-5002^{T}; Table 3.1). *Streptomyces* strains 22KA1, 25KA1, 26KA1, 40KA2, 41KA2, 58KA3 and 61KA3 had 100% similarity to the top hits and *Streptomyces* strains 39KA2, 42KA2 and 55KA3 had 99.9% similarity to the top hits (Table 3.1). In the maximum-likelihood phylogenetic tree based on 16S-rRNA gene sequences (Figure 3.4b), all the *Streptomyces* strains assigned to the orange potential-clone group clustered within the same 17-strain clade, which included *Streptomyces angustmyceticus* NRRL B-2347^{T}, *S. chattanoogensis* NRRL ISP-5002^{T}, *Streptomyces libani* subsp. *libani* NBRC 13452^{T}, *Streptomyces lydicamycinicus* NBRC 110027^{T}, *S. lydicus* ATCC 25470^{T}, *Streptomyces nigrescens* NBRC 12894^{T} and *Streptomyces sioyaensis* NRRL B-5408^{T}. This association had very strong bootstrap support (99%; Figure 3.4b) and was also seen in the NJ and MP trees (data not shown).

Streptomyces strains 39KA2 and 42KA2 formed a two-strain sub-clade, which formed a deep branch to the 15-strain clade that included all the other orange potential-clone group strains (Figure 3.4b). This association was seen in the NJ and MP trees (data not shown) but had weak bootstrap support (<70%; Figure 3.4b).

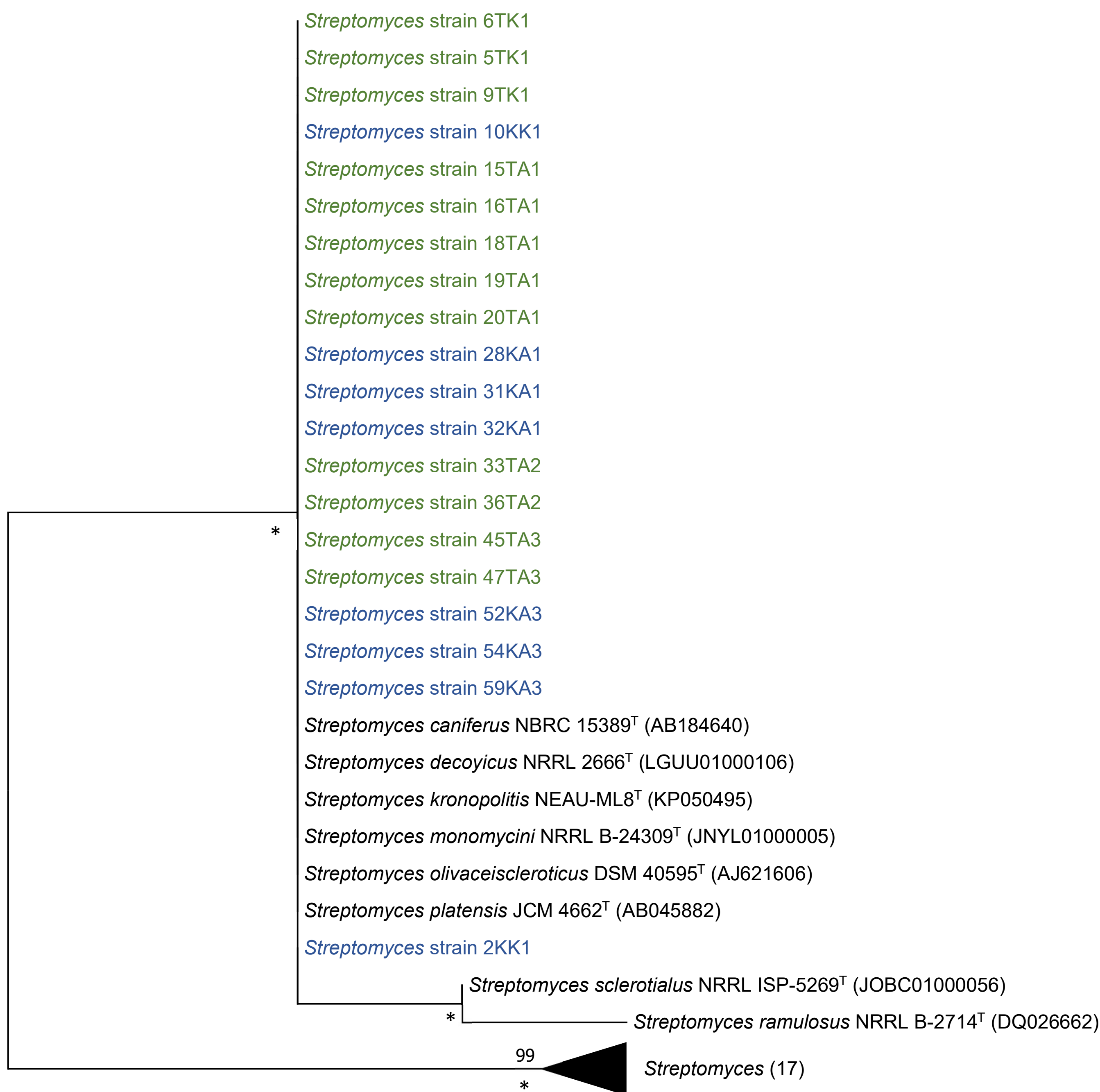

Figure 3.4a: Maximum likelihood phylogenetic tree of thirty *Streptomyces* isolates compared to the top five *Streptomyces* hits for each isolate by EzBioCloud analysis based on a partial 16S-rRNA gene sequence alignment of 545 bp (20 of the isolates can be seen in this part of the tree). The lower section of the tree is compressed and is shown in Figure 3.4b. The number in brackets indicates how many strains are in the compressed part of the tree. The value at each node represents the percentage bootstrap support calculated from 1000 resampled datasets. Only bootstrap values of ≥ 70% are displayed. The scale bar indicates 1 nucleotide changes per 1000 base pairs. "*" Indicates matching topology with the neighbour joining and maximum parsimony trees.

Figure 3.4b: Maximum likelihood phylogenetic tree of thirty *Streptomyces* isolates compared to the top five *Streptomyces* hits for each isolate by EzBioCloud analysis based on a partial 16S-rRNA gene sequence alignment of 545 bp (10 of the isolates can be seen in this part of the tree). The upper section of the tree is compressed and is shown in Figure 3.4a. The number in brackets indicates how many strains are in the compressed part of the tree. The value at each node represents the percentage bootstrap support calculated from 1000 resampled datasets. Only bootstrap values of ≥ 70% are displayed. The scale bar indicates 1 nucleotide changes per 1000 base pairs. "*" Indicates matching topology with the neighbour joining and maximum parsimony trees.

3.3.3.3.2 *Streptomyces* strains not assigned to a potential-clone group

An EzBioCloud analysis showed that *Streptomyces* strain 12TA1 had highest 16S-rRNA gene sequence similarity to *Streptomyces chartreusis* NBRC 12753^T (99.4% sequence similarity over 839 bp; Table 3.1). In the maximum-likelihood phylogenetic tree based on 16S-rRNA gene sequences (Figure 3.5a), *Streptomyces* strain 12TA1 formed part of a 14-strain clade, but the association had weak bootstrap support (<70%; Figure 3.5a) and was not supported by the NJ and MP trees (data not shown). *Streptomyces* strain 12TA1 and *Streptomyces olivochromogenes* DSM 40451^T formed a deep branch to a 12-strain clade which included *S. chartreusis* NBRC 12753^T (Figure 3.5a). The 12-strain clade also had weak bootstrap support (<70%; Figure 3.5a) and was and was not supported by the NJ and MP trees (data not shown).

An EzBioCloud analysis showed that *Streptomyces* strain 49TA3 had highest 16S-rRNA gene sequence similarity to *Streptomyces geldanamycininus* NRRL 3602^T (99.85% sequence similarity over 669 bp; Table 3.1). In the maximum-likelihood phylogenetic tree based on 16S-rRNA gene sequences (Figure 3.5a), *Streptomyces* strain 49TA3 clustered with *S. geldanamycininus* NRRL 3602^T. There was good bootstrap support for this association (86%; Figure 3.5a), which was also seen in the NJ and MP trees (data not shown).

An EzBioCloud analysis showed that *Streptomyces* strain 63KA3 had highest 16S-rRNA gene sequence similarity to *Streptomyces aureus* NBRC 100912^T (100% sequence similarity over 808 bp; Table 3.1). In the maximum-likelihood phylogenetic tree based on 16S-rRNA gene sequences (Figure 3.5a), *Streptomyces* strain 63KA3 formed part of a six-strain clade, which included *S. aureus* NBRC 100912^T. This association had very strong bootstrap support (98%; Figure 3.5a) and was also seen in the NJ and MP trees (data not shown). Phenotypic characterization showed *Streptomyces* strain 63KA3 to have bright yellow substrate mycelium and to produce a yellow diffusible pigment on ISP4 medium (Table A3; Figure A1). *S. aureus* makes a golden coloured diffusible pigment when grown on media lacking malt extract (such as ISP4) (Manfio *et al.*, 2003). This result suggests that *Streptomyces* strain 63KA3 may be a strain of *S. aureus*. However, additional testing (ANI and dDDH) would be required to confirm that strain 63KA3 belongs to *S. aureus*.

An EzBioCloud analysis showed that *Streptomyces* strain 60KA3 had highest 16S-rRNA gene sequence similarity to *Streptomyces adustus* WH-9^{T} (99.66% sequence similarity over 870 bp; Table 3.1) and that *Streptomyces* strain 23KA1 had highest 16S-rRNA gene sequence similarity to three top hits: *S. adustus* WH-9^{T}, *Streptomyces griseochromogenes* ATCC 14511^{T} and *Streptomyces lucensis* NBRC 13056^{T} (99.62% sequence similarity over 799 bp; Table 3.1). In the maximum-likelihood phylogenetic tree based on 16S-rRNA gene sequences (Figure 3.5a), *Streptomyces* strains 60KA3 and 23KA1 clustered together in a 12-strain clade, which included *S. adustus* WH-9^{T}, *S. griseochromogenes* ATCC 14511^{T} and *S. lucensis* NBRC 13056^{T}. This association had weak bootstrap support (<70%; Figure 3.5a) and was not supported by NJ and MP trees (data not shown). The association of strains 60KA3 and 23KA1 had strong bootstrap support (84%; Figure 3.5a) and was also seen in the NJ and MP trees (data not shown).

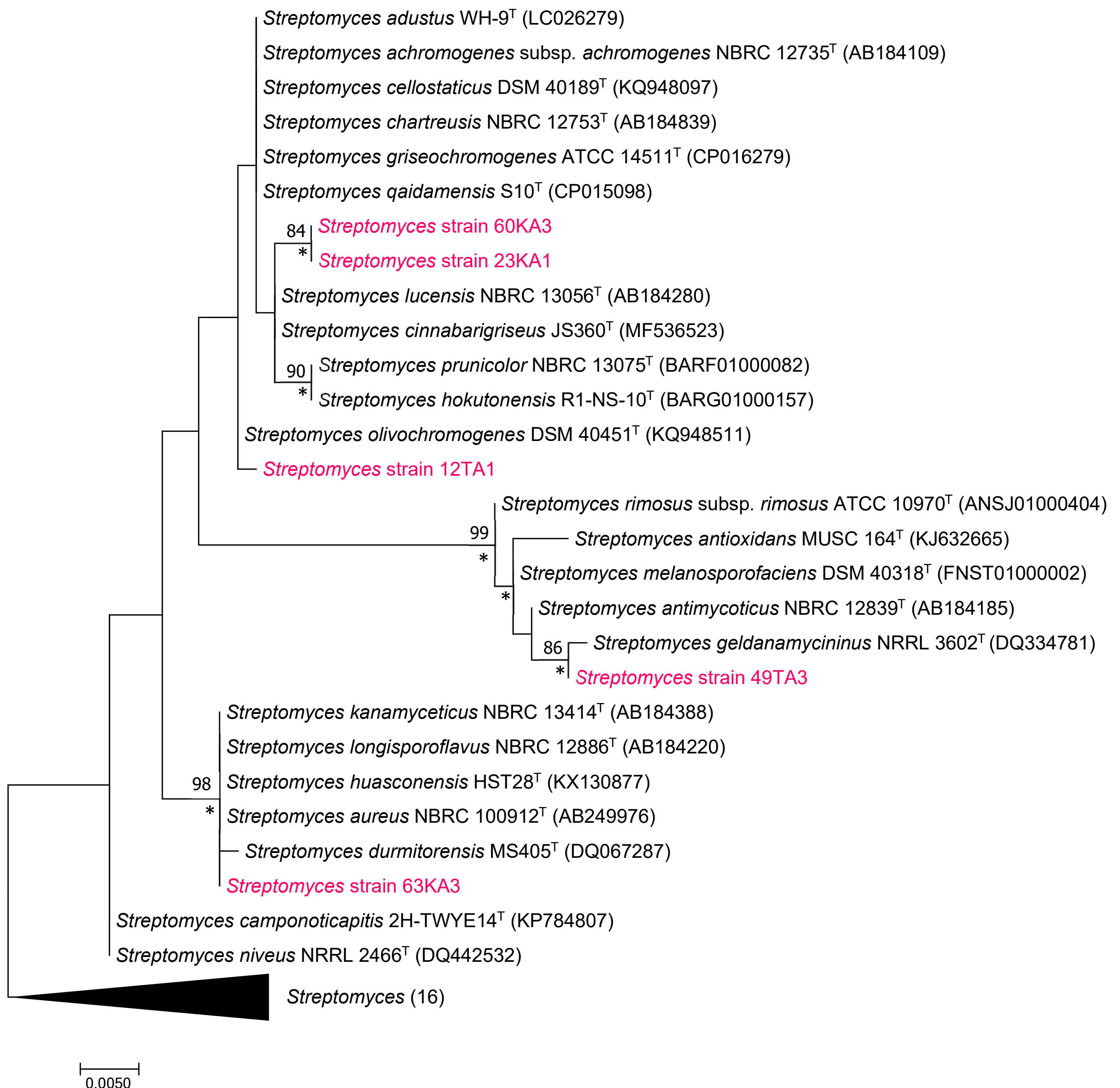

Figure 3.5a: Maximum likelihood phylogenetic tree of eight *Streptomyces* isolates compared to the top five hits of each isolate by EzBioCloud analysis based on a partial 16S-rRNA gene sequence alignment of 646 bp (only five of the isolates can be seen in this part of the tree). The lower section of the tree is compressed and is shown in Figure 3.5b. The number in brackets indicates how many strains are in the compressed part of the tree. The value at each node represents the percentage bootstrap support calculated from 1000 resampled datasets. Only bootstrap values of ≥ 70% are displayed. The scale bar indicates 5 nucleotide changes per 1000 base pairs. "*" Indicates matching topology with the neighbour joining and maximum parsimony trees.

An EzBioCloud analysis showed that *Streptomyces* strain 8KK1 had highest 16S-rRNA gene sequence similarity to *Streptomyces abietis* A191^T (100% sequence similarity over 828 bp; Table 3.1). In the maximum-likelihood phylogenetic tree based on 16S-rRNA gene sequences (Figure 3.5b), *Streptomyces* strain 8KK1 formed part of a three-strain clade, which included *Streptomyces alni* D65^T and *S. abietis* A191^T. There was weak bootstrap support for this association (<70%; Figure 3.5b), but the association was seen in the NJ and MP trees (data not shown).

An EzBioCloud analysis showed that *Streptomyces* strain 13TA1 had highest 16S-rRNA gene sequence similarity to *Streptomyces lannensis* TA4-8^T (100% sequence similarity over 842 bp; Table 3.1). In the maximum-likelihood phylogenetic tree based on 16S-rRNA gene sequences (Figure 3.5b), *Streptomyces* strain 13TA1 clustered with *S. lannensis* TA4-8^T. There was very strong bootstrap support for this association (95%; Figure 3.5b), which was also seen in the NJ and MP trees (data not shown).

An EzBioCloud analysis showed that *Streptomyces* strain 48TA3 had highest 16S-rRNA gene sequence similarity to *Streptomyces phaeoluteichromatogenes* NRRL 5799^T (99.88% sequence similarity over 864 bp; Table 3.1). In the maximum-likelihood phylogenetic tree based on 16S-rRNA gene sequences (Figure 3.5b), *Streptomyces* strain 48TA3 formed part of a six-strain clade, which included *S. phaeoluteichromatogenes* NRRL 5799^T. This association had very strong bootstrap support (98%; Figure 3.5b) and was also seen in the NJ and MP trees (data not shown).

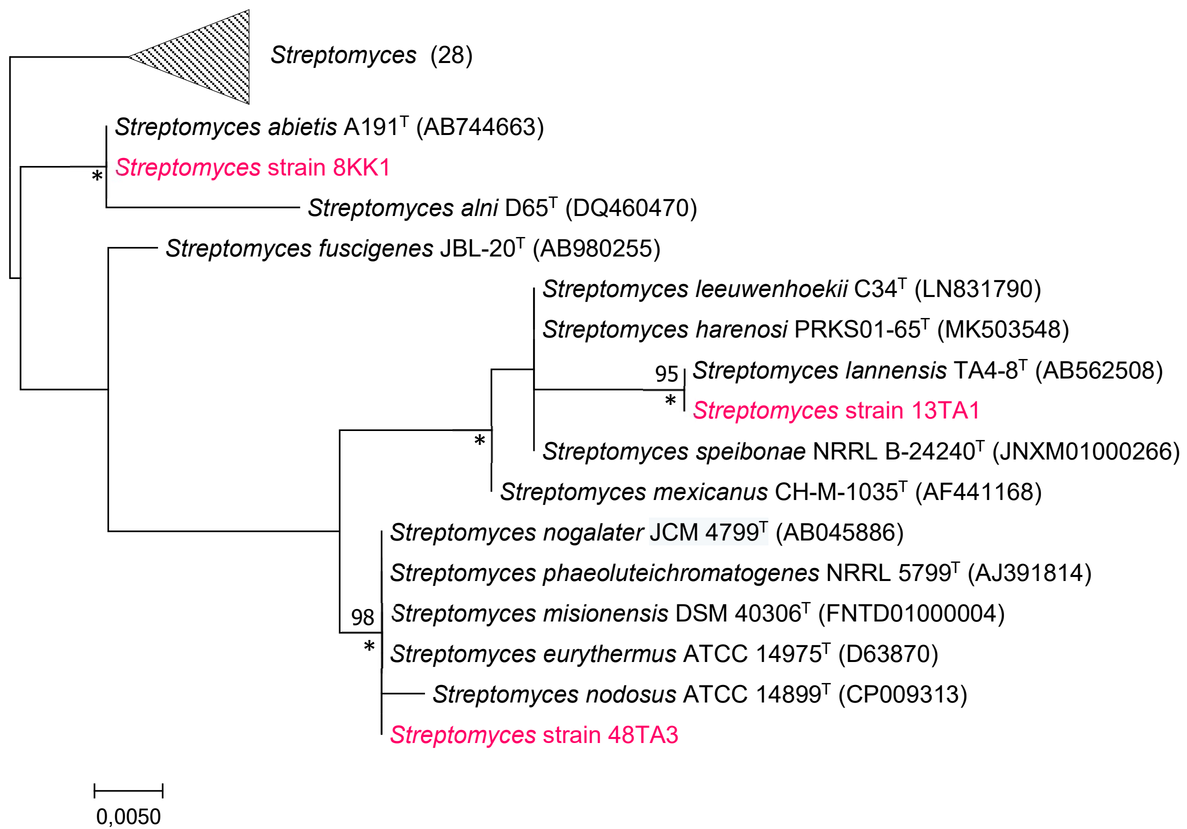

Figure 3.5b: Maximum likelihood phylogenetic tree of eight *Streptomyces* isolates compared to the top five hits of each isolate by EzBioCloud analysis based on a partial 16S-rRNA gene sequence alignment of 646 bp (only three of the isolates can be seen in this part of the tree). The upper section of the tree is compressed and is shown in Figure 3.5a. The number in brackets indicates how many strains are in the compressed part of the tree. The value at each node represents the percentage bootstrap support calculated from 1000 resampled datasets. Only bootstrap values of ≥ 70% are displayed. The scale bar indicates 5 nucleotide changes per 1000 base pairs. “*” Indicates matching topology with the neighbour joining and maximum parsimony trees.

3.3.4 Clone analysis

There were 29 *Streptomyces* isolates that had identical 16S-rRNA gene sequences to at least one other *Streptomyces* strain from the same sample (Table 3.1). Therefore some of the *Streptomyces* strains isolated could be clones of each other. Clones in this case being defined as isolates with identical genomic material originating from the same dividing bacterium. There were three potential-clone groups in this study. The K soil sample had two potential-clone groups: the blue potential-clone group and the orange potential-clone group. The blue potential-clone group is a group of eight *Streptomyces* strains (2KK1, 10KK1, 28KA1, 31KA1, 32KA1, 52KA3, 54KA3 and 59KA3) with the 16S-rRNA gene sequence top hit of *S. platensis* with 100% similarity in an EzBioCloud analysis (highlighted in blue) (Table 3.1). The orange potential-clone group is a group of ten *Streptomyces* strains (22KA1, 25KA1, 26KA1, 39KA2, 40KA2, 41KA2, 42KA2, 55KA3, 58KA3 and 61KA3) with the 16S-rRNA gene sequence top hit of *S. lydicus* with 100% similarity in an EzBioCloud analysis (highlighted in orange) (Table 3.1). The TC soil sample had one potential group of clones: the green potential-clone group. The green potential-clone group is a group of 12 *Streptomyces* strains (5TK1, 6TK1, 9TK1, 15TA1, 16TA1, 18TA1, 19TA1, 20TA1, 33TA2, 36TA2, 45TA3 and 47TA3) with the 16S-rRNA gene sequence top hit of *S. platensis* with 100% similarity in an EzBioCloud analysis (highlighted in green) (Table 3.1). All strains within potential-clone groups are highlighted in their distinct colour (blue, green or orange) in all figures and tables in this chapter.

Within the orange potential-clone group: *Streptomyces* strains 39KA2, 42KA2 and 55KA3 had identical 16S-rRNA gene sequences to each other, but had one nucleotide difference to the other strains within the orange clone group (Table 3.1). Within the green potential-clone group: *Streptomyces* strain 47TA3 had one nucleotide difference in its 16S-rRNA gene sequence to the other strains within the green potential-clone group (Table 3.1). These *Streptomyces* strains were included in their distinctive potential-clone groups to account for potential sequencing errors that may have occurred. The strains within the blue potential-clone group and the strains within the green potential-clone group had, by coincidence, identical 16S-rRNA gene sequences to each other. It is however impossible for a strain from the blue potential-clone group to be a clone of a strain from the green potential-clone group. This is because the blue and green potential-clone groups originated from different soil samples from

geographically distinct sampling locations a few hundred metres apart and therefore are expected to have had different parent bacteria.

It is important to determine whether isolated strains are clones of each other and to remove clones from further investigation. This is because, by dereplicating clone sets, you can reduce future work on the isolates by working on only one isolated strain from a group of clones. It should be noted that the 16S-rRNA gene has a very conserved sequence, so it is possible for strains to have identical partial 16S-rRNA gene sequences and be otherwise genetically different to each other.

The *gyrB* gene exhibits much more sequence variability than the 16S-rRNA gene and is, therefore, less conserved. This would mean that if two strains had identical *gyrB* gene sequences they would likely be genetically identical and, therefore, clones of each other. In order to determine which strains from the different potential-clone groups are indeed clones, the *gyrB* gene was sequenced for 23 of the 30 potential-clones. The *gyrB* gene sequences of the strains in each clone group were aligned in DNAMAN version 4.13. These multiple sequence alignments were used to create homology matrices for each potential-clone group, comparing the *gyrB* gene sequences to each other (Figure 3.7, Figure 3.8 and Figure 3.9). *gyrB* gene sequences of 589-669 bp in length were generated for the analyses (Table 3.1).

A maximum-likelihood phylogenetic tree was constructed in MEGA version 7.0 using all the *Streptomyces* strain *gyrB* gene sequences (Figure 3.6). The top five hits, with validly published names, of each *Streptomyces* strain based on a blastn analysis (https://blast.ncbi.nlm.nih.gov/Blast.cgi) of their *gyrB* gene sequences were included in the analysis (Figure 3.6).

In the maximum-likelihood phylogenetic tree based on *gyrB* gene sequences (Figure 3.6), each potential-clone group had all their strains clustered together and apart from strains of the other potential-clone groups (Figure 3.6). This shows the three clone groups to be distinct from each other with the members within the groups being closely related to each other.

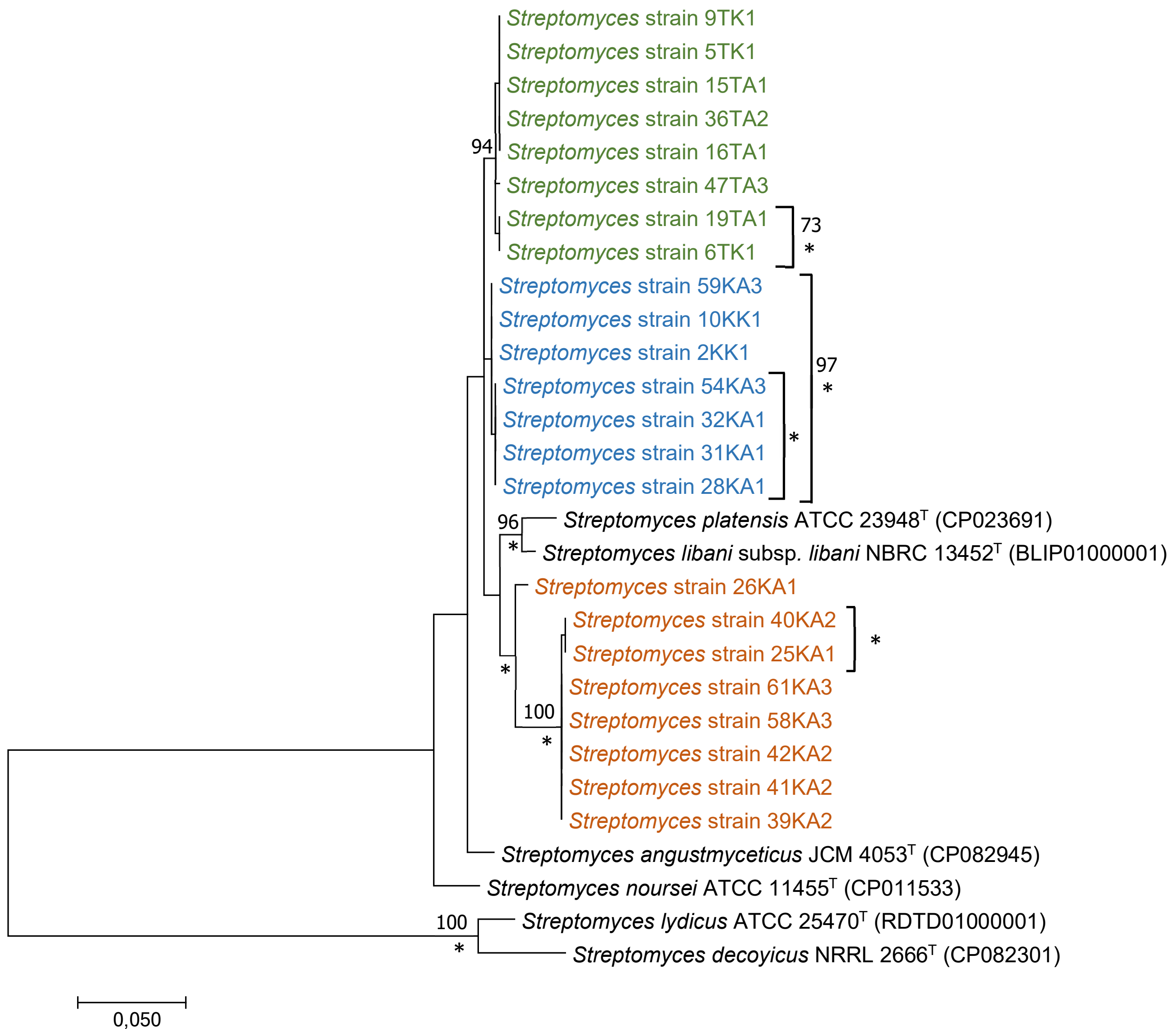

Figure 3.6: Maximum likelihood phylogenetic tree of 23 *Streptomyces* isolates compared to the top five hits (with validly published names) of each isolate by blastn analysis based on a partial *gyrB* gene sequence alignment of 576 bp. The value at each node represents the percentage bootstrap support calculated from 1000 resampled datasets. Only bootstrap values of ≥ 70% are displayed. The scale bar indicates 5 nucleotide changes per 100 base pairs. "*" Indicates matching topology with the neighbour joining and maximum parsimony trees.

3.3.4.1 The blue potential-clone group

Streptomyces 2KK1	100%						
Streptomyces 10KK1	100%	100%					
Streptomyces 59KA3	100%	100%	100%				
Streptomyces 28KA1	99.8%	99.7%	99.8%	100%			
Streptomyces 31KA1	99.8%	99.8%	99.8%	100%	100%		
Streptomyces 32KA1	99.8%	99.8%	99.8%	100%	100%	100%	
Streptomyces 54KA3	99.8%	99.8%	99.8%	100%	100%	100%	100%

Figure 3.7: Homology matrix based on the alignment of seven partial *gyrB* gene sequences of 589 bp from strains isolated from the Kramer soil sample within the blue potential-clone group. The homology matrix was generated using DNAMAN.

The *gyrB* gene sequences of seven strains in the blue potential-clone group from the K soil sample were analysed (Figure 3.6 and Figure 3.7). Based on the homology analysis, the blue potential-clone group appears to comprise two groups (Figure 3.7). *Streptomyces* strains 2KK1, 10KK1 and 59KA3 have identical *gyrB* gene sequences to each other, which suggests that they are clones of each other. However, growth of *Streptomyces* strains 2KK1, 10KK1 and 59KA3 on ISP4 shows *Streptomyces* strains 2KK1 and 10KK1 having pale yellow substrate mycelium, with *Streptomyces* strain 59KA3 differing in having burnt orange substrate mycelium (Table A3). This evidence disagrees with the homology analysis result and suggests that strain 59KA3 is not a clone of strains 2KK1 and 10KK1. *Streptomyces* strains 28KA1, 31KA1, 32KA1 and 54KA3 have identical *gyrB* gene sequences to each other and are, therefore, presumed to be clones of each other (Figure 3.7). Growth of *Streptomyces* strains 28KA1, 32KA1 and 54KA3 on ISP4 showed that they all had pale brown substrate mycelium with a white and grey spore mass (Table A3). Strain 31KA1 produced a pale-yellow substrate mycelium and white and grey spore mass on ISP4. Therefore, evidence suggests that strain 31KA1 is not be a clone of strains 28KA1, 31KA1 and 54KA3.

In the maximum-likelihood phylogenetic tree based on *gyrB* gene sequences (Figure 3.6), all of the strains from the blue potential-clone group clustered closely together. There was very strong bootstrap support for this association (97%; Figure 3.6) and the association was also seen in the NJ and MP trees (data not shown). This tree also shows that *Streptomyces* strains 28KA1, 31KA1, 32KA1 and 54KA3 formed a four-

strain sub-clade separate from the other strains in their potential-clone group. There was weak bootstrap support for this association (<70%; Figure 3.6), but this grouping was seen in the NJ and MP trees (data not shown).

Summarising the available data, strains 2KK1 and 10KK1 are likely to be clones of each other and strains 28KA1, 32KA1 and 54KA3 are likely to be clones of each other. Strain 59KA3, while very closely related to strains 2KK1 and 10KK1, was shown by its appearance on ISP4 to be a distinct strain, not a clone. Similarly, strain 31KA1 is very closely related to strains 28KA1, 32KA1 and 54KA3, but may be a distinct strain from these clones.

Streptomyces strain 52KA3 was the only strain from the potential blue clone group for which a *gyrB* gene sequenced was not determined. *Streptomyces* strain 52KA3 shows a pale brown substrate mycelium and a white and grey spore mass on ISP4 plates, showing the same colony characteristics as for *Streptomyces* strains 28KA1, 32KA1 and 54KA3 (Table A3). This suggests that *Streptomyces* strain 52KA3 may be a clone of *Streptomyces* strains 28KA1, 32KA1 and 54KA3.

3.3.4.2 The orange potential-clone group

Streptomyces 39KA2	100%							
Streptomyces 41KA2	100%	100%						
Streptomyces 42KA2	100%	100%	100%					
Streptomyces 58KA3	100%	100%	100%	100%				
Streptomyces 61KA3	100%	100%	100%	100%	100%			
Streptomyces 25KA1	99.8%	99.8%	99.8%	99.8%	99.8%	100%		
Streptomyces 40KA2	99.8%	99.8%	99.8%	99.8%	99.8%	100%	100%	
Streptomyces 26KA1	97.4%	97.4%	97.4%	97.4%	97.4%	97.3%	97.3%	100%

Figure 3.8: Homology matrix based on an alignment of eight partial *gyrB* gene sequences of 618 bp from strains isolated from the Kramer soil sample within the orange potential-clone group. The homology matrix was generated using DNAMAN.

The *gyrB* gene sequences of eight strains in the orange potential-clone group from the K soil sample were analysed (Figure 3.6 and Figure 3.8). *Streptomyces* strain 26KA1 had a unique *gyrB* gene sequence and therefore is not a clone (Figure 3.8). *Streptomyces* strains 39KA2, 41KA2, 42KA2, 58KA3 and 61KA3 have identical *gyrB* gene sequences to each other and are, therefore, presumed to be clones of each other (Figure 3.8). However, cultivation of *Streptomyces* strains 39KA2, 41KA2, 42KA2 and

61KA3 on ISP4 showed that *Streptomyces* strains 39KA2, 41KA2 and 42KA2 had a transparent white substrate mycelium, but *Streptomyces* strain 61KA3 had a pale brown substrate mycelium (Table A3). This evidence suggests that strain 61KA3 is not a clone. Furthermore, *Streptomyces* strain 58KA3 was unable to grow on ISP4 (Table A3). This evidence suggests that strain 58KA3 is also not a clone. *Streptomyces* strains 25KA1 and 40KA2 have identical *gyrB* gene sequences to each other and are, therefore, presumed to be clones of each other (Figure 3.8). Growth of *Streptomyces* strains 25KA1 and 40KA2 on ISP4 show *Streptomyces* strain 25KA1 having a transparent white substrate mycelium, but *Streptomyces* strain 40KA2 having a pale yellow/pale brown substrate mycelium (Table A3). This evidence suggests that strains 25KA1 and 40KA2 are not clones.

In the maximum-likelihood phylogenetic tree based on *gyrB* gene sequences (Figure 3.6), all of the strains from the orange potential-clone group clustered closely together. There was weak bootstrap support for this association (<70%; Figure 3.6) but the association was seen in the NJ and MP trees (data not shown). This tree also shows that *Streptomyces* strain 26KA1 branched separately from the other orange potential-clone group strains (Figure 3.6). The seven-strain clade was supported by a perfect bootstrap value (100%; Figure 3.6) and this association was also seen in the NJ and MP trees (data not shown).

In summary, strains 39KA2, 41KA2 and 42KA2 are likely to be clones of each other. Strains 61KA3 and 58KA3, while very closely related to strains 39KA2, 41KA2 and 42KA2, were shown by their appearance on ISP4 to be a distinct strains, not clones. Strains 25KA1 and 40KA2, while having identical partial *gyrB* gene sequences, are phenotypically distinct on ISP4 and are therefore not clones. Strain 26KA1, while having an identical partial 16S-rRNA gene sequence to the other member of the orange potential-clone group, has a distinct *gyrB* gene sequence and is therefore not a clone.

Streptomyces strains 55KA3 and 22KA1 are the only members of the orange potential-clone group for which *gyrB* gene sequences were not determined.

3.3.4.3 The green potential-clone group

Streptomyces 5TK1	100%							
Streptomyces 9TK1	100%	100%						
Streptomyces 15TA1	100%	100%	100%					
Streptomyces 16TA1	100%	100%	100%	100%				
Streptomyces 36TA2	100%	100%	100%	100%	100%			
Streptomyces 6TK1	99.7%	99.7%	99.7%	99.7%	99.7%	100%		
Streptomyces 19TA1	99.7%	99.7%	99.7%	99.5%	99.5%	100%	100%	
Streptomyces 47TA3	99.7%	99.7%	99.7%	99.7%	99.7%	99.7%	99.7%	100%

Figure 3.9: Homology matrix based on alignment of eight partial *gyrB* gene sequences of 591 bp from strains isolated from the Tennis Court soil sample within the green potential-clone group. The homology matrix was generated using DNAMAN.

The *gyrB* gene sequences of eight strains in the green potential-clone group from the TC soil sample were analysed (Figure 3.6 and Figure 3.9). *Streptomyces* strain 47TA3 had a *gyrB* gene sequence distinct from the other *gyrB* gene sequences and is therefore not a clone (Figure 3.9). *Streptomyces* strains 5TK1, 9TK1, 15TA1, 16TA1 and 36TA2 have identical *gyrB* gene sequences to each other and are, therefore, presumed to be clones (Figure 3.9). Cultivation of *Streptomyces* strains 5TK1, 9TK1, 15TA1, 16TA1 and 36TA2 on ISP4 showed that all of them had a pale yellow substrate mycelium and a white and grey spore mass, except strain 16TA1, which did not produce spores on ISP4 (Table A3). This evidence lends support to the suggestion that strains 5TK1, 9TK1, 15TA1, 16TA1 and 36TA2 are clones. *Streptomyces* strains 6TK1 and 19TA1 have identical *gyrB* gene sequences to each other and are, therefore, presumed to be clones (Figure 3.9). Growth of *Streptomyces* strains 6TK1 and 19TA1 on ISP4 showed that both strains had a pale yellow substrate mycelium and a white spore mass (Table A3). This evidence supports the conclusion that strains 6TK1 and 19TA1 are clones.

In the maximum-likelihood phylogenetic tree based on *gyrB* gene sequences (Figure 3.6), all of the strains from the green potential-clone group clustered closely together. There was very strong bootstrap support for this association (94%; Figure 3.6), but this grouping was not seen in the NJ and MP trees (data not shown). This tree also showed that *Streptomyces* strains 6TK1 and 19TA1 form a sub-clade within the green potential-clone group. There was moderate bootstrap support for this association (73%; Figure 3.6), which was also seen in the NJ and MP trees (data not shown).

Streptomyces strains 5TK1, 9TK1, 15TA1, 16TA1 and 36TA2 clustered separately from strains 6TK1 and 19TA1, but there was weak bootstrap support for this association (<70%; Figure 3.6) and the association was not supported by the NJ and MP trees (data not shown).

In summary, based on the available evidence, it is proposed that strains 5TK1, 9TK1, 15TA1, 16TA1 and 36TA2 are clones. Strains 6TK1 and 19TA1 are also proposed to be clones. The *gyrB* gene sequence of strain 47TA3 was distinct from those of the other strains in the green potential-clone group, so strain 47TA3 is not a clone (but it is closely related to the other seven strains).

Streptomyces strains 18TA1, 20TA1, 33TA2 and 45TA3 are the only members of the green potential-clone group for which *gyrB* gene sequences were not determined. Growth of *Streptomyces* strains 33TA2 and 47TA3 on ISP4 showed that both strains had a yellow substrate mycelium (Table A3), suggesting that they are clones. Growth of *Streptomyces* strains 18TA1, 20TA1 and 45TA3 on ISP4 showed that all of them had pale yellow substrate mycelium with a white and grey spore mass (Table A3), suggesting that *Streptomyces* strains 18TA1, 20TA1 and 45TA3 may be clones of *Streptomyces* strain 5TK1, 9TK1, 15TA1, 16TA1 and 36TA2.

The designation of strains as clones of each other is tentative, as limited testing was done. Whole-genome sequencing followed by ANI and dDDH analyses would be required to determine conclusively which strains are clones of each other and which are not clones. Nevertheless, the preliminary assessment of which strains are clones has provided sufficient information to guide future work on these strains. It is recommended that only one strain from each of the strain groups suggested to be clones be studied further (i.e. one of strains 2KK1 and 10KK1, one of strains 28KA1, 32KA1 and 54KA3, one of strains 39KA2, 41KA2 and 42KA2, one of strains 5TK1, 9TK1, 15TA1, 16TA1 and 36TA2 and one of strains 6TK1 and 19TA1).

3.4 References

Altschul, S. F. *et al*. (1997) 'Gapped BLAST and PSI-BLAST: a new generation of protein database search programs', Nucleic Acids Research, 25(17), pp. 3389–3402. doi: 10.1093/NAR/25.17.3389.

Barka, E. A. *et al*. (2016) 'Correction for Barka *et al*., Taxonomy, Physiology, and Natural Products of Actinobacteria', Microbiology and Molecular Biology Reviews, 80(4). doi: 10.1128/mmbr.00044-16.

Clarridge, J. E. (2004) 'Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases', Clinical Microbiology Reviews, pp. 840–862. doi: 10.1128/CMR.17.4.840-862.2004.

Connon, S. A. and Giovannoni, S. J. (2002) 'High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates', Applied and Environmental Microbiology, 68(8), pp. 3878–3885. doi: 10.1128/AEM.68.8.3878-3885.2002.

Cook, A. E. and Meyers, P. R. (2003) 'Rapid identification of filamentous actinomycetes to the genus level using genus-specific 16S rRNA gene restriction fragment patterns', International Journal of Systematic and Evolutionary Microbiology, 53(6), pp. 1907–1915. doi: 10.1099/ijs.0.02680-0.

Curtis, S. M. (2015). 'Enhanced phylogenetic analysis and targeted search for the genus *Kribbella*'. Ph.D. thesis. Department of Molecular and Cell Biology. University of Cape Town. Supervisor Paul Meyers. [online] Available at: https://open.uct.ac.za/bitstream/item/16849/thesis_sci_2015_curtis_sarah_maureen.pdf?sequence=1 [Accessed 12 Nov. 2019].

Curtis, S. M. *et al*. (2020) 'Development of a *Kribbella*-specific isolation medium and description of *Kribbella capetownensis* sp. nov. and *Kribbella speibonae* sp. nov., isolated from soil', Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology, 113(5), pp. 617–628. doi: 10.1007/s10482-019-01365-6.

Daniel, R. (2004) 'The soil metagenome - A rich resource for the discovery of novel natural products', Current Opinion in Biotechnology. Elsevier Current Trends, pp. 199–204. doi: 10.1016/j.copbio.2004.04.005.

Doolittle, W. F. (1999) ‘Phylogenetic classification and the universal tree’, Science. Science, pp. 2124–2128. doi: 10.1126/science.284.5423.2124.

Edgar, R. C. (2004) ‘MUSCLE: Multiple sequence alignment with high accuracy and high throughput’, Nucleic Acids Research, 32(5), pp. 1792–1797. doi: 10.1093/nar/gkh340.

Felsenstein, J. (1985) ‘Confidence Limits on Phylogenies: An Approach Using the Bootstrap’, Evolution, 39(4), p. 783. doi: 10.2307/2408678.

Goodfellow, M. (2010) ‘Selective Isolation of Actinobacteria’, in Manual of Industrial Microbiology and Biotechnology. John Wiley & Sons, Ltd, pp. 13–27. doi: 10.1128/9781555816827.ch2.

Kämpfer, P. and Glaeser, S. P. (2012) ‘Prokaryotic taxonomy in the sequencing era - the polyphasic approach revisited’, Environmental Microbiology, 14(2), pp. 291–317. doi: 10.1111/j.1462-2920.2011.02615.x.

Kim, O. S. *et al*. (2012) ‘Introducing EzTaxon-e: A prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species’, International Journal of Systematic and Evolutionary Microbiology, 62(PART 3), pp. 716–721. doi: 10.1099/ijs.0.038075-0.

Kimura, M. (1980) ‘A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences’, Journal of Molecular Evolution, 16(2), pp. 111–120. doi: 10.1007/BF01731581.

Kirby, B. M. *et al*.(2010) ‘Phylogenetic analysis of the genus *Kribbella* based on the *gyrB* gene: Proposal of a *gyrB*-sequence threshold for species delineation in the genus *Kribbella*’, Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology, 97(2), pp. 131–142. doi: 10.1007/s10482-009-9393-9.

Kumar, S. *et al*. (2016) ‘MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets’, Molecular biology and evolution, 33(7), pp. 1870–1874. doi: 10.1093/molbev/msw054.

Li, J. W. H. and Vederas, J. C. (2009) ‘Drug discovery and natural products: End of an era or an endless frontier?’, Science, pp. 161–165. doi: 10.1126/science.1168243.

Ludwig, W. and Klenk, H.-P. (2001) ‘Overview: A Phylogenetic Backbone and Taxonomic Framework for Procaryotic Systematics’, in Bergey’s Manual® of Systematic Bacteriology, pp. 49–65. doi: 10.1007/978-0-387-21609-6_8.

Manfio, G. P. *et al*. (2003) ‘Classification of novel soil streptomycetes as *Streptomyces aureus* sp. nov., *Streptomyces laceyi* sp. nov. and *Streptomyces sanglieri* sp. nov’, Antonie van Leeuwenhoek International Journal of General and Molecular Microbiology, 83(3), pp. 245–255. doi: 10.1023/A:1023332427794.

Meier-Kolthoff, J. P. *et al*. (2013) ‘When should a DDH experiment be mandatory in microbial taxonomy?’, Archives of Microbiology, 195(6), pp. 413–418. doi: 10.1007/s00203-013-0888-4.

Meyers, P. R. (2014) ‘Gyrase subunit B amino acid signatures for the actinobacterial family *Streptosporangiaceae*’, Systematic and Applied Microbiology, 37(4), pp. 252–260. doi: 10.1016/j.syapm.2013.12.002.

Meyers, P. R. (2015) ‘Analysis of recombinase A (*recA*/RecA) in the actinobacterial family *Streptosporangiaceae* and identification of molecular signatures’, Systematic and Applied Microbiology, 38(8), pp. 567–577. doi: 10.1016/j.syapm.2015.10.001.

le Roes, M. *et al*. (2008) ‘*Gordonia lacunae* sp. nov., isolated from an estuary’, Systematic and Applied Microbiology, 31(1), pp. 17–23. doi: 10.1016/j.syapm.2007.10.001.

Saitou, N. and Nei, M. (1984) ‘The Neighbour-joining Method: A New Method for Reconstructing Phylogenetic Trees’, Molecular Biology and Evolution, 97(4), pp. 406–425. Available at: http://igem.temple.edu/labs/nei/downloads/publications/1987 Publications/1987-saitou-nei.pdf.

Shirling, E. and Gottlieb, D. (1966) ‘Methods for characterization of *Streptomyces* species’, International journal of Systematic Bacteriology, 16(1946), p. 340.

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’, International Journal of Systematic Bacteriology. Microbiology Society, pp. 846–849. doi: 10.1099/00207713-44-4-846.

Tan, G. Y. A. *et al*. (2006) ‘Exploration of *Amycolatopsis* diversity in soil using genus-specific primers and novel selective media’, Systematic and Applied Microbiology, 29(7), pp. 557–569. doi: 10.1016/j.syapm.2006.01.007.

Tindall, B. J. *et al*. (2010) ‘Notes on the characterization of prokaryote strains for taxonomic purposes’, International Journal of Systematic and Evolutionary Microbiology, 60(1), pp. 249–266. doi: 10.1099/ijs.0.016949-0.

Wang, Y. *et al*. (1996) ‘A proposal to transfer *Microbispora bispora* (Lechevalier 1965) to a new genus, *Thermobispora* gen. nov., as *Thermobispora bispora* comb. nov.’, International Journal of Systematic Bacteriology, 46(4), pp. 933–938. doi: 10.1099/00207713-46-4-933.

Woese, C. R. *et al*. (1990) ‘Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eukarya’, Proceedings of the National Academy of Sciences of the United States of America, 87(12), pp. 4576–4579. doi: 10.1073/pnas.87.12.4576.

Yoon, S. H. *et al*. (2017) ‘Introducing EzBioCloud: A taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies’, International Journal of Systematic and Evolutionary Microbiology, 67(5), pp. 1613–1617. doi: 10.1099/ijsem.0.001755.

Chapter 4:

Screening of actinobacterial isolates for antimycobacterial activity

Summary

Forty (40) isolates from soil samples collected at UCT (Chapter 3) were screened for antimycobacterial activity on four different media: two presumptive actinobacterial isolates and 38 identified actinobacterial isolates. Isolates that showed strong or very strong antimycobacterial activity against *Mycobacterium aurum* strain A+ were further investigated for their antibiotic spectrum through testing for activity against *Escherichia coli* strain ATCC 25922 and *Staphylococcus aureus* strain ATCC 25923. A standard overlay method was used for all antibiotic screening. Over 50% of the isolates from both soil samples produced strong or very strong antimycobacterial activity, potentially making them good sources of novel antibiotics. A total of 35 strain-medium combinations (from 25 different isolates) showed strong or very strong antimycobacterial activity. None of the active actinobacterial strains showed activity against *E. coli* ATCC 25922, with just two strains (*Actinomadura* strain 1KK1 on R2A medium and *Streptomyces* strain 49TA3 on 7H9 and R2A media) showing strong or very strong activity against *S. aureus* ATCC 25923. Crude extracts of the most active actinobacterial strains were subjected to Global Natural Products Social (GNPS) molecular networking analyses to assess if the compounds produced are known compounds. Crude extracts of 26 of the 35 strain-medium combinations tested returned no matches on the GNPS database, with nine combinations having matches to know compounds. These no-match combinations could potentially be novel antibiotics and should be tested for activity against *Mycobacterium tuberculosis.*

Table of Contents

4.1 Introduction 91
4.2. Materials and methods 93
4.2.1 Stab inoculation 93
4.2.2 Agar overlay method 93
4.2.2.1 *Mycobacterium aurum* A+ 93
4.2.2.2 *S. aureus* ATCC 25923 and *E. coli* ATCC 25922 94
4.2.3 Determination of the zone of inhibition (ZOI) 94
4.2.4 Compound extraction 95
4.2.5 Spectroscopic analysis 96
4.2.6 GNPS molecular networking 96
4.3 Results and discussion 97
4.3.1 Screening for antimycobacterial activity 97
4.3.1.1 Antimycobacterial activity against *M. aurum* A+ 97
4.3.2 Characterization of antimycobacterial activity 100
4.3.2.1 Antibiotic activity against *S. aureus* and *E. coli* 100
4.3.2.2 GNPS molecular networking analysis 103
4.4 References 105

4.1 Introduction

Antibiotics started being used in the 1940s, which made the treatment of most types of infectious diseases possible (Martínková *et al.*, 2009). Recently, there has been a dramatic decrease in the discovery of novel antibiotics through the isolation and screening of actinobacteria, despite the growing incidence of antimicrobial-resistant pathogens (Niu, 2018). Antibiotic resistance is a major healthcare problem where microbial pathogens gain resistance to the antibiotic classes specifically recommended for their treatment. It can lead to the development of “superbugs” that have enhanced morbidity and mortality due to their resistance to antibiotics (Davies and Davies, 2010).

Tuberculosis (TB), caused by the bacillus *M. tuberculosis*, is a disease where “superbugs” have already developed. There were approximately 1.5 million deaths due to TB in 2020, showing that it remains a major healthcare problem worldwide (World Health Organization, 2021). Drug-susceptible TB can be treated with considerable success with a 6 month program of antibiotics consisting of a cocktail of the four first-line anti-TB drugs (isoniazid, rifampicin, ethambutol and pyrazinamide) (World Health Organization, 2021). However, antibiotic resistant *M. tuberculosis* strains have emerged: multidrug-resistant (MDR), extensively drug-resistant (XDR) and totally drug-resistant (TDR). MDR *M. tuberculosis* strains are resistant to isoniazid and rifampicin (World Health Organization, 2021), XDR *M. tuberculosis* strains are resistant to isoniazid and rifampicin, any fluoroquinolone and one or more of the injectable drugs (Sotgiu *et al.*, 2009) and TDR *M. tuberculosis* strains are resistant to isoniazid, rifampicin and all second-line drug classes (Velayati *et al.*, 2009). If left unchecked, spread of drug-resistant strains of *M. tuberculosis* will leave us unable to effectively treat the disease. The discovery of new antibiotic compounds with new mechanisms of action is one way to combat these antibiotic-resistant *M. tuberculosis* strains (Niu, 2018).

Filamentous actinobacteria have an extensive secondary metabolism (van Bergeijk *et al.*, 2020) and are renowned as the principal source of therapeutic pharmaceuticals (Miao and Davies, 2010). They have a unique ability to produce antibiotics, which are used for treating a variety of infections (Mahajan and Balachandran, 2012). Despite there being a decrease in the rate of discovery of novel antibiotics from actinobacteria, both *Streptomyces* species and non-*Streptomyces* species continue to be a good source of new secondary metabolites (Connon and Giovannoni, 2002; Li and Vederas, 2009).

This chapter presents the screening of the actinobacterial strains, isolated from the soil samples (described in Chapter 3), for antimycobacterial activity. The isolates were screened against *Mycobacterium aurum* A+ to test their antimycobacterial activity. *M. aurum* A+ was chosen as a test bacterium due to it being non-pathogenic and having a similar antibiotic susceptibility profile to that of *M. tuberculosis* (Chung *et al.*, 1995). Strains showing antibiotic activity against *M. aurum* A+, therefore, have the potential to also inhibit the growth of *M. tuberculosis* and the antibiotics they produce may be able to treat TB.

Additional screening was done for strains showing strong or very strong activity against *M. aurum* A+. These highly active strains were also screened for antibiotic activity against *Staphylococcus aureus* strain ATCC 25923 and *Escherichia coli* strain ATCC 25922 to determine the spectrum of antibiotic activity of the actinobacterial isolates. These highly active actinobacterial strains also underwent GNPS molecular networking analysis of the compounds that they produced to investigate if the antibiotic activity was due to previously discovered or unknown compounds.

4.2. Materials and methods

4.2.1 Stab inoculation

Thirty-eight (38) actinobacterial strains and two presumptive actinobacterial strains were investigated for production of antimycobacterial compounds (40 of the 42 strains from the soil isolations; Chapter 3).

Four media were inoculated with the 40 strains by taking a small amount of culture from a clean sub-culture plate and stabbing it into the test agar plate using a sterile toothpick. The four media chosen were: International *Streptomyces* Project medium No. 2 (ISP2) (Shirling and Gottlieb, 1966), Czapek solution agar (CZ) (Krzyśko-Lupicka *et al.*, 1997), Middlebrook 7H9 Broth (7H9) (Becton Dickinson, USA) and R2A medium (https://www.dsmz.de/microorganisms/medium/pdf/DSMZ_Medium830.pdf).

In the preliminary antibiotic screening, four strains were stabbed in the quadrants of a single plate. If a strain showed a zone of inhibition (ZOI) against the test bacterium, a confirmatory activity test was done in which that strain was stab inoculated at the centre of a plate on its own. The stab-inoculated agar plates were sealed in plastic bags and incubated for 10 days at 30°C.

4.2.2 Agar overlay method

4.2.2.1 *Mycobacterium aurum* A+

M. aurum A+ cultures used for overlays were inoculated 18-24 hours before the overlay assays were performed. Four or five loopfuls of *M. aurum* A+ cells from a pure culture on a 2YT plate (Miller, 1972) were inoculated into a sterile universal container containing 5 ml of 2YT liquid medium. The culture was vortexed and then incubated at 37°C with shaking overnight (16-18 hours). Each overnight test bacterial culture was inspected for contamination by Gram staining and streaking for single colonies on 2YT agar plates (1.5% agar). The test bacterial cell concentration for each round of overlays was kept constant by measuring the optical density (OD) of the *M. aurum* A+ cultures at 600 nm using a Beckman DU®-64 spectrophotometer. The appropriate volume of culture was added to the sloppy agar using the following empirical formula: 160 = volume (µl) x OD_{600}.

Sloppy 2YT medium, with 0.7% agar, was dispensed in 6-ml volumes into test tubes, autoclaved and then kept at 60°C. The 2YT sloppy agar test tubes were inoculated

with *M. aurum* A+ and vortexed to ensure an even distribution of the *M. aurum* A+ cells in the liquid sloppy agar medium. The *M. aurum* A+ inoculated sloppy medium was then poured gently over the stab inoculated test plates making sure to not disturb the spore mass of the stabbed-inoculated actinobacterial strains. Overlays were allowed to set and then incubated for 48 hours at 37°C .

4.2.2.2 *S. aureus* ATCC 25923 and *E. coli* ATCC 25922

S. aureus ATCC 25923 and *E. coli* ATCC 25922 cultures used for overlays were inoculated 18-24 hours before the assays were performed. One or two loopfuls of *S. aureus* ATCC 25923 or *E. coli* ATCC 25922 cells, from pure cultures on LB plates, were inoculated into a sterile universal containing 5 ml of Luria-Bertani (LB)(Miller, 1972) liquid medium. The culture was vortexed and then incubated at 37°C overnight (16-18 hours) with shaking. Each overnight test bacterial culture was inspected for contamination by Gram staining and streaking for single colonies on LB agar plates (1.5% agar). The test bacterial cell concentration for each round of overlays was kept constant by measuring the OD of the *S. aureus* ATCC 25923 and *E. coli* ATCC 25922 cultures at 600 nm using a Beckman DU®-64 spectrophotometer. The appropriate volume of culture was added to the sloppy agar using the following empirical formula: $4 = \text{volume } (\mu l) \times OD_{600}$.

Sloppy LB medium, with 0.7% agar, was dispensed in 6-ml volumes into test tubes, autoclaved and then kept at 60°C. The LB sloppy agar test tubes were inoculated with *S. aureus* ATCC 25923 or *E. coli* ATCC 25922 and vortexed to ensure an even distribution of the test bacterial cells in the liquid medium. The test bacterium inoculated sloppy medium was then poured gently over the stab inoculated test plates, making sure to not disturb the spores of the stab-inoculated actinobacterial strains. Overlays were allowed to set and then incubated for 18-24 hours at 37°C.

4.2.3 Determination of the zone of inhibition (ZOI)

After 18-24 hours (*S. aureus* and *E. coli*) or 48 hours of incubation (*M. aurum*), the overlaid plates were inspected and the ZOIs were measured. A positive result for activity of actinobacterial strains against the test bacteria was seen as a clear zone of growth inhibition of the test bacterium on the overlaid plate around an actinobacterial colony. A negative result for activity against the test bacterium being a lawn of test bacterial growth over the entire overlaid plate.

The colony diameter and ZOI diameter of actinobacterial strains showing activity against a test bacterium were measured. The area of the ZOI was calculated using the formula for the area of a circle: the area of the colony was subtracted from the area of the ZOI, giving just the area of bacterial growth inhibition surrounding the actinobacterial colony (the ZOI) in mm^2. ZOIs were classified according to an arbitrary assignment of antibiotic activity: very weak (<100 mm^2), weak (101-1000 mm^2), moderate (1001-2000 mm^2), strong (2001-3000 mm^2) and very strong (>3000 mm^2).

In this study, only actinobacterial strains that showed strong or very strong ZOIs (i.e. ≥2001 mm^2) against *M. aurum* A+ were tested against *S. aureus* ATCC 25923 and *E. coli* ATCC 25922. Furthermore, these actinobacterial strains were only tested against *S. aureus* ATCC 25923 and *E. coli* ATCC 25922 on the medium or media on which the strong or very strong ZOI against *M. aurum* A+ was recorded.

4.2.4 Compound extraction

Extraction of compounds was only carried out for strain-medium combinations that showed strong or very strong ZOIs against *M. aurum* A+. Compound extraction and mass spectrometric analyses were performed by Dr Daniel Watson, Division of Clinical Pharmacology, Faculty of Health Sciences, UCT.

In each case, the entire ZOI was excised by the author, using a spatula, as 1 cm x 1cm pieces of agar and placed into a plastic sterile 50 ml centrifuge tube. The centrifuge tube was stored at -20°C until the day of compound extraction. The pieces of agar were thawed at room temperature by Dr Watson and added to a mixture containing 5 ml methanol and 35 ml ethyl acetate (a mid-polarity solvent). The mixture was agitated on a multipurpose shaker at 120 rpm for one hour at room temperature. Liquid-liquid extraction was performed on the resultant liquid samples.

Each sample was poured into a separating funnel, 40 ml of water was added to the funnel and the contents were shaken to mix the aqueous and organic phases. The two phases were allowed to separate for 30 minutes at room temperature and then the bottom (aqueous) layer was discarded and the upper ethyl acetate (organic) layer was retained. The organic phase was then left overnight in a fume hood to dry through evaporation of the ethyl acetate.

Once completely dry, the extracted samples were redissolved in 1.5 ml of ethyl acetate. The samples were then transferred to 1.5 ml weighed, microcentrifuge tubes and the samples were then dried again using a sample concentrator. The microcentrifuge tubes were weighed and the mass of the sample was determined by subtracting the mass of the microcentrifuge tube from the mass of the tube plus the dry sample. Ethyl acetate was added to each dry sample to obtain a concentration of 10 mg/ml. The stock samples were then diluted 100-fold in methanol to a volume of 100 µl (100 ng/ml) for high pressure liquid chromatography-mass spectrometry (LC-MS).

4.2.5 Spectroscopic analysis

The spectroscopic analyses were carried out by Dr Daniel Watson, Division of Clinical Pharmacology, Faculty of Health Sciences, UCT.

The spectroscopic method was performed by Dr Watson as described in Watson *et al*. (2021) with the following changes: (1) for HPLC, the organic mobile phase was methanol with 0.1% formic acid, not methanol with 0.5% formic acid; (2) the HPLC column was run for 25 minutes with a hold of 2 minutes at the end rather than a run time of 18 minutes with no hold time and (3) the mass spectrometer was set to detect a mass range of 50-1500 Da, not 50-1300 Da.

4.2.6 GNPS molecular networking

The GNPS molecular networking method used in this study was performed by Dr Daniel Watson, Division of Clinical Pharmacology, Faculty of Health Sciences, UCT.

The method was as described in Watson *et al*. (2021) with the following changes: (1) when the network was created, edges were filtered to have more than five matched peaks, not six matched peaks; and (2) all matches kept between network spectra and library spectra were required to have at least five matched peaks, not six matched peaks.

4.3 Results and discussion

4.3.1 Screening for antimycobacterial activity

Actinobacterial strains isolated from the K and TC soil samples were screened for antimycobacterial activity against *M. aurum* A+ on four different media: ISP2, R2A, 7H9 and CZ. Forty (40) isolates were tested using the agar overlay method: 38 identified actinobacterial isolates (described in Chapter 3) and two presumptive actinobacterial isolates (strains 53KA3 and 57KA3). The agar overlay method was chosen as it allows for rapid screening against test bacteria in a semi-quantitative manner (Hockett and Baltrus, 2017). *Streptomyces* strains 12TA1 and 48TA3 were not tested for antimycobacterial activity, as they failed to grow from frozen stock cultures and could not be sub-cultured from old agar plates.

The antimycobacterial activity of the 40 isolates tested is summarised in Table 4.1. Activity against *M. aurum* A+ indicates that the actinobacterial strain may also have activity against *M. tuberculosis*, as *M. aurum* A+ has been shown to have a similar antibiotic susceptibility profile to that of *M. tuberculosis* (Chung *et al.*, 1995).

4.3.1.1 Antimycobacterial activity against *M. aurum* A+

The 40 isolates tested showed a wide range of antimycobacterial activity ranging from very weak to very strong based on the area of the ZOI. The antimycobacterial screening showed that: nine strains showed no, very weak or weak antimycobacterial activity (ZOI<1000 mm^2), six strains showed moderate antimycobacterial activity (1001 mm^2 < ZOI < 2000 mm^2), 12 strains showed strong antimycobacterial activity (2001 mm^2 < ZOI < 3000 mm^2) and 13 strains showed very strong antimycobacterial activity (ZOI>3000 mm^2) (Table 4.1).

The activity against *M. aurum* A+ differed across the four media tested (Table 4.1). This is not surprising as the four media have different nutrient compositions, providing the strains with different nutrient conditions during growth. Differing conditions have been shown to affect the antibiotic biosynthesis of microorganisms (Zhu *et al.*, 2014). On 7H9 (a defined medium), only *Streptomyces* strains 49TA3 and 61KA3 had strong antimycobacterial activity and no strains showed very strong antimycobacterial activity. On CZ (a defined medium), five strains showed strong antimycobacterial activity and no strains showed very strong antimycobacterial activity. On R2A medium,

six strains showed strong antimycobacterial activity and six strains showed very strong antimycobacterial activity. On ISP2 medium (a rich medium), 10 strains showed very strong antimycobacterial activity and six strains showed strong activity (Table 4.1). This result shows R2A and ISP2 media to be the most successful in eliciting antimycobacterial activity from the strains tested (Table 4.1). R2A and ISP2 are not defined media and their success is an unexpected result because antibiotic synthesis is usually activated under stressful conditions for bacteria, such as in the lower nutrient conditions of the CZ medium (Rigali *et al.*, 2008; Zhu *et al.,* 2014). One possible explanation for the strong antimycobacterial activity on the nutrient rich media (R2A and ISP2 media) is the high growth rate achieved when the strains were grown. This high growth rate would mean more cells present to produce antibiotics and, therefore, a stronger antibiotic response.

Of all the strains tested, nine strains showed strong or very strong antimycobacterial activity on more than one medium: *Streptomyces* strains 32KA1, 39KA2, 40KA2, 41KA2, 49TA3, 55KA3, 59KA3 and 61KA3 and *Actinomadura* 1KK1 (Table 4.1). *Streptomyces* strains 40KA2 and 41KA2 and *Actinomadura* 1KK1 were the only strains to show very strong antimycobacterial activity on two different media (ISP2 and R2A media) (Table 4.1). *Actinomadura* strain 1KK1 had arguably the best antimycobacterial activity of all the strains tested with the strain having very strong antimycobacterial activity (ZOIs > 6000 mm^2) on ISP2 and R2A media (Table 4.1). *Streptomyces* strain 49TA3 was the only strain to show strong or very strong antimycobacterial activity on three different media (7H9, CZ and R2A) and one of only two strains to show strong antimycobacterial activity on 7H9 medium (Table 4.1).

The TC sample had 53% (8 of 15 strains) of its strains showing strong or very strong antimycobacterial activity on one or more of the four media (Table 4.1) The K sample had 68% (17 of 25 strains) of its strains showing strong or very strong antimycobacterial activity on one or more of the four media (Table 4.1). This result indicates that the K sample had a slightly higher percentage of actinobacterial strains that produced significant antimycobacterial activity. Both soil samples however have been shown to be good sources of isolates with significant antimycobacterial activity.

It must be noted that the absence of antimycobacterial activity for a tested actinobacterial strain does not mean that that strain did not produce any bioactive

compounds on the given medium. It is possible that bioactive compounds were produced, but they were not active against *M. aurum* A+.

Table 4.1: Overlay results of actinobacterial isolates tested against *Mycobacterium aurum* A+ on four different media. Very strong activity, red shading (>3000 mm^2); strong activity, orange shading (2001-3000 mm^2); moderate activity, green shading (1001-2000 mm^2). Very weak activity (<100 mm^2) and weak activity (101-1000 mm^2) are not shaded.

Medium	**7H9**	**CZ**	**ISP2**	**R2A**
Isolate	ZOI area (mm^2)	ZOI area (mm^2)	ZOI area (mm^2)	ZOI area (mm^2)
Kramer sample				
Actinomadura strain 1KK1	1373	1237	6062	6054
Streptomyces strain 2KK1	0	587	1947	1335
Streptomyces strain 8KK1	0	0	0	0
Streptomyces strain 10KK1	0	1178	2714	1056
Streptomyces strain 22KA1	0	521	2529	1885
Streptomyces strain 23KA1	0	0	1103	691
Streptomyces strain 25KA1	75	0	2842	78
Streptomyces strain 26KA1	0	880	1178	0
Streptomyces strain 28KA1	0	160	1425	4873
Streptomyces strain 31KA1	66	1021	5881	1307
Streptomyces strain 32KA1	0	2104	2714	1583
Streptomyces strain 39KA2	63	307	5828	2488
Streptomyces strain 40KA2	0	939	5768	3695
Streptomyces strain 41KA2	0	0	3594	3478
Streptomyces strain 42KA2	0	1508	5466	511
Streptomyces strain 52KA3	0	0	0	0
Strain 53KA3	0	483	0	0
Streptomyces strain 54KA3	954	1206	1947	2011
Streptomyces strain 55KA3	829	0	4335	2714
Strain 57KA3	0	0	4913	0
Streptomyces strain 58KA3	0	0	0	75
Streptomyces strain 59KA3	1178	330	2749	2309
Streptomyces strain 60KA3	0	0	0	0
Streptomyces strain 61KA3	2714	939	120	2309
Streptomyces strain 63KA3	0	0	2529	0
Tennis Court sample				
Streptomyces strain 5TK1	0	2060	1374	1731
Streptomyces strain 6TK1	0	1228	4825	1696
Nonomuraea strain 7TK1	0	0	0	0
Streptomyces strain 9TK1	0	352	844	1583
Streptomyces strain 13TA1	0	0	0	0
Streptomyces strain 15TA1	0	687	1070	1598
Streptomyces strain 16TA1	352	587	864	552
Streptomyces strain 18TA1	0	869	776	1206
Streptomyces strain 19TA1	0	2815	1144	1731
Streptomyces strain 20TA1	0	2095	1178	1442
Streptomyces strain 33TA2	0	0	0	0
Streptomyces strain 36TA2	188	990	968	2529
Streptomyces strain 45TA3	1935	242	5928	433
Streptomyces strain 47TA3	449	687	0	4423
Streptomyces strain 49TA3	2368	2384	67	5881

4.3.2 Characterization of antimycobacterial activity

The strains that showed strong or very strong activity against *M. aurum* A+ could be producing potential antibiotics that have activity against *M. tuberculosis*. Therefore, their antibiotic spectrum was investigated through testing for activity against *S. aureus* ATCC 25923 (a Gram-positive bacterium) and *E. coli* ATCC 25922 (a Gram-negative bacterium). These tests were done using the agar overlay method on the medium or media on which the strains showed antimycobacterial activity. The results of the antibiotic overlays against *S. aureus* ATCC 25923 and *E. coli* ATCC 25922 are summarised in Table 4.2 along with the activity of that strain-medium combination against *M. aurum* A+ (data from Table 4.1).

Crude extracts of the most active actinobacterial strains (strong and very strong antimycobacterial activity) were prepared, followed by spectroscopic analysis (LC-MS). The mass spectra were subjected to Global Natural Products Social (GNPS) molecular networking analyses to assess if the compounds produced are known compounds. This rapid identification of strains producing known compounds reduces future work by allowing one to focus solely on "gifted" strains producing unknown compounds.

4.3.2.1 Antibiotic activity against *S. aureus* and *E. coli*

None of the active actinobacterial strains showed activity against *E. coli* ATCC 25922 (Table 4.2). This result indicates that the antibiotic activity of the strains tested is specific to Gram-positive bacteria. This result is not surprising given that Gram-negative bacteria, such as *E. coli,* are less susceptible to antibiotics because antibiotics have to cross both the outer membrane and the cytoplasmic membrane to reach their targets in the cell (Prescott *et al.,* 2002).

The activity against *S. aureus* ATCC 25923 was as follows: 26 strain-medium combinations showed no or weak activity (ZOI<1000 mm^2), six strain-medium combinations showed moderate activity (1001 mm^2 < ZOI < 2000 mm^2), *Actinomadura* strain 1KK1 on R2A medium showed strong activity (2001 mm^2<ZOI<3000 mm^2) and *Streptomyces* strain 49TA3 on 7H9 and R2A media was the only strain to show very strong activity (ZOI>3000 mm^2) (Table 4.2).

Actinomadura strain 1KK1 and *Streptomyces* strains 6TK1, 36TA3, 45TA3, 47TA3, 49TA3 and 55KA3 showed moderate, strong or very strong activity (ZOI>1001 mm^2) against *S. aureus* ATCC 25923 and strong or very strong activity (ZOI>2001 mm^2) against *M. aurum* A+ on one or more media (Table 4.2). This could mean that these strains are producing one compound that has antibiotic activity against both *S. aureus* ATCC 25923 and *M. aurum* A+. Alternatively, the strains are producing two or more compounds that are effective against different test bacteria. In order to determine if the antibiotic activity is produced by one compound or multiple compounds, bioautography could be used (i.e. separate the molecules in a crude extract by thin layer chromatography followed by direct application of the test bacterium to the developed chromatogram, incubation and detection of live bacterial cells using a tetrazolium salt solution, such as MTT [3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide]).

Strains 57KA3 and 63KA3 (on ISP2 medium) and *Streptomyces* strain 32KA1 (on CZ) were the only strain-medium combinations where there was strong or very strong activity (ZOI>2001 mm^2) against *M. aurum* A+ and no, or very weak (ZOI<100 mm^2), activity against *S. aureus* ATCC 25923 (Table 4.2). There were also 23 strain-medium combinations where strong or very strong activity (ZOI>2001 mm^2) against *M. aurum* A+ was coupled with weak activity (101 mm^2<ZOI<1000 mm^2) against *S. aureus* ATCC 25923 (Table 4.2). These 26 strain medium combinations show significant activity against *M. aurum* A+ and little (or no) activity against the other test bacteria. These strains could potentially be producing compounds with a narrow-spectrum of activity specific to the genus *Mycobacterium*. The use of narrow-spectrum antibiotics has advantages over the use of broad-spectrum antibiotics: narrow-spectrum antibiotics have a lower selection for resistance than broad-spectrum antibiotics and narrow-spectrum antibiotics do not alter the makeup of the gut microbiota as significantly as broad-spectrum antibiotics do (Melander *et al.*, 2018).

Table 4.2: Overlay results against *S. aureus* ATCC 25923 *and E. coli* ATCC 25922 compared to previously shown *M. aurum* A+ overlay results. The corresponding compound analysis result for each strain-medium combination is also shown. Very strong activity, red shading (>3000 mm^2); strong activity, orange shading (2001-3000 mm^2); moderate activity, green shading (1001-2000 mm^2). Very weak activity (<100 mm^2) and weak activity (101-1000 mm^2) are not shaded."*" indicates a weak compound match. $ Data determined by Dr Daniel Watson.

Isolate	Medium	*M. aurum* A+ ZOI (mm^2)	*S. aureus* ATCC 25923 ZOI (mm^2)	*E. coli* ATCC 25922 ZOI (mm^2)	Compound match$
Kramer sample					
Actinomadura strain 1KK1	ISP2	6062	1781	0	No match
Actinomadura strain 1KK1	R2A	6054	2622	0	No match
Streptomyces strain 10KK1	ISP2	2714	137	0	Puromycin
Streptomyces strain 22KA1	ISP2	2529	236	0	Antimycin
Streptomyces strain 25KA1	ISP2	2842	267	0	No match
Streptomyces strain 28KA1	R2A	4873	389	0	Lambdamycin*
Streptomyces strain 31KA1	ISP2	5881	829	0	No match
Streptomyces strain 32KA1	CZ	2104	0	0	No match
Streptomyces strain 32KA1	ISP2	2714	418	0	No match
Streptomyces strain 39KA2	ISP2	5828	176	0	No match
Streptomyces strain 39KA2	R2A	2488	507	0	No match
Streptomyces strain 40KA2	ISP2	5768	530	0	No match
Streptomyces strain 40KA2	R2A	3695	574	0	No match
Streptomyces strain 41KA2	ISP2	3594	160	0	Puromycin
Streptomyces strain 41KA2	R2A	3478	795	0	Puromycin
Streptomyces strain 42KA2	ISP2	5466	226	0	No match
Streptomyces strain 54KA3	R2A	2011	330	0	No match
Streptomyces strain 55KA3	ISP2	4335	339	0	No match
Streptomyces strain 55KA3	R2A	2714	1868	0	No match
Strain 57KA3	ISP2	4913	0	0	No match
Streptomyces strain 59KA3	R2A	2309	330	0	Mitomycin C analogue
Streptomyces strain 59KA3	ISP2	2749	628	0	No match
Streptomyces strain 61KA3	7H9	2714	378	0	No match
Streptomyces strain 61KA3	R2A	2309	547	0	No match
Streptomyces strain 63KA3	ISP2	2529	0	0	No match
Tennis Court sample					
Streptomyces strain 5TK1	CZ	2060	157	0	No match
Streptomyces strain 6TK1	ISP2	4825	1407	0	Puromycin
Streptomyces strain 19TA1	CZ	2815	511	0	Actinomycin D
Streptomyces strain 20TA1	CZ	2095	440	0	No match
Streptomyces strain 36TA2	R2A	2529	1096	0	No match
Streptomyces strain 45TA3	ISP2	5928	1457	0	No match
Streptomyces strain 47TA3	R2A	4423	1084	0	No match
Streptomyces strain 49TA3	R2A	5881	3402	0	No match
Streptomyces strain 49TA3	CZ	2384	643	0	No match
Streptomyces strain 49TA3	7H9	2368	6044	0	Antimycin

4.3.2.2 GNPS molecular networking analysis

Table 4.2 shows the compound matches to the GNPS database based on the mass spectra from the crude extracts prepared from the ZOIs excised from the agar plates of the most active actinobacterial strains. A compound on the GNPS database was only included as a match to the mass spectra if the cosine score was >0.7 and there were five or more matched peaks. The cosine value is a quantification of the alignment of the product mass spectrum to the mass spectrum of a known compound in the GNPS library. The larger the cosine value, the closer the match between the two spectra. Crude extracts of 26 strain-medium combinations tested returned no matches on the GNPS database and crude extracts of eight strain-medium combinations returned matches to known compounds in the GNPS database. The crude extract of *Streptomyces* strain 28KA1 (on R2A medium) returned a weak compound match to lambdamycin in the GNPS database (Table 4.2).

Streptomyces strain 6TK1 (on ISP2), *Streptomyces* strain 10KK1 (on ISP2) and *Streptomyces* strain 41KA2 (on both ISP2 and R2A) were found to be producing a compound that matched to the antibiotic puromycin (Table 4.2). Puromycin is produced by *Streptomyces alboniger* and structurally resembles the 3’ end of aminoacylated tRNA (aa-tRNA). This structure causes premature termination of translation through its addition to peptide chains. Puromycin is non-selective to its target and has a high toxicity (Aviner, 2020). *Streptomyces* strains 6TK1, 10KK1 and 41KA2 were shown to have strong to very strong activity against *M. aurum* A+ (Table 4.2), likely due to puromycin produced by the strains.

Streptomyces strain 19TA1 was found to be producing a compound that matched to actinomycin D when grown on CZ medium (Tale 4.2). Actinomycin D is a well-known antibiotic that consists of two lactone rings (Avendano and Menéndez, 2008) and has proven activity against *M. aurum* A+ (Jumpathong *et al.*, 2019). Furthermore, the antibiotic is often produced by members of the genus *Streptomyces* (Liu *et al.*, 2019). *Streptomyces* strain 19TA1 (on CZ medium) was shown to have strong antimycobacterial activity (Table 4.2), likely due to the presence of actinomycin D.

Streptomyces strain 28KA1 was found to be producing a compound that matched weakly to lambdamycin when grown on R2A medium (Table 4.2). Lambdamycin is a yellow-green pigment antibiotic produced by *Streptomyces glaucoachromogenes*.

Lambdamycin displays antimicrobial activity, particularly against Gram-positive bacteria, but also shows antiviral activity and cancerostatic action (Fleck *et al.*, 1976). *Streptomyces* strain 28KA1 (on R2A medium) was shown to have very strong antimycobacterial activity and weak activity against *S. aureus* ATCC 25923 (Table 4.2), potentially due to a lambdamycin-like molecule produced by the strain.

Streptomyces strain 49TA3 (on 7H9) and *Streptomyces* strain 22KA1 (on ISP2) were found to be producing a compound that matched to antimycin (Table 4.2). Antimycin has been shown to have little antibacterial activity against Gram-positive bacteria (Marquis, 1965). *Streptomyces* strain 22KA1 (on ISP2) was shown to have strong antimycobacterial activity and weak activity against *S. aureus* ATCC 25923 (Table 4.2). *Streptomyces* strain 49TA3 (on 7H9) was shown to have strong antimycobacterial activity and very strong activity against *S. aureus* ATCC 25923 (Table 4.2). It is therefore unlikely that antimycin is responsible for the antibiotic activity, but rather another active compound (or compounds) produced by the strains that had no structural match(es) in the GNPS database.

Streptomyces strain 59KA3 was found to be producing a compound that matched to a mitomycin C analogue when grown on R2A (Table 4.2). Mitomycin C is a compound active against a broad spectrum of bacteria, viruses and tumours. The compound’s mechanism of action is through DNA cross-linking and its main use is as an antitumour drug (Crooke and Bradner, 1976). *Streptomyces* strain 59KA3 (on R2A) was shown to have strong activity against *M. aurum* A+ and weak activity against *S. aureus* ATCC 25923 (Table 4.2), potentially due to the production of the mitomycin C analogue by the strain.

The crude extracts of the strain-medium combinations for which there were no matches indicate that these strains could be producing unknown compounds that have antimycobacterial activity. These “gifted” strains should be prioritised for testing against *M. tuberculosis* to assess if they might be candidates for development as new drugs to treat TB.

4.4 References

Avendano, C. and Menéndez, J.C. (2008) ‘DNA Intercalators and topoisomerase inhibitors’. Medicinal Chemistry of Anticancer Drugs, pp.199-228.

Aviner, R. (2020) ‘The science of puromycin: From studies of ribosome function to applications in biotechnology’, Computational and Structural Biotechnology Journal. Elsevier B.V., pp. 1074–1083. doi: 10.1016/j.csbj.2020.04.014.

Becton Dickinson (2017) Middlebrook 7H9 Broth, DifcoTM & BBLTM Manual, 2nd Edition. Available at: https://www.bd.com/europe/regulatory/Assets/IFU/Difco_BBL/212352.pdf (Accessed: 12 November 2019).

Chung, G. A. C. *et al*. (1995) ‘High-throughput screen for detecting antimycobacterial agents’, Antimicrobial Agents and Chemotherapy, 39(10), pp. 2235–2238. doi: 10.1128/AAC.39.10.2235.

Connon, S. A. and Giovannoni, S. J. (2002) ‘High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates’, Applied and Environmental Microbiology, 68(8), pp. 3878–3885. doi: 10.1128/AEM.68.8.3878-3885.2002.

Crooke, S.T. and Bradner, W.T.(1976) ‘Mitomycin C: a review’. Cancer treatment reviews, 3(3), pp.121-139.

Davies, J. and Davies, D. (2010) ‘Origins and evolution of antibiotic resistance.’, Microbiology and Molecular Biology Reviews, 73(3), pp. 417–433. doi: 10.1128/mmbr.00016-10.

Fleck, W. *et al*. (1976) ‘Lambdamycin, ein Antibioticum aus dem Stamm IMET 31118 von Streptomyces glaucoachromogenes PRAUSER’, Zeitschrift für allgemeine Mikrobiologie, 16(7), pp. 521–528. doi: 10.1002/jobm.3630160704.

Hockett, K. L. and Baltrus, D. A. (2017) ‘Use of the soft-agar overlay technique to screen for bacterially produced inhibitory compounds’, Journal of Visualized Experiments, 2017(119), p. e55064. doi: 10.3791/55064.

Jumpathong, J. *et al*. (2019) ‘Thin layer chromatography-bioautography assay for antibacterial compounds from *Streptomyces* sp. TBRC 8912, a newly isolated actinomycin D producer’, Chiang Mai Journal of Science, 46(5), pp. 839–849. Available at: https://www.thaiscience.info/Journals/Article/CMJS/10990679.pdf (Accessed: 29 August 2022).

Krzyśko-Lupicka, T. *et al*. (1997) ‘The ability of soil-borne fungi to degrade organophosphonate carbon-to-phosphorus bonds’, Applied Microbiology and Biotechnology, 48(4), pp. 549–552. doi: 10.1007/s002530051095.

Li, J. W. H. and Vederas, J. C. (2009) ‘Drug discovery and natural products: End of an era or an endless frontier?’, Science, pp. 161–165. doi: 10.1126/science.1168243.

Liu, M. *et al*. (2019) ‘Identification of the actinomycin D biosynthetic pathway from marine-derived *Streptomyces costaricanus* SCSIO ZS0073’, Marine Drugs, 17(4). doi: 10.3390/md17040240.

Mahajan, G. B. and Balachandran, L. (2012) ‘Antibacterial agents from actinomycetes’, Frontiers in Bioscience, (4), pp. 240–253. doi: 10.1109/epqu.2011.6128966.

Marquis, R. E. (1965) ‘Nature of the Bactericidal Action of Antimycin A for *Bacillus megaterium*’, Journal of bacteriology, 89(6), pp. 1453–1459. doi: 10.1128/JB.89.6.1453-1459.1965.

Martínková, L. *et al*. (2009) ‘Biodegradation potential of the genus *Rhodococcus*’, Environment International, 35(1), pp. 162–177. doi: 10.1016/j.envint.2008.07.018.

Melander, R. J. *et al.* (2018) ‘Narrow-spectrum antibacterial agents’, MedChemComm, pp. 12–21. doi: 10.1039/c7md00528h.

Miao, V. and Davies, J. (2010) ‘Actinobacteria: The good, the bad, and the ugly’, Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology, pp. 143–150. doi: 10.1007/s10482-010-9440-6.

Miller, J. H. (1972) ‘Experiments in molecular genetics’, Astrophysics and Space Science, pp. 203–209. (Accessed: 18 August 2022).

Niu, G. (2018) ‘Genomics-Driven Natural Product Discovery in Actinomycetes’, Trends in Biotechnology, 36(3), pp. 238–241. doi: 10.1016/j.tibtech.2017.10.009.

Prescott, L. M. *et al.* (2002) MICROBIOLOGY, fifth edition.

Rigali, S. *et al*. (2008) ‘Feast or famine: The global regulator DasR links nutrient stress to antibiotic production by *Streptomyces*’, EMBO Reports, 9(7), pp. 670–675. doi: 10.1038/embor.2008.83.

Shirling, E. and Gottlieb, D. (1966) ‘Methods for characterization of *Streptomyces* species’, International journal of Systematic Bacteriology, 16(1946), p. 340.

Sotgiu, G. *et al*. (2009) ‘Epidemiology and clinical management of XDR-TB: A systematic review by TBNET’, European Respiratory Journal, pp. 871–881. doi: 10.1183/09031936.00168008.

van Bergeijk, D. A. *et al*. (2020) ‘Ecology and genomics of Actinobacteria: new concepts for natural product discovery’, Nature Reviews Microbiology, 18(10), pp. 546–558. doi: 10.1038/s41579-020-0379-y.

Velayati, A. A. *et al*. (2009) ‘Emergence of new forms of totally drug-resistant tuberculosis bacilli: Super extensively drug-resistant tuberculosis or totally drug-resistant strains in Iran’, Chest, 136(2), pp. 420–425. doi: 10.1378/chest.08-2427.

Watson, D. J. *et al*. (2021) ‘Discovery of novel cyclic ethers with synergistic antiplasmodial activity in combination with valinomycin’, Molecules, 26(24). doi: 10.3390/molecules26247494.

World Health Organization (2021) Global Tuberculosis Report 2021, Licence: CC BY-NC-SA 3.0 IGO.

Zhu, H. *et al.* (2014) ‘Triggers and cues that activate antibiotic production by actinomycetes’, Journal of Industrial Microbiology and Biotechnology, pp. 371–386. doi: 10.1007/s10295-013-1309-z.

Chapter 5:

General discussion

Despite there being a decrease in the rate of discovery of novel antibiotics from actinobacteria, they remain a good source of new secondary metabolites (Connon and Giovannoni, 2002; Li and Vederas, 2009). The isolation and characterization of poorly studied actinobacterial taxa allows access to potentially novel actinobacterial strains and any novel bioactive compounds they produce (Goodfellow, 2010). The discovery of new antibiotic compounds is needed to combat the emergence of “superbugs” that have arisen as a result of antibiotic resistance (Davies and Davies, 2010). Although less than 1% of the microbial diversity in the terrestrial environment can be isolated in pure culture (Arjun and Harikrishnan, 2011), the use of selective media for culture-based isolations can promote the growth of unique populations of actinobacteria in laboratory culture (Connon and Giovannoni, 2002).

This study had three primary aims: (1) a metagenomic analysis of the actinobacterial diversity in two soil samples, (2) a metagenomic-guided investigation of the culturable actinobacterial diversity from the same two soil samples and (3) the screening of the isolated actinobacterial strains for antimycobacterial activity. The soil samples, taken from two sites at the UCT, underwent culture-independent metagenomic analysis through the construction of actinobacteria-specific 16S-rRNA gene libraries. The results of the metagenomic analyses guided the choice of selective isolation media for the second part of the project. Four media were chosen, targeting the genera *Amycolatopsis* and *Kribbella*, for the isolation of actinobacteria from the original soil samples using standard plate cultivation techniques. The isolated bacteria were identified to the genus level using their 16S-rRNA gene sequences and phylogenetic trees were constructed to investigate their affiliations within their genera. The isolates were screened for antimycobacterial antibiotic activity using the agar overlay method and *M. aurum* strain A+ as the test bacterium. Isolates showing strong or very strong antimycobacterial activity were also screened against *S. aureus* ATCC 25923 and *E. coli* ATCC 25922 to assess their spectrum of antimicrobial activity. These highly active

actinobacterial strains also underwent GNPS molecular networking analyses to investigate if the compounds produced matched to known compounds in the database. Thus, all the aims of the project were achieved.

Rarefaction curves were generated for the metagenomic clone library sequences for the K and TC soil samples (Figure 2.1). Both the blue rarefaction curve (K sample) and the orange curve (TC sample) showed that the full actinobacterial biodiversity of both samples had not been captured. It would be beneficial in future to generate enough clone sequences such that the rarefaction curves reach an asymptote and the complete actinobacterial biodiversity of the sample is captured. Capturing the full actinobacterial biodiversity would give one a better idea of the actinobacterial make-up of the soil and allow a more informed decision on the choice of selective media. The author would have liked to generate an extra 100 clone sequences for each soil sample in addition to the 55 clone sequences that were generated in an attempt to capture the full actinobacterial biodiversity. This was not done due to time lost as a result of the lockdown in 2020 during the Covid-19 pandemic. If a similar study was attempted in future, the author would suggest using NGS for the metagenomic analysis of chosen samples. NGS has a much higher throughput and a simpler library preparation compared to genetic library preparation. NGS also requires no cloning (Pagani *et al.*, 2012). Although more expensive, for these reasons NGS would be a more time effective method for genomic analysis and provide comprehensive results of the microbial diversity of the samples.

Despite not capturing the entire biodiversity of the samples during the culture-independent portion of the project, both samples showed a wide biodiversity. The K and TC samples both had clone sequences that represented 10 different orders, with eight orders common to both clone libraries. The orders *Pseudonocardiales, Propionibacteriales* and *Acidimicrobiales* made up a large proportion (>7% for each order) of clone sequences in both clone libraries (Figures 2.2 and 2.3). This result showed there to be a large actinobacterial biodiversity in both soil samples. The actinobacterial biodiversity of the culture-based isolates (Table 3.1), in comparison to the culture-independent actinobacterial biodiversity identified (Figures 2.2 and 2.3), showed how the actinobacterial diversity detected is narrowed and skewed when isolations are performed on agar media.

For the metagenomic analysis the actinobacteria specific primer set Com 2xf-Ac118r was chosen for the amplification of eDNA. The actinobacteria specific primer set was effective with no non-actinobacterial DNA amplified (Table A1 and A2) however it did result in the amplification of short segments of DNA (about 270 bp) to be analysed. This meant that no accurate species information could be deduced from the very short 16S rRNA gene segments amplified from the eDNA. As a result, the author could not compare the species identified in the metagenomic section to the species isolated in the culture-based isolations. If the study was to be repeated, the author would attempt to amplify longer actinobacterial specific segments using different primers so that species information could be deduced from the metagenomic data.

For the culture-based isolations, the aim was to promote the isolation of rare (non-*Streptomyces*) actinobacterial strains, particularly from the genera *Amycolatopsis* and *Kribbella*, which had been identified in the metagenomic analyses. The media ASM1, ASM2 and ASM3 were chosen to be selective for the *Amycolatopsis* strains and KSM1 medium was chosen to be selective for *Kribbella* strains. Forty (40) of 42 isolates were identified to the genus level with 38 isolates belonging to the genus *Streptomyces*, one isolate belonging to the genus *Actinomadura* (strain 1KK1) and one isolate belonging to the genus *Nonomuraea* (strain 7TK1; Table 3.1). No *Amycolatopsis* or *Kribbella* strains were isolated in the culture-based isolations (Table 3.1), despite *Amycolatopsis* strains being shown to be present in both samples (Table 2.1 and 2.2) and a *Kribbella* strain being detected in the K sample (Table 2.1). *Streptomyces* species were the dominant genus in the culture-based isolations despite the use of selective media. This is, however, not surprising as *Streptomyces* is known to dominate in culture-based studies, because they grow very well on laboratory media. If the author were to repeat the study, he would want to isolate fewer *Streptomyces* strains and more of the rarer actinobacterial strains. One way to do this would be to eliminate *Streptomyces* strains from the soil sample suspension prior to plating on isolation plates through the addition of *Streptomyces*-specific bacteriophages. Following the addition of the *Streptomyces*-specific bacteriophages to kill the *Streptomyces* cells, the soil suspension would be plated onto selective media agar plates. Even if the isolations were repeated with the addition of *Streptomyces*-specific bacteriophages, it is likely that there would still be some *Streptomyces* colonies on the subsequent isolation plates. This being because the bacteriophages would only be able to eliminate

vegetative *Streptomyces* cells and not be effective in killing *Streptomyces* spores. Thus, *Streptomyces* spores and vegetative *Streptomyces* cells resistant to the bacteriophages would form colonies on the isolation plates. Five different *Streptomyces*-specific bacteriophages were used by (Kurtböke and French, 2007) in their selective isolation of bacteria from the gut of termites. In their study, some *Streptomyces* strains *were* isolated, but there was a significant reduction in their numbers (Kurtböke and French, 2007). This bacteriophage treatment would, therefore, reduce the number of *Streptomyces* strains present on the isolation plates and further bias, in conjunction with the selective media, the isolation of rare actinobacteria.

All of the bacteria identified to the genus level had top hits with a sequence similarity of >99.15% over partial 16S-rRNA gene sequence lengths ranging from 605 to 905 bp (Table 3.1). This suggests that all the isolates that were identified likely do not represent new species. There was, however, little data on the isolates so further investigation may reveal some of them to be novel species. In this regard, *Actinomadura* strain 1KK1 and *Nonomuraea* strain 7TK1 would be worth investigating further, which would entail more detailed phenotypic characterisation of the isolates to compare their characteristics to those of their closest relatives (published in the original descriptions of these relatives). If one or both of strains 1KK1 and 7TK1 show distinct characteristics, it would then be worth sequencing their genome(s). dDDH and ANI measurements would then be done to determine if these isolates represent unique genomic species (i.e. <70% DNA relatedness by dDDH and ANI values <95-96% between each isolate and the type strains of their closest phylogenetic relatives).

There were 30 *Streptomyces* isolates that were grouped into three potential-clone groups coloured blue, orange and green (Table 3.1). Clone analysis was carried out for 23 of the 30 *Streptomyces* potential-clone strains through phenotypic analysis on International *Streptomyces* Project medium No. 4 (ISP4) agar plates (Table A3) and analyses using partial *gyrB* gene sequences. For the *gyrB* gene sequences, homology matrix construction (Figures 3.7, 3.8 and 3.9) and phylogenetic analyses (Figure 3.6) were carried out. These analyses divided the 23 strains from the three potential-clone groups into five clone groups and eight distinctive *Streptomyces* strains. It is recommended that in future only one strain from each of the clone groups be studied

further (i.e. one of strains 2KK1 and 10KK1; one of strains 28KA1, 32KA1 and 54KA3; one of strains 39KA2, 41KA2 and 42KA2; one of strains 5TK1, 9TK1, 15TA1, 16TA1 and 36TA2 and one of strains 6TK1 and 19TA1).

Forty (40) of the isolates were tested against *M. aurum* strain A+ using the agar overlay method on four different media to test for antimycobacterial activity. The isolates tested showed many active strains with significant antimycobacterial activity: there were 35 strain-medium combinations that showed strong (2001 mm^2 < ZOI < 3000 mm^2) or very strong (ZOI>3000 mm^2) antimycobacterial activity (Table 4.1). These active strain-medium combinations were also tested against *E. coli* strain ATCC 25922 and *S. aureus* strain ATCC 25923 and had compounds extracted from their ZOIs subjected to GNPS molecular networking analyses to assess if the compounds produced are known compounds.

Based on the GNPS molecular analysis: *Streptomyces* strains 22KA1 and 49TA3 were found to produce antimycin, *Streptomyces* strain 28KA1 was found to produce a lambdamycin-like compound (with a weak match), *Streptomyces* stains 6TK1, 10KK1 and 41KA2 were found to produce puromycin, *Streptomyces* strain 59KA3 was found to produce a mitomycin C analogue and *Streptomyces* strain 19TA1 was found to produce actinomycin D (Table 4.2). 16S-rRNA phylogenetic trees (Figures 3.4a, 3.4b and 3.5a) showed the closely related species to these *Streptomyces* strains, with none of the closely related strains being known producers of any of the above identified antibiotics. The finding of potentially novel producers of lambdamycin, puromycin and mitomycin C was unexpected, as these antibiotics are known to be produced primarily by just one *Streptomyces* species (Fleck *et al.*, 1976; Martin *et al.*, 2002; Khadka *et al.*, 2022). The finding of producers of antimycin and actinomycin D is, however, more likely as both compounds are produced by many *Streptomyces* species (Toumatia et al., 2015; McLean et al., 2016).

The antibiotic overlays against *E. coli* ATCC 25922 showed no activity for all strain-medium combinations tested (Table 4.2). Therefore, all the compounds produced were ineffective against this Gram-negative bacterium. Twenty six (26) off the 35 strain-medium combinations tested showed no or weak activity (ZOI<1000 mm^2) against *S. aureus* ATCC 25923 and nine strain-medium combinations tested showed moderate

to very strong activity against *S. aureus* ATCC 25923 (Table 4.2). Those strains showing little activity against *S. aureus* ATCC 25923 may be producing compounds with a narrow spectrum of activity that are specific to the genus *Mycobacterium.* Strains that showed good activity against *S. aureus* ATCC 25923 are potentially producing one compound that has antibiotic activity against both *S. aureus* ATCC 25923 and *M. aurum* A+ or the strains are producing two or more compounds that are effective against different test bacteria. To determine if one or multiple compounds are responsible for the activity, bioautography could be done in future on the compounds extracted from liquid cultures of the active strains.

The crude extracts revealed there to be 26 strain-medium combinations of the 35 tested that returned no matches on the GNPS database (Table 4.2), meaning that these "gifted" actinobacterial strains are producing unknown compounds that result in strong and very strong activity against *M. aurum* A+ on specific media. In future, these "gifted" strains should be prioritized for further study and first sent for testing against *M. tuberculosis* (in the Faculty of Health Sciences, UCT) to assess if they might be candidates for development as new drugs to treat TB. There is also a chance that these "gifted" strains are producing compounds that could be effective in the treatment of other diseases, such as malaria, a disease mainly caused by *Plasmodium falciparum* that results in over 600 000 deaths annually (https://www.who.int/teams/global-malaria-programme). The author suggests the screening of the compounds produced by the "gifted" strains against *P. falciparum* to assess if they could be effective antimalarials (these assays require erythrocyte cultivation of the *P. falciparum* cells in a Biosafety Level 2 laboratory and would be conducted by Dr Daniel Watson in the Faculty of Health Sciences, UCT).

References

Arjun, J.K. and Harikrishnan, K. (2011) 'Metagenomic analysis of bacterial diversity in the rice rhizosphere soil microbiome', *Biotechnology Bioinf. Bioengineering*, 1(3), pp. 361–367.

Connon, S.A. and Giovannoni, S.J. (2002) 'High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates', *Applied and Environmental Microbiology*, 68(8), pp. 3878–3885. Available at: https://doi.org/10.1128/AEM.68.8.3878-3885.2002.

Davies, J. and Davies, D. (2010) 'Origins and evolution of antibiotic resistance.', *Microbiology and Molecular Biology Reviews*, 73(3), pp. 417–433. Available at: https://doi.org/10.1128/mmbr.00016-10.

Fleck, W. *et al.* (1976) 'Lambdamycin, ein Antibioticum aus dem Stamm IMET 31118 von Streptomyces glaucoachromogenes PRAUSER', *Zeitschrift für allgemeine Mikrobiologie*, 16(7), pp. 521–528. Available at: https://doi.org/10.1002/jobm.3630160704.

Goodfellow, M. (2010) 'Selective Isolation of Actinobacteria', in *Manual of Industrial Microbiology and Biotechnology*. John Wiley & Sons, Ltd, pp. 13–27. Available at: https://doi.org/10.1128/9781555816827.ch2.

Khadka, S. *et al.* (2022) 'Genetic depletion of de novo coenzyme A biosynthesis exacerbates puromycin toxicity', *bioRxiv* [Preprint]. Available at: https://doi.org/10.1101/2022.09.06.506844.

Kurtböke, D.I. and French, J.R.J. (2007) 'Use of phage battery to investigate the actinofloral layers of termite gut microflora', *Journal of Applied Microbiology*, 103(3), pp. 722–734. Available at: https://doi.org/10.1111/j.1365-2672.2007.03308.x.

Li, J.W.H. and Vederas, J.C. (2009) 'Drug discovery and natural products: End of an era or an endless frontier?', *Science*, pp. 161–165. Available at: https://doi.org/10.1126/science.1168243.

Martin, T.W. *et al.* (2002) 'Molecular Basis of Mitomycin C Resistance in *Streptomyces*: Structure and Function of the MRD Protein', *Structure*, 10, pp. 933–942.

McLean, T.C. *et al.* (2016) 'Coordinate Regulation of Antimycin and Candicidin Biosynthesis', *mSphere*, 1(6). Available at: https://doi.org/10.1128/msphere.00305-16.

Pagani, I. *et al.* (2012) 'The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata', *Nucleic Acids Research*, 40(D1), pp. D571–D579. Available at: https://doi.org/10.1093/NAR/GKR1100.

Toumatia, O. *et al.* (2015) 'Antifungal properties of an actinomycin D-producing strain, Streptomyces sp. IA1, isolated from a Saharan soil', *Journal of Basic Microbiology*, 55(2), pp. 221–228. Available at: https://doi.org/10.1002/jobm.201400202.

Appendices

Table A1: All EzBioCloud results for all Kramer sample metagenomic clone sequences. Entries in red indicate duplicate clones that were removed before the biodiversity was analysed.

Clone no.	Sequence length	Closest relative (GenBank Accesion Number)	% Similarity	Order	Family	Genus
1	272	Uncultured Bacterial Clone (AY792234)	96.7	*Acidimicrobiales*	*Acidimicrobiaceae*	*Aciditerrimonas*
2	277	*Amycolatopsis azurea* (ANMG01000133)	97.8	*Pseudonocardiales*	*Pseudonocardiaceae*	*Amycolatopsis*
3	273	*Actinomycetospora chiangmaiensis* (ARBI01000033)	100	*Pseudonocardiales*	*Pseudonocardiaceae*	*Actinomycetospora*
4	278	Uncultured Bacterial Clone (PAC000666)	95.2	*Streptosporangiales*	-	-
5	270	*Nocardioides dilutus* (EF466121)	98.9	*Propionibacteriales*	*Nocardioidaceae*	*Nocardioides*
6	271	*Naasia aerilata* (JQ639051)	98.5	*Micrococcales*	*Microbacteriaceae*	*Naasia*
7	272	Uncultured Bacterial Clone (FJ478778)	98.2	*Acidimicrobiales*	*Ilumatobacteraceae*	*Desertimonas*
8	271	Uncultured Bacterial Clone (JF179655)	99.3	*Acidimicrobiales*	*Iamiaceae*	-
9	272	*Kineosporia rhamnosa* (AB003935)	100.0	*Kineosporiales*	*Kineosporiaceae*	*Kineosporia*
10	271	Uncultured Bacterial Clone (PAC000340)	99.3	*Streptosporangiales*	-	-
11	272	Uncultured Bacterial Clone (EU753662)	96.3	*Acidimicrobiales*	*Iamiaceae*	-
12	273	*Pseudolysinimonas kribbensis* (EF466129)	100	*Micrococcales*	*Microbacteriaceae*	*Pseudolysinimonas*
R	274	*Pseudolysinimonas kribbensis* (EF466129)	100	*Micrococcales*	*Microbacteriaceae*	*Pseudolysinimonas*
13	272	Uncultured Bacterial Clone (FJ592837)	97.1	*Acidimicrobiales*	*Acidimicrobiaceae*	-
14	278	*Streptomyces fragilis* (AY999917)	99.3	*Streptomycetales*	*Streptomycetaceae*	*Streptomyces*
15	279	Millisia brevis (BCRN01000038)	95.6	*Mycobacteriales*	*Nocardiaceae*	*Millisia*
16	272	*Microlunatus phosphovorus* (AP012204)	100	*Propionibacteriales*	*Propionibacteriaceae*	*Microlunatus*
17	271	*Aeromicrobium ginsengisoli* (AB245394)	100	*Propionibacteriales*	*Nocardioidaceae*	*Aeromicrobium*
18	272	*Diaminobutyricimonas aerilata* (JQ639052)	100	*Micrococcales*	*Microbacteriaceae*	*Diaminobutyricimonas*
19	273	*Actinophytocola corallina* (AB511316)	99.6	*Pseudonocardiales*	*Pseudonocardiaceae*	*Actinophytocola*
R	273	*Actinophytocola corallina* (AB511316)	99.6	*Pseudonocardiales*	*Pseudonocardiaceae*	*Actinophytocola*
R	273	*Actinophytocola corallina* (AB511316)	99.6	*Pseudonocardiales*	*Pseudonocardiaceae*	*Actinophytocola*
20	276	*Arthrobacter humicola* (AB279890)	99.6	*Micrococcales*	*Micrococcaceae*	*Arthrobacter*
21	271	Uncultured Bacterial Clone (AJ241005)	96.3	*Acidimicrobiales*	*Microthrixaceae*	-
R	271	Uncultured Bacterial Clone (AJ241005)	96.3	*Acidimicrobiales*	*Microthrixaceae*	-
22	272	Uncultured Bacterial Clone (HQ910322)	98.2	*Frankiales*	-	-
23	273	*Pseudonocardia spinosispora* (AJ249206)	98.5	*Pseudonocardiales*	*Pseudonocardiaceae*	*Pseudonocardia*

Table A1: (continued)

Clone no.	Sequence length	Closest relative (GenBank Accesion Number)	% Similarity	Order	Family	Genus
24	272	*Hamadaea tsunoensis* (AUAX01000066)	97.8	*Micromonosporales*	*Micromonosporaceae*	*Hamadaea*
25	273	*Actinomycetospora lutea* (AB514515)	98.5	*Pseudonocardiales*	*Pseudonocardiaceae*	*Actinomycetospora*
26	272	Uncultured Bacterial Clone (FJ478790)	97.4	*Acidimicrobiales*	-	-
27	272	*Nocardioides iriomotensis* (AB544079)	97.4	*Propionibacteriales*	*Nocardioidaceae*	*Nocardioides*
28	270	*Nocardioides daejeonensis* (JF937066)	98.2	*Propionibacteriales*	*Nocardioidaceae*	*Nocardioides*
29	272	Uncultured Bacterial Clone (JF135243)	99.3	*Streptosporangiales*	-	-
30	298	Uncultured Bacterial Clone (JF202912)	99.6	*Propionibacteriales*	*Nocardioidaceae*	*Nocardioides*
31	273	*Amycolatopsis mediterranei* (AJ293754)	99.3	*Pseudonocardiales*	*Pseudonocardiaceae*	*Amycolatopsis*
32	272	Uncultured Bacterial Clone (FJ592837)	97.8	*Acidimicrobiales*	*Acidimicrobiaceae*	-
33	272	Uncultured Bacterial Clone (DQ129383)	97.4	*Acidimicrobiales*	*Iamiaceae*	-
34	272	Uncultured Bacterial Clone (EF516495)	99.6	*Acidimicrobiales*	*Iamiaceae*	-
35	273	*Mycolicibacterium confluentis* (LQOQ01000043)	98.2	*Mycobacteriales*	*Mycobacteriaceae*	*Mycolicibacterium*
36	272	Uncultured Bacterial Clone (JX098497)	98.9	*Acidimicrobiales*	*Acidimicrobiaceae*	-
37	272	*Micromonospora rifamycinica* (LRMV01000349)	99.3	*Micromonosporales*	*Micromonosporaceae*	*Micromonospora*
38	273	*Mycobacterium cookii* (AF480598)	100	*Mycobacteriales*	*Mycobacteriaceae*	*Mycobacterium*
39	273	*Pseudonocardia antarctica* (AJ576010)	98.5	*Pseudonocardiales*	*Pseudonocardiaceae*	*Pseudonocardia*
40	277	*Cryobacterium psychrophilum* (AJ544063)	95.2	*Micrococcales*	*Microbacteriaceae*	*Cryobacterium*
41	271	Uncultured Bacterial Clone (EU132518)	99.6	*Frankiales*	-	-
42	273	*Amycolatopsis mediterranei* (AJ293754)	99.3	*Pseudonocardiales*	*Pseudonocardiaceae*	*Amycolatopsis*
43	273	Uncultured Bacterial Clone (PAC000665)	98.2	*Acidimicrobiales*	*Iamiaceae*	-
44	273	*Naasia aerilata* (JQ639051)	95.2	*Micrococcales*	*Microbacteriaceae*	*Naasia*
45	274	*Actinophytocola sediminis* (KJ013500)	97.1	*Pseudonocardiales*	*Pseudonocardiaceae*	*Actinophytocola*
46	280	*Kribbella jiaozuonensis* (MK817655)	99.6	*Propionibacteriales*	*Kribbellaceae*	*Kribbella*
47	274	*Pseudonocardia alaniniphila* (EU722519)	100	*Pseudonocardiales*	*Pseudonocardiaceae*	*Pseudonocardia*
R	274	*Pseudonocardia alaniniphila* (EU722519)	100	*Pseudonocardiales*	*Pseudonocardiaceae*	*Pseudonocardia*
48	273	Uncultured Bacterial Clone (PAC002681)	97.1	*Acidimicrobiales*	*Ilumatobacteraceae*	-
49	275	*Leifsonia lichenia* (AB278552)	97.1	*Micrococcales*	*Microbacteriaceae*	*Leifsonia*
50	273	Uncultured Bacterial Clone (JX079124)	96.7	*Propionibacteriales*	*Nocardioidaceae*	*Nocardioides*

Table A2: All EzBioCloud results for all Tennis Court sample metagenomic clone sequences. Entries in red indicate duplicate clones that were removed before the biodiversity was analysed.

Clone no.	Sequence length	Closest relative (GenBank Accesion Number)	% Similarity	Order	Family	Genus
1	278	*Streptomyces haliclonae* (AB473554)	99.6	*Streptomycetales*	*Streptomycetaceae*	*Streptomyces*
2	272	*Streptomyces regensis* (LFVR01000689)	100	*Streptomycetales*	*Streptomycetaceae*	*Streptomyces*
3	274	*Lysinimonas yzui* (MG934573)	98.2	*Micrococcales*	*Microbacteriaceae*	*Lysinimonas*
4	272	Uncultured Bacterial Clone (GQ397040)	96.7	*Acidimicrobiales*	*Ilumatobacteraceae*	-
5	271	*Nocardioides halotolerans* (KE383922)	98.5	*Propionibacteriales*	*Nocardioidaceae*	*Nocardioides*
6	272	Uncultured Bacterial Clone (PAC003044)	99.6	*Acidimicrobiales*	*Iamiaceae*	-
7	272	Uncultured Bacterial Clone (PAC000665)	98.5	*Acidimicrobiales*	*Iamiaceae*	-
8	273	Uncultured Bacterial Clone (JF216358)	98.2	*Acidimicrobiales*	*Iamiaceae*	-
9	273	*Pseudonocardia bannensis* (FJ817375)	96.7	*Pseudonocardiales*	*Pseudonocardiaceae*	*Pseudonocardia*
10	273	Uncultured Bacterial Clone (JQ407953)	97.1	*Acidimicrobiales*	*Iamiaceae*	-
11	279	Uncultured Bacterial Clone (PAC000666)	93.4	*Streptosporangiales*	-	-
12	273	Uncultured Bacterial Clone (EF018761)	99.3	*Mycobacteriales*	*Mycobacteriaceae*	*Mycobacterium*
13	273	*Actinokineospora enzanensis* (ARFV01000009)	98.2	*Pseudonocardiales*	*Pseudonocardiaceae*	*Actinokineospora*
14	273	*Pseudonocardia tetrahydrofuranoxydans* (AJ249200)	99.6	*Pseudonocardiales*	*Pseudonocardiaceae*	*Pseudonocardia*
15	273	*Amycolatopsis helveola* (AB327254)	99.6	*Pseudonocardiales*	*Pseudonocardiaceae*	*Amycolatopsis*
16	274	Uncultured Bacterial Clone (JF173550)	98.5	*Pseudonocardiales*	*Pseudonocardiaceae*	*Pseudonocardia*
17	272	*Actinoallomurus amamiensis* (AB364583)	97.1	*Streptosporangiales*	*Thermomonosporaceae*	*Actinoallomurus*
R	273	*Actinoallomurus amamiensis* (AB364583)	97.1	*Streptosporangiales*	*Thermomonosporaceae*	*Actinoallomurus*
18	273	*Mycolicibacterium mucogenicum* (AY457074)	97.4	*Mycobacteriales*	*Mycobacteriaceae*	*Mycolicibacterium*
19	273	*Amycolatopsis azurea* (ANMG01000133)	100	*Pseudonocardiales*	*Pseudonocardiaceae*	*Amycolatopsis*
R	273	*Amycolatopsis azurea* (ANMG01000133)	100	*Pseudonocardiales*	*Pseudonocardiaceae*	*Amycolatopsis*
20	272	AY234624_s (AY234624)	100	*Acidimicrobiales*	*Iamiaceae*	-
21	272	*Catenulispora acidiphila* (CP001700)	100	*Catenulisporales*	*Catenulisporaceae*	*Catenulispora*
22	273	*Amycolatopsis pigmentata* (AB327253)	97.8	*Pseudonocardiales*	*Pseudonocardiaceae*	*Amycolatopsis*
23	272	*Jatrophihabitans endophyticus* (jgi.1107904)	97.1	*Jatrophihabitantales*	*Jatrophihabitantaceae*	*Jatrophihabitans*
24	271	*Jatrophihabitans endophyticus* (jgi.1107904)	100	*Jatrophihabitantales*	*Jatrophihabitantaceae*	*Jatrophihabitans*
25	272	Uncultured Bacterial Clone (JF179655)	98.8	*Acidimicrobiales*	*Iamiaceae*	-
26	273	*Curtobacterium citreum* (X77436)	95.2	*Micrococcales*	*Microbacteriaceae*	*Curtobacterium*

Table A2: (continued)

Clone no.	Sequence length	Closest relative (GenBank Accesion Number)	% Similarity	Order	Family	Genus
27	273	Uncultured Bacterial Clone (DQ129564)	99.6	*Pseudonocardiales*	*Pseudonocardiaceae*	*Pseudonocardia*
28	272	Uncultured Bacterial Clone (EF516583)	98.2	*Acidimicrobiales*	*Microthrixaceae*	-
29	272	Uncultured Bacterial Clone (PAC000665)	96.7	*Acidimicrobiales*	*Iamiaceae*	-
30	272	Uncultured Bacterial Clone (PAC000665)	98.5	*Acidimicrobiales*	*Iamiaceae*	-
31	273	*Mycolicibacterium tokaiense* (AF480590)	97.1	*Mycobacteriales*	*Mycobacteriaceae*	*Mycolicibacterium*
32	262	*Streptomyces cinnabarinus* (JNXQ01000585)	99.6	*Streptomycetales*	*Streptomycetaceae*	*Streptomyces*
33	273	*Actinophytocola corallina* (AB511316)	100	*Pseudonocardiales*	*Pseudonocardiaceae*	*Actinophytocola*
R	273	*Actinophytocola corallina* (AB511316)	100	*Pseudonocardiales*	*Pseudonocardiaceae*	*Actinophytocola*
34	272	Uncultured Bacterial Clone (JF216358)	96.3	*Acidimicrobiales*	*Iamiaceae*	-
35	272	Uncultured Bacterial Clone (EF516495)	97.8	*Acidimicrobiales*	*Iamiaceae*	-
36	280	*Streptomyces fragilis* (AY999917)	100	*Streptomycetales*	*Streptomycetaceae*	*Streptomyces*
37	271	*Nocardioides daejeonensis* (JF937066)	97.1	*Propionibacteriales*	*Nocardioidaceae*	*Nocardioides*
38	277	Uncultured Bacterial Clone (JQ978888)	96.0	*Acidimicrobiales*	*Iamiaceae*	-
39	272	Uncultured Bacterial Clone (EF516495)	98.5	*Acidimicrobiales*	*Iamiaceae*	-
40	271	*Jatrophihabitans endophyticus* (jgi.1107904)	95.9	*Jatrophihabitantales*	*Jatrophihabitantaceae*	*Jatrophihabitans*
41	272	Uncultured Bacterial Clone (PAC000665)	98.5	*Acidimicrobiales*	*Iamiaceae*	-
42	259	AP019371_s (AP019371)	96.1	*Micromonosporales*	*Micromonosporaceae*	*Actinoplanes*
43	272	Uncultured Bacterial Clone (GQ397040)	97.1	*Acidimicrobiales*	*Ilumatobacteraceae*	-
44	269	*Streptomyces niveiscabiei* (AF361786)	98.5	*Streptomycetales*	*Streptomycetaceae*	*Streptomyces*
45	273	*Actinophytocola corallina* (AB511316)	99.6	*Pseudonocardiales*	*Pseudonocardiaceae*	*Actinophytocola*
46	282	*Nocardioides albus* (AF004988)	96.7	*Propionibacteriales*	*Nocardioidaceae*	*Nocardioides*
47	258	*Desertimonas flava* (QKYK01000032)	100	*Acidimicrobiales*	*Ilumatobacteraceae*	*Desertimonas*
48	272	*Nocardioides allogilvus* (MG800321)	98.2	*Propionibacteriales*	*Nocardioidaceae*	*Nocardioides*
49	260	*Demequina lutea* (AB639016)	98.9	*Demequinales*	*Demequinaceae*	*Demequina*
50	272	Uncultured Bacterial Clone (PAC003044)	95.6	*Acidimicrobiales*	*Iamiaceae*	-
51	274	*Amycolatopsis mediterranei* (AJ293754)	99.6	*Pseudonocardiales*	*Pseudonocardiaceae*	*Amycolatopsis*
52	272	Uncultured Bacterial Clone (JQ427076)	98.5	*Acidimicrobiales*	*Microthrixaceae*	-

Table A3: Colony characteristics of isolated strains grown on ISP4 medium for 14 days at 30°C. '-' indicates that no spore mass was present. 'n/a' indicates that the strain failed to grow from frozen stock cultures and could not be sub-cultured from old agar plates. 'NG' (no growth) indicates that the strains were unable to grow on ISP4 plates.

Strain	Substrate mycelium colour	Spore mass colour
Kramer sample		
Actinomadura 1KK1	Translucent white	-
Streptomyces 2KK1	Pale yellow	White and Grey
Streptomyces 8KK1	Pale tan	Not present
Streptomyces 10KK1	Pale yellow	White
Streptomyces 22KA1	Pale brown	White and Grey
Streptomyces 23KA1	Dull orange	White
Streptomyces 25KA1	Transparent white	Grey
Streptomyces 26KA1	Pale yellow	White and Grey
Streptomyces 28KA1	Pale brown	White, Grey and Black
Streptomyces 31KA1	Pale yellow	White and Grey
Streptomyces 32KA1	Pale brown	White and Grey
Streptomyces 39KA2	Transparent white	White
Streptomyces 40KA2	Pale yellow or Pale brown	White and Grey
Streptomyces 41KA2	Transparent white	White and Grey
Streptomyces 42KA2	Transparent white	White
Streptomyces 52KA3	Pale brown	White and Grey
53KA3	NG	-
Streptomyces 54KA3	Pale brown	White and Grey
Streptomyces 55KA3	Pale yellow	White and Grey
57KA3	NG	-
Streptomyces 58KA3	NG	NG
Streptomyces 59KA3	Burnt orange	White
Streptomyces 60KA3	pale yellow	White
Streptomyces 61KA3	Pale brown	White and Grey
Streptomyces 63KA3	Bright yellow	Yellow
Tennis Court Sample		
Streptomyces 5TK1	Pale yellow	White and Grey
Streptomyces 6TK1	Pale yellow	White
Nonomuraea 7TK1	NG	-
Streptomyces 9TK1	Pale yellow	White and Grey
Streptomyces 12TA1	n/a	n/a
Streptomyces 13TA1	Dark grey	White
Streptomyces 15TA1	Pale yellow	White and Grey
Streptomyces 16TA1	Pale yellow	-
Streptomyces 18TA1	Pale yellow	White and Grey
Streptomyces 19TA1	Pale yellow	White
Streptomyces 20TA1	Pale yellow	White and Grey
Streptomyces 33TA2	Yellow	White
Streptomyces 36TA2	Pale yellow	White and Grey
Streptomyces 45TA3	Pale yellow	White, Grey and Black
Streptomyces 47TA3	Yellow	-
Streptomyces 48TA3	n/a	n/a
Streptomyces 49TA3	Pale Yellow	White/Grey

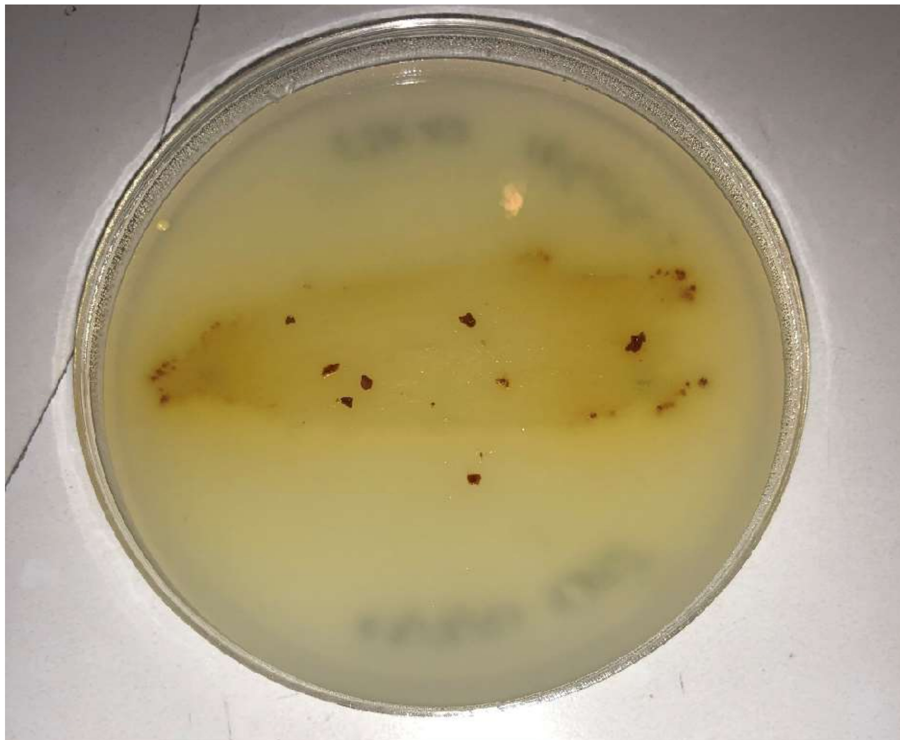

Figure A1: Image of *Streptomyces* strain 63KA3 grown on ISP4 medium for 14 days at 30°C showing the production of a yellow diffusible pigment.