Phenotypic and genomic insights into ultraviolet resistance of *Arthrobacter* and *Pseudarthrobacter* isolated from desert soil

Elizabeth Buckley

A thesis submitted to

Auckland University of Technology

in fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

2020

School of Science

Abstract

Soil microbial communities are important contributors to the desert ecosystem. Top soil is vital for biological activity but is often exposed to large amounts of DNA damaging ultraviolet (UV) radiation. In this thesis, soil bacteria from the Antarctic Dry Valleys and the Namib Desert were investigated for their resistance and sensitivity to UV radiation. The primary objective of this research was to provide genomic insights into how soil bacteria in arid deserts may survive UV radiation. To achieve this objective, abiotic drivers of the bacterial community were investigated, UV resistant bacteria were identified via a rapid screening technique, and whole genome analysis using comparative genomic approaches and predicted protein tertiary structures was carried out.

In the first part of this study, the environmental DNA (eDNA) of each desert was investigated to examine the overall bacterial community. This was done using Illumina based 16S rRNA gene-defined community diversity; for this analysis the V3 and V4 hypervariable regions of the 16S rRNA gene were targeted. Actinobacteria was the most abundant phylum in most of the desert locations. The soil chemistry for the Dry Valleys and the Namib Desert was mostly within the range of published values, except for cation-exchange capacity in the Namib Desert. This range appeared to be influenced by the high concentration of calcium found in the Namib Desert soil.

Bacteria isolated from desert soil were exposed to UV radiation using a modified plate drop method for rapid screening of UV resistance or sensitivity. A total of 285 bacterial isolates were tested on solid growth media. All isolates survived exposure to UVA and 35 also survived 10 minutes of 15 W/m^2 UVB radiation. In addition, through this method, 16 of the 285 isolates were deemed resistant to 5 W/m^2 UVC radiation and 10 were deemed sensitive to 5 W/m^2 UVC radiation. Sanger sequencing of the 16S rRNA region using the 27F and 1492R primers identified four genera; *Arthrobacter*, *Pseudarthrobacter*, *Pseudomonas* and *Stenotrophomonas*, that contained isolates both resistant and sensitive to 5 W/m^2 of UVC radiation. As Actinobacteria was the most abundant phylum identified in the 16S rRNA gene-defined communities, the UV resistance and sensitivity of *Arthrobacter* and *Pseudarthrobacter* was further investigated at the genomic level.

Comparative analyses of the draft genomes of *Arthrobacter* and *Pseudarthrobacter* isolated in this study indicate that the isolates were able to reduce nitrate to nitrite, indicating that they may play a role in the nitrogen cycle. The genomes were also predicted to reduce sulphate to sulphite, indicating the isolates may contribute to the sulphur cycle within soil. Computational genome comparative tools such as OrthoANI and *in silico* DNA-DNA hybridisation (DDH) indicate that all four isolates selected in this study are genetically distinct from the *Arthrobacter* and *Pseudarthrobacter* reference genomes and may be new species.

Comparative genome analysis of four isolates revealed the two UVC resistant bacteria shared the gene cytochrome P450 (CYP), a gene that was absent from both the UVC sensitive isolates. Analysis of the genome of the UVC sensitive isolates revealed that the CYP gene was absent. It is theorised in this thesis that the presence of CYP may help the UVC resistant isolates identified in this study survive all three types of UV radiation by activating CYP as a DNA repair enzyme. Conversely, several DNA repair genes involved in base excision, nucleotide excision and recombinational repair pathways were present in both the UV resistant and UVC sensitive genomes, as well as several reference genomes. This indicates the presence or absence of the gene products do not have a role in expression of the UV resistant phenotype; however, the expression and function of the gene products may have a role. To infer function of these proteins, the predicted secondary and tertiary structures were compared within the *Arthrobacter* and *Pseudarthrobacter* isolates from this study. Protein alignment and analysis of the predicted tertiary structure of the isolates' UvrABC proteins revealed that, while the UvrA1, UvrA2a and UvrB proteins appeared similar, the UvrC protein for one of the sensitive isolates appeared to be non-functional.

Overall, this study has found that Actinobacteria were one of the most abundant phyla in desert soil from Namibia and the Dry Valleys, however, the bacterial community did not appear to be driven by soil chemistry. This study has also produced cultured bacterial isolates that can survive 10 minutes of 5 W/m^2 UVC, and therefore represent an important platform for more in-depth studies of UV resistance and sensitivity within Actinobacteria. Comparative genomics showed differences in UV resistance genes between the UVC resistant and UVC sensitive isolates, most notably the CYP gene. Finally, the predicted tertiary protein structures showed that the UvrA and UvrB proteins appeared to be conserved between the isolates, while the UvrC protein appeared to be less conserved, with this protein appearing to be non-functional in one UVC sensitive isolate.

The draft genome sequences of these isolates provide a resource for further investigations into *Arthrobacter* and *Pseudarthrobacter* and possible physiological attributes that enable their survival in arid desert locations. Furthermore, this study provides further avenues of investigation into UV resistance genes in bacteria isolated from harsh environments.

Table of Contents

List of Figures

List of Tables

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning

Signed

Elizabeth Buckley

Date: 21 January 2020

Acknowledgements

It is my pleasure to acknowledge and extend my gratitude to those that have supported me throughout my thesis.

Firstly, I would like to express my sincerest gratitude to my primary supervisor Dr Brent Seale. Thank you for accepting me for this project and for encouraging me through the many, many tough times. Even though this project wasn’t quite hedgehogs, I am so appreciative for all the time and effort you have invested into me these past several years. Without your guidance I would never have made it through. Your calm demeanour never failed to alleviate my lack of confidence in my own abilities. Slowly but surely you have shown me what I am capable of. I am so appreciative of your help and kind ear.

I would also like to thank my secondary supervisor Dr Colleen Higgins. Thank you for your support, patience and motivation. You have always been so encouraging and it has been such an honour to be welcomed into your fold when I had nowhere else to go. Your confidence in me has helped me to flourish and improve my skills, even when I was a little resistant at times. In addition to all this, you have taken the time to help me grow as a person, and for that I am truly humbled and grateful.

I would like to express my deepest thanks to my third supervisor Dr Kevin Lee. Your knowledge on everything bioinformatics is truly mind boggling. I have learnt so much from you in such a short space of time. Thank you for always making time for me and for patiently explaining things again and again until I understood. I am so grateful for all your experimental support and help. This project would not have been able to float without you.

My sincere appreciation goes to our Applied Ecology New Zealand team for the collection of the soil for this thesis. Thank you to Dr Stephen Archer and Dr Barbara Breen for bringing these precious samples to me. Thank you to the AUT staff and technicians, especially Timothy Lawrence for running my bacterial community on the MiSeq. I would like to extend special thanks to AUT for funding this research project.

I would like to thank the Molecular Genetics Research Group: you have all given me a home within molecular research. Thank you to Priya and Dr Gardette Valmonte-Cortes for always providing valuable suggestions and ideas. Thank you to Zoe, Shweta, Lee and Toni for keeping me sane, letting me vent my feelings and then pop me back on my feet with words of kindness and encouragement. Your support has been invaluable, and I cannot express how much this has meant to me.

Thank you to my friends outside of university. Kate, Alicia and Shannon, I would have lost my mind long ago without you all!

Thank you to my family for your endless support. Thank you to my brother Nicholas for words of encouragement when I needed them. Thank you, mum and dad, you are two amazing humans. You have supported me at every turn and continuously reminded me that hard work is always hard but is also always rewarding. I could not have done this without you.

Finally, thank you to Jamie. You have been so supportive through this whole journey. You have encouraged me to keep going when I wanted to stop. You have sat through every break down and celebrated every victory with me. I am forever grateful for your endless patience, love and understanding. I know it has been hard, but I am so happy to have had you along for the ride.

Abbreviations

aa	Amino acid
bp	Base pair
°C	Degrees Celsius
dNTP	Deoxynucleotide
ddNTP	Dideoxynucleotide
g	Grams
g	Gravity
Gb	Gigabase
kg	Kilograms
km	Kilometres
M	Molar
Mb	Mega base pair
mg	Milligrams
min	Minute/s
mL	Millilitre
mM	Millimolar
MPa	Megapascal
ng	Nanograms
nm	Nanometres
nt	Nucleotide
sec	Second/s
µg	Microgram
µL	Microlitre
W/m^2	Watts per metre squared
6-4PP	Pyrimidine (6-4) pyrimidone photoproducts
ANI	Average nucleotide identity
AP	Apurinic/apyrimidinic
APS	Adenosine 5'-phosphosulfate
ASV	Amplicon sequence variant
ATP	Adenosine triphosphate
BER	Base excision repair
BLAST	Basic Local Alignment Search Tool
CDS	Coding sequences
CEC	Cation exchange capacity
CFU	Colony forming units
COG	Cluster of Orthologous Genes
CPD	Cyclobutane pyrimidine dimers
CRISPR	Clustered regularly interspaced short palindromic repeats
CYP	Cytochrome P450
DDH	DNA-DNA hybridisation
DGDG	Digalactosyldiacylglycerol
DMG	Dimannosylglyceride
DNA	Deoxyribonucleic acid
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
EC	Enzyme commission
eDNA	Environmental DNA
EMBL	European Molecular Biology Laboratory

GGDC	Genome-to-Genome Distance Calculator
HhH	Helix-hairpin-helix
HMM	Hidden Markov model
IMViC	Indole, methyl red, Voges-Proskauer and citrate
KEGG	Kyoto Encyclopaedia of Genes and Genomes
LB	Luria-Bertani
LCB	Locally collinear blocks
MCL	Markov Clustering Algorithm
MCM	Mauve Contig Mover
MGDG	Monogalactosyldiacylglycerol
ML	Maximum likelihood
MMR	Mismatch repair
NA	Nutrient agar
NB	Nutrient broth
NCBI	National Center for Biotechnology Information
NER	Nucleotide excision repair
NIWA	National Institute of Water and Atmospheric Research
NGS	Next generation sequencing
NJ	Neighbour joining
NTC	No template control
OG	Orthologous groups
ONT	Oxford Nanopore Technologies
OTU	Operational taxonomic unit
PCoA	Principal coordinates analysis
PCR	Polymerase chain reaction
PI	Phosphatidylinositol
PL1-5	Unidentified phospholipid
PP_i	Inorganic pyrophosphate
R2A	Reasoner’s 2A agar
RAST	Rapid Annotation using Subsystems Technology
Res	Resistant [to radiation]
RDA	Redundancy analysis
ROS	Reactive oxygen species
rRNA	Ribosomal ribonucleic acid
Sen	Sensitive [to radition]
SMRT	Single molecular real-time
sp./spp.	Species (singular)/species (plural)
TCA	Tricarboxylic acid cycle
TLS	Translesion DNA synthesis
TMDG	Trimannosyldiacylglycerol
TSA	Tryptic soy agar
WGS	Whole genome shotgun
UNEP	United Nations Environment Programme
UV	Ultraviolet
UVA	Ultraviolet A
UVB	Ultraviolet B
UVC	Ultraviolet C
ZMW	Zero-mode waveguide

Chapter 1: General Introduction

1.1 Rationale and Significance of the study

Solar radiation is a major mutagen that causes deoxyribose nucleic acid (DNA) damage through the formation of photoproducts between adjacent thymine, or cytosine bases (García-Gómez et al. 2012; Matallana-Surget et al. 2008; Matallana-Surget and Wattiez 2013). Solar radiation is comprised of a broad range of damaging radiations, including gamma radiation and ultraviolet (UV) radiation. Gamma rays do not reach Earth's surface (Matallana-Surget and Wattiez 2013), however, larger wavelengths of UV radiation do. Exposure of microorganisms to high frequency UV radiation leads to direct and indirect damage to the cell (García-Gómez et al. 2012; Matallana-Surget and Wattiez 2013). To survive this damage, microorganisms require mechanisms to remain active during, and following, UV exposure. Some microorganisms are known to produce UV-protective pigments, which lessens the UV that reaches the cell (Cary et al. 2010; Makhalanyane et al. 2015; Pointing and Belnap 2012), while others possess damage repair pathways (García-Gómez et al. 2012; Matallana-Surget et al. 2008; Matallana-Surget and Wattiez 2013; Sinha and Häder 2002). Irreparable DNA damage will occur when the capacity of the microorganism's mechanisms are strained, absent, or insufficient to reverse the damage caused by UV (García-Gómez et al. 2012; Matallana-Surget and Wattiez 2013). The biological effect of UV radiation depends on the rate of radiation-induced damage, the microorganism's efficiency of cell protection, and the efficiency of DNA damage reparation (Matallana-Surget and Wattiez 2013). It is therefore essential for the microorganism to repair DNA damage efficiently to ensure survival (García-Gómez et al. 2012; Morales-Ruiz et al. 2006; Roldán-Arjona and Ariza 2009; Schimel et al. 2007).

Microorganisms that have adapted to grow optimally in environmental conditions that are considered "extreme" are collectively known as extremophiles. Extremophiles are able to survive harsh environmental conditions such as high and low pressure (up to 110 MPa), extreme acidic (pH 0) and basic conditions (pH 12.8), high and low temperatures (below 15°C or above 45°C), ionising (gamma) and non-ionising (UV) radiations (Rampelotto 2013). Extremophiles in these conditions must have efficient repair mechanisms to deal with the constant stress of the environment. "Stress" to a microorganism is considered to be a physiological challenge that adversely affects microbial growth or survival (Schimel et al. 2007). Under damaging UV exposure, microorganisms must adapt to sudden stress by reallocating resources for growth into survival pathways (Matallana-Surget and Wattiez 2013; Schimel et al. 2007). Some microorganisms may be able to inherently resist minor stress but must actively adapt as this stress increases in intensity (Schimel et al. 2007; Walker et al. 2006; Wallenstein and Hall 2012).

Bacteria present a wide variety of tolerance mechanisms to damaging radiation, and are the simplest model organisms for studying stress responses in terms of gene regulation (Matallana-

Surget and Wattiez 2013). Understanding bacterial radiation responses is an important component of many research areas, ranging from microbial ecology, decontamination of water using UV, and astrobiology for future exploratory space missions (Matallana-Surget and Wattiez 2013). Developing a stronger understanding of how microorganisms have adapted to high stress environments, such as exposure to DNA damaging UV, will lead to greater insight of physiological responses to stress.

Research into the environmental effects of UV on terrestrial ecosystems has been widely investigated since the discovery of the ozone layer depletion over Antarctica (Dib et al. 2008). The resulting increase of UV reaching the Earth's surface has fuelled interest into the various effects this might have on microorganisms. The prevalence of UV increases with elevated altitudes, as well as in desert environments (Cordero et al. 2014). This consequently produces an environment that is restrictive to the growth of most living organisms (García-Gómez et al. 2012). Ozone depletion coupled with arid environments results in high levels of UV radiation and an adverse environment for microbial growth. An increase in surface UV radiation may lead to damaging effects on terrestrial ecosystems (Cordero et al. 2014).

1.1.2 Previous studies

Previous studies have isolated microorganisms, and identified community structures, from several global desert locations (Deng et al. 2014; Dib et al. 2008; Lacap et al. 2011; Pointing et al. 2009). However, there is limited research into the isolation and UV resistance screening of organisms from soil (Table 1.1). The development of the rapid UV resistance screening method in this thesis, using a modified drop plate approach, will help to quickly identify UV resistant organisms for future research.

As shown in Table 1.1, there is a limited number of studies that have isolated and screened specific isolates directly from soil for UV resistance.

Table 1.1: Previous studies related to UV radiation exposure of soil bacteria.

Bacterium	Location	Radiation tested	Reference
Nesterenkonia sp. Act20	Laguna Socompa soil, Andean Lakes, Chile	UVB	Albarracín et al. (2013)
Chroococcidiopsis sp.	Negev Desert, Israel	UVC	Cockell et al. (2008)
Bacillus sp. *Pseudomonas stutzeri*	Atacama Desert, Chile	UVC	Paulino-Lima et al. (2013)
Deinococcus sp.	Dry Valleys, Antarctica	UVC	Hirsch et al. (2004)
Curtobacterium flaccumfaciens *Frigoribacteria* sp. *Arthrobacter* sp. *Geodermatophilus obscurus* *Cellulomonas* sp.	Whipple Mountains, California, USA	UVC	Kuhlman et al. (2005)
Deinococcus gobiensis I-0	Gobi Desert, China	UVC	Yuan et al. (2012)
Deinococcus deserti VCD115	Sahara Desert	UVC	de Groot et al. (2009)
Hymenobacter sp. DG25A	Seoul, South Korea	UVC	Srinivasan et al. (2017)
Hymenobacter sp. A9A5 (A) *Chryseobacterium* sp. A9A5 (A)	King George Island, Antarctica	UVC	Órdenes-Aenishanslins et al. (2016)
Arthrobacter sp. MN05-02	Sonoran Desert, Arizona, USA	UVC	Ii et al. (2019)

Expanding our current understanding of the global distribution of UV resistant isolates will highlight genetic differences within isolated strains. Investigations into UV resistance and DNA repair in microorganisms from arid locations is important to continue our understanding the diversity and adaptation of repair systems within desert environments. Research in this area will provide further insight into the relationship between microbial repair systems and cellular resistance to DNA damage.

1.2 Aims of this study

Two arid locations have been selected for this thesis; the Antarctic Dry Valleys, and the Namib Desert. Several studies have been conducted into the microbial community structure of these desert locations (Cary et al. 2010; Cowan et al. 2014; Wei et al. 2016), and UV radiation has been highlighted as a major stressor for these communities (Chan et al. 2012; Cowan et al. 2010; Pointing and Belnap 2012). Despite this, very few studies have screened desert soil isolates for UV resistance (Table 1.1). To the authors knowledge, UV screening has not been conducted on soil from the Namib Desert previously.

The purpose of this thesis is to identify soil microorganisms from arid environments that can survive high levels of UV radiation, with special interest in short wavelength resistant organisms. The following are the aims of this project:

Aim 1: Investigate bacterial communities present at each sample site and potential abiotic drivers for each community.

Objective 1.1 Determine chemical composition of the soil and investigate if the chemistry composition is an abiotic driver for community structure.

Objective 1.2 Investigate the community structure of the sample sites using 16S rRNA gene-defined community diversity.

Objective 1.3 Investigate abiotic drivers of community structure of the Dry Valleys and Namib Desert

Aim 2: Identification and analysis of UV resistant organisms.

Objective 2.1 Isolate aerobic heterotrophic bacteria from soil from the Dry Valleys and the Namib Desert using culture-based methods.

Objective 2.2 Develop and use a rapid screening method to identify UV resistant bacteria using different wavelengths, intensities and exposure times.

Objective 2.3 Use Sanger sequencing to identify UVC resistant and UVC sensitive organisms based on their 16S rRNA gene.

Objective 2.4 Analyse the phylogenetic relationships between the UVC resistant and sensitive organisms, and related species using the 16S rRNA gene.

Aim 3: Whole genome analysis of the UVC resistant and sensitive organisms identified in Aim 2 (Chapter 5).

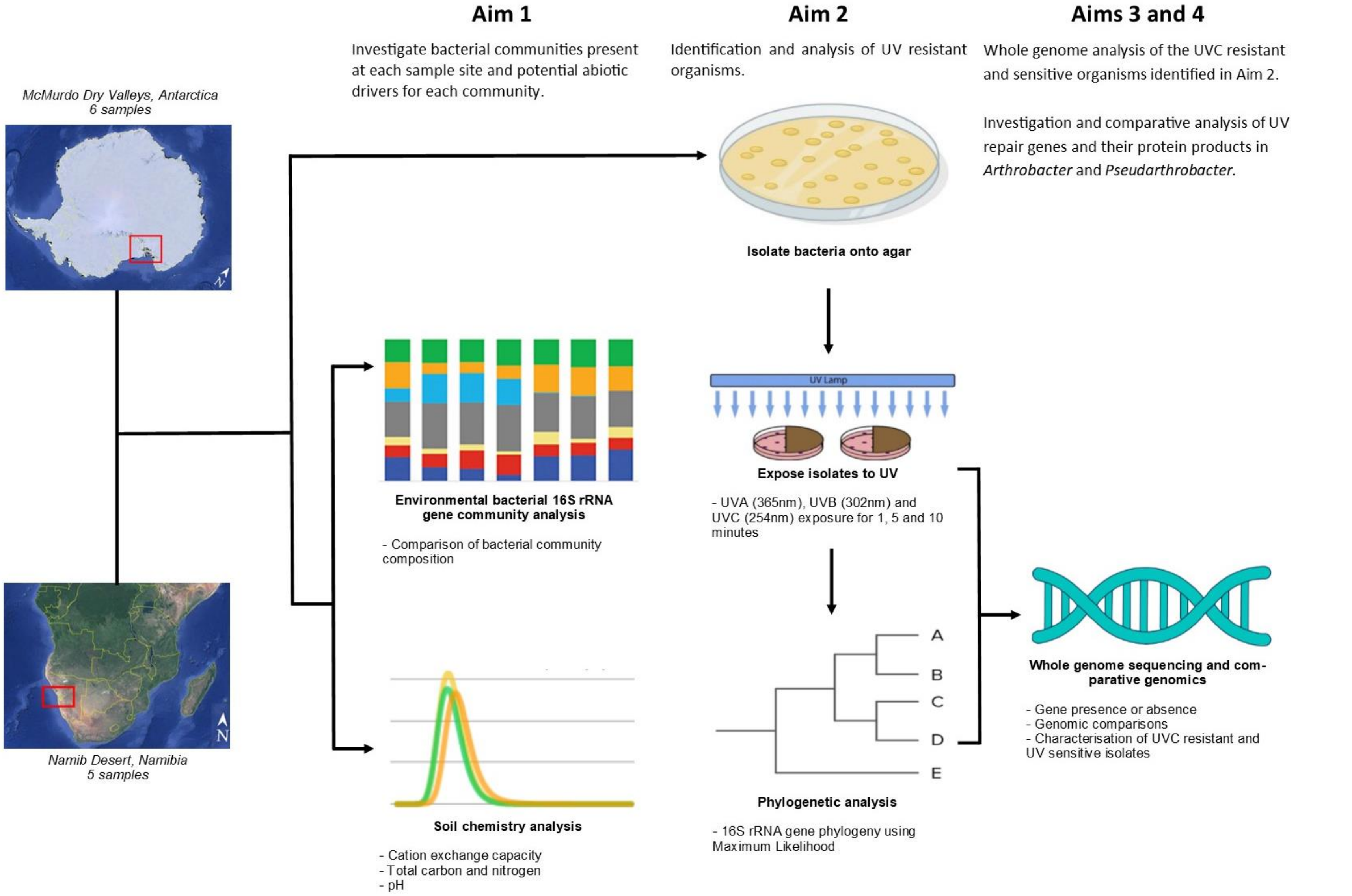
Objective 3.1 Extract genomic DNA from closely related UVC resistant and UVC sensitive organisms selected in Objective 2.3, and whole genome sequence using the Illumina HiSeq™.

Objective 3.2	*De novo* assembly of the genome and functional annotation.
Objective 3.3	Characterisation of organisms using comparative genomic approaches with publicly available reference genomes.

Aim 4: Investigation and comparative analysis of UV repair genes and their protein products in *Arthrobacter* and *Pseudarthrobacter* (Chapter 6).

Objective 4.1	Identify gene elements associated with UV resistance.
Objective 4.2	Comparative analysis of predicted tertiary protein structures encoded by elements identified in Objective 4.1.

The schematic overview for this thesis can be seen below in Figure 1.1.



Aim 1
Aim 2
Aims 3 and 4
McMurdo Dry Valleys, Antarctica
6 samples
N
Namib Desert, Namibia
5 samples
N
Isolate bacteria onto agar
UV Lamp
Expose isolates to UV
Environmental bacterial 16S rRNA gene community analysis
Soil chemistry analysis
A
B
C
D
E
Phylogenetic analysis
Whole genome sequencing and comparative genomics

Investigate bacterial communities present at each sample site and potential abiotic drivers for each community.

Identification and analysis of UV resistant organisms.

Whole genome analysis of the UVC resistant and sensitive organisms identified in Aim 2.

Investigation and comparative analysis of UV repair genes and their protein products in *Arthrobacter* and *Pseudarthrobacter*.

- Comparison of bacterial community composition

- Cation exchange capacity
- Total carbon and nitrogen
- pH

- UVA (365nm), UVB (302nm) and UVC (254nm) exposure for 1, 5 and 10 minutes

- 16S rRNA gene phylogeny using Maximum Likelihood

- Gene presence or absence
- Genomic comparisons
- Characterisation of UVC resistant and UV sensitive isolates

Figure 1.1: Schematic overview of thesis structure and Aims. Aim 1 will be covered in Chapter 3, Aim 2 will be covered in Chapter 4, Aim 3 will be covered in Chapter 5 and Aim 4 will be covered in Chapter 6.

Chapter 2: Literature Review

Ultra violet (UV) is an important abiotic stressor in arid environments (Makhalanyane et al. 2015; Pointing and Belnap 2012). This review starts with a brief introduction into arid environments and a description of the importance of UV as an abiotic stressor to microorganisms in the two sampling locations for this thesis. This review then describes what is currently known about bacterial DNA repair mechanisms, as well as the limitations of our knowledge due to traditional culturing restrictions. Finally, metagenomic processes are discussed with emphasis on how whole genome comparisons will improve current understanding of microbial stress adaptation and survival.

2.1 Global hyper-arid environments

Deserts are terrestrial regions that become arid due to little rainfall, and are often restrictive environments due to many abiotic stressors such as limited water availability and lack of plant life (Lacap et al. 2011). Approximately one third of Earth's biomes are desert environments, covering a total land surface area of approximately 3.4 x 10^7 km^2 (Chan et al. 2012; Makhalanyane et al. 2015). Due to continuous trends in climate change, it is estimated that this area will increase, leading to potentially severe environmental consequences on soil fertility and biogeochemical cycles of carbon, nitrogen and phosphorus (Makhalanyane et al. 2015).

The definition of a desert is complicated, as there is no agreement on what classifies a desert. A key basis for desert sub-classification is the low level of precipitation within the environment (Makhalanyane et al. 2015). A desert can be defined using direct meteorological observations to determine aridity; a ratio of precipitation to potential evapotranspiration (P/PET) of less than 1 constitutes a desert environment (UNEP 1992). From this, a further four zones of aridity can be characterised, as seen in Table 2.1. Figure 2.1 shows the global distribution of hot arid regions.

Table 2.1: Subtypes of deserts classified by aridity as defined by UNEP (1992).

Subtype	Aridity Index (P/PET)
Dry sub-humid	0.50-0.65
Semi-arid	0.20-0.50
Arid	0.05-0.20
Hyper-arid	<0.05

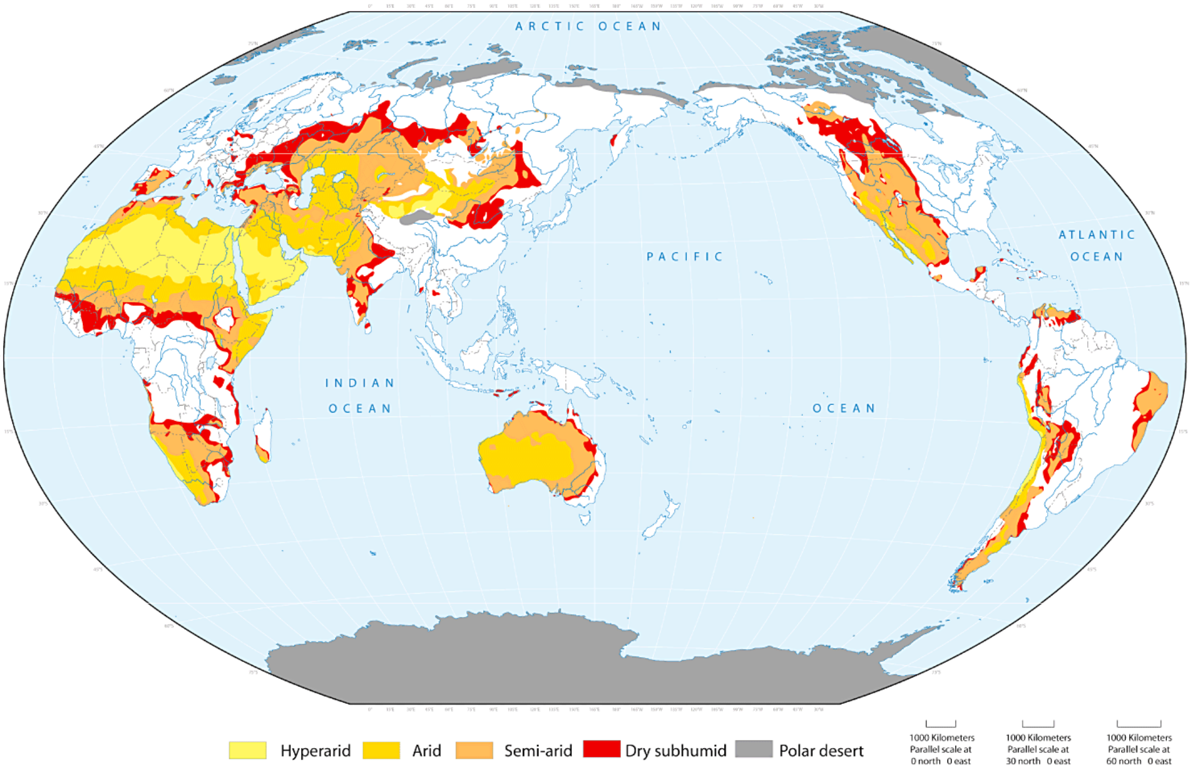

ARCTIC OCEAN
PACIFIC
OCEAN
ATLANTIC OCEAN
INDIAN OCEAN
Hyperarid
Arid
Semi-arid
Dry subhumid
Polar desert
1000 Kilometers Parallel scale at 0 north 0 east
1000 Kilometers Parallel scale at 30 north 0 east
1000 Kilometers Parallel scale at 60 north 0 east

Figure 2.1: Global distribution of arid regions (Chan et al. 2012).

As environmental aridity increases, the abundance of plant life decreases, and microbial communities in soil become essential for carbon, nitrogen and phosphorus cycling (Chan et al. 2012; Johnson et al. 2017; Lacap et al. 2011; Makhalanyane et al. 2015; Pointing and Belnap 2012). The surface layer of soil in desert landscapes is considered a critical zone for biological activity (Pointing and Belnap 2012), because biodiversity and activity rapidly decline in subsurface layers (Makhalanyane et al. 2015; Stomeo et al. 2013).

Soil microbial communities are strongly influenced by environmental factors, such as temperature, desiccation and UV (Pointing and Belnap 2012; Rampelotto 2013). It is predicted that global surface temperatures will rise by 2 – 6°C over the coming decades (Meehl et al. 2007), which could lead to dramatic shifts in the structure and function of terrestrial microbial communities. Several studies have indicated that increases in temperature due to climate change may directly affect the microbial community composition in arid regions (Garcia-Pichel et al. 2013; Reed et al. 2012). This can lead to large biological soil crust community alterations, and a loss in microbial diversity and soil functionality as a result (Hooper et al. 2012; Mace et al. 2012; Maestre et al. 2013; Philippot et al. 2013). The disturbance of arid microbial communities may therefore be regarded as an important contributor to the desertification process (Makhalanyane et al. 2015; Pointing and Belnap 2012). From an ecological perspective, it is therefore essential to improve understanding of microbial community systems in arid soils, due to our rapidly changing climate. Understanding microbial diversity and adaptation in arid environments may improve conservation and rehabilitation efforts, as well as sustainable land management practices (Makhalanyane et al. 2015).

2.1.1 UV as an abiotic stressor in arid environments

Harsh environmental conditions such as large temperature fluctuations (freeze-thaw cycles), low water availability and non-ionising UV radiations can have a severe impact on microorganisms (Pointing and Belnap 2012; Rampelotto 2013). Abiotic stressors, such as exposure to UV radiation, can cause microbial DNA to suffer considerable and often irreparable damage (García-Gómez et al. 2012). There are three types of UV rays; UVA (315-400 nm), UVB (280-315 nm), and UVC (100-280 nm). Most hyper-arid deserts have a high incidence of UVA and UVB reaching the desert surface, due to the absence of cloud cover (Cordero et al. 2014). UVC is mostly blocked by the troposphere and does not reach the Earth's surface (Diffey 2002). Earth's surface UV depends on the total ozone column, local aerosols, altitude and Sun-Earth distance (Cordero et al. 2014; Rampelotto 2013).

The McMurdo Dry Valleys (hereafter Dry Valleys) in Antarctica are classified as a hyper-arid desert, and the numerous environmental stressors include extremely low temperatures, oligotrophic soils, desiccation, high levels of salinity, and high irradiation incidence (Cary et al. 2010; Chan et al. 2013; Pointing et al. 2009; Van Goethem et al. 2016; Wei et al. 2016). A relatively high occurrence of biodiversity and productivity occur at the Antarctic lakes and melt water ponds (Archer et al. 2015; Wei et al. 2016), however, it is the terrestrial landscape that supports the greatest microbial ecosystem in the Dry Valleys (Cary et al. 2010; Cowan et al. 2014; Van Goethem et al. 2016; Wei et al. 2016). The presence of the ozone hole above Antarctica has fuelled interest into the environmental effects that increased UV may have on microorganisms (Dib et al. 2008). The restrictiveness of this environment makes it an ideal model system for assessing how adverse conditions help shape microbial community diversity and functional processes (Cowan et al. 2014; Van Goethem et al. 2016).

It has previously been noted that the microbial community structure in the Dry Valleys develops in such a way as to avoid UV radiation (Cary et al. 2010; Cowan et al. 2010; Pointing and Belnap 2012). Cowan et al. (2010) previously noted that bacterial hypolithic communities colonise in such a way as to reduce the amount of UV radiation that is transmitted through the rocks. Cowan et al. (2010) also noted that nearly all UVB was filtered out of the hypolithic communities, due to effective exclusion of damaging UVB radiation being a driver for community colonisation. UVB irradiance in Antarctica has previously been recorded as reaching up to 4 W/m^2 (Lud et al. 2002), while the UVB reaching hypolithic communities is between 0-0.1% of outside UVB (Cowan et al. 2010). Previous reports have not mentioned the presence of UVC reaching the surface of Antarctica. However, during soil collection in January 2017 for the present study, UVC irradiance was recorded as 1.07 W/m^2 (UVC radiometer, UVP Inc, Upland, CA).

The Namib Desert is another hyper-arid region that extends 2,000 km along the West Coast of Namibia and is one of the oldest and driest deserts in the world (Goudie and Viles 2014; Johnson et al. 2017; Stomeo et al. 2013). Due to the hyper-aridity of the environment, many studies have focused on microbial communities in hypolithic niches (Büdel et al. 2009; Johnson et al. 2017; Makhalanyane et al. 2013; Prestel et al. 2008; Ronca et al. 2015; Scola et al. 2017). However, the survivability of isolates to UV radiation in the Namib Desert has not yet been investigated (Stomeo et al. 2013). Rainfall in the Namib Desert is extremely limited, however, the presence of fog throughout the year is thought to be the main contributing source of biologically accessible water for soil microorganisms in this location (Stomeo et al. 2013). While there is no ozone hole over the Namib Desert region, data collected by the NASA EOS Aura spacecraft shows that UVB radiation reaching the Western Coast of Africa is much higher than many other global locations (Beckmann et al. 2014; Lucas et al. 2016). The total UVB radiation reaching the surface of the Namib Desert fluctuates between 1 W/m^2 and 3.4 W/m^2 throughout the year (Lucas et al. 2016; NIWA 2016).

A summary of the general characteristics of the Dry Valleys and the Namib Desert can be seen below in Table 2.2.

Table 2.2: General characteristics of the Dry Valleys and the Namib Desert.

Name/Location	Approx. size (km^2)	Topography	Approx. temperature range (°C)	Approx. precipitation (mm/year)/ classification	Soil characteristics	UV irradiance range (W/m^2)	References
Dry Valleys/ Antarctica	4,800 [a, b]	Ice covered lakes, arid, rocky soils, ice-cemented soils [a]	-53.7 – 10 [a, b]	7-100 hyper-arid/arid [a]	pH: 6.5 – 9.4 [c, d] C (%): 0.01 – 0.03 [b] N (%): 0 – 0.12 [b, d]	UVA: 0 – 33.73 [e] UVB: 0 – 5.15 [e, f, g] UVC: 0 – 1.07 [h]	[a] Doran et al. (2002) [b] Goordial et al. (2016) [c] Cowan and Tow (2004) [d] Pointing et al. (2009) [e] NIWA (2016) [f] Lud et al. (2002) [g]Obryk et al. (2018) [h] Archer (2017)
Namib Desert/ Southwestern Africa	81,000 [h]	Gravel plains, sand dunes and plains [j]	5 – 45 [j, k]	5-100 hyper-arid/arid [k, l]	pH: 7.9-8.5 [j, k] C (%): 0.1-0.3 [k, l] N (%): 0.03-0.05 [k, l]	UVA: 39.35 – 65.72 [e] UVB: 1 – 3.4 [e, j] UVC: ND	[j] Lucas et al. (2016) [k] Makhalanyane et al. (2015) [l] Stomeo et al. (2013)

ND = no data

UV wavelengths: UVA = 315-400nm; UVB = 280-315nm; UVC = 100-280nm

2.1.2 Microbial composition of Dry Valley and Namib Desert soil

Desert microbiomes have been shown to be distinct from other soils in terms of microbial function and community composition (Fierer et al. 2012; Makhalanyane et al. 2015). While the diversity of soil communities within desert systems is lower than non-desert soil, the taxonomic diversity of desert soil is more diverse than previously thought (Fierer et al. 2012; Makhalanyane et al. 2015).

The soil in the Dry Valleys has been previously shown to be microbiologically distinct from other soils worldwide (Fierer et al. 2012). The Dry Valleys have a specific community structure that cycles carbon and nitrogen; an essential process due to the lack of plant life (Wei et al. 2016). The most abundant microorganisms found in the proposed sample sites of the Dry Valleys and the Namib Desert are bacteria of the Proteobacteria, Actinobacteria and Acidobacteria phyla (Cary et al. 2010; Chan et al. 2013; Makhalanyane et al. 2013; Pointing et al. 2009). Cyanobacteria are also traditionally found at the proposed sites (Chan et al. 2013; Makhalanyane et al. 2013; Pointing et al. 2009; Yung et al. 2014). Previous research in the Dry Valleys has primarily focused on the soil microbial biodiversity, and prominent bacterial populations of Actinobacteria, Proteobacteria, Acidobacteria and Cyanobacteria have been observed (Cary et al. 2010; Chan et al. 2013; Pointing et al. 2009; Yung et al. 2014).

Previous studies into the microbial ecology of soil from the Namib Desert have observed distinct populations dominated by the phyla Actinobacteria and Proteobacteria (Makhalanyane et al. 2013). The phyla Acidobacteria, Cyanobacteria, Bacteroidetes and Chloroflexi have also been detected, but in lower occurrence (Makhalanyane et al. 2013).

2.1.3 Drivers of desert microbial community structure

Edaphic microbial communities have been identified as important ecosystem drivers of the carbon and nitrogen biogeochemical cycles (Pointing and Belnap 2012; Wei et al. 2016; Yergeau et al. 2012). As such, investigations into the microbial communities of desert locations has been the focus of a considerable number of studies (Archer et al. 2017; Armstrong 2014; Pointing and Belnap 2012; Wei et al. 2016; Yergeau et al. 2012). Important factors that appear to influence soil diversity are organic carbon and soluble salts (Fierer et al. 2012; Johnson et al. 2017; Pointing et al. 2009). Water content, phosphorus, nitrogen and soil pH have also been shown to be influential factors in the shaping of soil communities (Johnson et al. 2017; Pointing and Belnap 2012).

Fierer et al. (2012) suggests that external environmental drivers, such as soil pH and organic carbon, have more influence over bacterial community structure than biological drivers, such as microbial competition. In contrast, Caruso et al. (2011) and Lee et al. (2019) both note that desert community structures rely on a combination of deterministic and stochastic forces. Lee et al. (2019) goes on to note that while topography and soil temperature had direct effects on species richness in soil communities, biotic interactions within the soil were also important contributors to microbial diversity. This highlights the need for continued investigations into desert community structure, and the drivers of those communities.

More recently, it has been demonstrated that the manganese/iron intracellular ratio correlates with UVC resistance in isolated cultures of *Arthrobacter* (Paulino-Lima et al. 2016). Intracellular manganese however, does not correlate with UVC resistance (Paulino-Lima et al. 2016). The intracellular manganese/iron ratio within the soil may be an important community driver within these high UV incidence environments.

2.1.4 Soil chemistry of Dry Valley and Namib Desert soil

Previous studies into the soil chemistry of soil from the Dry Valleys have found that the soils usually have low levels of carbon and nitrogen and that the pH range of the soil varies depending on whether the soil is inland or costal (Bockheim 2008; Cannone et al. 2008). Soil pH in the Dry Valleys is variable (Aislabie et al. 2008). Soil pH is influenced by the available salts in the soil. Soils near the coastline will have chlorides, while inland soils contain more nitrate and sulphate. While the ornithogenic soils of the Dry Valleys display high values for carbon, nitrogen and phosphorus, mineral soil of the Dry Valley region shows low levels of these chemicals (Aislabie et al. 2008; Ugolini and Bockheim 2008).

The Namib desert has high variation of nutrient and chemicals within the soil (Ronca et al. 2015). Dunes have a higher sodium content due to marine fog deposits; a common source of moisture in the Namib Desert (Ronca et al. 2015). Gravel planes in the Namib Desert have low organic carbon availability and usually have a slightly alkaline pH (Frossard et al. 2015). Calcium and potassium levels are usually found at high levels within gravel planes (Armstrong 2014; Frossard et al. 2015).

Current research around desert microbial communities is primarily focused on hypoliths (Chan et al. 2012; Lacap-Bugler et al. 2017; Makhalanyane et al. 2013; Pointing 2016; Wei et al. 2016). Hypolithic colonisation is viewed as a UV avoidance strategy, where the overlying mineral substrate provides protection from UV radiation (Chan et al. 2012; Cowan et al. 2010; Pointing et al. 2009). However, isolates from open soil provide a better understanding of how microorganisms survive UV radiation. It is suspected that the synthesis of pigments, such as scytonemin, carotenoids and mycosporine-like amino acids by Cyanobacteria, provide a greater barrier against UV induced damage and protect other microorganisms in the community (Cary et al. 2010; Makhalanyane et al. 2015; Pointing and Belnap 2012). Hypo- and endolithic communities often form a layered community whereby the photoprotective Cyanobacteria form an upper layer, which protects the underlying layers which are less UV tolerant (Pointing and

Belnap 2012). This thesis will focus only on microbes found in open soil environments, not hypo or endolithic communities.

Microbial stress responses to UV have been extensively reviewed in model organisms such as *Escherichia coli* and *Deinococcus radiodurans*, but the microbial tolerance to UV from isolates in open soil desert environments has been largely overlooked (Pointing and Belnap 2012). This requires further research to improve our understanding of high-UV tolerant organisms from arid locations, and their overall abundance within the community structure.

The Dry Valleys and the Namib desert are two ecologically important locations for desert studies. Both locations offer a unique insight into how microorganisms can survive arid and highly restrictive environments. A comparative study of these two arid locations and the biotic and abiotic drivers of the community structure will help to improve understanding of how desert microbial communities develop. Climate change can impact on the carbon, nitrogen and phosphorus biogeochemical cycles, which in turn can damage the delicate community balance (Makhalanyane et al. 2015). Loss of important microbial community members such carbon and nitrogen cyclers, as well as UV ‘blocker’ microorganisms, would have a dramatic impact on the desert.

2.2 UV radiation

Ozone-depleting chemicals, such as chlorofluorocarbons, along with increased greenhouse gas concentrations have the potential to alter the spatial distribution of ozone within the stratosphere and troposphere. This influences the amount of UV radiation reaching the Earth's surface (Coelho et al. 2013). There are three major subgroups of UV rays; UVA (315-400 nm), UVB (280-315 nm), and UVC (100-280 nm) (Diffey 2002).

UVA most often reaches the Earth's surface, accounting for approximately 95% of UV at ground level while UVB constitutes the remaining 5% (Diffey 2002; Murray et al. 2015). UVC is absorbed by stratospheric gasses, primarily oxygen and ozone, therefore often fails to reach the troposphere (Coelho et al. 2013; Diffey 2002; Horneck et al. 2010; Kurth et al. 2015). While negligible amounts of radiation shorter than 290 nm reach the Earth’s surface, interest in the field of astrobiology has lead several studies to test the resistance of specific organisms against high doses of UVC radiation (Abrevaya et al. 2011; Schuerger et al. 2003). The values of UV radiation reaching the surface of Earth, compared to the Earth’s orbit layer is shown in Table 2.3.

Table 2.3: Average solar irradiances at the Earth’s surface and Earth’s orbit

Spectral ranges (nm)	Earth’s surface W/m^2	Earth’s orbit W/m^2
UVA (315 – 400)	40.68 – 65.9* [a,b]	89.280 [c,d]
UVB (280 – 315)	0 – 3.4* [a,b]	19.490 [c,d]
UVC (200 – 280)	0	7.390 [c,d]

*UVA and UVB averages fluctuate largely depending on season, altitude and hemisphere (Cordero et al. 2014).
[a] Pehnec et al. (2009)
[b] NIWA (2016)
[c] Schuerger et al. (2003)
[d] Abrevaya et al. (2011)

UV radiation damages DNA, with detrimental or lethal effects on the life of a microorganism. UVB and UVC are the most biologically damaging to living cells due to their shorter wavelengths being absorbed more readily by DNA (Albarracín et al. 2013; Diffey 2002). This action can result in single strand breaks in the sugar-phosphate backbone or the formation of cyclobutane pyrimidine dimers (CPDs) (Baliga et al. 2004; Daly 2009; Kurth et al. 2015; Murray et al. 2015; Winter et al. 2001). Of the DNA bases, thymine is the most sensitive to UV radiation, followed by cystine (Yang and Li 2015).

While shorter wavelengths are filtered by the ozone layer, a significant amount of damaging UVB radiation reaches Earth’s surface. The prevalence of UV is typically higher in desert areas due to the absence of cloud cover (Cordero et al. 2014). In addition to increased UV exposure, microorganisms living in desert environments also deal with large temperature fluctuations and low water availability (Cordero et al. 2014; Pulschen et al. 2015; Rampelotto 2013).

2.3 Microbial DNA damage from UV radiation

Solar UV radiation is a major mutagen that causes DNA damage through the formation of photoproducts between adjacent thymine or cytosine bases. (García-Gómez et al. 2012; Matallana-Surget et al. 2008; Matallana-Surget and Wattiez 2013; Rastogi et al. 2010). UV radiation is non-ionising radiation and can cause oxidative stress, resulting in the formation of reactive oxygen species (ROS). ROS are chemically reactive species that contain oxygen, and can oxidise DNA, lipids and proteins, and can cause cell damage (Matallana-Surget and Wattiez 2013). UV radiation induced damage often involves the nucleic acid bases of the DNA strand. The most common types of DNA damage from UV radiation are the formation of CPDs, pyrimidine (6-4) pyrimidone photoproducts (6-4PPs) and Dewar isomers. (García-Gómez et al. 2012; Ikehata and Ono 2011; Matallana-Surget and Wattiez 2013; Rastogi et al. 2010). There are 12 possible photoproducts that can be induced, however, not all are produced at the same frequency, and differ according to the GC content of the genome (Matallana-Surget et al. 2008; Matallana-Surget and Wattiez 2013).

2.3.1 Reactive oxygen species

Reactive oxygen species induce DNA lesions include base and nucleotide lesions and strand breaks. Most DNA damage caused by ROS is through the modification of nitrogenous bases (García-Gómez et al. 2012; Ikehata and Ono 2011). ROS are initiated by activating smaller molecules such as riboflavin or tryptophan which can activate cellular oxygen (Ikehata and Ono 2011). ROS can damage DNA, causing strand breaks, and are able to cause oxidative base damage. ROS also produce oxidised nucleotides such as 8-hydroxy deoxyguanosine-triphosphate, which are still able to act as nucleotide precursors for DNA synthesis (Ikehata and Ono 2011; Suzuki and Nishizawa 2014).

While UVA, UVB and UVC are all able to induce ROS, generally most ROS in microorganisms are induced by UVA due to the prevalence of this wavelength at the earth’s surface (García-Gómez et al. 2012; Ikehata and Ono 2011).

2.3.2 Cyclobutane pyrimidine dimers

CPDs are formed when two adjacent pyrimidines on the same DNA strand form a covalent bond with one another, resulting in a non-linear sugar-phosphate backbone, as shown in Figure 2.2. During DNA replication, non-linear DNA will cause a block which DNA polymerase is unable to pass, leading to replication disruption (Courcelle et al. 1999).

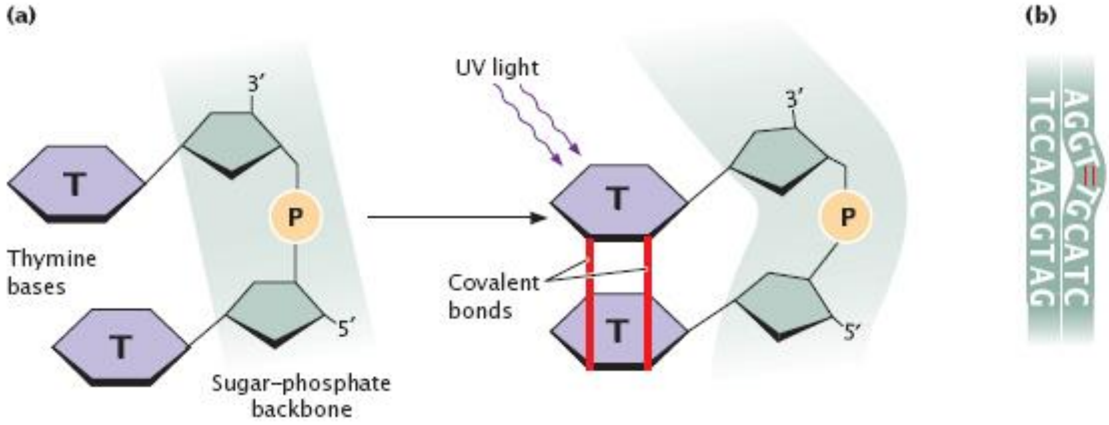


Figure 2.2: The formation of cyclobutane pyrimidine dimers between two adjacent pyrimidines. Dimers are formed between two adjacent pyrimidines. (A) shows the formation of a thymine dimer. (B) shows the distorted DNA backbone (UCSI University 2016).

CPDs prevent DNA synthesis by impeding DNA polymerase from passing over them when they exist on the template strand during DNA replication (Ikehata and Ono 2011). Although cells try to remove these errors through excision repair mechanisms, failure to remove the damage before the replication fork passing would lead to a stall. This causes the fork to collapse at the damaged site and leads to a potential double strand break, usually resulting in cell death (Ikehata and Ono 2011). UVA, UVB and UVC all induce CPDs in microorganisms (Ikehata and Ono 2011).

However, UVA also plays an important role in the removal of CPDs through photoreactivation; a process that eliminates UV-induced photoproducts and restores DNA integrity (García-Gómez et al. 2012). Photoreactivation requires the presence of wavelengths between 300 nm and visible light. UVA is within this range, and can therefore assist the cell with photoreactivation initiation (García-Gómez et al. 2012; Murray et al. 2015).

2.3.3 Pyrimidine (6-4) pyrimidone photoproducts

Pyrimidine (6-4) pyrimidone photoproducts are formed by two adjacent pyrimidines, similar to the formation of CPDs. However, conversely to CPDs, during the formation of 6-4PPs, one pyrimidine base will rotate and cross-link with the second pyrimidine base. Figure 2.3 below shows the difference between CPD formation and 6-4PP formation.

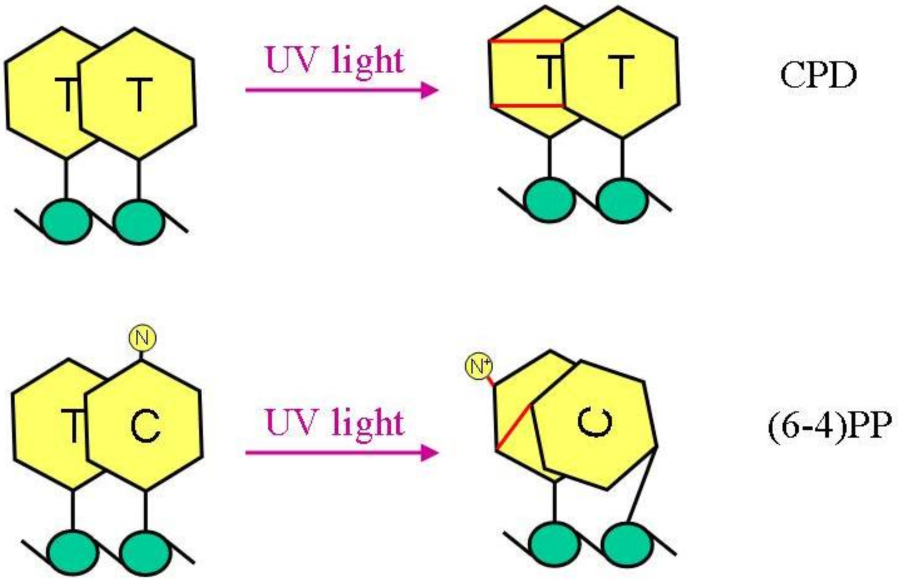


Figure 2.3: The formation of cyclobutane pyrimidine dimers compared with the formation of 6-4PPs (Nouspikel 2019).

6-4PPs prevent cell-cycle progression by inhibiting cell division due to the synthesis of cellular components required for cell growth and maintenance being disrupted (Daly 2009; García-Gómez et al. 2012). As mentioned, the presence of non-linear, damaged DNA will cause a block, thereby preventing the synthesis of essential cellular components by DNA polymerase. 6-4PPs are induced by either UVB or UVC radiation (Crowley et al. 2006; Demple and Harrison 1994; Ikehata and Ono 2011; Kurth et al. 2015; Rastogi et al. 2010; van der Veen and Tang 2015).

2.3.4 Dewar photoproducts

An additional photoproduct of excessive UV radiation is a Dewar photoproduct. A Dewar photoproduct is the result of an unstable 6-4PP forming a cross-linked bond between N4 and C6 of the 3'-pyrimidone ring (Taylor 2005). The additional presence of UVA is the catalyst for 6-4PPs to convert into a Dewar photoproduct. Dewar photoproducts destabilise DNA and cause issues with DNA replication (Ikehata and Ono 2011; Lee et al. 2000; Rastogi et al. 2010). Figure 2.4 demonstrates the different DNA damage that may accumulate through UV exposure; CPD formation, 6-4PPs and Dewar photoproducts.

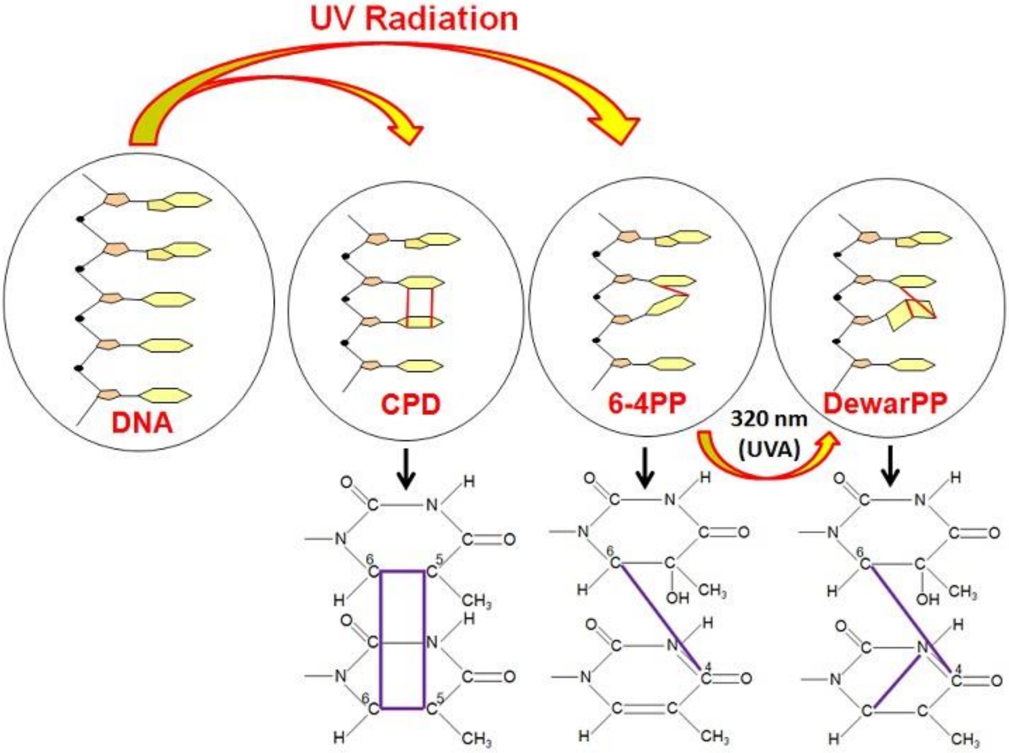


Figure 2.4: The formation of the main UV radiation photoproducts; CPDs, 6-4PP and Dewar photoproducts (Schuch 2014).

Dewar photoproducts are less mutagenic than 6-4PPs and are more easily bypassed by DNA polymerase. Due to this, they are considered less of a block during DNA replication than their 6-4PP counterpart (Lee et al. 2000).

2.4 DNA repair processes in Prokaryotes

Bacteria demonstrate a broad diversity of tolerance and repair mechanisms to DNA-damaging radiation and are the simplest model organisms for studying defence response strategies with regard to gene regulation (Matallana-Surget and Wattiez 2013). Microorganisms capable of surviving in hyper-arid regions provide valuable insight into how single-celled organisms respond to environmental stress (Boor 2006; Matallana-Surget and Wattiez 2013). The survival of a microorganism depends on the ability to sense and respond to changes in the surrounding environment with appropriate changes in gene expression and protein activity (Baliga et al. 2004; Boor 2006). Biological systems have evolved mechanisms to duly respond to environmental stresses that can damage their proteins and DNA.

As previously discussed, there are multiple ways in which microbial DNA can be damaged (Eisen and Hanawalt 1999). Microorganisms therefore have a variety of DNA repair mechanisms (Eisen and Hanawalt 1999; Matallana-Surget and Wattiez 2013). The functions on these repair mechanisms are also diverse, whereby some pathways such as photoreactivation are only able to repair a single type of DNA damage, while others such as recombinational repair can repair many DNA damage types (Eisen and Hanawalt 1999). Following UV irradiation, DNA repair mechanisms include repair or removal of UV photoproducts, cell-cycle arrest checkpoints, and damage tolerance mechanisms that allow cells to replicate even when damage remains (Crowley et al. 2006; Matallana-Surget and Wattiez 2013; Rastogi et al. 2010; Sinha and Häder 2002). Figure 2.5 shows the different biological responses to DNA damage.

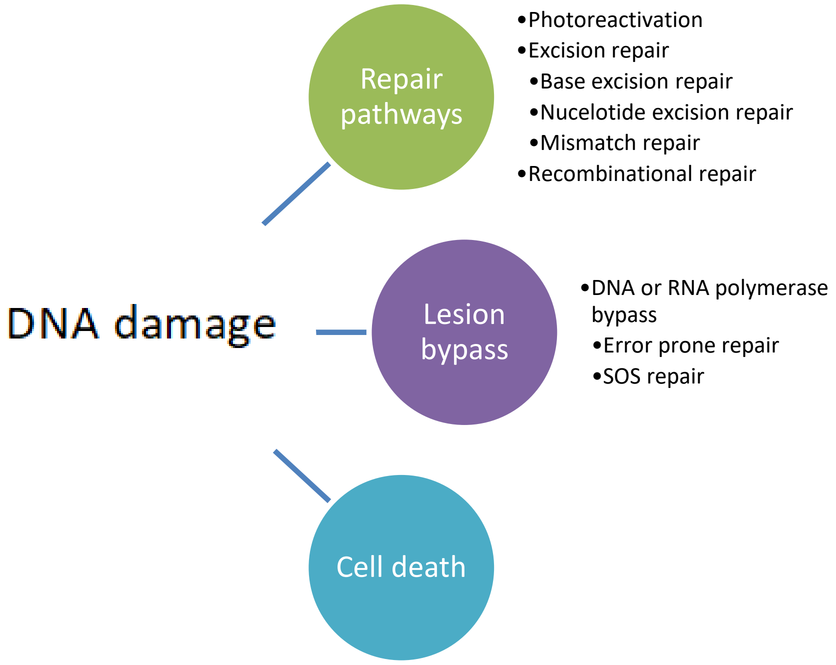


Figure 2.5: Biological responses to DNA damage. Adapted from Matallana-Surget and Wattiez (2013).

The survival of the microorganism during UV irradiation depends on the rate of radiation-induced damage, and the efficiency of microbial DNA repair and conservation (Matallana-Surget and Wattiez 2013). Additionally, it has been observed that the resistance of microorganisms to

UV is influenced by their GC content also (Eisen and Hanawalt 1999; Matallana-Surget et al. 2008).

Stressed cells stimulate a complex network of specific protein phosphorylation and dephosphorylation reactions, resulting in the activation or deactivation of specific gene groups (Crowley et al. 2006; García-Gómez et al. 2012; Klelner 2005; Kyriakis and Avruch 2001; Sutherland 1981). The cell must repair the DNA damage, or it will die. The two most common methods for microorganisms to remove damaged DNA is through photoreactivation or light independent excision repair (Baliga et al. 2004; van der Veen and Tang 2015). The damaged DNA is removed by helicases and nucleases, while being repaired by DNA polymerase and ligase (Baliga et al. 2004).

2.4.1 Gene regulation

Alternations in bacterial gene expression are generally controlled at the transcriptional level through changes between the catalytic core of the RNA polymerase enzyme and the various sigma factors present in the bacterial cell (Boor 2006; Campbell et al. 2008). Bacteria only contain one form of core RNA polymerase, as opposed to archaea and eukaryotes which contain multiple forms (Ghosh et al. 2010). RNA polymerase is a multi-subunit enzyme responsible for creating mRNA transcripts to be translated into new proteins, as well as for identifying relevant genes under specific environmental conditions (Boor 2006; Ghosh et al. 2010). RNA polymerase is able to recognise non-specific DNA but requires an additional sigma factor to identify and attach to specific promoter regions.

Sigma factors are detachable subunits of the prokaryotic RNA polymerase. Bacteria have at least one essential sigma factor that serves to transcribe the genes required for cell viability. Most bacteria also contain alternative sigma factors that transcribe operons in response to specific stimuli (Campbell et al. 2008). The RNA polymerase holoenzyme is formed when a sigma factor associates with a core RNA polymerase. The holoenzyme is then directed to recognised conserved DNA promoter regions which precede gene sequences (Boor 2006; Campbell et al. 2008). Associations between different alternative sigma factors and core RNA polymerase causes the RNA polymerase holoenzyme to recognise different promoter regions and leads to the expression of entirely new sets of target genes. The genes controlled by a single group of sigma factors can reach into the hundreds, meaning that sigma factors can provide effective mechanisms for simultaneously regulating large numbers of prokaryotic genes (Boor 2006; Campbell et al. 2008). Additionally, sigma factors assist with DNA strand separation; an important step in transcription and maintaining the genetic cohesion of the microbial DNA (Boor 2006; Campbell et al. 2008).

2.4.2 Photoreactivation

A critical repair mechanism for microorganisms that encounter high levels of UV radiation in their environment is photoreactivation. During photoreactivation, the enzyme photolyase binds to CPDs in the DNA. Following this, photolyase absorbs and utilises the energy of visible light to reverse UV induced CPDs (Crowley et al. 2006; Matallana-Surget and Wattiez 2013; Rastogi et al. 2010). This mechanism can be seen in Figure 2.6.

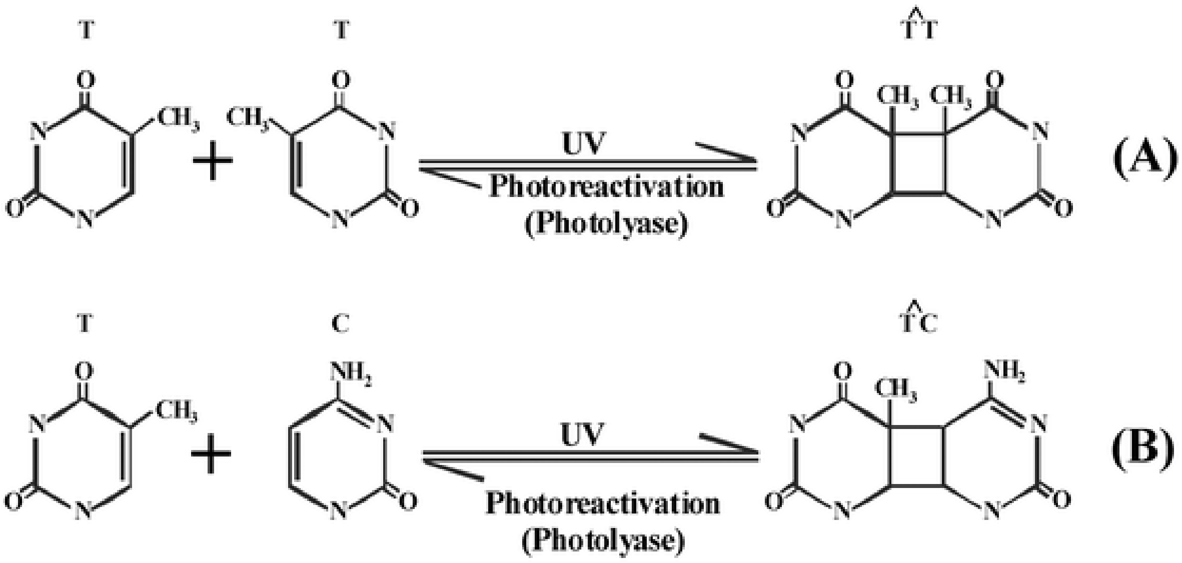


Figure 2.6: Formation of the most detrimental DNA lesion, cyclobutane-pyrimidine dimers by UV radiation. Seen here, (A) is a thymine-thymine CPD, and (B) is a thymine-cytosine dimer. Both CPDs are separated through photoreactivation by a photolyase enzyme when light is present (Sinha and Häder 2002).

Photolyases search the microbial genome for UV induced lesions and bind to the photoproduct before using a light-detecting chromophore to absorb a blue-light photon (450nm) of visible light (García-Gómez et al. 2012; Sinha and Häder 2002). Known photolyase enzymes include deoxyribose pyrimidine photolyase and spore photoproduct lyase.

Deoxyribose pyrimidine photolyase is an enzyme that can remove 6-4PPs and CPDs. Sancar et al. (1983) determined that deoxyribose pyrimidine photolyase binds UV induced legions and will repair damage caused once the cell is exposed to near UV light (300-500 nm). Lorence et al. (1990) also notes that deoxyribose pyrimidine photolyase will bind to DNA in the dark before catalysing the dimer split upon absorption of a photon within the near UV light range.

Spore photoproduct lyase (SPL) repairs thymine dimers that are also known as the spore photoproduct in germinating endospores. SPL is a radical *S*-adenosylmethionine enzyme, which uses an available 5′-deoxyadenosyl radical to cleave the two thymines (Yang and Li 2015). For SPL enzymes to be activated, a particular type of photoproduct must be present. While a CPD is formed by a cross-link bridge between two adjacent pyrimidines, spore photoproducts form a 5-thyminyl-5,6-dihydrothymine, which is a unique dimer (Yang and Li 2015). The repair of 5-thyminyl-5,6-dihydrothymine can be seen in Figure 2.7 below.

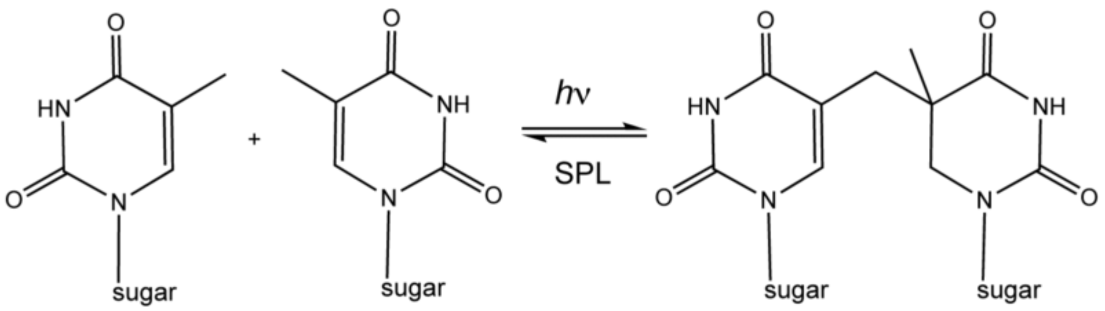


Figure 2.7: The creation of 5-thyminyl-5,6-dihydrothymine and cleavage by spore photoproduct lyase (SPL) (Wikimedia Commons 2019).

Photoreactivation is an efficient repair mechanism for removing damaged DNA; approximately one dimer is split for every blue-light photon incorporated (Sinha and Häder 2002). While photoreactivation must take place in the presence of sunlight, photolyases are able to bind to CPDs in the dark. Once the cell is exposed to visible light, the photoreactivation mechanism will occur. If the cell is not exposed to visible light, the binding of photolyase to damaged DNA acts as a beacon for other repair mechanisms, such as excision repair systems (Sinha and Häder 2002).

2.4.3 Excision repair

In contrast to photoreactivation, light independent pathways are more complex and do not reverse DNA damage but instead replace the damaged DNA with new nucleotides (Rastogi et al. 2010; Sinha and Häder 2002). The three major categories of excision repair pathways involve replacing either the damaged base (base excision repair), replacing the whole damaged nucleotide (nucleotide excision repair) in the DNA fragment or through general excision repair of mismatches bases (mismatch repair). The repair of mutagenic DNA lesions is essential to prevent DNA mutations and cell death, allowing the cells to maintain their genetic integrity (Baliga et al. 2004; Coelho et al. 2013; García-Gómez et al. 2012; Ponferrada-Marín et al. 2010).

2.4.3.1 Base excision repair

Base excision repair (BER) is an important repair mechanism that removes oxidative and UV radiation induced DNA damage (García-Ortiz et al. 2001; Kurth et al. 2015). This type of repair is triggered when the amount of damage is more than what photoreactivation methods can cope with (Fernández Zenoff et al. 2006b). BER is a template based repair mechanism which is initiated by the repair enzyme DNA glycosylase (van der Veen and Tang 2015). DNA glycosylase splits the N-glycosidic bond between the dexoyribose and the target base, thus removing the damaged base, but leaving the phosphodiester backbone intact (Crowley et al. 2006; García-Gómez et al. 2012; García-Ortiz et al. 2001; Rastogi et al. 2010; Roldán-Arjona and Ariza 2009; Seeberga et al. 1995; Sinha and Häder 2002; van der Veen and Tang 2015). Once the target base is removed, the apurinic/apyrimidinic (AP) site can either be removed and replaced using AP endonuclease and DNA polymerase (Roldán-Arjona and Ariza 2009; Seeberga et al. 1995; Sinha and Häder 2002) or the AP site is replaced temporarily using an unmethylated base (García-Gómez et al. 2012; Rastogi et al. 2010; Sinha and Häder 2002; van der Veen and Tang 2015). Because deaminated base derivatives have different coding properties, mutations in the genome will occur as a direct result unless the base is removed (Seeberga et al. 1995).

Different DNA glycosylases remove different types of damage, meaning that the specificity of the repair pathway is determined by the type of glycosylase involved (Sinha and Häder 2002). A frequently occurring hydrolysis reaction is the deamination of cytosine to uracil (Sinha and Häder 2002). Uracil is highly mutagenic when present in the DNA strand and so must be removed by uracil specific DNA glycosylase (Seeberga et al. 1995; Sinha and Häder 2002). Known BER genes include Uracil-DNA glycosylase (*ung*), G/U mismatch-specific DNA glycosylase (*mug*), ultraviolet N-glycosylase AP lyase (*nth*), nucleoside diphosphate kinase (*ndk*), formamidopyrimidine-DNA glycosylase 1 (*mutM*), A-G mismatch DNA glycosylase (*mutY*), exodeoxyribonuclease (*exoIII*) and 3-methyladenine DNA glycosylase (*alkA*). The mechanism for BER genes is variable, however, the primary target for all removals starts at the N-glycosylic bond (Krwawicz et al. 2007; Sinha and Häder 2002). The mechanisms for these BER genes are shown in Table 2.4.

Table 2.4: BER genes and their mode of action.

Gene	Mechanism	Reference
ung	Removes uracil from DNA by cleaving the N-glycosylic bond	Krokan et al. (1997)
mug	Removes thymine or uracil which has mismatched with guanine	Kurth et al. (2015)
nth	Cleaves the N-glycosylic bond of CPD damaged DNA and creates an incision of the phosphodiester backbone at the abasic site	Piersen et al. (1995)
ndk	AP-lyase, 3'- phosphodiesterase, 3'-phosphatase and 3' - 5' exonuclease activities	Krwawicz et al. (2007)
mutM	Removes oxidised guanine (8-oxoG) from damaged DNA	Serre et al. (2002)
mutY	Removes adenine which has mismatched with 8-oxoG	Krwawicz et al. (2007)
exoIII	Repair of AP sites by cleavage in the 3'- to 5'-direction	Demple et al. (1983)
alkA	Removal of 3-methyladenine, an alkylation product from double and single stranded DNA	Krwawicz et al. (2007)

2.4.3.2 Nucleotide excision repair

Nucleotide excision repair (NER) removes a large range of DNA errors caused by UV and chemical mutagens (Crowley et al. 2006; Rastogi et al. 2010; Sinha and Häder 2002). Where BER pathways remove the damaged base and leave the phosphodiester backbone intact with an AP site, NER pathways remove a section of the phosphodiester backbone around the damaged base (García-Gómez et al. 2012; Goosen and Moolenaar 2008). NER can be initiated in two ways: (i) DNA damage is detected by UvrA, which then recruits other enzymes to repair damage, or (ii) DNA damage is detected by transcription-repair coupling factor *mfd* when RNA polymerase stalls at a DNA lesion during transcription. The Mfd enzyme will remove RNA polymerase from the DNA strand and recruits the UvrABC enzymes for DNA repair (Deaconescu et al. 2012; Kiskeer et al. 2013).

The UvrABC system uses the proteins UvrA, UvrB and UvrC to locate and form incisions on both sides of DNA lesions through a complex series of reactions (Crowley et al. 2006; Demple and Harrison 1994). The mechanism of the UvrABC system is as follows: two UvrA proteins form a dimer with one UvrB protein. This UvrAB complex detects DNA damage and binds to the damaged site (Kiskeer et al. 2013; Stracy et al. 2016). The UvrA dimer will dissociate from the lesion once UvrB begins downstream NER reactions, and UvrC protein will bind to UvrB, creating a UvrBC complex (Stracy et al. 2016). UvrC then cleave the phosphodiester bond four nucleotides downstream and eight nucleotides upstream from the DNA damage, excising a 12-nucleotide segment. Subsequently, UvrD helicase will remove the damaged DNA, and DNA polymerase I and ligase then refill and seal the resulting DNA gap using the other DNA strand as a template (Demple and Harrison 1994; Sinha and Häder 2002; Stracy et al. 2016). Figure 2.8 shows how excision repair occurs.

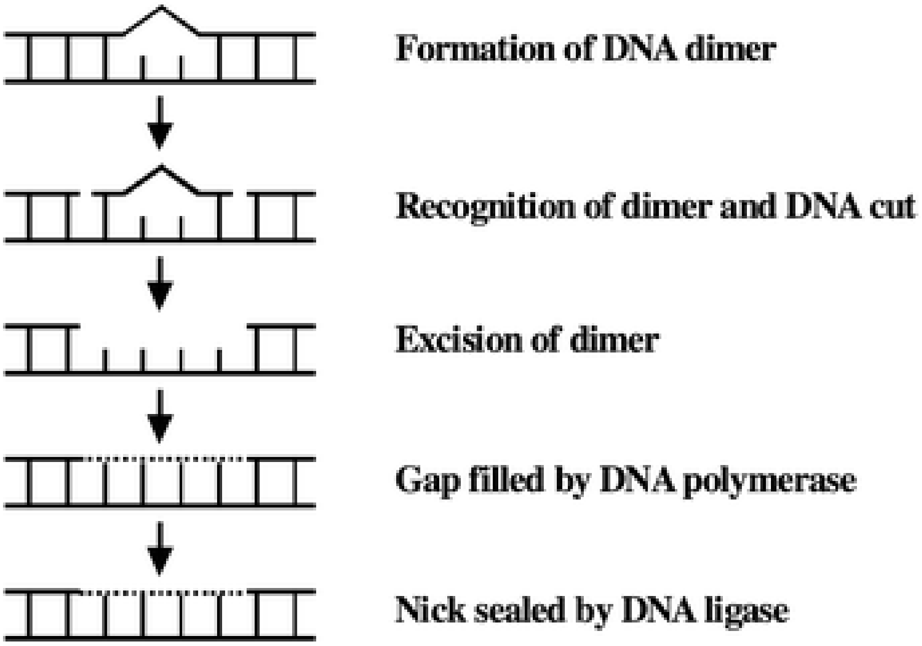

Formation of DNA dimer
Recognition of dimer and DNA cut
Excision of dimer
Gap filled by DNA polymerase
Nick sealed by DNA ligase

Figure 2.8: Diagram representation of the nucleotide excision repair pathway (Sinha and Häder 2002).

It has previously been shown that the copy number of UvrA and UvrB significantly increased when the cell is exposed to DNA damaging agents compared to ‘normal’ cell conditions (Janion 2001; Van Houten et al. 2005). This has led to the conclusion that the *uvrA*, *uvrB* and *uvrD* genes, but not the *uvrC* gene, is also under the control of the SOS response system when DNA damaging agents are present (Van Houten et al. 2005). In addition, it has been demonstrated that UvrA and UvrB also have a role in preventing illegitimate recombination from double strand DNA breaks following UV radiation (Hanada et al. 2000).

2.4.4 Mismatch repair

DNA mismatch repair (MMR) is a highly conserved excision repair mechanism that contributes towards maintaining genome stability by correcting mismatched base pairs (Fukui 2010). Mismatched base pairs can arise as a result of replication error, or through exposure to mutagens, such as UV. To detect DNA mismatches, MMR proteins scan for and recognise of the methylation pattern of DNA at specific recognition sites, which is the palindromic GATC in *E. coli* (Fukui 2010). MMR in most bacteria directs the repair towards the error-containing strand by identifying strand inconsistencies in the methylation pattern of DNA bases (Eisen and Hanawalt 1999; Fukui 2010). During MMR initiation, the MutS protein binds to a mismatch and recruits MutL and MutH to indicate the region is targeted for excision repair. MutL and MutH form a complex, and MutL will identify the correctly methylated DNA strand and cause this strand to form a loop (Fukui 2010). MutH will then cut the unmethylated strand at GATC sites close to the mismatch. Following this, exonucleases such as UvrD helicase will complete the excision, removing a large patch of DNA which is then repaired by DNA polymerase I using the intact strand as a template for replication (Eisen and Hanawalt 1999).

2.4.5 Recombinational repair

Recombination is an important DNA repair process which ensures the transmission of the correct genetic information (Sinha and Häder 2002). Single or double strand DNA breaks are lethal to bacteria and must be repaired efficiently for the survival of the organism (Ikehata and Ono 2011; Morimatsu and Kowalczykowski 2003). Recombinational repair can use a complementary strand from a homologous region of DNA to repair the daughter strand gap (Sinha and Häder 2002). However, in the case of a double strand break, the process is more complicated (Singleton et al. 2004). Recombinational repair systems in prokaryotes usually falls under the RecBCD, RecFOR and SbcBCD pathways (Eisen and Hanawalt 1999; Rice and Cox 2001). A complete set of genes is usually found more frequently in the RecFOR pathway than in the RecBCD and SbcBCD pathways (Rocha et al. 2005). During recombinational repair, the homologous region of a strand complementary to the damaged region is used as a template for DNA repair. The following sections describe the different known pathways for recombinational repair mechanisms.

2.4.5.1 RecBCD pathway

The RecBCD enzyme is made up of the RecB, RecC and RecD subunits. RecBCD initiates recombinational repair from double-strand breaks in DNA, which have usually occurred because of UV damage, ionising radiation or replication errors (Singleton et al. 2004; Smith 2012; Spies and Kowalczykowski 2005). The RecBCD enzyme is able to both unwind and make a single strand incision in the DNA (Singleton et al. 2004; Smith 2012; Spies and Kowalczykowski 2005). The damaged site of DNA is then digested until the recombinational site Chi (5'-GCTGGTGG-3') is reached (Singleton et al. 2004). Once the Chi site is reached, the RecBCD enzyme loads the enzyme RecA onto the 3' tail of the DNA to be repaired (Singleton et al. 2004; Smith 2012; Spies and Kowalczykowski 2005). Figure 2.9 shows the process of double strand break repair by the RecBCD enzyme with RecA assistance.

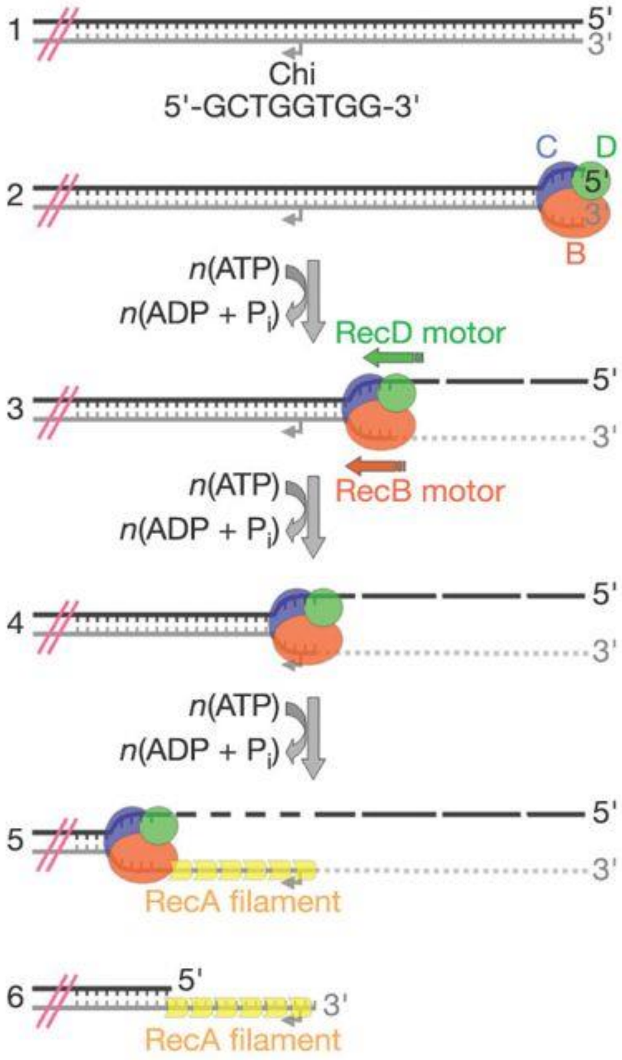


Figure 2.9: Double strand break repair using the RecBCD enzyme (Singleton et al. 2004). RecB is coloured orange, RecC is blue and RecD is green. The process is as follows: 1) a double strand break is created as the result of DNA damage. 2) RecBCD forms a complex and binds to the DNA. The DNA strands are unwound .3) ATP-dependent DNA unwinding continues, and the DNA of both strands is digested. 4) RecBCD encounters the Chi site (grey arrow) and 3' digestion ceases while 5' digestion becomes more frequent. 5) RecA is loaded onto the 3' tail. 6) RecBCD dissociates, leaving a RecA-coated 3' tail which can initiate homologous recombination.

The RecBCD enzyme is made up of three subunits. The RecB subunit is a 3'-5' helicase and nuclease, RecC contains the Chi recognition site and RecD is a 5'-3' helicase (Singleton et al. 2004).

2.4.5.2 RecFOR pathway

The RecFOR pathway is another double-strand break repair pathway that has been observed in known UV resistant bacteria such as *Deinococcus radiodurans* (Bentchikou et al. 2010). Bentchikou et al. (2010) and Morimatsu and Kowalczykowski (2003) note that the RecFOR pathway primarily repairs single-strand breaks or gaps but will repair double-strand breaks when there are mutations in the RecBCD enzyme or the RecBCD pathway is absent.

Michel et al. (2004) note that following UV radiation damage, the RecFOR and RecA complexes are essential for DNA replication restart.

The reconstruction of an intact genome in *Deinococcus radiodurans* following DNA damage is through an extended synthesis-dependent strand annealing process, and occurs in much the same way as the RecBCD pathway (Bentchikou et al. 2010). A model of the RecFOR pathway for a double strand break can be seen below in Figure 2.10.

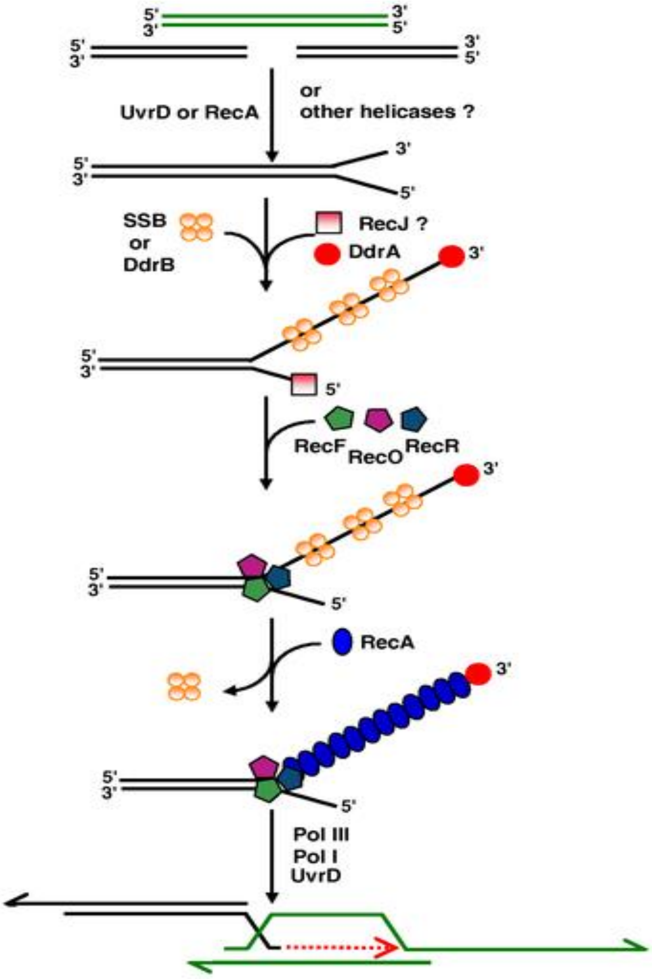


Figure 2.10: Double strand break repair using the RecFOR pathway (Bentchikou et al. 2010).

As seen above, the RecFOR pathway is activated by either UvrD or RecA. RecJ and DdrA attach to the 5' and 3' ends respectively. The RecF, RecO and RecR enzymes attach to the DNA strand break region and facilitate RecA filamentation to help repair the break. Pol III, Pol I or UvrD then initiate recombinational repair as previously described (Figure 2.9).

The RecBCD pathway in *Deinococcus radiodurans* has previously been reported as being absent, meaning that the RecFOR pathway becomes the primary recombinational repair pathway for this organism (Bentchikou et al. 2010). The RecFOR pathway is comprised of the enzymes RecF, RecO, RecR, RecN, RecJ, RecQ, RecG, RecA, RuvA, RuvB and RuvC (Morimatsu and Kowalczykowski 2003). Their functions can be seen below in Table 2.5.

Table 2.5: RecFOR pathway genes and their function. Adapted from Martins-Pinheiro et al. (2007).

Gene	Function
recF	DNA strand break (single or double) binding and assistance with RecA filamentation
recO	DNA strand binding and *recF* assistance
recR	ATP-binding and *recF* assistance
recN	ATP-binding
recJ	DNA exonuclease in 5' – 3' direction
recQ	ATP-dependent DNA helicase
recG	Resolvase in 3' – 5' Holliday junctions
recA	Recombinase and DNA renaturation
ruvA	5' – 3' Holliday junction helicase
ruvB	5' – 3' Holliday junction helicase
ruvC	Holliday junction endonuclease

In mutants devoid of either the RecBCD or the RecFOR pathways, organisms typically became radiosensitive where they were previously radiation resistant (Bentchikou et al. 2010). The repair of double-stand breaks through either the RecBCD or RecFOR pathways is therefore important for organism survival following treatment by radiation (Bentchikou et al. 2010; Morimatsu and Kowalczykowski 2003).

2.4.5.3 SbcBCD pathway

The SbcBCD enzyme is made up of the SbcB, SbcC and SbcD subunits. The SbcB protein is a DNA-specific 3' – 5' exonuclease, while SbcC and SbcD make up an exonuclease that cleaves hairpin structures (Eisen and Hanawalt 1999; Martins-Pinheiro et al. 2007; Rocha et al. 2005). When the SbcBCD system is defective, the RecFOR pathway is activated (Martins-Pinheiro et al. 2007; Rocha et al. 2005). This is due to the SbcB nuclease supressing RecFOR action by removing the 3' end on which RecA is loaded by the RecFOR system (Rocha et al. 2005).

2.4.6 Lesion bypass

In the presence of very high UV doses, excision repair mechanisms may not be sufficient to repair all DNA lesions (Michel et al. 2004). DNA lesions that remain unrepaired can lead to a halt in DNA replication. Lesion bypass is the only way a microorganism can survive in a situation where DNA repair cannot occur. UV induced DNA damage is usually repaired by RecFOR and RecA through either recombinational repair, the induction of SOS response enzymes, or through lesion bypass using DNA polymerase V (Michel et al. 2004). DNA polymerase V is able to insert nucleotides opposite DNA lesions in a process termed translesion DNA synthesis (Ikehata and Ono 2011; Lee et al. 2000; Patel et al. 2010; Sinha and Häder 2002). Lesion bypass can be achieved in two ways: error-prone repair or SOS repair. Both pathways often result mutagenic events and as a result are often the last mechanisms employed to survive irradiation (Dib et al. 2008; Lee et al. 2000).

2.4.6.1 Error-prone repair

Under extreme UV stress microorganisms increase mutational events, known as error-prone repair, in an effort to survive. Mutational events are increased during this time due to the amount of DNA damage that has been accumulated, and efforts by the cell to quickly remove this damage. This system enables lesion bypass during DNA replication and allows DNA polymerase V to bypass UV induced lesions, resulting in mutations (Dib et al. 2008; Lee et al. 2000). This process is known as translesion replication or bypass synthesis, and mutations are caused by the tendency of DNA polymerase V to insert an incorrect nucleotide opposite the lesion during translesion DNA synthesis (TLS) (Lee et al. 2000). However, there is evidence to suggest that lesion bypass methods are accurate when bypassing CPDs (Sinha and Häder 2002; Swanson et al. 2012). DNA polymerase V will generally insert adenine bases in both places opposite a CPD, leading to a high base pair efficiency due to most CPDs being formed between thymine dimers (Sinha and Häder 2002). When a cytosine dimer is produced, the stability of the cytosine drastically reduces and will often deaminate to uracil, which can be removed by the BER pathway (Ikehata and Ono 2011). When encountering 6-4PPs however, DNA polymerase V may insert G instead of A leading to the term 'error-prone repair' (Sinha and Häder 2002).

TLS error-prone repair involves several TLS polymerases which work by insertion and extension of mismatched nucleotides opposite UV lesions (Ikehata and Ono 2011; Patel et al. 2010). TLS polymerases are specialised DNA polymerases which work to prevent replication blockages by restarting DNA synthesis once it has stalled at damaged bases on the template strand. When a replicative assembly encounters a miscoding lesion, DNA synthesis halts and TLS will usually insert an incorrect nucleotide opposite the lesion so that synthesis will continue (Ikehata and Ono 2011; Patel et al. 2010; Sinha and Häder 2002). Figure 2.11 demonstrates how TLS errors occur.

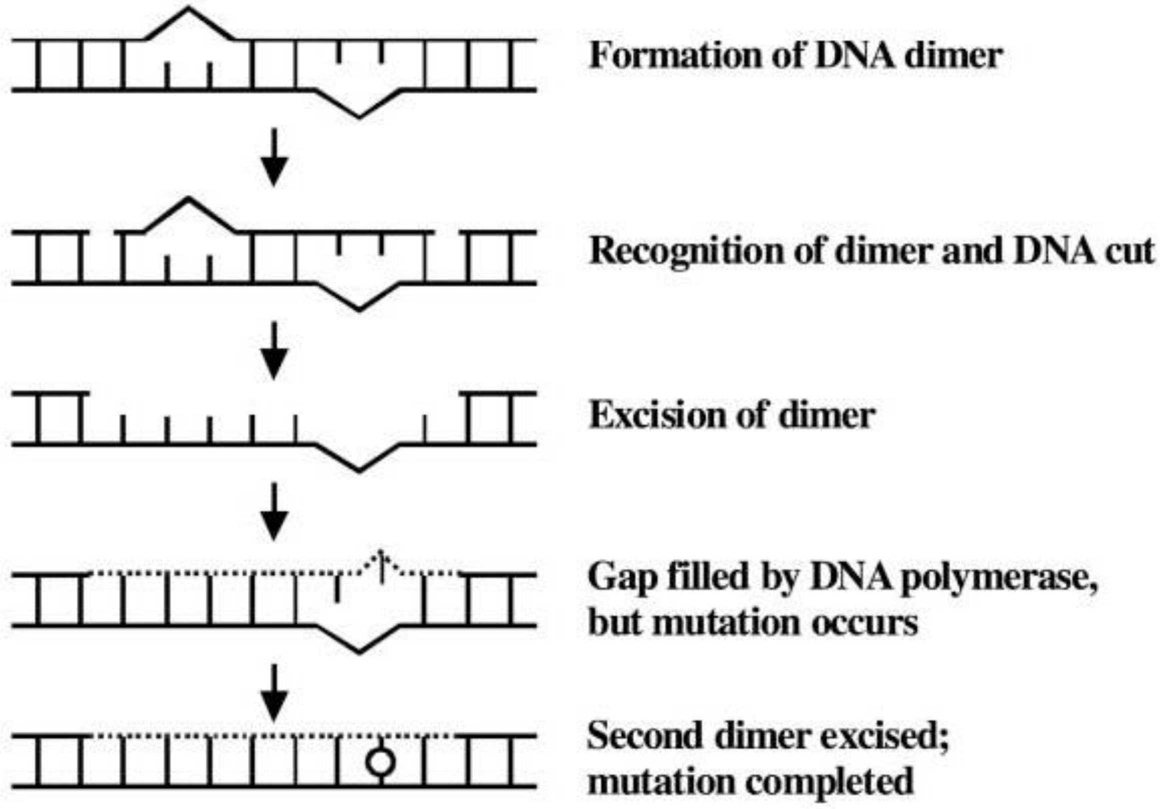

Formation of DNA dimer
Recognition of dimer and DNA cut
Excision of dimer
Gap filled by DNA polymerase, but mutation occurs
Second dimer excised; mutation completed

Figure 2.11: Diagram representation of the mutagenic error-prone repair pathway (Sinha and Häder 2002).

DNA synthesis using TLS polymerases is error-prone due to the conventional base pairing rule of A-T and G-C often being ignored, meaning that there is a high frequency of cellular mutations (Ikehata and Ono 2011). When dimers are formed on opposite strands, only one strand may be repaired at a time. If the cell were to attempt to repair both concurrently, a double strand break would occur (Ikehata and Ono 2011). When the strand being repaired encounters the dimer of the template strand, a random base is inserted due to TLS. When the second strand is being repaired, the complementary base for the incorrect base that was dropped opposite the dimer site will be inserted, causing a mutation in the DNA strand (Ikehata and Ono 2011). TLS error-prone repair involves several TLS polymerases which work by insertion and extension of mismatched nucleotides opposite UV lesions (Ikehata and Ono 2011; Sinha and Häder 2002). As a result, DNA polymerase V is strictly regulated in the cell so as to avoid genomic mutation overload (Patel et al. 2010).

2.4.6.2 SOS repair

Double strand breaks along the DNA strand can occur during exposure to DNA damaging agents. Following DNA damage, the gaps produced are repaired by enzymes of the recombinational repair pathway. RecA is a multifunctional protein that is involved in a number of cellular processes, including the coordination of cell division, homologous recombination, DNA repair and SOS response (Aranda et al. 2011; Ayora et al. 2011; Benedict and Kowalczykowski 1988; Jungfer et al. 2007; Mirshad and Kowalczykowski 2003). Due to its importance in microbial SOS responses, the RecA protein is well conserved and omnipresent in a range of prokaryotes, and is activated by DNA-damaging agents, such as UV (Jungfer et al. 2007). The presence of the RecA protein and UV damaged single-strand DNA promotes the cleavage of the LexA repressor (Ayora et al. 2011; Jungfer et al. 2007). Following this, SOS genes are activated to assist with repair of the damaged strand. The interaction of RecA and LexA is shown in Figure 2.12.

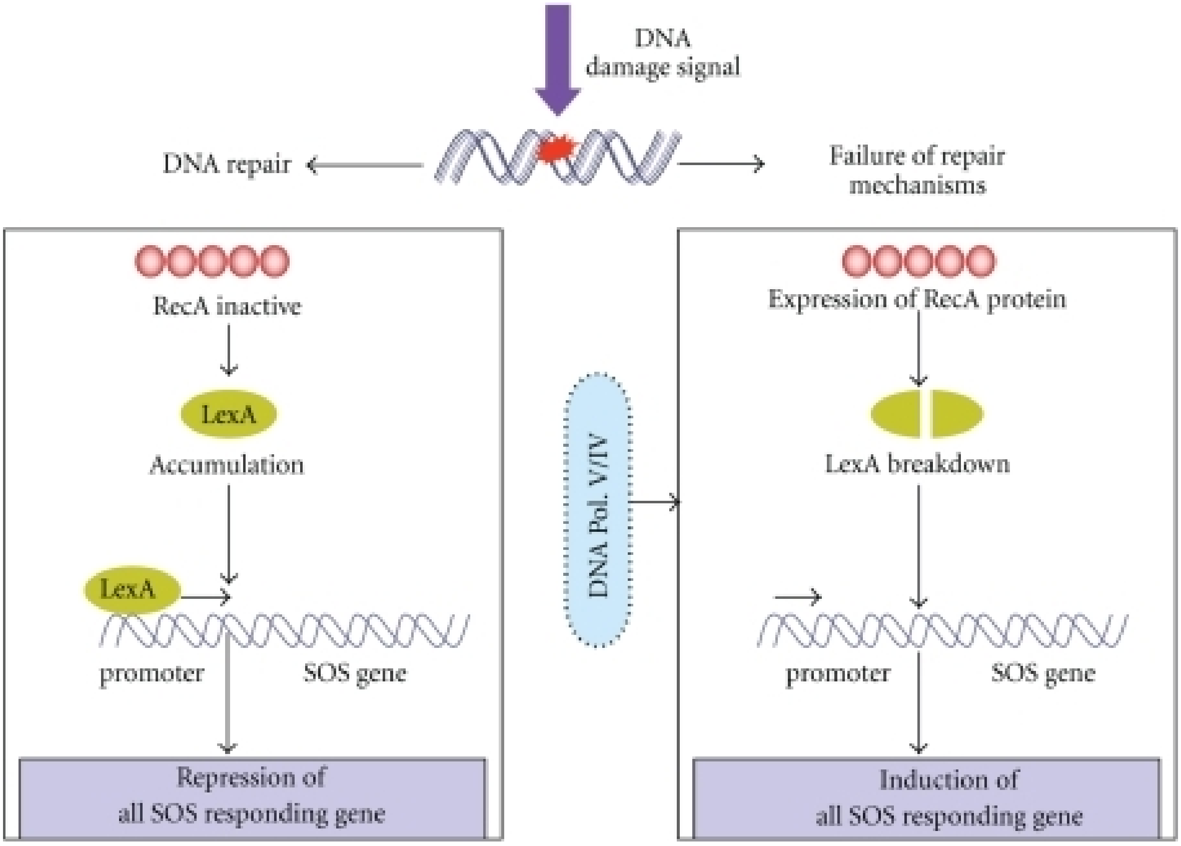


Figure 2.12: DNA SOS repair by RecA (Rastogi et al. 2010).

SOS response enzymes tend to be highly error prone, however, they allow the replication fork to proceed over damaged DNA sites, thereby guaranteeing DNA replication and improving the chance of cell survival (Aranda et al. 2011). This process is complex and is broken into five steps; (i) recognition of the break site and formation of a repair centre; (ii) end-processing of the broken ends to generate a 3'-tailed duplex; (iii) loading RecA onto single stranded DNA; (iv) branch migration and resolution of the recombination intermediates; (v) disassembly of the recombination apparatus (Ayora et al. 2011).

RecA has a protective role within bacteria. Studies have shown that the removal or alteration of the RecA enzyme in microorganisms causes a decrease in the organism's ability to survive DNA damaging agents (Aranda et al. 2011; Lin et al. 2006; Mirshad and Kowalczykowski 2003).

2.4.7 DNA repair summary

Irradiation of DNA with UV produces a variety of photoproducts that inhibit DNA replication. Therefore, microorganisms must have several repair mechanisms in place to remove lesions to ensure cell survival. One strategy for removing lesions is through photoreactivation, where blue light photons are used to split CPDs. Excision repair is another mechanism that can assist in the removal of more extensive damage. Finally, SOS repair mechanisms are activated when the amount of DNA damage is too great for photoreactivation and excision repair to remove alone. Identifying which repair processes are activated under varying levels of UV exposure is an important step in understanding how microorganisms in arid conditions can survive.

2.5 Microbial identification techniques

Bacterial identification through DNA based techniques is one of the most important platforms for environmental biology (Ronaghi 2001). Currently, pure culture is still the most widely used technique for studying microbial physiology with regard to the roles of genes, proteins, and metabolic pathways (Mandlik et al. 2016). Accurate microorganism identification is essential for improving our understanding of microbial systems and niche function. Traditional culture methods of bacterial identification relies on phenotypical characteristics and biochemical reactions, however, this is completely reliant on *in vitro* pure culture (Mandlik et al. 2016).

More recently, understanding of microbial communities has largely improved due to the development of culture-independent next generation sequencing (NGS) techniques (Sanschagrin and Yergeau 2014; Shokralla et al. 2012). NGS targeting of the 16S rRNA gene have led researchers to new fastidious species and phyla, which have previously been uncultured (Mandlik et al. 2016; Yarza et al. 2014). Current analyses indicate that there are 52 bacterial phyla, of which 26 are candidate phyla, which are uncultivable and only known through gene sequences (Mandlik et al. 2016; Rappé and Giovannoni 2003). This has revealed that microbial communities contain tremendous diversity, and are more complex than previously thought (Mandlik et al. 2016; Zengler 2009).

2.5.1 Culturing methods

2.5.1.1 Traditional culture media for microbes isolates from arid soil

The environments from which soil samples will be collected are arid soil that lack large quantities of nutrients. Previous studies focusing on low nutrient environments have recommended the use of the following non-selective general media: tryptic soy agar (Cockell et al. 2008; Goordial et al. 2016; Kim et al. 2012; Kuhlman et al. 2005), nutrient agar (Dsouza et al. 2015; Kim et al. 2012), Reasoner's 2A agar (Ball et al. 2014; Dsouza et al. 2015; Kim et al. 2012) and Luria-Bertani agar (Albarracín et al. 2014; Di Capua et al. 2011; Dib et al. 2008; Fernández Zenoff et al. 2006a; Kim et al. 2012; Kurth et al. 2015). Growing microorganisms on solid agar under laboratory conditions requires knowledge of their growth requirements (Mandlik et al. 2016). Cultivation success from low nutrient environments is usually inhibited by a large concentration of nutrients provided in the media. Previous research has shown that media containing high concentrations of nutrients can have an inhibitory effect (Zengler 2009). As previously mentioned, culturing environmental microorganisms onto agar plates is challenging due to an estimated 99% of environmental microorganisms currently being 'unculturable' to date (McKay 2008; Navarro-González et al. 2003).

Cultivation methods allow for comprehensive study of microbial physiology (Mandlik et al. 2016). Despite this, the cultivation-based identification approach has limitations: primarily, it can take several days or weeks for fastidious bacteria to grow (Mandlik et al. 2016). Culture-based methods also often have low sensitivity and specificity, as well as the added cost and difficulty of identifying a large number of isolates (Mandlik et al. 2016).

It has previously been noted that microorganisms in pure culture experiments undergo genetic changes over time when cultivated under the same conditions (Zengler 2009). This genetic ‘drift’ results in genetically diverse subpopulations, leading to questions of how ‘pure’ the culture really is (Zengler 2009).

Microorganisms have two physiological states; viable and non-viable. Viable cells can be actively dividing or non-dividing (Puspita et al. 2012; Stewart 2012). Problems begin to arise with 'uncultivable' cells from environmental samples when i) microbial groups have no previous cultivated representatives, so appropriate laboratory conditions for growth have yet to be identified, and ii) microorganisms which have previously been cultured, but whose cells are currently in a non-growing state (Puspita et al. 2012). Typically, the lowest percentages of cultivable cells are from low nutrient environments, such as deserts (Navarro-González et al. 2003; Puspita et al. 2012). It is thought that uncultivable microorganisms from soil environments require several rounds of enrichment and natural-environment simulation within diffusion chambers before appearing on petri dishes in sufficient numbers for isolation (Pham and Kim 2012). Mimicking natural environmental conditions, little or no added nutrients to agar media, longer incubation periods, and co-culturing techniques have been applied to visualise previously unculturable organisms on agar (Mandlik et al. 2016; Pham and Kim 2012; Stewart 2012).

The modernisation of traditional culturing techniques such as simulating natural environments and co-culturing have allowed researchers to observe previously unculturable organisms within a laboratory setting. Improved culture media formulation that more closely resemble natural conditions are required to study and identify new roles and functions of some microorganisms within soil samples. Currently, pure culture is still the most widely used technique for studying microbial physiology with regard to the roles of genes, proteins, and metabolic pathways (Pham and Kim 2012).

2.5.2 Molecular approaches for microbial identification

Molecular approaches for microbial identification require a target gene region that reveals species-unique information about the microorganism’s identity. The use of rRNA molecules are widely used for comprehensive microbial identification, characterisation and classification (Mandlik et al. 2016; Yarza et al. 2014; Zengler 2009). Small subunit sequencing has therefore become one of the most applied techniques in microbial identification. Marker genes, such as the

23S and 16S rRNA gene, can be used for single, or community, bacterial identification (Woo et al. 2008; Yarza et al. 2014; Zengler 2009). The 16S rRNA gene is present in almost all bacterial species, providing a target gene for a wide range of organisms in an environmental sample.

2.5.2.1 16S rRNA gene-based identification

The 16S rRNA gene sequence to study bacterial taxonomy and phylogeny has been one the most common genetic markers used, for several reasons. Before 16S sequencing, DNA-DNA hybridisation was used to determine species. This method is labour-intensive, time consuming and often expensive to perform (Janda and Abbott 2007; Sanschagrin and Yergeau 2014). The 16S rRNA gene has approximately 1540 nucleotides, with 10 variable regions throughout (Mandlik et al. 2016; Rappé and Giovannoni 2003). These variable regions contain unique information about the genus and species (Mandlik et al. 2016; Zengler 2009). While there are variable regions within the 16S rRNA gene, the gene itself is highly conserved, and is long enough to be highly informative, yet short enough to be easily targeted by Sanger sequencing (Janda and Abbott 2007; Mandlik et al. 2016; Roberts and Darveau 2015; Větrovský and Baldrian 2013). These properties make the 16S rRNA gene an efficient way of identifying an organism (Janda and Abbott 2007). Following the inclusion of 16S sequencing, the recognised species of microorganisms increased drastically due to the ease of this sequencing compared with DNA-DNA hybridisation (Janda and Abbott 2007).

Using 16S rRNA sequencing when identifying a new species has some shortcomings. One of the most appealing uses of 16S rRNA sequencing is the ability to provide genus and species identification for isolates that do not fit any recognised biochemical profiles. It has been found that 16S sequencing is able to provide accurate genus identification in most cases, but accurate species identification is unreliable (Janda and Abbott 2007; Woo et al. 2008; Zengler 2009). These issues are most likely encountered when species shares similar or identical 16S rRNA data (Janda and Abbott 2007). The *Bacillus* strains *B. globisporus* and *B. psychrophilus* demonstrate this issue; the two species have 99.5% 16S rRNA similarity, but only share 23-50% relatedness when analysed through hybridisation reactions (Janda and Abbott 2007). Additionally, if the nearest neighbour exhibiting a similarity score of <97%, the organism is a new species. A similarity score of >97% does not have a clear meaning, however, as this could indicate either a new species or clustering within a previously defined taxon (Janda and Abbott 2007). In this scenario, DNA-DNA hybridisation may be required. While 16S rRNA sequencing has a large range of uses, there is no defined 'threshold values', above which there is a collective agreement on what constitutes a conclusive identification of a new species (Janda and Abbott 2007).

2.5.2.2 Sanger Sequencing

Sanger sequencing is a method of DNA sequencing which incorporates amplifying the target sequence using the polymerase chain reaction (PCR). There are several methods for Sanger sequencing, including classical and dye-terminator based. Dye-terminator based Sanger sequencing is more commonly used due to its greater speed and convenience (Ferrari et al. 2005; Sanschagrin and Yergeau 2014; Zhang et al. 2012b). During PCR, the target DNA or RNA code is denatured, annealed and then extended so as to make many copies of the target code (Ferrari et al. 2005; Sanschagrin and Yergeau 2014; Zhang et al. 2012b). To perform Sanger sequencing, DNA is isolated from a sample and a sequence target region is amplified. The sample DNA, target sequence primer, deoxynucleotide (dNTPs), chain terminating dideoxynucleotides (ddNTPs) and DNA polymerase are combined as a mixture. A PCR is then performed on the mixture and the final PCR product is then sequenced to determine the genetic code.

Typical dye-based PCR employs labelling of the chain terminator ddNTP to complete the PCR in one reaction (Heather and Chain 2016). Each ddNTP is labelled with fluorescent dyes which emit light at different wavelengths (Venter et al. 2001). The PCR fragments are passed through capillary gel electrophoresis where a laser illuminates their fluorescence. The fluorescence of each fragment is detected and recorded on a computer (Heather and Chain 2016). During Sanger sequencing, a product of between 500-900 bases is expected.

A common issue with DNA sequencing via Sanger's is the poor quality of the first 20-50 bases of the sequence due to issues with primer binding. The quality of sequencing also declines after 700-900 bases (Venter et al. 2001). This makes Sanger sequencing ideal for short sequences, such as partial sequencing of the 16S rRNA gene, with the potential to cover the entire 16S rRNA gene with multiple primers and sequencing runs. However, Sanger sequencing has largely been replaced by NGS methods for large scale genome analyses.

2.6 Metagenomics

Metagenomics, also called community sequencing, is the study of genetic material from environmental samples through genomic identification (Handelsman 2004). Traditional genetic identification techniques rely on cultivated organisms, while community gene sequencing focus on specific genes, such as the 16S rRNA gene, to produce a community profile of the sample (Handelsman 2004; Mandlik et al. 2016). Through these methods, it is possible to identify a large majority of the organisms within a sample, as opposed to only those that are culturable (Handelsman 2004; Mandlik et al. 2016). The application of metagenomic techniques to soil has become very widespread (Riesenfeld et al. 2004; Streit and Schmitz 2004; Tringe et al. 2005).

2.6.1 Environmental bacterial 16S rRNA gene-defined community diversity

As previously mentioned, it is estimated that approximately only 1% of the microorganisms on Earth are culturable using current methods (Cowan et al. 2015; McKay 2008; Navarro-González et al. 2003). Species diversity in an environmental sample is therefore a fundamental piece of information regarding the established community structure. It is critical to know the community structure of an environment to continue improving our understanding of microbial ecological systems, and microbiology in general. Incorporating modern sequencing techniques into studies regarding bacterial communities has allowed researchers to study the prokaryotic diversity of an environment within a laboratory setting (Mandlik et al. 2016). Microorganisms that have unusual phenotypic profiles , have low abundance in an environmental sample, are slow growing, or are uncultivable can be identified using 16S rRNA sequencing (Sanschagrin and Yergeau 2014; Woo et al. 2008). Several studies have utilised 16S rRNA gene-based information to investigate specific prokaryotic community structures, which has aided in-depth comparative investigations into the relationship between physical, chemical and biological aspects of microbial ecology function (Archer et al. 2017; Armstrong et al. 2016; Caruso et al. 2011; Fierer et al. 2012; Hill et al. 2016). From these studies, an understanding of the mechanisms of community assembly is beginning to develop, although remains limited (Fierer et al. 2012; Van Goethem et al. 2016).

The current common practice for studying bacterial diversity in an environmental sample is through a 16S rRNA gene analysis pipeline. This involves the recovery of 16S rRNA sequences from the sample, clustering the sequences at 97% nucleotide identity, and finally, a comparison of the resulting operational taxonomic units (OTUs) which are used as an estimation of α-diversity, or species richness (Armstrong et al. 2016; Rodriguez-R et al. 2018). The 97% threshold was proposed as a lower boundary to screen isolates for species assignment, with higher resolution than DNA-DNA hybridisation, as previously discussed (Janda and Abbott 2007; Rodriguez-R et al. 2018).

Due to high sequence conservation of the 16S rRNA gene at the genus level, organisms belonging to closely related yet different species may be grouped together under the same OTU. It is estimated that the 16S rRNA gene-based approach will underestimate the number of OTUs by around 12% (Rodriguez-R et al. 2018). This underestimation can rise to around 50% for some taxa where the 16S rRNA gene has a very high degree of similarity for distinct species, such as the genera *Pseudomonas*, *Campylobacter* and *Citrobacter* (Rodriguez-R et al. 2018). Rodriguez-R et al. (2018) recommends that using 97% nucleotide identity for 16S rRNA gene-defined community diversity should be considered a 'lower threshold' estimate.

However, Rodriguez-R et al. (2018) also notes that previous studies have typically aimed to cover overall phylogenetic diversity, instead of investigations into closely related species within a community. In addition, many studies do not typically aim for high-quality 16S rRNA gene sequences, but instead use 16S rRNA gene-defined diversity methods as an opportunity to observe bacterial diversity within an established environmental community (Armstrong et al. 2016; Rodriguez-R et al. 2018). While current 16S rRNA gene-defined diversity thresholds provide an insight into the community structure of an environmental sample, there is indication that 16S rRNA gene universal primers may not be as 'universal' as anticipated (Zengler 2009). This implies that there may be sequences that are not detected by current universal primers.

As previously discussed, furthering current understanding of desert microbial communities is an important venture due to our changing climate. Dramatic and sudden changes in water availability has been shown to drastically alter the microbial soil community (Aislabie et al. 2008; Aslam et al. 2016; Barnard et al. 2014; Cregger et al. 2012). The community structure of both the Dry Valleys and the Namib Desert is vitally important as the soil microbes are the main carbon and nitrogen cyclers of these environments. An alteration to these structures may result in the loss of microbial diversity, which could in turn have an impact on the larger environment. While numerous studies have described the community structure of the Dry Valleys (Cary et al. 2010; Lee et al. 2018; Pointing et al. 2009; Stomeo et al. 2012; Wei et al. 2016) and the Namib Desert (Makhalanyane et al. 2013; Scola et al. 2017; Stomeo et al. 2013), continuing to track environmental changes of these environments over time will provide a clearer understanding of the impact of climate change on these environments.

2.6.2 Next generation sequencing methods

More recently, next-generation sequencing (NGS) platforms have been widely utilised for genetic research. Sequencing technology has evolved rapidly in recent years, with several sequencing platforms being released (Quail et al. 2012). NGS platforms can sequence many DNA templates simultaneously, resulting in a large amount of data being produced (Hodkinson and Grice 2015; Méndez-García et al. 2018). Additionally, the cost per base under NGS is much cheaper when compared with technologies such as Sanger sequencing (Caporaso et al. 2012). The application of NGS methods in bacterial research has led to a more comprehensive understanding of bacterial diversity and community structures through metagenomics. NGS technologies include short-read sequencing such as pyrosequencing, Ion Torrent, SOLiD and Illumina, and long-read sequencing such as PacBio.

2.6.2.2 Ion Torrent

The Ion Torrent sequencing is a sequencing-by-synthesis platform and detects H^+ ions that are released as dNTPs and are incorporated into template DNA (Hodkinson and Grice 2015; Méndez-García et al. 2018). Like pyrosequencing, Ion Torrent DNA templates are loaded onto beads within microwells on a plate. When the H^+ ion is released into the solution, the pH change is detected by microdetectors inside the wells (Hodkinson and Grice 2015; Méndez-García et al. 2018). The change in pH is proportional to the number of nucleotides incorporated. Quail et al. (2012) notes that Ion Torrent displays a large bias when sequencing known AT-rich genomes. Regardless, Ion Torrent has become a widely-applied technology in a wide range of bacterial metagenomic studies exploring soil (Buscardo et al. 2018; Kaplan et al. 2019; Li et al. 2019; Ma et al. 2018; Oka and Uchida 2018), the human gut microbiome (Milani et al. 2013), periodontitis (Jünemann et al. 2012) and rumen (de la Fuente et al. 2014).

2.6.2.3 SOLiD

Sequencing by Oligo Ligation Detection (SOLiD) is a sequencing platform that is sequence-by-ligation method, whereby a labelled two-base-encoded probe binds to a DNA template sequence (Méndez-García et al. 2018). While this technology has a low error rate and high throughput, it has not achieved the same popularity as pyrosequencing or Illumina sequencing (Hodkinson and Grice 2015; Méndez-García et al. 2018). However Mitra et al. (2013) demonstrated that SOLiD sequencing is able to be utilised for accurate bacterial community metagenomics.

2.6.2.4 Illumina

The technology behind Illumina sequencing was first developed in 1998 and commercialised in 2006 (Balasubramanian 2015; Méndez-García et al. 2018). Illumina sequencing is a sequencing-by-synthesis process and begins with the construction of short, fragmented libraries. Adaptors are ligated at each end of the fragmented sequence and are added to an Illumina flow cell. Each flow cell contains up to eight lanes and the fragments are immobilised with oligonucleotides complementary to the adaptor sequences (Méndez-García et al. 2018). This will begin the process of amplification of double-stranded DNA (dsDNA) resulting in what is known as bridge amplification. The dsDNA is then denatured to form single-stranded DNA (ssDNA). Millions of ssDNA sequences are then amplified within the flow cell and sequencing will then begin. Illumina sequencing is based on all four nucleotides being labelled with different coloured fluorophores at the 3' end. When a nucleotide is incorporated, the specific colour of fluorescence is recorded, and a sequence is then generated based on the colour order emitted (Hodkinson and Grice 2015). Figure 2.13 below shows the overall process of Illumina sequencing.

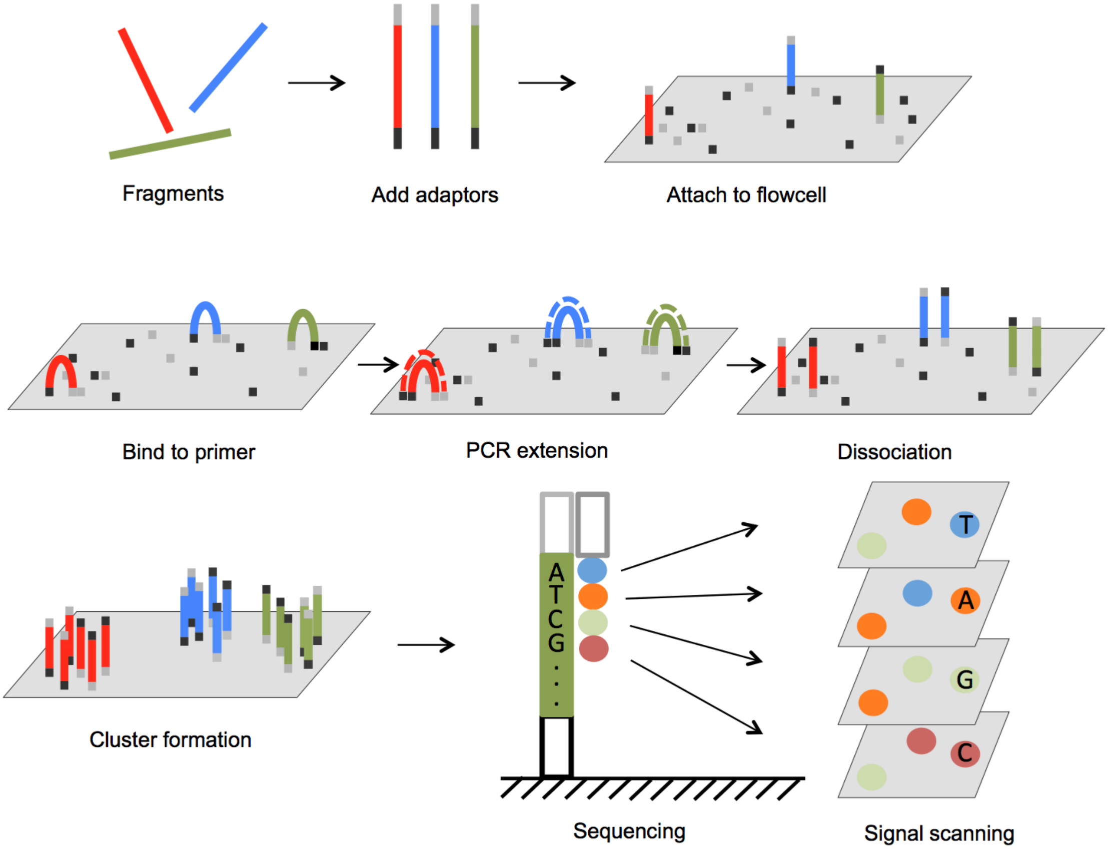


Figure 2.13: Outline of Illumina sequencing processes. The process is as follows: (1) Adaptors are annealed to each end of the DNA fragments. (2) Fragments bind to primer loaded flow cell and bridge PCR reactions amplify each bound fragment, fragments then dissociate, producing a cluster of fragments with the same sequence. (3) During each sequencing cycle, one fluorophore attaches to the bound DNA sequence and is detected by a laser and recorded. The fluorophore at the 3' end is removed before the next sequencing cycle starts (Lu et al. 2016b)

The observed raw error rate of Illumina sequencing is 0.26-0.8%, compared with 1.71% for Ion Torrent and 12.86% for PacBio (Quail et al. 2012). Illumina instruments will read up to 150 bases at a time, giving a total sequence yield of 1.5 to 2Gb per run (Quail et al. 2012; Shokralla et al. 2012). Current sequencing on Illumina platforms such as the Illumina HiSeq™ can achieve up to 600 Gb of data per run (Quail et al. 2012). Illumina sequencing also provides a cost efficient and higher resolution alternative to other DNA sequencing methods (Lange et al. 2014).

The limitations of community sequencing though Illumina platforms are mostly technical. Two common issues include PCR primer biases and differential DNA extraction efficiency from organisms within a complex community (Caporaso et al. 2012; Shokralla et al. 2012). Quail et al. (2012) also note that while Illumina sequencing displays near perfect coverage of GC-rich, neutral and moderately AT-rich genomes, there is large bias when sequencing very AT-rich genomes, leaving around 30% of the genome un-sequenced. However, Luo et al. (2012) found that because Illumina sequencing returned high sequence coverage for AT-rich genomes, this issue was resolved.

Given that the error rate is much lower than other sequencing platforms, Illumina sequencing has become the industry standard for bacterial whole genome sequencing, and as a means for identifying fastidious and non-cultivable organisms within a soil sample (Fadrosh et al. 2014; Forsberg et al. 2014; Hong et al. 2015; Lu et al. 2016b; Rajasekar et al. 2018; Shokralla et al. 2012; Vasileiadis et al. 2012).

2.6.2.5 Pacific Biosciences single molecular real-time sequencing

Pacific Biosciences (PacBio) is a third-generation sequencing platform that does not require PCR amplification, and is sequenced in real time using a single molecular real-time (SMRT) sequencing technology (Méndez-García et al. 2018). SMRT uses a flow cell with thousands of wells containing many zero-mode waveguide (ZMW) detectors which only allow light to pass through the bottom of the wells (Hodkinson and Grice 2015; Méndez-García et al. 2018). Each nucleotide is labelled with a different fluorophore. The polymerase for the sequencing reaction is fixed to the bottom of the well, so that when a nucleotide is incorporated into the target DNA sequence, the resulting fluorescence is detected by the ZMW (Méndez-García et al. 2018). The colour and duration of the emitted light is visualised and recorded continuously with a laser system.

While the PacBio SMRT system has a fast run time (2 hours) compared to other sequencing platforms, the observed error rate is 12-15% (Méndez-García et al. 2018; Quail et al. 2012). Longer read lengths accumulate more errors through DNA sequencing platforms, therefore shorter read lengths are more desirable (Shokralla et al. 2012). However, this error rate can be overridden by high coverage, but this then becomes a limitation of itself due to the sequencing cost of PacBio SMRT (Méndez-García et al. 2018). Despite this, several studies have utilised PacBio SMRT sequencing for *de novo* genome assemblies (Brown et al. 2014; Liao et al. 2015; Satou et al. 2014) and even for microbial community profiling (Pootakham et al. 2017).

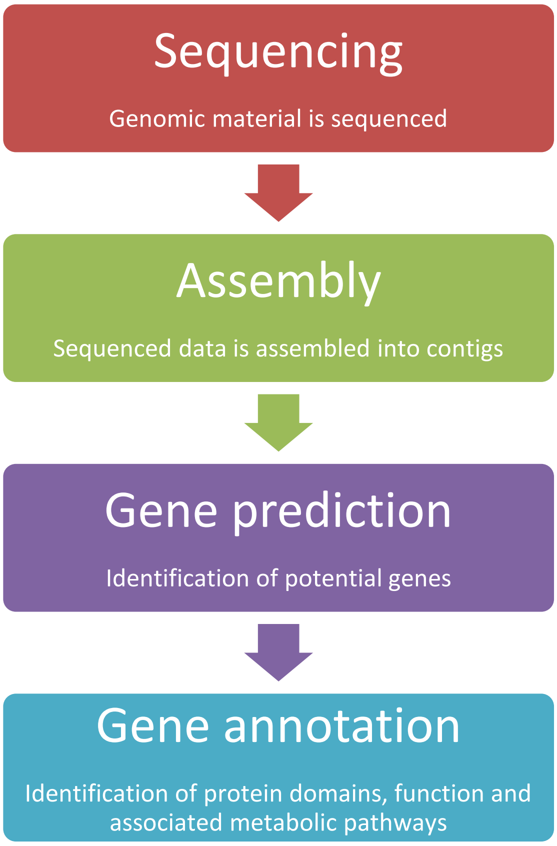
2.6.2.6 Oxford Nanopore Technologies

Oxford Nanopore Technologies (ONT) offer real-time sequencing platforms that uses nanopore technology for sequencing long DNA and RNA molecules (Jain et al. 2016; Tyson et al. 2017). These platforms are a third-generation sequencing method that can increase read lengths of NGS platforms by up to 1000-fold. Large DNA repeat regions can cause fragmentation of the genome during assembly when using NGS platforms, but this issue is largely resolved in third-generation sequencing methods. (Karlsson et al. 2015; Lu et al. 2016a; Tyson et al. 2017; Wick et al. 2017a). The ONT system works by applying the nanopore strand sequencing method, similar to PacBio SMRT. However, instead of using ZMW detectors, ionic current changes within the flow cell are detected when a nucleotide is added. The changes in ionic current are specific to each nucleotide (Lu et al. 2016a; Tyson et al. 2017).

The ONT systems have been used in conjunction with other sequencing methods, such as Illumina sequencing, to produce a complete *de novo* assembled genome (Karlsson et al. 2015; Tyson et al. 2017; Wick et al. 2017a). However, as mentioned previously, with increasing read length, the error rate of the technology increases. Early versions of ONT sequencing had error rates of 22-35%, while more recently this error rate has been improved to 5-10% (Jain et al. 2016; Lu et al. 2016a; Tyson et al. 2017). However, Karlsson et al. (2015) notes that when mapping the ONT reads back onto the Illumina assembled contigs, the overall consensus accuracy was 99.8%. Wick et al. (2017a) noted that while the combination of mapping ONT reads onto Illumina scaffolds resolved hybrid assemblies of *Klebsiella pneumonieae*, the error rate of ONT only reads was too high (1 error every 287 bp) to generate a reliable genome. The combined usage of ONT and NGS methods would therefore help to resolve genomes with high levels of fragmentation.

2.7 Bioinformatics

The advancements in sequencing capacity have, in turn, led to the rapid development of software tools for interpreting this data. Once assembled and analysed, genomic sequences can provide information regarding genes that encode proteins, RNA genes, protein structural motifs and repetitive and regulatory sequences (Overbeek et al. 2013). Comparing genes within a genus or species can highlight similarities between protein function in each organism. A brief description of the bioinformatic pipelines and programmes used in this thesis can be found in Table 2.6, Section 2.9. Figure 2.14 below shows a flow diagram of a typical bioinformatic pipeline.



Sequencing

Genomic material is sequenced

Assembly

Sequenced data is assembled into contigs

Gene prediction

Identification of potential genes

Gene annotation

Identification of protein domains, function and associated metabolic pathways

Figure 2.14: Overview of a typical bioinformatics pipeline Following genome sequencing, the data is assembled using assembly software to produce contigs. Potential genes then predicted and annotated from the sequence.

2.7.1 Genomic assembly

As discussed, most NGS techniques, such as Illumina sequencing, produce reads of DNA ranging from 100 – 200 bp in length. Assembling overlapping raw reads into contiguous sequences (contigs) will produce a consensus sequence of DNA in that region. When the ends of contigs overlap and are aligned correctly by a genome assembly programme, a complete genome can be reconstructed. While there are numerous computational programmes for *de novo* genomic assembly, the programmes used in this thesis include SPAdes (Bankevich et al. 2012), Unicycler

(Wick et al. 2017b) and Bandage (Wick et al. 2015). A brief description of these programmes can be found in Table 2.5, Section 2.9.

2.7.2 Gene prediction and annotation

Gene annotation is usually comprised of running an automated annotation pipeline with subsequent manual curation (Richardson and Watson 2013). Annotation pipelines use the homologous genes of closely related annotated reference genomes to infer information onto the new genome. This can lead to poor annotation and errors, as if there is no homologue identified on the reference genome, the protein is usually labelled a 'hypothetical protein'. Manual curation can correct these errors, however, it is time consuming and no longer feasible with the number of genomes now available (Richardson and Watson 2013). Inferring protein function from a closely related annotated genome may also leave some undesirable gaps in the annotation if the areas of interest on genome Y are not present on genome X (Richardson and Watson 2013).

The two genome annotation programmes used in this study are Rapid Annotation using Subsystems Technology (RAST) (Aziz et al. 2008) and Prokka (Seemann 2014). RAST utilises the GLIMMER3 software for gene prediction, while Prokka primarily uses BLAST+ to search multiple databases.

Overbeek et al. (2013) notes that RAST employs a multistep process for bacterial annotation, with frequent rechecking of uncalled genes, or genes without a subsystem. While RAST primarily uses GLIMMER3 for initial prediction, BLAST+ is also utilised in many steps of the RAST annotation process. Similarly, Prokka follows a hierarchy of annotation that ranges from highly specific annotation to general annotation. The hierarch is as follows: 1) a set of user-provided annotated proteins; 2) bacterial proteins available on UniProt (Apweiler et al. 2004); 3) proteins from a completed RefSeq genome within the specified genus, and; 4) searching for hidden Markov models within databases. If no matches are found, the protein is labelled 'hypothetical protein' (Seemann 2014).

Both the RAST and Prokka pipelines have been used extensively in conjunction for functional annotation of bacterial genomes (Arahal et al. 2016; Ee et al. 2015; Horn et al. 2016; Miranda et al. 2019; Sayed et al. 2017).

2.8 Comparative genomics

Comparative genomics is an important tool for understanding gene function through sequence comparisons, gene order and regulation. Comparative genomics often involves comparing large parts of different genomes to study biological similarities and differences (Bhasin and Raghava 2006; Haubold and Wiehe 2004). This enables us to improve our understanding of the genetic basis of diversity in organisms (Bhasin and Raghava 2006; Haubold and Wiehe 2004). Therefore, comparative genomic approaches will often start with a search for orthologous sequences within a genus, followed by a genomic alignment with a reference organism (Yao et al. 2015a).

2.8.1 Orthology analysis

Homologous genes and their relationships can be identified through numerous bioinformatic tools. Within homology, evaluation of orthologous sequences is important for genomic annotation and phylogenetic inference (Nichio et al. 2017; Wang et al. 2015). Orthologous genes are gene clusters, within different species, that originated from a single gene in the last common ancestor. Orthology analysis and comparison can help to improve understanding of genomic structure and protein function (Wang et al. 2015). Establishing orthologous relationships between a well-studied organism and a new or less-studied organism helps to improve our understanding of phylogeny. As more genomic data becomes publicly available, whole-genome orthologous comparisons across different genera becomes more achievable.

2.8.1.1 Clusters of Orthologous Genes

The Cluster of Orthologous Genes (COG) database was established in 1997 as a means for phylogenetic classification of the proteins within bacteria, archaea and eukaryotes (Tatusov et al. 2000). COGs were established on the all-against-all basis of analysis of genome-specific best hits through the BLAST algorithm. The premise of COGs is that proteins that are observed to be similar to each other belong to the same orthologous family (Huerta-Cepas et al. 2016; Tatusov et al. 2000). The original COG database is no longer maintained, however, numerous comparative bioinformatic programmes are still based on the COG principle. Three such programmes used in this thesis are Kyoto Encyclopaedia of Genes and Genomes (KEGG), eggNOG v4.5.1 (http://eggnogdb.embl.de/) and OrthoVenn (http://www.bioinfogenome.net/OrthoVenn/).

2.8.1.1.1 KEGG

The KEGG database is a tool for biological interpretation of gene molecular function. Gene function is recorded by the KEGG orthology database, and can be displayed as a KEGG pathway map, BRITE hierarchy or KEGG modules (Kanehisa et al. 2016).

The KEGG database was established in 1995 and is an open-access resource for biological interpretation of genomic sequences. KEGG pathway maps are a representation of experimental knowledge on metabolism and other cellular functions. Each pathway map contains a network of known molecular interactions that link specific genes to protein products through the use of enzyme commission (EC) numbers (Kanehisa et al. 2006; Kanehisa et al. 2016). This allows users to compare the gene content of the query genome with the KEGG pathway database, allowing for examination of encoded genomic pathways and their associated functions (Kanehisa et al. 2006). A limitation of the KEGG database is that some EC numbers are updated by the official KEGG database but are not updated by servers that employ the KEGG software for annotation. This can lead to confusion surrounding if the gene encoding the enzyme is present, and double checking of the annotation is required.

2.8.1.1.2 EggNOG

EggNOG v4.5.1 is a publicly available web-based resource that identifies and assigns protein sequences to established COGs (Huerta-Cepas et al. 2016; Nichio et al. 2017). EggNOG employs a graph-based clustering algorithm to produce genome wide orthology relationships (Huerta-Cepas et al. 2016). EggNOG extracts protein sequences from sequence data and applies an all-against-all pairwise similarity to determine COG relationships using Smith-Waterman alignments (Huerta-Cepas et al. 2016). Following this, functional annotations and descriptions are assigned to each orthologous group. Functional annotations are collected from multiple sources to establish appropriate COG assignment based on shared ortholog descriptions, conserved protein domains and gene ontology assignments (Huerta-Cepas et al. 2016). EggNOG groups orthologous genes based on the single-letter functional categories established by the COG database.

2.8.1.1.3 OrthoVenn

OrthoVenn is a web-based programme that relates orthologous clusters across multiple species and identifies the overlap between species. Similar orthologous sequences will be identified by OthroVenn as an overlap and will be reported by the software (Wang et al. 2015). The relationship between the organisms is illustrated in a Venn diagram. The orthologous clusters that are unique to each organism in the search, as well the clusters shared by all query searches, are displayed in an interactive interface. OrthoVenn also provides information regarding the number of clusters within each uploaded genome, as well as additional resources for further protein structure investigation (Wang et al. 2015).

The OrthoVenn software applies the methods of “all vs. all”, BLAST, MCL and the OrthAgogue tool to identify orthologous relationships (Nichio et al. 2017; Wang et al. 2015). Six protein sequences can be compared at one time using OrthoVenn.

2.8.2 Whole genome comparisons

While 16S rRNA gene comparison provides an objective method for ‘identifying’ a specific organism, there are critical limitations in its use at the species level (Yoon et al. 2017). Several studies have noted that almost identical 16S rRNA gene sequences do not ensure that the two strains belong to the same species (Fox et al. 1992; Kim et al. 2014; Yoon et al. 2017). The implementation of whole genome analysis for bacteria helps to achieve a more complete taxonomic classification that only relies on reference genes such as the 16S rRNA gene.

2.8.2.1 *In silico* DNA-DNA hybridisation

DNA-DNA hybridisation (DDH) is a wet-lab based method that is currently used as the taxonomic benchmark for new bacterial species description. DDH is employed when the 16S rRNA gene similarity is above 97% with closely related organisms (Meier-Kolthoff et al. 2013). However, DDH is largely considered to be error-prone, as well as tedious, laborious and time consuming (Klenk and Göker 2010; Schleifer 2009). Several experimental repetitions are required to establish statistical confidence in wet-lab DDH methods, and the technique is only conducted by specialised laboratories (Meier-Kolthoff et al. 2013). These issues have recently given rise to computerised programmes to replace wet-lab DDH.

In silico DDH software programmes such as Genome-to-Genome Distance Calculator 2.1 (GGDC) (http://ggdc.dsmz.de/ggdc.php) have gained popularity as a simple method for mimicking wet-lab DDH analysis (Meier-Kolthoff et al. 2013). Under DDH criteria, if the genomic DNA of two genomes is below 70%, the two respective organisms are regarded as distinct species (Meier-Kolthoff et al. 2013). Computerised DDH methods aim to resolve the issues currently surrounding DDH to ensure that the results are reproducible, accurate and in line with the species boundaries of wet-lab DDH (Auch et al. 2010; Meier-Kolthoff et al. 2013).

During *in silico* DDH analysis using GGDC, genome X is compared using BLAST against genome Y. The alignment of genome X and Y is completed in a single pass using all available sequence information from each genome. The resulting matches between the genomes is called a ‘high-scoring segment pair’ and represents local alignments that are statistically significant if the e-value is below 10^{-2} (Meier-Kolthoff et al. 2013). Following this, the high-scoring segment pairs are converted into a distance value. To alleviate segment-overlap due to paralogous genes, three different filtering approaches are used: ‘greedy’, ‘greedy-with-trimming’ and ‘coverage’. The ‘greedy-with-trimming’ formula is generally recommended by GGDC, as this formula preserves more information and is more valuable when attempting to infer phylogenetic relationships (Meier-Kolthoff et al. 2013).

Recently, several studies have utilised *in silico* DDH methods when establishing a new species (Dunlap et al. 2017; El Houmami et al. 2017; Miller et al. 2016; Murra et al. 2018; Tong et al. 2015).

2.8.2.2 Average Nucleotide Identity

The original approach to overcome the 16S rRNA problem was to use DDH as a complement to the 16S rRNA gene. However, this method is often inaccurate and tedious (Goris et al. 2007; Lee et al. 2016) and more reliable overall genome relatedness indices were proposed as appropriate programmes for new species identification (Goris et al. 2007; Lee et al. 2016; Yoon et al. 2017). Computational programmes such as average nucleotide identity (ANI) have become widely used in place of DNA-DNA hybridisation (Lee et al. 2016; Yoon et al. 2017). ANI is calculated from two genome sequences and is expressed as a percentage identity (Lee et al. 2016; Yoon et al. 2017). During the ANI calculation, the genomes sequence of the query strain is divided into 1,020 bp-long fragments. Each fragment is then searched against the whole genome sequence of reference strain genome using the NCBI BLASTn algorithm (Lee et al. 2016; Yoon et al. 2017). The BLASTn programme calculates the nucleotide identity of fragments between the query and reference genome. The ANI output percentage is the mean of the nucleotide identity values. Recently, more refined ANI programmes that search for orthologous fragments between the two sequences have been developed to show reciprocal best hit BLASTn searches (Yoon et al. 2017). This is shown in Figure 2.15 below.

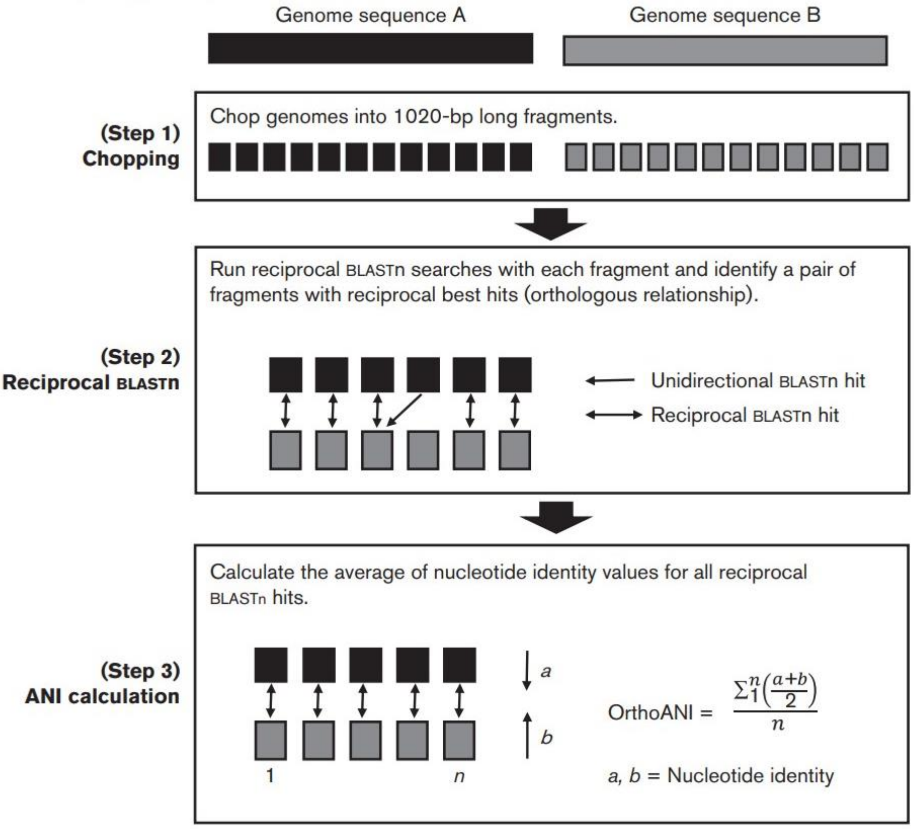

Genome sequence A
Genome sequence B
(Step 1)
Chopping
Chop genomes into 1020-bp long fragments.
(Step 2)
Reciprocal BLASTN
Run reciprocal BLASTn searches with each fragment and identify a pair of fragments with reciprocal best hits (orthologous relationship).
Unidirectional BLASTn hit
Reciprocal BLASTn hit
(Step 3)
ANI calculation
Calculate the average of nucleotide identity values for all reciprocal BLASTn hits.
a
b
1
n
OrthoANI = $\frac{\sum_1^n\left(\frac{a+b}{2}\right)}{n}$
a, b = Nucleotide identity

Figure 2.15: Schematic diagram for the OrthoANI algorithm (Yoon et al. 2017).

The proposed ANI value for identification of a possible new species is <95% similarity (Lee et al. 2016; Yoon et al. 2017).

2.8.2.3 Mauve

Other computational programmes such as Mauve (Darling et al. 2004) have been developed for visualising evolutionary events such as genome rearrangement or inversion. The Mauve software is able to identify conserved genomic regions, as well as identify breakpoints, rearrangement events and inversions across several genomes (Darling et al. 2004). Mauve uses locally collinear blocks (LCB) to identify homologous regions of sequences shared by the selected genomes. The weight of an LCB provides a measure of confidence that the identified section is a true genome rearrangement instead of a false match (Darling et al. 2004). Using a high LCB weight allows the user to identify highly likely genome rearrangements, while selecting a lower LCB weight allows the user to observe a more sensitive alignment for smaller genome rearrangements (Darling et al. 2004). Figure 2.16 shows how the Mauve algorithm identifies and displays LCBs in genomic regions.

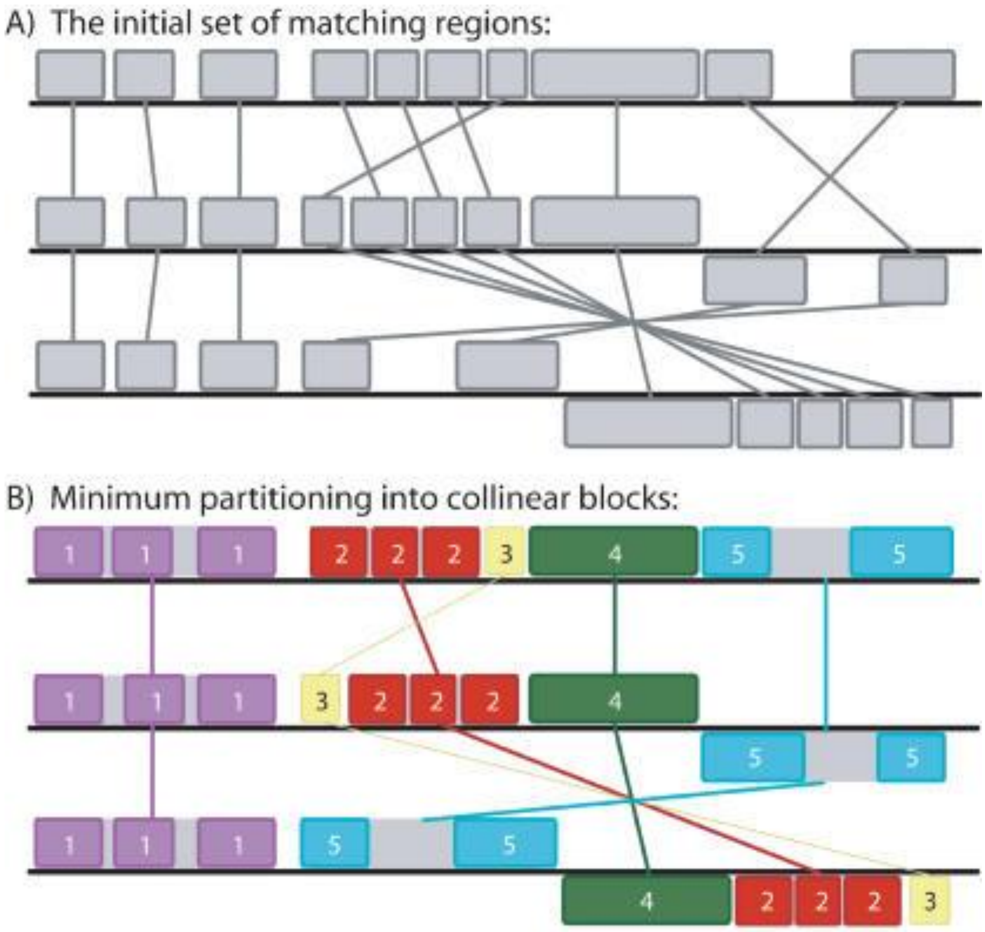

A) The initial set of matching regions:
B) Minimum partitioning into collinear blocks:
1
2
3
4
5

Figure 2.16: Mauve algorithm for identifying LCBs. A) Genome regions are identified and located on each genome. B) The matching genomic regions are partitioned into collinear blocks. As seen above, block 3 has been rearranged in the second genome, while block 5 has been rearranged in the third genome when compared to the first genome as a reference (Darling et al. 2004).

The next iteration of Mauve, called progressiveMauve, builds on the strengths of the original algorithm. ProgressiveMauve can align more diverse genomes than the original application, allowing for more comprehensive comparisons to be conducted (Darling et al. 2010). Other additions to the earlier versions of Mauve, such as the Mauve Contig Mover allow for the reordering of contigs relative to the reference sequence, resulting in a more powerful exploration of Mauve alignments (Rissman et al. 2009).

2.8.2.3 PhyloPhlAn

Other methods for whole genome taxonomic classification include multi-gene approaches. Using multi-gene analysis is more robust that using single gene approaches, as the resulting phylogenetic analysis offers a more complete evolutionary relationship between species, as well as taxonomic identification.

PhyloPhlAn is a phylogenetic tool that uses over 400 conserved proteins to establish accurate phylogenetic relationships between species (Segata et al. 2013). PhyloPhlAn can discriminate between closely related species with a high level of confidence, and as such, has been able to create a high-resolution phylogeny of microbial genomic sequences. Ubiquitously conserved bacterial proteins were detected during the construction of the PhyloPhlAn software, with the most accurate tree of life being produced by >4,600 aligned amino-acid positions from 400 sampled proteins (Segata et al. 2013).

Several studies have utilised the PhyloPhlAn software for identifying novel species of closely related strains (Kougias et al. 2017; Sangal et al. 2015) and to help with the reclassification of misclassified species (Sedlar et al. 2017; Segata et al. 2013).

2.8.3 Protein tertiary structure prediction

Protein tertiary structure prediction involves uploading the amino acid (aa) sequence of a protein to a database for structure modelling. Matching the sequence of interest to a library of known structures is often more appropriate for tertiary structure prediction than error-prone simulated folding (Kelley et al. 2015). Current protein tertiary structure prediction is based on using large databases of sequences to construct an evolutionary profile and use this to model the query sequence (Kelley et al. 2015).

Phyre2 (http://www.sbg.bio.ic.ac.uk/) is a web-based tool for protein prediction. Phyre2 protein prediction is broken up into four stages: 1) homologous sequence gathering; 2) fold library scanning; 3) loop modelling, and; 4) side-chain placement. During stage 1, the query protein sequence is scanned against other library sequences using HHblits. The secondary structure of the query protein is then predicted via PSIPRED using multiple sequence alignment. A query hidden Markov model (HMM) is then generated. Stage 2 involves scanning HMMs of known proteins to predict the query protein structure. Stage 3 involves correcting the indels in the predicted protein model. Stage 4 involves adding aa side chains to generate the final predicted Phyre2 model (Kelley et al. 2015). The result of Phyre2 is a 3D protein structure. A more robust protein prediction can be achieved using the 'intensive mode' of Phyre2. 'Intensive mode' is aimed at maximising the confidence and coverage of the query sequence (Kelley et al. 2015).

One limitation of the Phyre2 server is that the protein models are based on homology. If no homology is detected, the modelling of the query protein will be unreliable (Kelley et al. 2015). A second limitation of Phyre2 is that it is difficult to predict the structural effects of point mutations. While it is possible for the Phyre2 server to predict the phenotypic effects of point mutations, it is difficult to estimate the wider structural effect of this mutation (Kelley et al. 2015). Despite these limitations, Phyre2 is a widely used protein prediction software that can predict the 3D structure of a submitted sequence, as well as investigate model quality, function and mutation effects using Phyre Investigator (Kelley et al. 2015).

2.9 Summary of literature review

This literature review has discussed methods of analysis that are widely used for 16S rRNA gene defined community analysis, whole genome comparative genomics and the function of UV repair mechanisms. Through this literature review, current methods of analysis have been identified that will be appropriate to achieve the Aims of this study.

As many bacteria are unable to be cultured onto agar using current isolation methods. In order to observe the entire microbial communities of the soil, a gene-defined community analysis will be conducted to investigate abiotic community drivers. A comparison of the Dry Valleys and the Namib Desert is important for continuing our understanding of the role that bacteria play within these desert communities. Current accepted methods in soil chemistry analysis will be used (Blakemore et al. 1981; Olsen et al. 1954; Rayment and Higginson 1992; Sader et al. 2004). For the community analysis, amplicon sequence variants, alpha and beta diversity and redundancy analysis will be used. These methods are in line with current research (Archer et al. 2019; Lee et al. 2018).

While there are many ways in which to study bacteria, the nature of this thesis relies on the culture of viable bacterial isolates. Therefore, bacteria will be isolated onto agar for UV exposure. These isolates will be exposed to UV using a modified plate drop method that was developed for this thesis. Isolates found to be resistant or sensitive to UV will be identified using Sanger sequencing of the 16S rRNA gene. This will allow for phylogenetic relationships between the UV resistant and UV sensitive bacteria to be established.

Finally, the Illumina HiSeq platform will be used for whole genome sequencing. A wide range of techniques will be used to investigate the whole genome of bacteria in this study. The assembly programmes selected for this thesis are SPAdes (Bankevich et al. 2012) and Unicycler (Wick et al. 2017b). Bandage (Wick et al. 2015) will be used to visualise the assembly. The genome annotations RAST (Aziz et al. 2008) and Prokka (Seemann 2014) will be used for automated annotation. Several comparative genomics tools will be used to investigate different aspects of the assembled and annotated genomes. These tools have been selected for molecular function, orthologous clustering, genome similarity, genome alignment, taxonomy and predicted protein modelling. A brief description of the computational and web-based tools used in this thesis can be seen in Table 2.6. The specific application of these programmes and their parameters will be discussed in Chapters 5 and 6.

Table 2.6: Description of computational programmes and online tools used for genome assembly and annotation in this thesis.

Programme	Description
Genomic assembly	
SPAdes	A genome assembly algorithm for bacterial data sets (Bankevich et al. 2012)
Unicycler	A tool for assembling bacterial genomes from a combination of long and short reads. Unicycler works in conjunction with SPAdes to produce a highly accurate genome assembly (Wick et al. 2017b)
Bandage	A tool for visualising assembly graphs generated by genomic assembly programmes. Bandage can be used to visualise the assembly graph produced by Unicycler (Wick et al. 2015)
Genome annotation	
RAST	A web-based server (http://rast.nmpdr.org) for automated annotations of bacterial genomes (Aziz et al. 2008)
Prokka	A software package for rapidly annotating prokaryotic genomes (Seemann 2014)
Comparative genomics	
KEGG	A database for biological interpretation of genome sequences and the molecular functions of orthologous genes (Kanehisa et al. 2016)
eggNOG	A web-based server (http://eggnogdb.embl.de/) for establishing hierarchical orthology through functional COG annotations (Huerta-Cepas et al. 2016)
OrthoVenn	A web-based server (http://www.bioinfogenome.net/OrthoVenn/) for establishing relationships between orthologous clusters across multiple species. Results are displayed as a Venn diagram (Wang et al. 2015)
In silico DDH	A web-based server (http://ggdc.dsmz.de/) for digital DDH accurate delineation of prokaryotic species (Meier-Kolthoff et al. 2013)
OrthoANI	A web-based server (https://www.ezbiocloud.net/tools/ani) for calculating average nucleotide identity for taxonomic classification (Lee et al. 2016)
Mauve	A multiple alignment tool for identification of conserved genomic sequences within query organisms, and analysis of rearrangement events (Darling et al. 2004)
PhyloPhlAn	A phylogenetic tool for taxonomic classification using a multi-gene approach (Segata et al. 2013)
Phyre2	A web-based server (http://www.sbg.bio.ic.ac.uk/) for automated tertiary protein structure prediction (Kelley et al. 2015)

Chapter 3: Analysis of soil chemistry and bacterial community structure from the McMurdo Dry Valleys and the Namib Desert

3.1 Introduction

Soil microorganisms play a critical role in deserts, functioning as the main cyclers of carbon and nitrogen in these extreme environments (Cowan et al. 2011; Johnson et al. 2017; Makhalanyane et al. 2013). In addition to the characteristic low precipitation, seasonal temperature fluctuations and oligotrophy, microorganisms in these environments also have to contend with high UV irradiance (Fierer et al. 2012; Kouskoumvekaki et al. 2014; McKenna et al. 2010). Recent advances in high-throughput sequencing platforms and bioinformatics tools have facilitated in-depth comparative studies of desert microbial ecology and function (Fierer et al. 2012; Kouskoumvekaki et al. 2014; McKenna et al. 2010). These advances have allowed microbial ecologists to characterise and compare the phylogenetic, functional and taxonomic diversity of microbial soil communities across broad geographical gradients (Fierer et al. 2012; Soon et al. 2013; Subramanian et al. 2013). From these studies, an understanding of how microbial diversity is related to physical, chemical and biological characteristics of these ecosystems is beginning to develop (Fierer et al. 2012) and how these characteristics influence microbial diversity. Despite these advances, comparatively little is known regarding the effect of specific environmental stressors on microbial community diversity and function (Van Goethem et al. 2016).

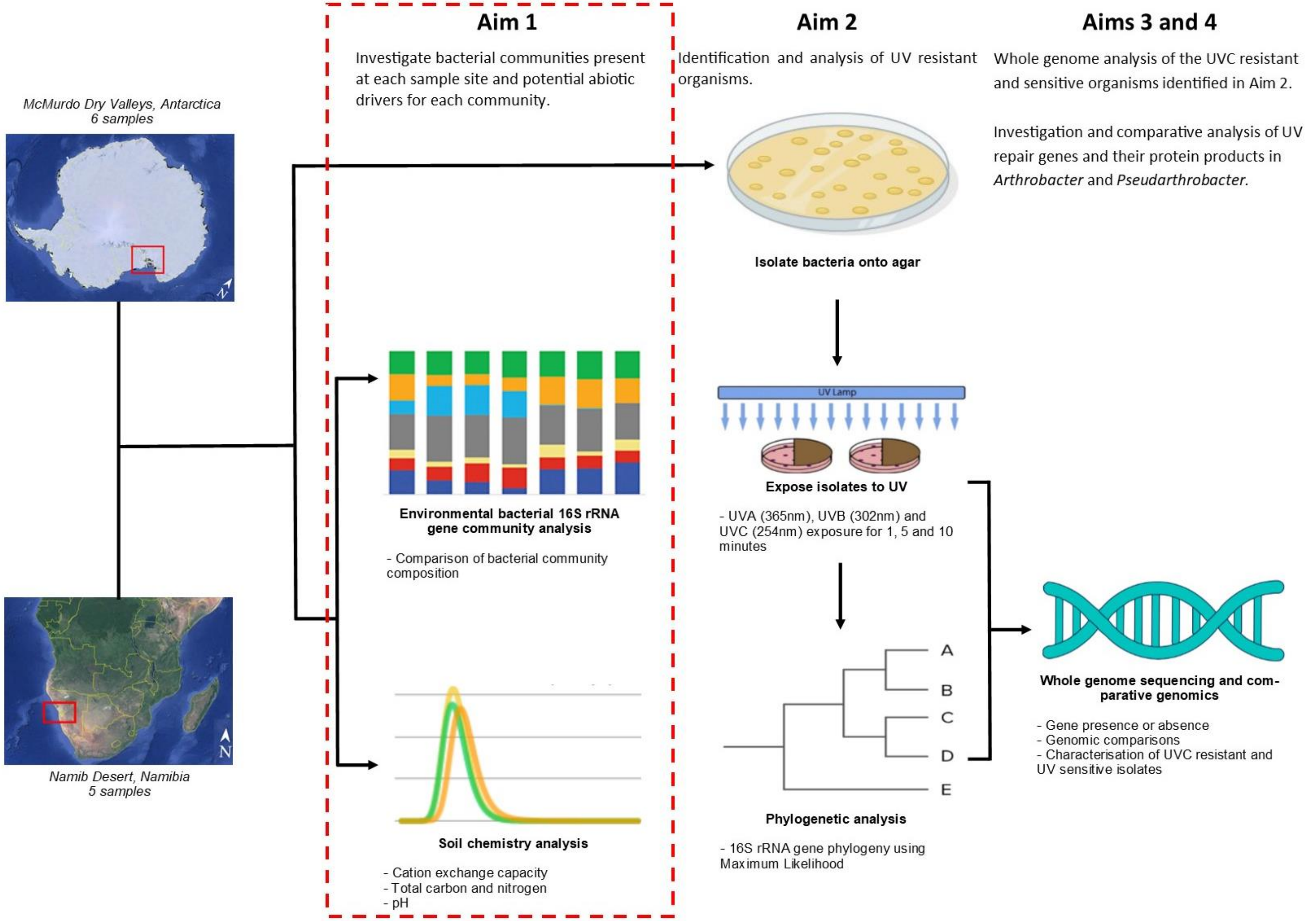
Factors influencing microbial community structure extend from local (Pointing et al. 2009) to environmental gradient scales (Langenheder et al. 2011; Lindström and Langenheder 2012; Van Goethem et al. 2016). While understanding of the mechanisms influencing beta diversity in soil communities composition remains limited (Makhalanyane et al. 2013), it is thought that drivers of beta diversity exist in both spatial (Martiny et al. 2011) and temporal scales (Langenheder et al. 2011; Lindström and Langenheder 2012; Makhalanyane et al. 2013).

As previously mentioned, the Dry Valleys provide an ideal environment for assessing how restrictive conditions may shape microbial diversity. Similarly, analysis of the Namib Desert also provides valuable insight into how microbial communities are shaped in hot arid deserts. Recent studies have shown significant differences in soil community compositions and species richness between different global locations (Fierer et al. 2012). There is evidence to suggest that abiotic factors, such as soil temperature, moisture, and nutrient availability have also been shown to influence microbial community soil structure (Makhalanyane et al. 2013; Van Goethem et al. 2016; Van Horn et al. 2013).

Previous studies into the bacterial communities of the Dry Valley soils have observed prominent populations of Actinobacteria, Proteobacteria, Acidobacteria and Cyanobacteria (Cary et al. 2010; Chan et al. 2013; Pointing et al. 2009; Yung et al. 2014), while notable populations of Actinobacteria and Proteobacteria have been observed in the Namib Desert (Makhalanyane et al. 2013) (Section 2.1.2). External factors such as pH, ion concentrations and total carbon have been noted as drivers for community structure in both the Dry Valleys and the Namib Desert (Johnson et al. 2017; Pointing et al. 2009). Continuing investigations into drivers of community diversity will help to improve understanding of the communities that UV resistant bacteria can arise from. To date, little information is available on the desert community structures that give rise the UV resistant bacteria.

In this Chapter analysis into the microbial soil communities from arid deserts and their drivers is presented. To achieve this, the use of traditional culture media, soil chemistry analysis and environmental bacterial 16S rRNA gene-defined community analysis methods were used. Culture media were selected from literature to assist in identifying aerobic heterotrophic organisms (Di Capua et al. 2011; Fernández Zenoff et al. 2006a; Kim et al. 2012; Kurth et al. 2015). Organisms not cultivated based on the method used were not tested for UV resistance, and will be a limitation of this study. By directing the focus to aerobic heterotrophic organisms, this investigation does not aim to be an exhaustive search; further investigations into currently uncultured organisms will be required in future work. In addition, autotrophic organisms were not selected for using culture media during this thesis and would also require further investigation.

The aim of this investigation was to investigate bacterial communities present at each sample site and potential abiotic drivers for each community. To achieve this Aim, three objectives were outlined: (1) determine soil chemical composition and investigate if this is a driver for community structure; (2) investigate the community structure of sample sites using 16S rRNA gene-defined community diversity; (3) investigate abiotic drivers of bacterial community structure of the Dry Valleys and Namib Desert.



Aim 1

Investigate bacterial communities present at each sample site and potential abiotic drivers for each community.

Aim 2

Identification and analysis of UV resistant organisms.

Aims 3 and 4

Whole genome analysis of the UVC resistant and sensitive organisms identified in Aim 2.

Investigation and comparative analysis of UV repair genes and their protein products in *Arthrobacter* and *Pseudarthrobacter*.

McMurdo Dry Valleys, Antarctica
6 samples

Namib Desert, Namibia
5 samples

Environmental bacterial 16S rRNA gene community analysis

- Comparison of bacterial community composition

Soil chemistry analysis

- Cation exchange capacity
- Total carbon and nitrogen
- pH

Isolate bacteria onto agar

Expose isolates to UV

- UVA (365nm), UVB (302nm) and UVC (254nm) exposure for 1, 5 and 10 minutes

Phylogenetic analysis

- 16S rRNA gene phylogeny using Maximum Likelihood

Whole genome sequencing and comparative genomics

- Gene presence or absence
- Genomic comparisons
- Characterisation of UVC resistant and UV sensitive isolates

Figure 3.1: Breakdown of schematic overview of this thesis. Aim 1 will be covered in this Chapter.

3.2 Materials and methods

3.2.1 Sample collection

Top soil was collected from six locations along two transects in the Wright Valley and along Bull Pass in the McMurdo Dry Valleys by Dr Stephen Archer (The Institute for Applied Ecology New Zealand, Auckland University of Technology) (Table 3.1, Figure 3.2). After soil pavement pebbles were removed, approximately 50 g of surface to 2 cm depth of soil were collected at each location over a one-week period in January 2017. Sites were selected to provide a broad range of habitats including the highly productive valley and low productivity Bull Pass.

Namib Desert soils were collected from five locations along road C14 within the Namib-Naukluft National Park by Dr Barbara Breen (The Institute for Applied Ecology New Zealand, Auckland University of Technology), Dr Jean-Baptiste Raymond (University of Pretoria Centre for Microbial Ecology and Genomics, South Africa) and Professor Don Cowan (University of Pretoria Centre for Microbial Ecology and Genomics, South Africa) (Table 3.1, Figure 3.3). Approximately 70 g of top soil was collected at four points in each location during a two-day period in April 2017.

Table 3.1: Summary of site details.

Sample ID	**Location**	**GPS Coordinates**		**Elevation (asl)**
		Latitude	Longitude	
BP1	Dry Valleys	77.70416667	161.95583333	549.7 m
BP2	Dry Valleys	77.53638889	161.87972222	649.7 m
V1	Dry Valleys	77.54944444	161.80194444	153.0 m
V2	Dry Valleys	77.64861111	161.80972222	98.3 m
BW	Dry Valleys	77.56861111	161.99638889	167.4 m
Camp	Dry Valleys	77.56861111	161.89388889	141.4 m
T2	Namib Desert	23.25527778	14.74833333	117.7 m
T4	Namib Desert	23.02111111	15.00194444	312.8 m
T6	Namib Desert	23.30000000	15.14916667	529.2 m
T8	Namib Desert	23.28722222	15.36055556	613.4 m
T10	Namib Desert	22.43166667	15.53416667	748.6 m

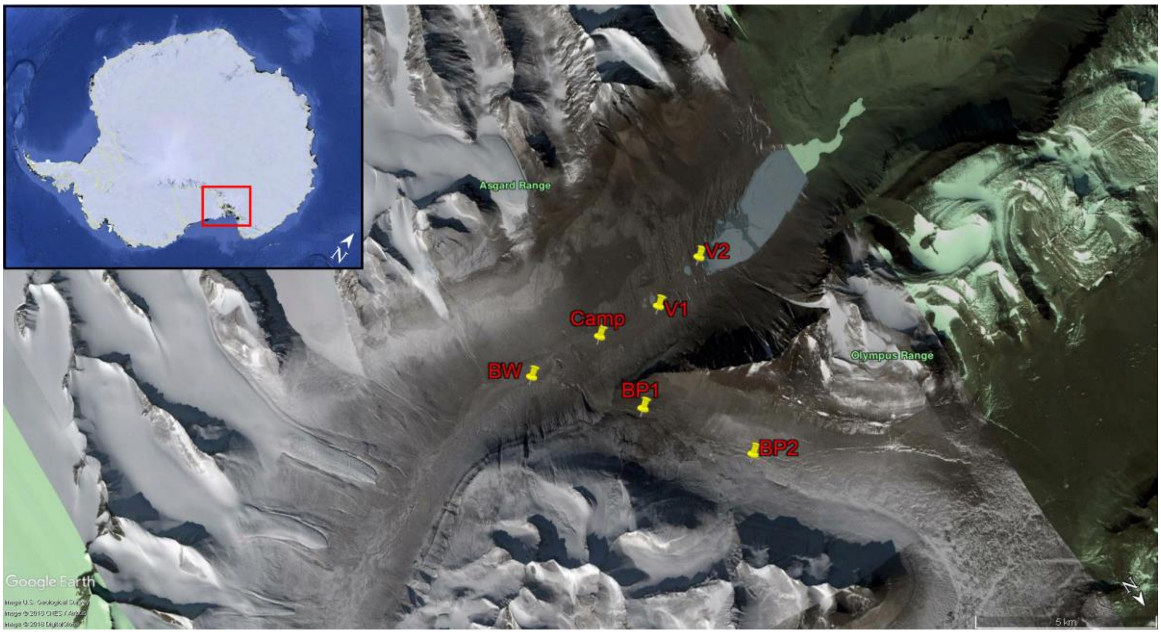


Figure 3.2: Dry Valley soil sample sites for January 2017 (Google Earth 2018a).

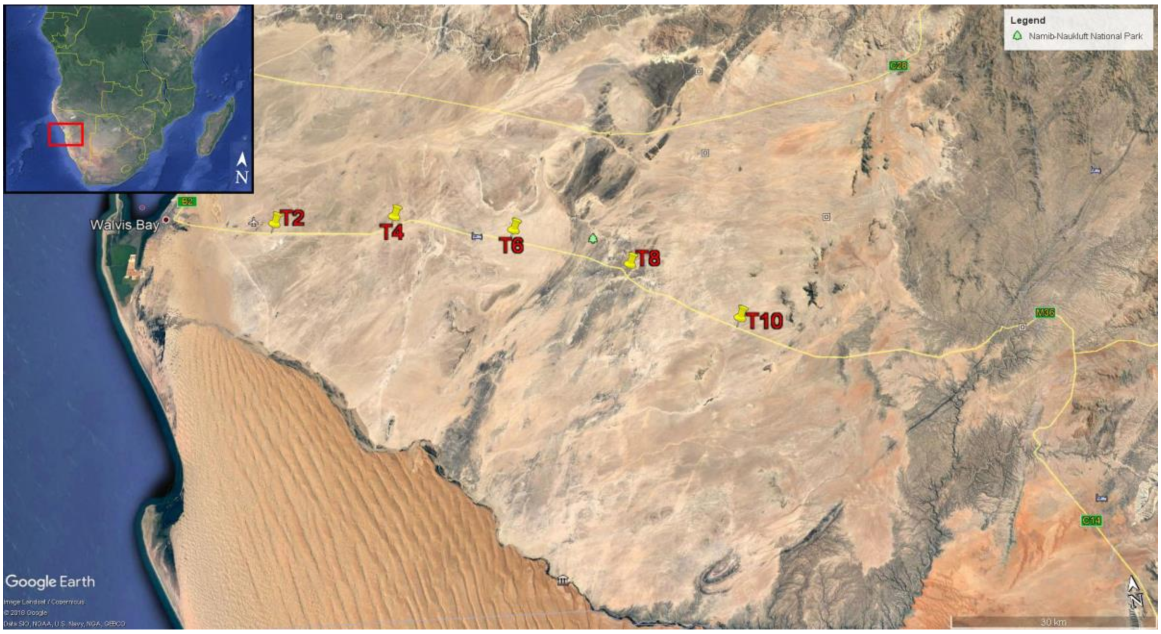


Figure 3.3: Namib Desert soil sample sites for April 2017 (Google Earth 2018b).

All samples were collected in sterile 50 mL centrifuge tubes using all reasonable precautions to prevent sample contamination. Samples were transported and stored at 4°C until analysis.

3.2.2 Soil chemistry analysis

Chemical analysis of the soil samples was carried out by Ravensdown Ltd. (Christchurch, New Zealand) using standardised industry procedures. The pH was measured using the slurry technique, which consists of a standard ratio of 1:2.5 soil:deionised water and allowing samples to settle for between 1 and 4 hours (Blakemore et al. 1981). Olsen phosphorus was measured by extracting air-dry soil with 0.5 M $NaHCO_3$, with an adjusted pH of 8.5 (Olsen et al. 1954). The soil extraction was dried for 30 min at a soil:solution ratio of 1:20. Soil cations were determined by equilibrating the soil with 1 M ammonium acetate at pH 7 for 30 min, followed by mechanical shaking at a soil:solution ratio of 1:20 (Rayment and Higginson 1992). The cations were measured on a microwave plasma-atomic emission spectrometer (Agilent, Santa Clara, CA, USA). Total carbon and nitrogen were analysed by a combustion method using an Elementar Vario Max Cube Analyser (Elementar, Langenselbold, Germany) (Sader et al. 2004).

A Welch t-test was performed on the soil physicochemical data using R version 3.5.0 (R Core Team, Vienna, Austria) to identify statistically significant differences in soil chemistry between sampling locations.

3.2.3 DNA extraction for environmental community analysis

Total eDNA was extracted from 0.75 g ±0.025 of soil using a modified bead-beating method (Warren-Rhodes et al. 2019). The soil was added to 0.5 g of 0.1 mm silica-zirconia beads. To each sample, 320 µL of phosphate buffer (100 mM NaH_2PO_4) and 320 µL of SDS lysis buffer (100 mM NaCl, 500 mM Tris pH 8.0, 10% SDS) were added and samples were homogenised using the FastPrep®-24 (MP Biomedicals, Ohio, USA) at setting 5.5 for two 30 sec runs. Samples were centrifuged at 18,506 x *g* for 3 min (20°C) using a 5810 Centrifuge (Eppendorf, Hamburg, Germany). Following this, 230 µL of cetyltrimethylammonium bromide-polyvinylpyrrolidone (CTAB) extraction buffer (100 mM Tis-HCl, 1.4 M NaCl, 20 mM EDTA, 2% CTAB, 1% polyvinylpyrrolidone and 0.4% β-mecaptoethanol) was added to each sample. Samples were vortexed for 10 sec before incubation at 60°C and 300 rpm on a Incubated Shaker SI-300R (Acorn Scientific, New Zealand) for 30 min. Samples were centrifuged at 18,506 x *g* for 1 min and then 400 µL of chloroform:isoamyl alcohol (24:1) was added. Samples were then vortexed for 10 sec and centrifuged at 18,506 x *g* for 5 min. The upper aqueous layer was pipetted into a new 1.5 mL microcentrifuge tube and 550 µL of chloroform:isoamyl alcohol (24:1) was added to the layer. Samples were vortexed for 15 sec and centrifuged at 18,506 x *g* for 5 min. The upper aqueous phase was pipetted into a new 1.5 mL microcentrifuge tube. Following this, 10 M ammonium acetate was added to each sample to achieve a final concentration of 2.5 M. Samples were vortexed for 10 sec and centrifuged at 18,506 x *g* for 5 min. The aqueous layer was pipetted into a new tube and 0.54 volume of isopropanol was added and mixed by inversion. Samples were left for 48 hours at -20°C and then centrifuged for 10 mins at 18,506 x *g* (4°C). The supernatant was discarded, and the pellet was washed with 1 mL 70% ethanol and centrifuged for a further 10 min at 18,506 x *g*. Ethanol was removed using a pipette, and then further evaporated using the Concentrator plus speed vacuum for 4 min (Eppendorf, Hamburg, Germany). DNA was re-suspended in 20 µL of sterile DNase free water (ThermoFisher Scientific, MA, USA).

A Qubit® 2.0 fluorometer (ThermoFisher Scientific, Massachusetts, USA) was used to determine the DNA concentration according to the manufacturer's instructions. The Qubit® working solution was prepared by diluting the Qubit® DNA BS Reagent to a 1:200 dilution in Qubit® DNA BS Buffer. Qubit® working solution (190 µL) was added to both Standard 1 and Standard 2 (10 µL) and vortexed for three seconds to mix. For each sample, 198 µL of Qubit® working solution was combined with 2 µL of the sample and vortexed for three sec. All samples, including the standards, were incubated at room temperature for 2 min to allow the binding of the dye. Standard 1 had a final concentration of 0 ng/µL in TE buffer and Standard 2 had a final concentration of 5 ng/µL in TE buffer. Both were inserted into the sample chamber to calibrate the fluorometer. Samples were inserted into the sample chamber sequentially and the concentration of the DNA was determined from the standard curve. Samples were then stored at -20°C until required.

3.2.4 Bacterial community profiling using 16S rRNA sequences

3.2.4.1 16S rRNA sequencing

The sequencing library for each soil was prepared as described in Lee et al. (2018). Extracted DNA was adjusted to 5 ng/µL where possible before the Illumina MiSeq library was constructed, as detailed by the manufacturer (16S Metagenomic Sequencing Library Preparation Part # 15044223 Rev. B; Illumina, San Diego, CA, USA). A PCR was conducted with the primer set targeting the V3 and V4 regions of the bacterial and archaeal 16S rRNA gene. Primers used were PCR1F (5′ TCGTCGGCAG CGTCAGATGT GTATAAGAGA CAGCCTACGG GNGGCWGCAG 3′) and PCR1R (5′ GTCTCGTGGG CTCGGAGATG TGTATAAGAG ACAGGACTAC HVGGGTATCT AATCC 3′) with KAPA HiFi Hotstart Readymix (Kapa Biosystems, Wilmington, MA, USA) and the following thermocycling parameters: (1) 95°C for 3 min, (2) 25 cycles of 95°C for 30 sec, 55°C for 30 sec, 72°C for 30 sec, 72°C for 5 min, and (3) holding the samples at 4°C. The amplicons were then indexed and sequenced on an Illumina MiSeq (Illumina) as outlined in (Lee et al. 2018).

3.2.4.2 Analysis of environmental community structure

Sequencing data for 16S rRNA gene amplicons was processed as previously described (Archer et al. 2019) using the DADA2 version 1.8 pipeline (Callahan et al. 2016). Cutadapt (Martin 2011) was used to remove the forward and reverse primer sequences and the reads were then uniformly trimmed (280 bp for forward reads and 250 bp for reverse reads) and filtered to remove reads exceeding the maximum expected errors (2 for forward reads and 5 for reverse reads). Forward and reverse reads were merged, and chimeric sequences were removed. ASVs were assigned bacterial taxa using DADA2 with the SILVA nr version 132 database (Quast et al. 2012).

3.2.4.3 Alpha and Beta diversity and ordination

The sequences of the resulting Amplicon Sequence Variants (ASVs) were then used for further analysis as previously described (Archer et al. 2019). ASVs were processed using the R packages phyloseq, DESeq2, vegan and ggplot2 for downstream analysis and visualisation of the alpha and beta community diversity and ordination plots. Alpha diversity was investigated using Chao1 and Shannon’s index visualised in a box and whisker plot as well as a principle coordinate analysis plot. Beta diversity was investigated by analysing the amplicon sequence variants (ASVs) for each sample site and visualising these in a heatmap created using the heatmap.2 function in the R gplots package v. 3.5.2 (Warnes et al. 2015) with R software environment (R Core Team 2017) and visualised in RStudio (v 1.1.463).

3.2.4.4 Drivers of community diversity

To identify abiotic drivers of the bacterial community in the Dry Valleys and the Namib Desert, a redundancy analysis (RDA) was carried out. The RDA plot was generated using the soil chemistry data and ASV data in the phyloseq (McMurdie and Holmes 2013), DESeq2 (Love et al. 2014) and ggplot2 (Wickham 2009) packages for R software environment (R Core Team 2017). The RDA was visualised in RStudio (v 1.1.463).

3.3 Results

3.3.1 Soil chemistry

The average chemical properties for the Dry Valleys and the Namib Desert are shown in Table 3.2 while the data for each sample site is shown in Table 3.3. The mean pH of the Dry Valley soil was 7.83 while the mean pH of the Namib Desert soil was 8.6 (Table 3.2). There was a significant difference in Olsen phosphorus (t(df)= -3.0834, *p* = 0.034, where df = 4.2369), calcium (t(df)= -3.1412, *p* = 0.035, where df = 4.0073) and cation exchange capacity (CEC) (t(df) = -3.1382, *p* = 0.035, where df = 4.0131) values using an unpaired Welch Two Sample t-test between the Dry Valleys and Namib Desert samples (Table 3.2). No other statistically significant differences (*p* = 0.05) in the chemistry between the two locations was observed; however, the *p* value for potassium between the two locations was *p* = 0.052 under the 95th percentile. Table 3.2 shows the average values for the sample locations. For each sample site, the chemical component values were generally higher for the Namib Desert soil, except for magnesium and sodium (Table 3.3).

Table 3.2: Average chemical properties of soil from the Antarctic Dry Valleys (six sites) and the Namib Desert (five sites). Numbers with an * indicate average values are significantly different between the two locations (p < 0.05).

	Units	Dry Valleys (SE value)	Namib Desert (SE value)
pH	-	7.83 (0.19)	8.06 (0.23)
Olsen phosphorus	µg/mL	1.17* (0.17)	4.2* (0.97)
Potassium	mg/kg	42.25 (10.77)	103.74 (22.73)
Calcium	mg/kg	350* (146.04)	15,576* (4845.01))
Magnesium	mg/kg	48 (20.65)	35.04 (11.74)
Sodium	mg/kg	113.85 (62.26)	50.60 (16.43)
Sulphur	µg/g	82.17 (40.62)	167.8 (50.58)
Total carbon	% w/w	0.057 (0.037)	0.098 (0.023)
Total nitrogen	% w/w	0.021 (0.002)	0.038 (0.007)
Cation exchange capacity (CEC)	cmol(+)/kg	3.17* (0.029)	79.4* (0.023)

Table 3.3: individual chemical properties from each sample site.

Sample	pH	Olsen P	K	Ca	Mg	Na	S	Total C	Total N	CEC
	-	µg/mL	mg/kg				µg/g	w/w%		cmol(+)/kg
BP1	8.4	<1	19.5	100	6	4.6	>250	0.07	0.02	1
BP2	8.0	1	19.5	100	7.2	4.6	154	0.04	0.02	1
V1	8.3	2	78	1,020	50.4	50.6	9	0.08	0.02	6
V2	7.4	<1	66.3	380	116.4	305.9	58	0.27	0.02	5
BW	7.6	<1	19.5	100	6	4.6	4	0.03	0.03	1
Camp	7.3	<1	50.7	400	102	312.8	18	0.05	0.02	5
T2	7.6	5	124.8	24,640	15.6	105.8	>250	0.09	0.03	125
T4	7.9	2	163.8	26,080	36	71.3	>250	0.1	0.05	132
T6	7.6	5	70.2	18,940	12	25.3	>250	0.08	0.03	96
T8	8.6	2	35.1	2,940	33.6	25.3	60	0.04	0.02	16
T10	8.6	7	124.8	5,280	78	25.3	29	0.18	0.06	28

Previous studies into the Dry Valleys and the Namib Desert have shown that the soil chemical properties of the two locations varies widely (Lee et al. 2012; Stomeo et al. 2012; Stomeo et al. 2013). Comparisons of the chemical results obtained in this study with published data for both locations are shown in Table 3.4 (Dry Valleys) and Table 3.5 (Namib Desert).

Tables 3.4 and 3.5 show that there is a large variation between different studies investigating the chemistry of soil from the Dry Valleys and the Namib Desert. The soil chemistry results obtained in this study generally fall within previous ranges of observed literature results. The Dry Valley soil chemistry generally falls close to previously observed results, except for potassium, magnesium, sodium and sulphur. The values for potassium, magnesium and sodium are generally lower in the Dry Valley sample sites used in this study than they are in literature (Lee et al. 2012; Pointing et al. 2009; Stomeo et al. 2012). This indicates that the Dry Valley soil is more nutrient poor than that that has been used in other studies. The sulphur in three sites (BP1, BP2 and V2) was higher than previous ranges of observed literature results, indicating that fresh sulphur deposits may have recently occurred. The Namib Desert soil chemistry also follows closely with previously observed results, except for CEC, which is much higher than previously published values (Armstrong 2014; Frossard et al. 2015; Scola et al. 2017). CEC is a measure of the number of cations can be retained on the surface of soil particles. Calcium (Ca^{2+}) is one of the cations that is retained by soil particle surfaces. As seen in Table 3.3, the amount of calcium observed in most of the Namib Desert samples was very high. In particular, sites T2, T4 and T6 had very high levels of calcium, while sites T8 and T10 had much lower calcium, but still higher than any Dry Valley sites.

Table 3.4: Chemical property range from Dry Valleys sample locations compared with literature ranges.

Chemistry properties (unit)	Dry Valleys range from this study	Dry Valleys range from literature
pH	7.3 – 8.4	7.6 – 9.1[a] 8.5 – 8.9[b] 6.5 – 8.3[c] 6.96 – 8.62[d] 7.1 – 8.2[e]
Olsen phosphorus (µg/mL)	<1 – 2	0.020 – 0.023[a] 1 – 40[e]
Potassium (mg/kg)	19.5 – 78	513 – 2,579[d] 136 – 209[e] 500 – 1,200[g]
Calcium (mg/kg)	100 – 1,020	88 – 1,070[b] 1 – 140[c] 2,747 – 12,337[d] 6,700 – 14,000[g]
Magnesium (mg/kg)	6 – 116.4	25 – 262[b] 2,762 – 20,345[d] 3,800 – 6,600[g]
Sodium (mg/kg)	4.6 – 312.8	200 – 4,890[b] 1,400 – 3,884[d] 1,200 – 6,100[g]
Sulphur (µg/g)	4 – >250	3 – 15[c] 12.4 – 14.26[f]
Total carbon (% w/w)	0.04 – 0.27	0.03 – 0.06[a] 0.02 – 0.38[c] 0.1 – 0.46[d] 0.05 – 0.17[g]
Total nitrogen (% w/w)	0.02 – 0.03	0 – 0.01[a] 0.04 – 0.12[d] 0.04 – 0.12[g]
CEC (cmol(+)/kg)	1 – 6	2.5 – 9[h]

[a] Aislabie et al. (2008)
[b] McLeod et al. (2008)
[c] Cowan and Tow (2004)
[d] Lee et al. (2012)
[e] Stomeo et al. (2012)
[f] Bao and Marchant (2006)
[g] Pointing et al. (2009)
[h] Cameron et al. (1970)

Table 3.4 shows that the Dry Valley chemistry values obtained for pH, phosphorus, calcium, total carbon, total nitrogen and CEC all fell within published ranges for the Dry Valleys. However, some the values of potassium (BP1, BP2 and BW), magnesium (BP1, BP2 and BW) and sodium (BP1, BP2, BW and V1) were lower than the range of values obtained by literature.

Table 3.5: Chemical property range from Namib Desert sample locations compared with literature ranges.

Chemistry properties (unit)	Namib Desert range from this study	Namib Desert range from literature
pH	7.6 – 8.6	7.0[a] 7.44 – 9.46[b] 6.70 – 8.75[c] 7.93 – 8.48[d] 8.0 – 8.8[e] 8.4 – 9.5[f]
Olsen phosphorus (µg/mL)	2 – 7	4.4 – 9[e] 0.6 – 28.3[f]
Potassium (mg/kg)	35.1 – 163.8	164.21[a] 191.54 – 356.94[b] 70.3 – 522[c] 109.57 – 549.10[d] 101.1 – 311.1[e] 16.4 – 1,014.2[f]
Calcium (mg/kg)	2,940 – 26,080	2,793.44[a] 1,160.66 – 2,399.66[b] 700 – 2,740[c] 2,042.50 – 14,926[d] 724 – 2,235.7[e] 9.9 – 34,126.7[f]
Magnesium (mg/kg)	12 – 78	93.22[a] 59.3 – 113[b] 23.7 – 460[c] 63.30 – 198.60[d] 49.9 – 69.7[e] 1.3 – 694.6[f]
Sodium (mg/kg)	25.3 – 105.8	160.93[a] 16.63 – 1,364.63[b] 14.6 – 1,589[c] 20.17 – 1,382.98[d] 112 – 371.5[e] 0.05 – 21,600.2[f]
Sulphur (µg/g)	29 – >250	2.7 – 16,436[c] 19.5 – 110.5[e]
Total carbon (% w/w)	0.04 – 0.18	0.09[a] 0.06 – 0.22[b] 0.13 – 0.32[d] 0.04 – 0.07[e]
Total nitrogen (% w/w)	0.02 – 0.06	0.016[a] 0.03 – 0.05[d]
CEC (cmol(+)/kg)	16 – 132	6.29 – 39.07[b] 3.78 – 8.23[c] 3 – 3.5[e]

[a] Makhalanyane (2012) (mean values of five locations)
[b] Armstrong (2014)
[c] Scola et al. (2017)
[d] Stomeo et al. (2013)
[e] Frossard et al. (2015)
[f] Johnson et al. (2017)

The variability in the published ranges for the Namib Desert can be explained by the area of the Namib Desert (81,000 km^2), meaning that studies will cover a much larger area than the

Dry Valleys (4,800 km^2). The Dry Valleys are also isolated from a large amount of outside influence on the soil, whereas the Namib Desert has frequent fog events which deposit sea minerals (Eckardt and Spiro 1999). Armstrong (2014) observed the variability of soil chemistry over the course of a year and found that the soil chemistry fluctuated frequently during this time. As seen in Table 3.5, the amount of sodium and calcium in sites T2, T4 and T6 were up to 5x higher than sites T8 and T10.

3.3.2 Community structure

The Dry Valley and Namib Desert microbial community structure was investigated using 16S rRNA gene-defined community diversity. The sequencing depth of the amplicon libraries from the eleven sites was checked; the raw number of reads for each sample site can be seen in Table 3.6.

The sequencing depth of the Namib Desert sites T2, T4 and T6 was 884, 77 and 17, respectively. These values were too low to accurately capture the community structure (Bates et al. 2010; Smith and Peay 2014). As such, community data for sites T2, T4 and T6 were removed from further analyses.

Table 3.6: Raw number of reads for community analysis for each location. Samples in bold and with an * show the sequencing depth was insufficient for further analysis.

Sample site code	Location	Sequencing depth
BP1	Dry Valleys	48,903
BP2	Dry Valleys	48,344
V1	Dry Valleys	80,869
V2	Dry Valleys	60,327
BW	Dry Valleys	41,515
Camp	Dry Valleys	42,409
T2	Namib Desert	**884***
T4	Namib Desert	**77***
T6	Namib Desert	**17***
T8	Namib Desert	60,209
T10	Namib Desert	41,207

The relative abundance of phyla present in each of the soil communities can be seen below in Figure 3.4. The soil microbial communities from the Dry Valleys were often dominated by the phyla Actinobacteria, Proteobacteria and Bacteroidetes, with populations of Gemmatimonadetes and Acidobacteria also being present. The community structure of soil from the Namib Desert is dominated by the phyla Actinobacteria and Proteobacteria. Populations of the phyla Acidobacteria, Cyanobacteria, Bacteroidetes and Chloroflexi are also present in these soils.

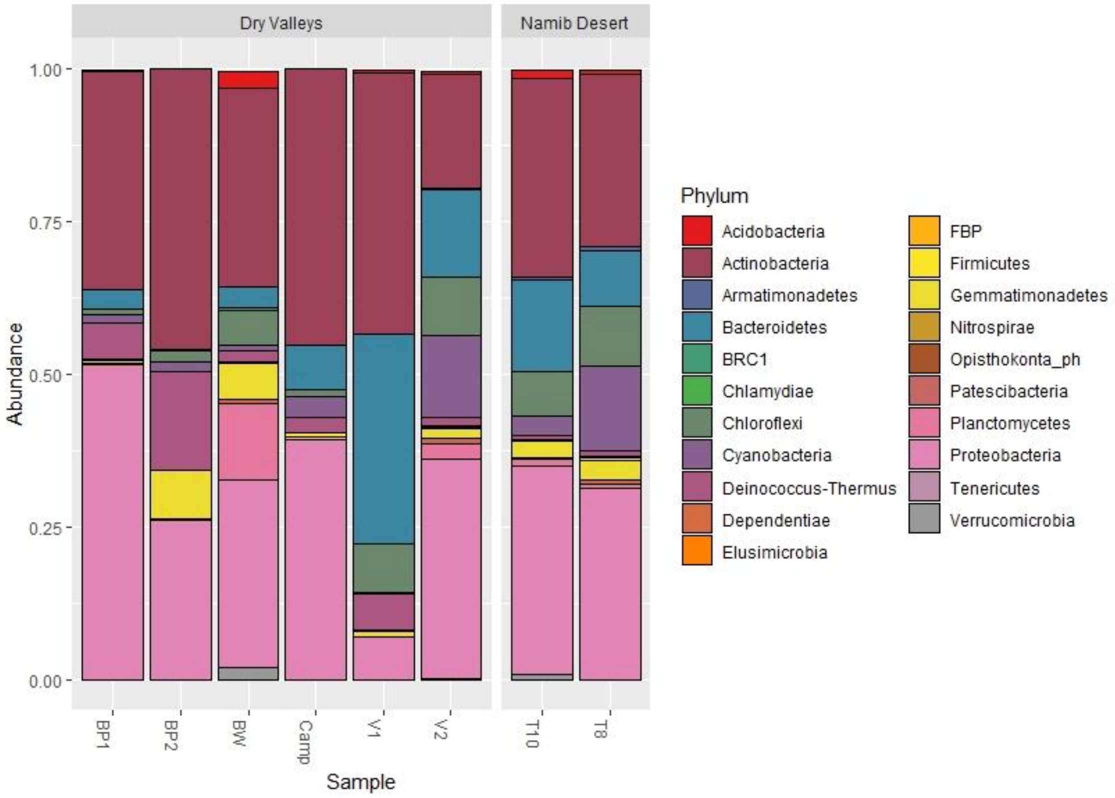


Figure 3.4: Stacked bar plot of phylum classification of sample sites from the Dry Valleys and the Namib Desert used in this study.

The phyla richness of the Dry Valleys was variable, but which phyla were present was similar in each location (Figure 3.4). Actinobacteria and Proteobacteria appear to be the dominant phyla in all soil samples studied here, which is typical of arid soils (Delgado-Baquerizo et al. 2018). Sites V1 and V2 showed slightly different community structures when compared with the other sample sites. Site V2 was located on the edge of Lake Vanda (Figure 3.2, Section 3.2); the community of the lake may have some influence on the surrounding soil community structure. Previous studies have found that water from Lake Vanda has large populations of Bacteroidetes and Proteobacteria (Zaikova et al. 2019), with smaller, but distinguished populations of Cyanobacteria: a trend that can be seen in the soil of site V2 which is closer to the lake. Both sites V1 and V2, the Bacteroidetes and Chloroflexi were more abundant compared to other sites.

The structures of the Namib Desert communities show similar trends across both sites. The dominant phyla are Actinobacteria and Proteobacteria, with large populations of Bacteroidetes and Chloroflexi also observed. The populations of Actinobacteria, Bacteroidetes, Cyanobacteria and Firmicutes appear to increase in relative abundance as the sites move further away from the coastline (Figure 3.3, Section 3.2).

Both the community profiles of the Dry Valleys and the Namib Desert follow the expected general structure found in other studies (Archer et al. 2019; Makhalanyane 2012; Ronca et al. 2015). Further analysis to the family level of the phyla Actinobacteria (Figure 3.5), Proteobacteria

(Figure 3.6) and Cyanobacteria (Figure 3.7) was investigated to provide more information regarding the overall community structure.

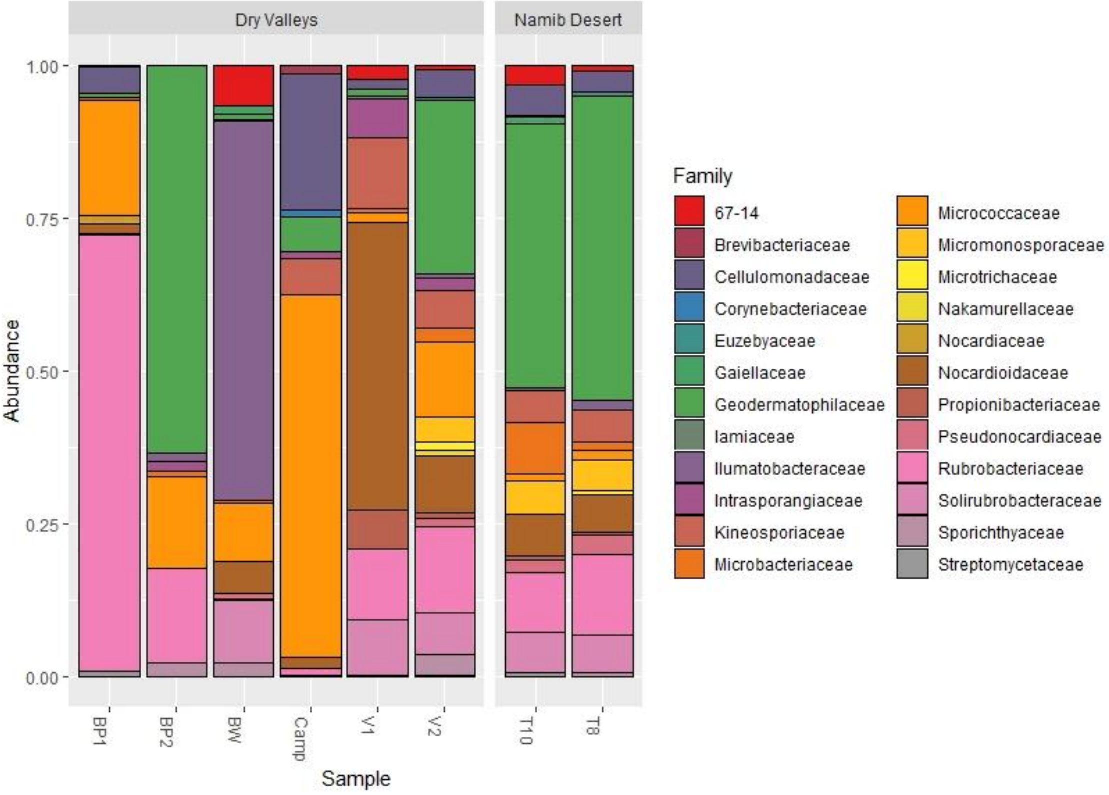


Figure 3.5: Stacked bar plot of family classification of the Actinobacteria phylum of sample sites from the Dry Valleys and the Namib Desert.

The family distribution of the Dry Valley sites is highly variable between locations. Figure 3.4 showed that the Actinobacteria phylum was one of the dominant phyla found in each of the sample sites. Site BW, Camp and V1 had a very similar relative Actinobacteria distribution at the phylum level but the family-level classification of these sites shows that the Actinobacteria in these sites have a different family distribution. Site BW has a large population of *Ilumatobacteraceae*, Camp has a large population of *Micrococcaceae* and V1 has a large population of *Nocardioidaceae*.

Sites T8 and T10 from the Namib show very similar Actinobacterial family communities, with *Geodermatophilaceae* as the largest family. Site V2 from the Dry Valleys shows a similar family distributed Actinobacteria community to sites T8 and T10. Site V2 was taken from near Lake Vanda, while sites T8 and T10 were close to dried riverbeds (Figure 3.3, Section 3.2). The family distribution of Actinobacteria may be influenced by proximity to fresh water. In general, the family distribution of Actinobacteria varies between the Dry Valleys and the Namib Desert.

The family distribution of Proteobacteria was also investigated. The Proteobacteria family classification of the Dry Valleys and the Namib Desert can be seen in Figure 3.6 below.

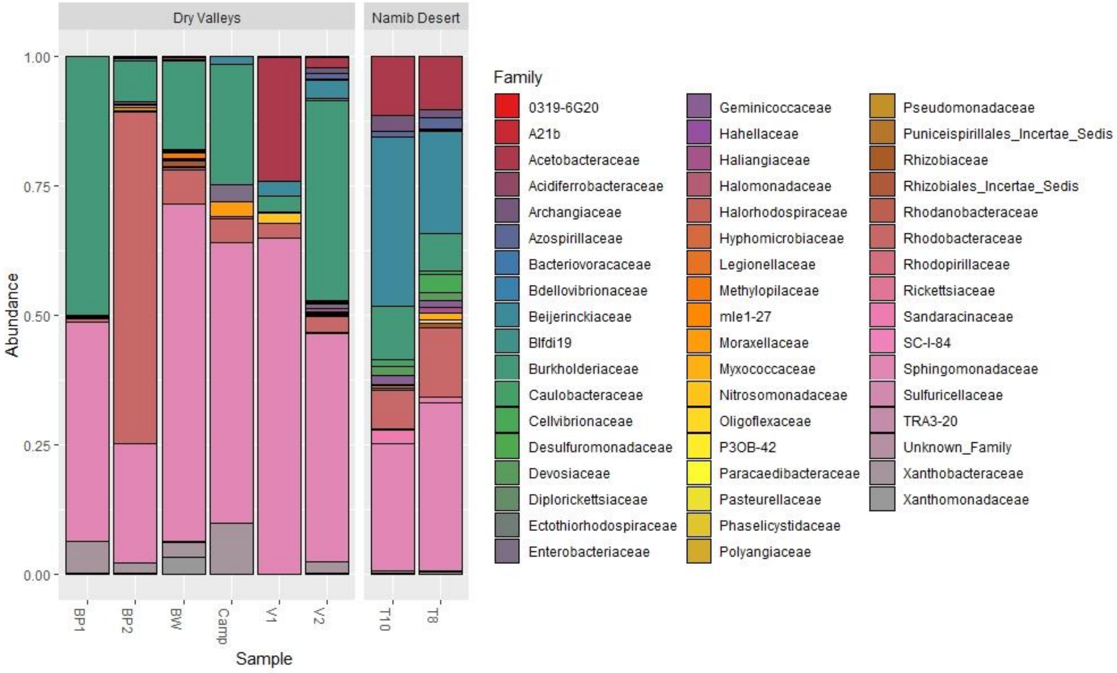


Figure 3.6: Stacked bar plot of family classification of the Proteobacteria phylum of sample sites from the Dry Valleys and the Namib Desert.

Proteobacteria was another dominant phylum in the sample sites used in this study (Figure 3.4). Sites V1 and V2 have populations of the *Acetobacteraceae* family, which was not seen in the other Dry Valley communities. *Acetobacteraceae* have previously been isolated from soils in Antarctica (Pulschen et al. 2017; Valdespino-Castillo et al. 2018). Site BP2 has a much larger population of *Rhodanobacteraceae* than the other Dry Valley locations; species in this family are noted for their ability to degrade hydrocarbons as their sole source of carbon and energy (Gutierrez 2017). This would be beneficial to organisms living in arid locations where plant life is scarce, and the soil is nutrient poor.

As seen in Figure 3.6, the proteobacterial families are very similar in T8 and T10. Both T8 and T10 had similar populations of *Acetobacteraceae*, while T8 had a higher distribution of *Sphingomonadaceae* and T10 had a higher distribution of *Beijerinckiaceae*.

Both the Dry Valleys and the Namib Desert had populations of *Burkholderiaceae* and the Namib Desert had a large population of *Beijerinckiaceae*. The Dry Valleys had either none or a very small population of *Beijerinckiaceae*, suggesting that this family has a more important role in the Namib than in the Dry Valleys. All communities observed have a large *Sphingomonadaceae* population. *Acetobacteraceae* and *Sphingomonadaceae* are *Alphaproteobacteria* and some members of these families are photoheterotrophs (Glaeser and Kämpfer 2014; Valdespino-Castillo et al. 2018). The presence of both *Acetobacteraceae* and *Sphingomonadaceae* would be beneficial to the Proteobacterial community structures of these arid locations as plant life in these deserts are almost non-existent. Therefore, the presence of photoheterotrophs could help with cycling carbon within these communities.

Finally, the family distribution of Cyanobacteria was also investigated. The Cyanobacteria family profiles of the Dry Valleys and the Namib Desert can be seen in Figure 3.7 below.

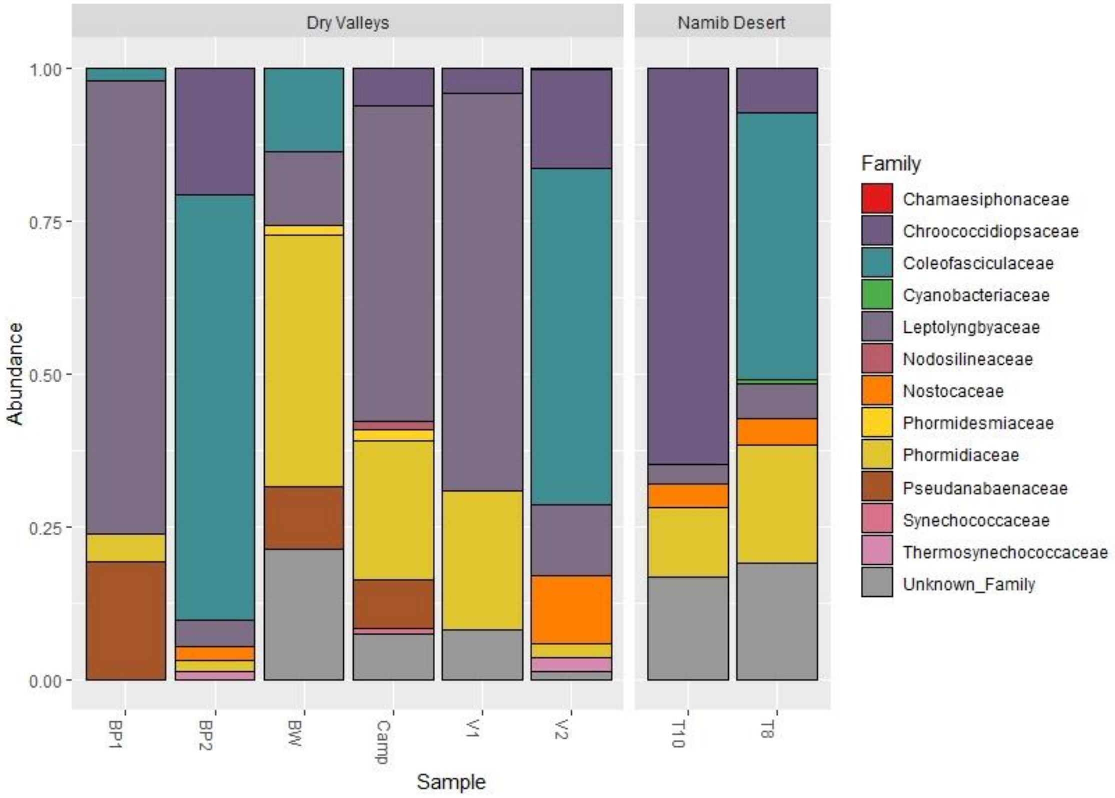


Figure 3.7: Stacked bar plot of family classification of the Cyanobacterial phylum of sample sites from the Dry Valleys and the Namib Desert.

Figure 3.7 shows the family abundance of Cyanobacteria in each sample site used in this study. It is important to note that samples, V1, BW, BP2 and BP1 had very low cyanobacterial abundance (Figure 3.4). The distribution of cyanobacterial taxa within these sample sites should be observed with caution.

A wide range of Cyanobacterial families were observed in the sites, including *Coleofasciculaceae*, *Letolyngbyaceae* and *Phormidiaceae*. Interestingly, site T8 (Namib Desert) and site V2 (Dry Valleys) appear to have a similar Cyanobacterial family community structure. Observation of the T8 site shows that the region appears to have numerous dried riverbeds. The V2 site was taken from soil next to Lake Vanda. The presence of water or higher moisture content of the soil in these two locations may have influenced a similar community structure.

3.3.2.1 Amplicon sequence variants

Amplicon sequence variants (ASVs) were used to assign bacterial taxa (Callahan et al. 2017). The use of ASVs is currently used to resolve individual DNA sequences within a sample without the use of clustering methods, such as those used in operational taxonomic units (OTUs). The distribution of ASVs provides an indication regarding the β-diversity of the samples and allows for a direct comparison across all samples without clustering similar reads into one OTU (Porter and Hajibabaei 2018).

A total of 4,588 ASVs were observed across all eleven sample sites analysed in this study. The individual ASVs for each sample site ranged from 233 (BP2) to 1,134 (V2), although these ASV sequences were observed multiple times within the communities. For example, ASV00001 occurred 1,586 times in BP1 and 76 times in T10. The number of ASVs observed in each of the sample sites can be seen in Table 3.7 below.

Table 3.7: Observed ASVs for each location. An * indicates that the observed ASVs were too low for further analysis.

Sample site code	Location	Observed ASVs
BP1	Dry Valleys	350
BP2	Dry Valleys	233
V1	Dry Valleys	473
V2	Dry Valleys	1,334
BW	Dry Valleys	970
Camp	Dry Valleys	318
T2	Namib Desert	**36***
T4	Namib Desert	**4***
T6	Namib Desert	**3***
T8	Namib Desert	889
T10	Namib Desert	789

As seen in Table 3.7 above, sites T8 and T10 had similar numbers of ASVs, while sites T2, T4 and T6 had 36, 4 and 3 ASVs, respectively. This is likely due to the sub-optimal sequencing depth of these sample sites. The ASVs observed for the Dry Valleys ranged between 233 and 1,334. Sites V1 and BW had the highest number of ASVs for the Dry Valleys, while sites BP1, BP2 and Camp had the lowest. The mean number of ASVs for the Dry Valleys was 613 while the mean for the Namib Desert was 839, excluding sites T2, T4 and T6.

The 1,000 most abundant bacterial ASVs captured 77.87% of the overall sequencing reads. The percentage of the community when using the top 2,000, 3,000, and all 4,588 ASVs captured 91.43%, 97.61% and 100% of each community, respectively. As such, the top 2,000 ASVs were used to capture a representation of the community of each sample site. The distribution of ASVs within the sample sites used in this study can be seen in Figure 3.8 below.

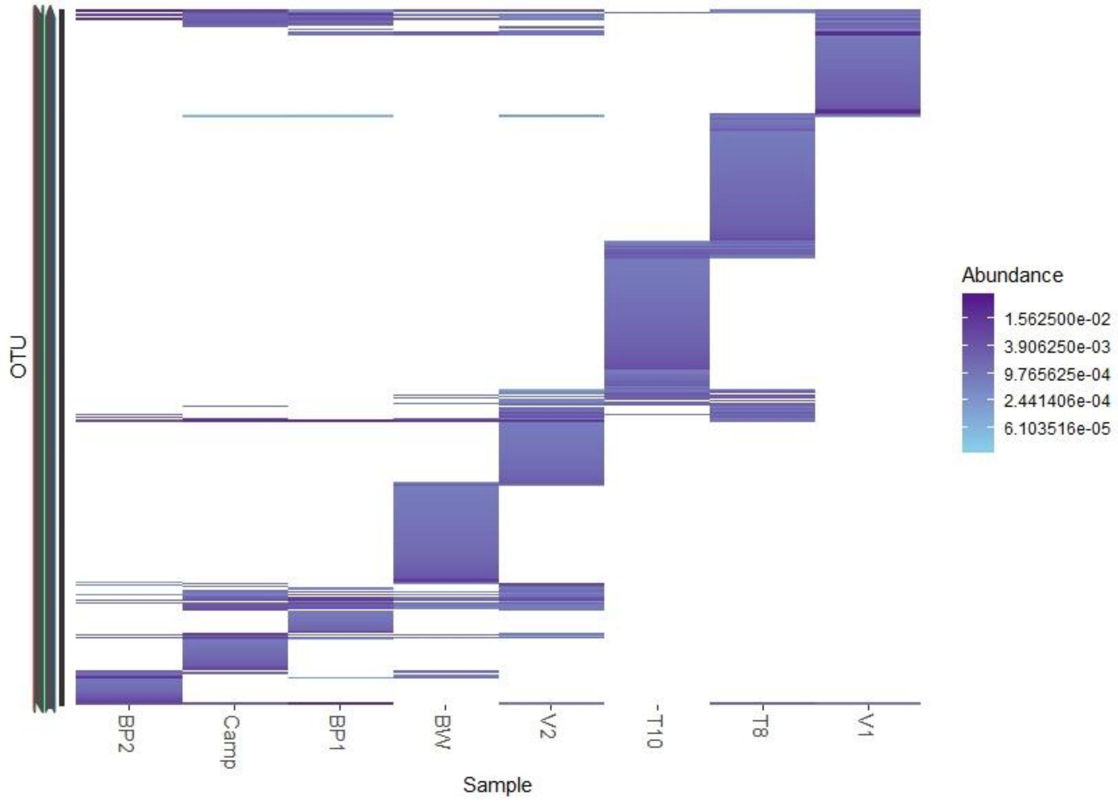


Figure 3.8: Distribution and relative abundance of the total bacterial ASVs for each sample location. ASVs are distributed based on the NMDS method and Bray-Curtis distance. Dark purple indicates high relative abundance, white indicates absence of that ASV within the sample site.

As seen in Figure 3.8, all sample sites have ASVs that are unique to that site. The distribution of ASVs appears to be similar across the Dry Valleys sites, while the Namib Desert shows a wider range of diversity between the sites. Sites T2, T4 and T6 were omitted from this analysis because of their low observed number of ASVs (Table 3.7).

While some ASVs were found in both the Namib Desert and the Dry Valleys, overall there appears to be a large difference in the ASV presence both between the two deserts and also between sites.

3.3.3 Alpha diversity

Alpha diversity is a measure of the average species diversity within a location. Two different α-diversity measures were used to analyse the Dry Valley and Namib Desert samples: Chao1 and Shannon's index. Chao1 is an estimation of species abundance while Shannon's index is a measure of species diversity. Alpha diversity analysis showed that the species abundance and species diversity varied between the Dry Valleys and the Namib Desert (Figure 3.9).

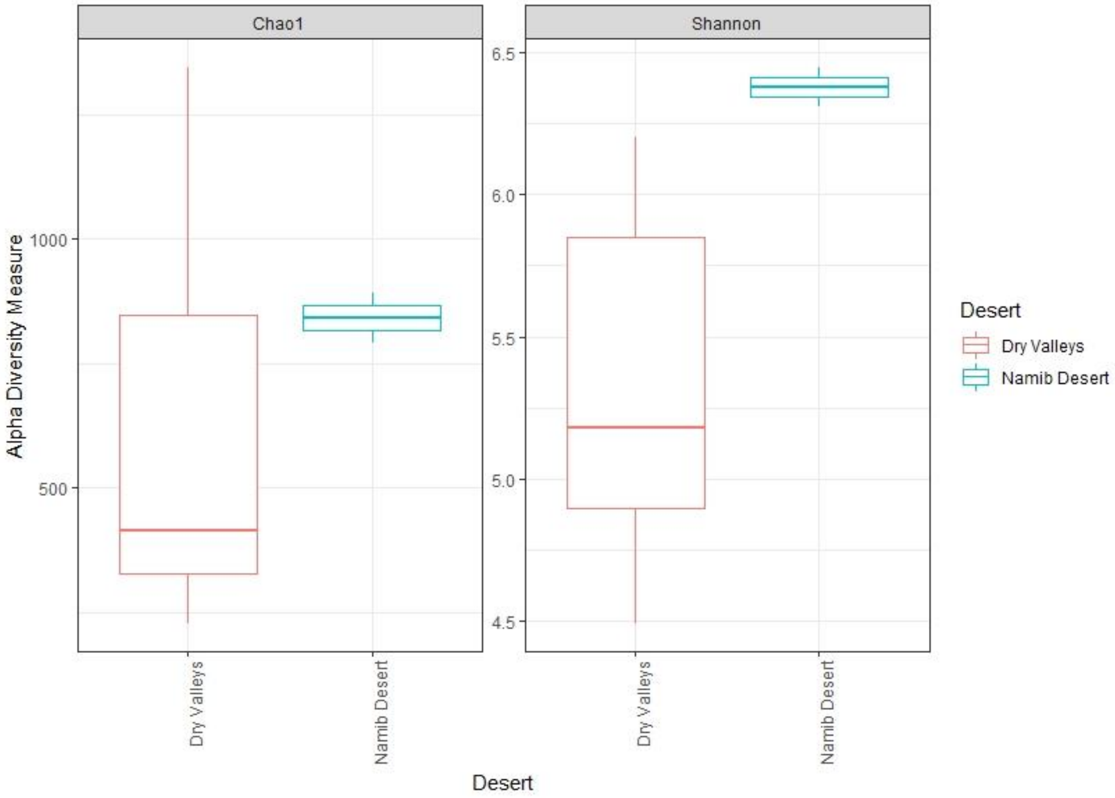


Figure 3.9: Bacterial alpha diversity estimates using Chao1 and Shannon's index. Individual data points are not shown. Box plot whiskers represent the highest and lowest observations outside the upper and lower quartiles.

As seen in Figure 3.9, the Namib Desert samples showed short tail whiskers for both Chao1 and Shannon's index diversity, indicating that the Chao1 and Shannon's indices for both the Namib sites was very similar. The Dry Valleys had greater sequence abundance (Chao1), although the median of this diversity was lower than that for the Namib Desert. The Dry Valley sites had more variable species abundance (Chao1) compared to species abundance of the Namib Desert samples. Sites V2, BW and V1 had the highest Chao1 scores (data not shown). Overall, the species diversity (Shannon's index) of the Namib Desert was very similar between sites, while the Dry Valleys had more variable species diversity. However, the results from the Namib Desert should be interpreted with caution, due to only two sample sites meeting the threshold for use in this analysis.

3.3.4 Principal Coordinates Analysis

Principal Coordinates Analysis (PCoA) was used to visualise β-diversity, which is the diversity between samples within the same location. PCoA can show the relative species abundance between sample sites without the use of a phylogenetic tree (Anderson and Willis 2003). Figure 3.10 below shows the PCoA of the Dry Valleys and the Namib Desert samples.

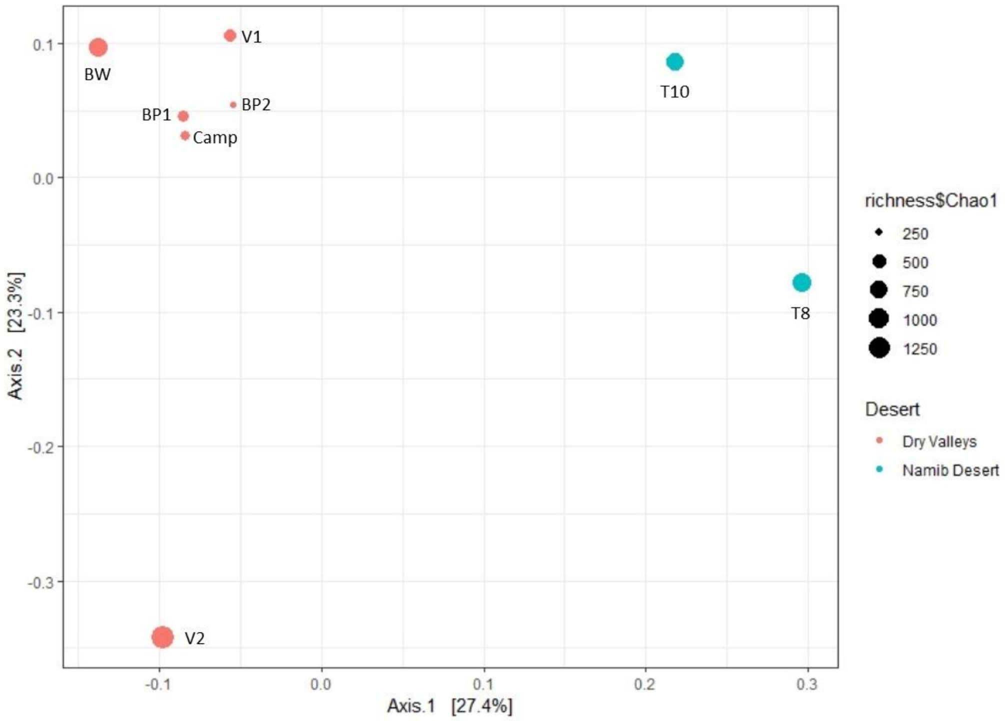


Figure 3.10: Visualisation of bacterial community dissimilarity using principal coordinate analysis plot with Bray-Curtis dissimilarity. The size of each circle represents the estimated species richness at the site.

Figure 3.10 shows that PCoA 1 (axis 1) explains 27.4% of the variance in community structure between the Dry Valleys and the Namib Desert, while PCoA 2 (axis 2) explains 23.3% of the variance. PCoA 1 shows that the phyla community of the Dry Valleys are distinct from the Namib Desert while PCoA 2 does not distinguish the Dry Valleys and the Namib Desert phyla communities from one another. Figure 3.10 also indicates species richness (Chao1) of each sample site. Sites V2, V1 and BW show the most species richness for the Dry Valleys, while BP1, BP2 and Camp show the lowest species abundance. Sites T10 and T8 from the Namib Desert show high species richness.

3.3.5 Redundancy analysis

To determine if any aspect of the chemical analysis correlates with community profiles in each location, a redundancy analysis (RDA) was conducted to observe the abiotic drivers of community structure. RDA was used over constrained correspondence analysis due to the overall relationship between the soil chemistry and the ASVs found in the sample sites being linear. Linear ordination methods such as RDA are recommended for heterogeneous data sets (Legendre and Gallagher 2001). Figure 3.11 shows the RDA analysis with type III scaling.

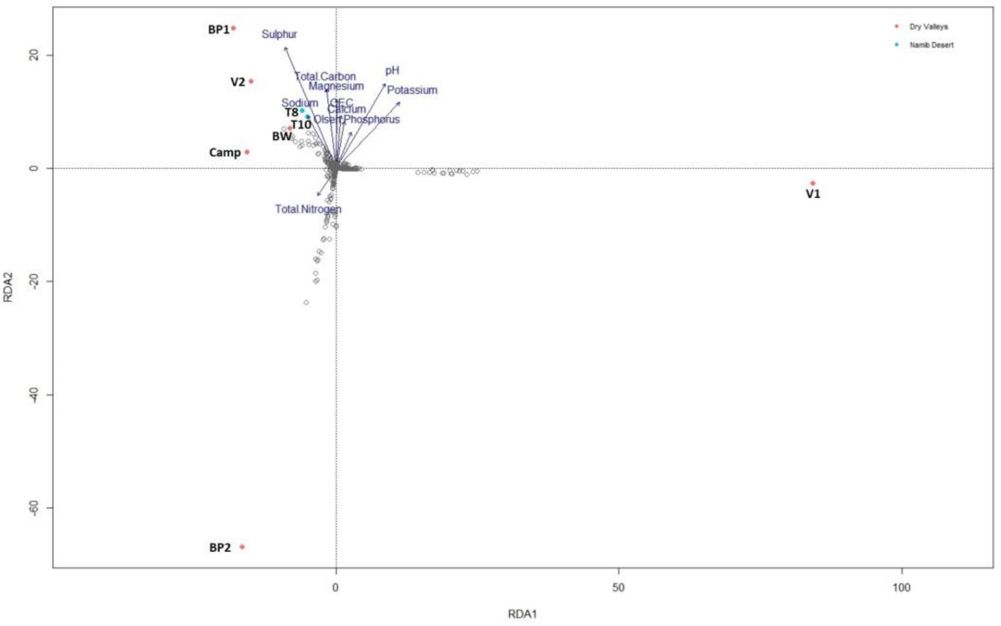


Figure 3.11: Redundancy analysis (RDA) biplot with type III scaling of bacterial ASVs and microenvironmental parameters using non-metric multidimensional scaling and Bray-Curtis dissimilarity. All ASVs are shown on the plot (grey circles). Sample sites are indicated in red (Dry Valleys) and cyan (Namib Desert). All environmental variables tested are fitted to the ordination arrows.

As seen in Figure 3.11, the difference between the Namib Desert and the Dry Valley ASV communities are not separated based on RDA1 or RDA2. The Dry Valley sites appear to cluster together on the RDA, with the exception of sites V1 and BP2. The Namib ASV communities cluster close together. No significant variability in microbial community structure was observed due to environmental variables when checked by ANOVA.

The ASV distribution of Figure 3.11 appears to be non-redundant within the explanation of the soil chemistry. That is, soil chemistry does not explain the differences in community structure between the sample sites of the Dry Valley or the Namib Desert. This indicates that all tested soil chemical properties have more of an influence over the communities of BP1, V2, Camp, BW, T8 and T10 than BP2 and V1, with the exception of nitrogen on BP2. Olsen phosphorus appears to be a less important driver of the communities, as seen by the length and direction of the arrow. Sodium and magnesium appear to be important drivers in the Namib Desert communities. Overall, bacterial communities were more similar within deserts than between individual sample locations. This is consistent with the concept of habitat filtering, where a group of closely related species share a trait or traits that allow them to exist within a specified habitat (Horner-Devine and Bohannan 2006; Scola et al. 2017).

3.4 Discussion

In this Chapter, the soil chemistry, bacterial 16S rRNA gene defined community and the culturable soil bacteria of the Namib Desert and the Dry Valleys were investigated. The 16S rRNA gene-defined community showed the distribution of bacterial phyla based on geography (Figure 3.4). It was also evident that the soil chemistry of the Namib Desert and the Dry Valleys differed in the amounts of some chemicals found within the soil (Tables 3.2 & 3.3), but these were not found to significantly influence the community structure of these deserts (Figure 3.11).

3.4.1 Soil chemistry

Soil chemistry analysis was conducted to compare the sample sites with previously published values, and to be used in conjunction with the ASVs for each location to determine drivers of the microbial community of these deserts.

The soil chemistry of the sample sites was diverse. Significant differences in phosphorus, calcium and CEC were observed between the Dry Valleys and the Namib Desert soil (Table 3.2). On average, the Namib Desert soil was more alkaline, had more potassium, sulphur, total carbon and total nitrogen. The Dry Valleys on average had higher deposits of sodium and magnesium, and the soils were more neutral than the Namib Desert soil.

The soil chemistry was variable amongst sites. The variability in soil chemistry of the Dry Valleys is most likely due to different sampling locations (Lee et al. 2012; Pointing et al. 2009). Overall, sulphur had the greatest variability within the Dry Valleys (Table 3.3). Recently, Obryk et al. (2018) reported that there has been an upwards trend of sulphur dioxide in the atmosphere over Antarctica. This increased sulphur dioxide may have caused the increased levels of sulphur observed in the Dry Valleys at sites BP1, BP2 and V2. In addition, Graf et al. (2010) notes that melting snow around Antarctica is often a source for sulphur deposition. These phenomena may explain the variability of sulphur amounts between sites. This means that the levels of sulphur within the Dry Valley soil will most likely vary depending on sampling time and season.

The chemistry of the Namib soil was more variable between the sample sites compared to the Dry Valleys. However, this variation is within the probability of normal distribution to the 95th percentile. While most of the chemistry values obtained from the Namib Desert fell within the expected range as outlined by previous research (Frossard et al. 2015; Johnson et al. 2017; Makhalanyane 2012), both the obtained calcium and CEC values were very high. CEC is calculated using a combination of cations, including calcium, explaining the elevated values of both within the Namib soil. Calcium fell within the range of Johnson et al. (2017), but the amount of calcium across the Namib Desert samples varied greatly. Calcium deposits, such as gypsum ($CaSO_4.2H_2O$), are commonly found in soil from the Namib Desert (J-B. Raymond, email message, 19 March 2019), and tend to be found in greater abundance near the coast and less so further inland into the Namib Desert (Eckardt and Drake 2010). Gypsum could have caused the dramatic differences observed in calcium levels with sites closer to the coast compared to those sites further inland. In addition, the CEC observed in sites T2, T4 and T6 was much higher than previous literature values (Armstrong 2014; Frossard et al. 2015; Scola et al. 2017). The amount of calcium in the soil at sites T2, T4 and T6 is higher than values obtained by the same authors (Armstrong 2014; Frossard et al. 2015; Scola et al. 2017). It is therefore theorised that the high amount of calcium in the soil was a result of gypsum, and this in turn dramatically increased the CEC of sites T2, T4 and T6 to levels that have not previously been reported.

Additionally, the average sulphur observed in the Namib Desert was higher than the Dry Valleys. Eckardt and Spiro (1999) notes that the amount of sulphur in the Namib Desert soil is brought in by frequent fog events. The sulphur in these fog events is generally hydrogen sulphide, which is produced by anaerobic bacteria from the ocean floor (Eckardt and Spiro 1999; Jørgensen and Boetius 2007). It has been theorised that this influx of sulphur into the desert helps to create the gypsum crusts in the Namib plains (Viles and Goudie 2013). The sulphur recorded in the Namib Desert samples is within the range of published values (Frossard et al. 2015; Scola et al. 2017), and high sulphur has previously been reported as part of the natural environment of this desert (Eckardt and Spiro 1999; Viles and Goudie 2013).

3.4.3 16S rRNA gene defined bacterial community

An obvious disadvantage of cultivation-based methods is that not all microorganisms in the soil can be isolated onto solid growth media. The estimated number of microbes isolated onto agar is around 1% of the total microbial community in the soil (Cowan et al. 2015; McKay 2008; Navarro-González et al. 2003). Environmental community defined 16S rRNA diversity allows for a more robust investigation into the community of soil, as more of the community can be observed (Handelsman 2004; Mandlik et al. 2016). Previous studies into bacterial communities in the Dry Valleys have used OTUs to compare sample sites (Goordial et al. 2016; Lee et al. 2012; Stomeo et al. 2012; Tiao et al. 2012). While OTUs and ASVs are not directly comparable, the observed ASVs for the Dry Valleys in this research appear to be within an expected phyla range based on previous OTU observations in this region.

Overall, the community profile of the Dry Valley sample sites follows an expected trend of Actinobacteria and Proteobacteria as the dominant phyla, followed by Bacteroidetes, Chlorofexi and Cyanobacteria (Cary et al. 2010; Chan et al. 2013; Makhalanyane et al. 2013; Pointing et al. 2009). Bacteroidetes had more established populations in sites V1 and V2. The V1 and V2 sites were located near Lake Vanda, which has previously been shown to have large populations of Bacteroidetes (Aislabie et al. 2006; Zaikova et al. 2019). The phyla Firmicutes was observed in all Dry Valleys sites, but larger populations were observed in sites BP2 and BW. Firmicutes are known to harbour survival strategies such as endospore and biofilm formation capabilities, which may in turn help them to survive harsh environments such as deserts (Filippidou et al. 2016; Malard et al. 2019).

The Dry Valleys are a harsh environment and are slow to respond to change (Stomeo et al. 2012). A combination of this slow response and the long term preservation of DNA in Antarctic soils can lead to 16S rRNA gene community profiles that reflect both current and historical microbial presence (Stomeo et al. 2012). As such, the community of the Dry Valleys observed in the current study may show both current and historical microbial communities.

The Namib Desert samples produced between 41,207 – 60,209 reads. Two sample sites, T4 and T6 were not able to provide sequencing depths that were sufficient for establishing an accurate community profile for these locations. Armstrong (2014) notes that substances such as humic and fulvic acids are often extracted along with DNA and inhibit the PCR process. This in turn leads to low sequencing reads, making extracting sufficient DNA for 16S rRNA gene community analysis difficult. Often, substances such as salt and gypsum can cause issues when extracting DNA (J-B. Raymond, email message, 19 March 2019). It is theorised that the PCR reaction for these two sites was inhibited by the high quantities of calcium in the soil as discussed previously.

After omitting sites T2, T4 and T6, the community profile of sites T8 and T10 follow an expected trend of Actinobacteria as the dominant phyla, followed by Proteobacteria and Bacteroidetes (Makhalanyane et al. 2013). Populations of Cyanobacteria were observed in the Namib communities, with a large community of this phyla being observed in T8 (Figure 3.4). The area in which the T8 site sample was taken appears to have numerous dried riverbeds (Google Earth 2018b), which may explain the increased Cyanobacteria population in this area. On average, the Namib Desert receives between 5 – 100 mm of rainfall per year. However, during or following these rain events rivers can refill and water can flow. On the 20 February 2017, before the T8 collection, the Namib Desert Lodge (24.07833, 15.93361) recorded the highest precipitation since November 2009 (when recording started) of 55.9 mm (Namibia Weather 2019). Frossard et al. (2015) notes that the intensity of water deposit has more impact on the bacterial soil community in the Namib Desert than frequent water deposits. In addition, Stomeo et al. (2013) found that water regimes impacted on microbial community structure, including the structure of hypolithic Cyanobacterial communities. Cyanobacteria are ubiquitous in most terrestrial habitats (Makhalanyane 2012) and the frequent fog events of the Namib Desert, which bring moisture and water activity to the soil, mean that Cyanobacterial populations are expected (Valverde et al. 2015). The increased population of Cyanobacteria at site T8 may therefore have had a population boom induced by recent rainfall. As with the Dry Valleys, the Namib Desert is a restrictive environment, and the observed Cyanobacterial population may be historical.

Previous studies into bacterial communities in the Namib Desert have used OTUs to compare sample sites (Armstrong 2014; Johnson et al. 2017; Ronca et al. 2015; Scola et al. 2017;

Stomeo et al. 2013). Once again, OTUs and ASVs are not directly comparable, but the observed number of ASVs for the Namib Desert in this thesis appear to be much higher than previous reports for sites T8 and T10. This may be due to the sites selected for sample collection, the season of collection or the recent rainfall event. As previously discussed, due to the low sampling depth, sites T2, T4 and T6 were excluded from 16S rRNA gene defined community analysis. Further steps to analyse the 16S rRNA communities of all the Namib sites would involve resequencing sites T2, T4 and T6 to achieve better sequencing depth. However, due to insufficient soil for resequencing, this is not possible for this study. While some ASVs were found in both the Namib Desert and the Dry Valleys, overall there appears to be a large difference in the ASV presence both between the two deserts and also between sites.

3.4.4 Drivers of community diversity

RDA analysis revealed that there was limited variation in community composition that could be explained by the chemical composition (Figure 3.11). ANOVA revealed that the tested environmental variables did not significantly explain variation in the microbial community structure. This indicates that other factors may be more influential in shaping the community structures of the Namib Desert and the Dry Valleys. This further suggests that stochasticity may play a role in the bacterial community assembly of both the Namib Desert and the Dry Valleys. This is supported by previous studies, which found that a combination of spatial and environmental parameters explained a low percentage of the observed variation within these communities (Makhalanyane 2012; Scola et al. 2017). However, other studies have observed a strong correlation between edaphic parameters and community structure of sand dunes within the Namib Desert (Ronca et al. 2015). While it appears that sodium and magnesium are important drivers in the Namib Desert communities, further research into a wider range of Namib Desert open soil communities may help to resolve the degree of influence these factors have on the overall bacterial community. Nitrogen appears to be a driver of the Dry Valley community BP2, although this was not statistically significant.

A greater in-depth analysis of the abiotic factors that may influence community structure in these two locations is required. Including abiotic analyses such as soil moisture, relative humidity, temperature and solar radiation may further explain bacterial community differences between the Namib Desert and the Dry Valleys. Additionally, further analysis into the biotic drivers and interactions within the sample sites is required. It has recently been suggested that biotic factors, such as microbial interactions, may help control the bacterial community structure of these locations (Lee et al. 2019).

The work outlined in this Chapter provides an insight into phyla that are present in the soil of the Dry Valleys and the Namib Desert. While both locations are harsh environments, the bacterial community structure was different between the two deserts. This information will be used in the following Chapter where bacteria from these soils were isolated onto solid growth media in an attempt to identify UV resistance.

Chapter 4: Screening and characterisation of UV resistant bacteria from the Dry Valleys and the Namib Desert

4.1 Introduction

Ultraviolet radiation has been acknowledged as an important stressor for microorganisms found in desert environments (Cordero et al. 2014; Pointing and Belnap 2012; Rampelotto 2013). Both UVB (280-315 nm) and UVC (100-280 nm) can be detrimental to microbial cell survival due to the strong absorption by DNA at these wavelengths resulting in mutations. Alternatively, UVA (315-400 nm) often causes indirect DNA damage through the formation of ROS (Albarracín et al. 2013; Matallana-Surget and Wattiez 2013).

Previous studies into UV resistance of soil microorganisms have revealed a large range of UV resistant genera (Albarracín et al. 2013; Cockell et al. 2008; de Groot et al. 2009; Hirsch et al. 2004; Kuhlman et al. 2005; Srinivasan et al. 2017; Yuan et al. 2012). Microbial stress responses to UV have been extensively reviewed in model organisms such as *Escherichia coli* and *Deinococcus radiodurans*, but the microbial tolerance to UV from isolates in desert environments has been largely overlooked (Pointing and Belnap 2012). Improving understanding of the diversity of UV resistant microorganisms within extreme environments, such as the Dry Valleys and the Namib Desert, will help advance our understanding of DNA repair mechanisms, and the overall abundance of these isolates within the community structure. To the author's knowledge, UV screening has not previously been conducted on soil isolates from the Namib Desert.

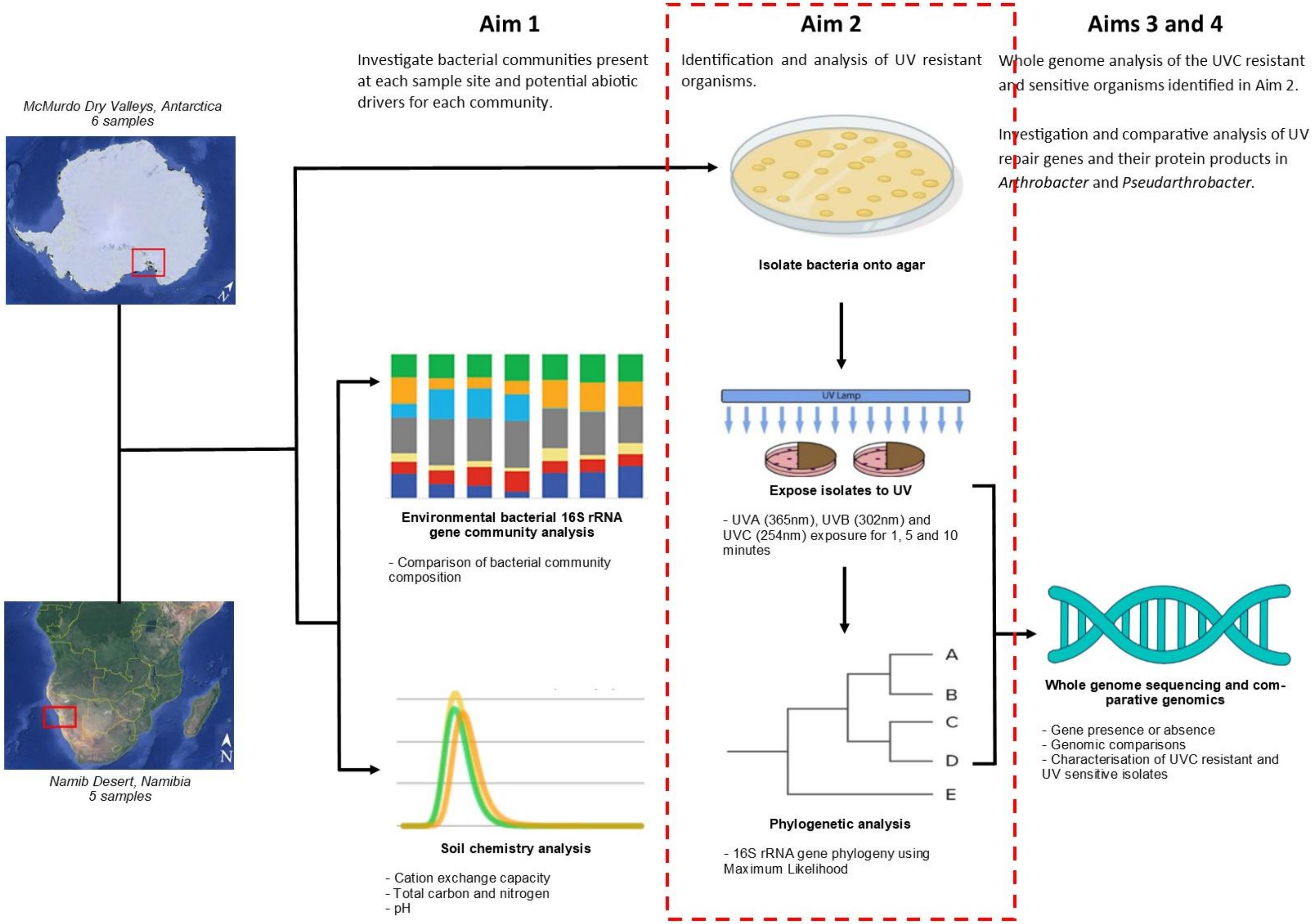
The intensity for UVA (30 W/m^2) was selected to screen for any isolates that may be highly sensitive to UVA radiation. The amount of UVA reaching the Earth's surface ranges between 40.68 and 65.9 W/m^2 (NIWA 2016; Pehnec et al. 2009); isolates that did not survive UVA 365 nm at 30 W/m^2 were considered highly sensitive to UV radiation.

The two UVB intensities used (10 and 15 W/m^2) were selected to screen for isolates that are resistant to more damaging radiation. The amount of UVB estimated to be reaching the Earth's surface ranges between 0 and 3.4 W/m^2 (NIWA 2016; Pehnec et al. 2009), although values of up to 8.15 W/m^2 have previously been recorded in the Andes (Cabrol et al. 2014). Organisms were considered 'UVB resistant' in this thesis if they survived 302 nm at 15 W/m^2 for 10 minutes.

The UVC intensities used (1 and 5 W/m^2) were selected to screen for isolates that were resistant to the most damaging wavelength of UV. While it is typically assumed that no UVC reaches Earth's surface, 1.07 W/m^2 was reported during the 2017 Antarctic soil collection (Archer 2017). Therefore, soil organisms that are resistant to UVC are of great interest and require further investigation. Organisms were considered 'UVC resistant' in this thesis if they survived 254 nm at 5 W/m^2 for 10 minutes. Organisms that survived 1 minute, but not 5 minutes of UVC 254 nm at 1 or 5 W/m^2 or UVB 302 nm at 10 or 15 W/m^2 exposure were deemed 'UVC sensitive' isolates. Genera with both UVC resistant and UVC sensitive isolates were selected for further analysis. The UVB and UVC intensities were selected to deliberately screen for bacteria resistant to radiation that they would not often be encountered in their natural environments. This was to determine if isolates were able to survive radiation that the bacteria had most likely never encountered before, and therefore indicate that they may have important DNA repair mechanisms for UV repair. Experiments were carried out in triplicate.

Previous studies have isolated UV resistant bacteria such as *Halomonas* sp. (Musilova et al. 2015), *Deinococcus* sp. (Hirsch et al. 2004), *Hymenobacter* sp. and *Chryseobacterium* sp. (Órdenes-Aenishanslins et al. 2016) from Antarctica. To the author's knowledge, UV resistant bacteria have not previously been reported in the Namib Desert. Here, aerobic heterotrophic bacteria were isolates from soil and exposed to different UV wavelengths and intensities to categorise organisms as resistant or sensitive to UV radiation. The overall abundance of these organisms was then compared with the community profiles described in Chapter 3 (Section 3.2). The goal of this was to identify UV resistant organisms for further molecular analysis to better understand bacterial UV resistance.

This Chapter will address Aim 2 (Figure 4.1): identification and analysis of UV resistant organisms. To achieve this aim, four objectives were defined: (1) isolate aerobic heterotrophic bacteria from soil from the Dry Valleys and the Namib Desert using culture-based methods; (2) develop and utilise a rapid screening method to quickly identify UVC resistant and UVC sensitive bacteria; (3) use Sanger sequencing to identify UVC resistant and UVC sensitive organisms based on their 16S rRNA gene; (4) use phylogenetics to understand the relationships between the 16S rRNA gene of UVC resistant and UVC sensitive organisms of the same genus.



Investigate bacterial communities present at each sample site and potential abiotic drivers for each community.

Identification and analysis of UV resistant organisms.

Whole genome analysis of the UVC resistant and sensitive organisms identified in Aim 2.

Investigation and comparative analysis of UV repair genes and their protein products in *Arthrobacter* and *Pseudarthrobacter*.

McMurdo Dry Valleys, Antarctica
6 samples
N
Namib Desert, Namibia
5 samples
N
Isolate bacteria onto agar
UV Lamp
Expose isolates to UV
Environmental bacterial 16S rRNA gene community analysis
Soil chemistry analysis
Phylogenetic analysis
A
B
C
D
E
Whole genome sequencing and comparative genomics

- UVA (365nm), UVB (302nm) and UVC (254nm) exposure for 1, 5 and 10 minutes

- Comparison of bacterial community composition

- Cation exchange capacity
- Total carbon and nitrogen
- pH

- 16S rRNA gene phylogeny using Maximum Likelihood

- Gene presence or absence
- Genomic comparisons
- Characterisation of UVC resistant and UV sensitive isolates

Figure 4.1: Breakdown of schematic overview of this thesis. Aim 2 will be covered in this Chapter.

4.2 Methods and materials

4.2.1 Isolation of culturable microorganisms

Aerobic heterotrophic bacteria were targeted for this study. Isolates were obtained by suspending 10 g of soil in 90 mL of sterile peptone water. The suspension was homogenised by gentle hand inversion for 2 minutes. Each soil suspension were grown via spread plate (0.1 mL) onto each of the four culture media to enrich representative bacteria from the sample; culture media used to isolate organisms were 10% tryptic soy agar (TSA) (BD Difco, USA), 10% nutrient agar (NA) (BD Difco, USA), Reasoner's 2A agar (R2A) (Neogen, UK) and Luria-Bertani agar (LB) (BD Difco, USA). Triplicates of soil suspensions on each culture media was incubated aerobically at 4°C (3 weeks), 15°C (96 hours), 20°C (72 hours), 35°C (48 hours) and 45°C (48 hours). Presumed bacterial colonies were transferred and grown on fresh agar of the same type as the agar they were isolated on. Pure isolates were picked and transferred to nutrient broth (NB) and grown until the OD_{600} reached 0.5. Isolates were stored in 1.5 mL 1:1 25% v/v glycerol stocks at -80°C until required.

4.2.2 UV survival evaluation via modified drop plate method

Dry Valley and Namib Desert isolates were exposed to UV radiation from the UVP 3UV™ Lamp (Analytik Jena, Upland, CA, USA). To test the resistance of culturable bacteria to UV radiation, the isolates were exposed to three wavelengths and five intensities of UV. This is shown in Table 4.1.

Table 4.1: UV type, wavelength and irradiation intensity (W/m2) of UV used.

UV type	Wavelength (nm)	Intensities (W/m^2)
UVA	365	30
UVB	302	10 and 15
UVC	254	1 and 5

For exposure, one glycerol stock of each organism was grown in 5 mL of nutrient broth (NB) until stationary phase of OD_{600} 0.5. This broth was further subcultured by inoculating 0.5 mL of this culture into three new 5 mL NB. The three new broths were grown until stationary phase of OD_{600} 0.5. The incubation times were 48 hours (35°C and 45°C), 72 hours (20°C), 96 hours (15°C) or 3 weeks (4°C). This is outlined in Figure 4.2 below.

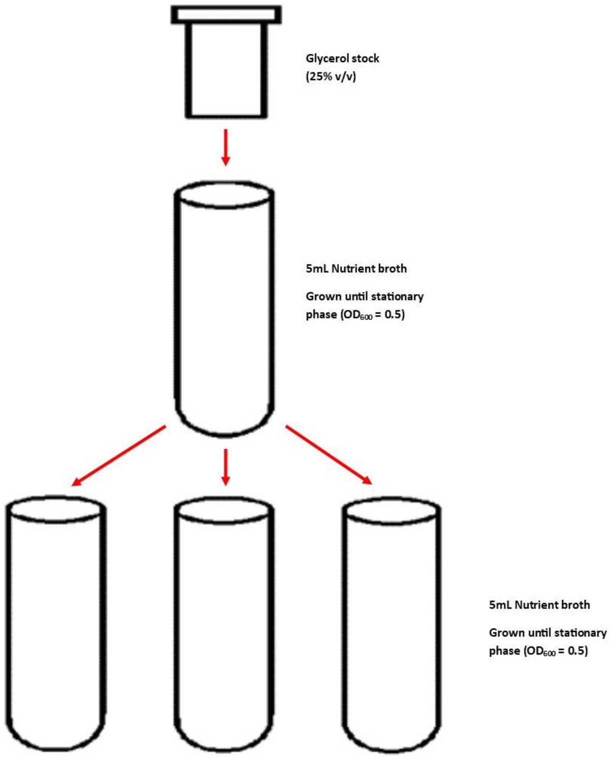

Glycerol stock
(25% v/v)
5mL Nutrient broth
Grown until stationary phase (OD_{600} = 0.5)
5mL Nutrient broth
Grown until stationary phase (OD_{600} = 0.5)

Figure 4.2: Outline of bacterial growth strategy prior to UV exposure in this thesis (Parker 2019).

After subculturing into the three NB as outlined above in Figure 4.2, the isolates were exposed to UV radiation using a modified drop plate method (Figure 4.3). Using 90 mm split agar plates (Techno Plas, St Marys, South Australia, Australia), 3 µL of each isolate was plated to the same position on each half of the agar, creating a mirror plate. One half was covered with cardboard to serve as the control, while the other half was exposed to UV radiation (Figure 4.3a & b). Up to 12 isolates were assigned to each plate. Each plate contained organisms isolates from any of the sample sites investigated in this thesis (Figures 3.2 and 3.3), however, isolates on the same plate were isolated on the same culture media (NA, TSA, LB or R2A) and at the same incubation temperature (4°C, 15°C, 20°C 35°C or 45°C). Plates were exposed to UVA (365 nm), UVB (302 nm) and UVC (254 nm) at selected intensities (Table 4.1) for 1, 5 and 10 minutes. UVA exposure was at 30 W/m^2 (3 cm above plate surface), UVB was at both 10 W/m^2 (14 cm above the plate surface) and 15 W/m^2 (10 cm above plate surface), and UVC was at both 1 W/m^2 (43 cm above the plate surface) and 5 W/m^2 (16 cm above the plate surface). Intensities were measured using the Solarmeter® (Solarlight Company, Glenside, PA, USA) Model 4.0, Model 6.0 and Model 8.0 for UVA, UVB and UVC, respectively. The UV intensity for each wavelength was adjusted by moving the UVP 3UV™ Lamp closer to, or further away from, the agar plates. Exposure times were 1 min, 5 min and 10 mins for each UV wavelength and intensity (Table 4.1). Plates were incubated under light exposure to allow for photoreactivation pathways to engage. Plates were incubated for 48 hours (35°C and 45°C), 72 hours (20°C), 96 hours (15°C) or 3 weeks (4°C) after exposure. Plates were photographed using ProtoCOL 2 (Synbiosis LTD, SDI Group, Cambridge, UK).

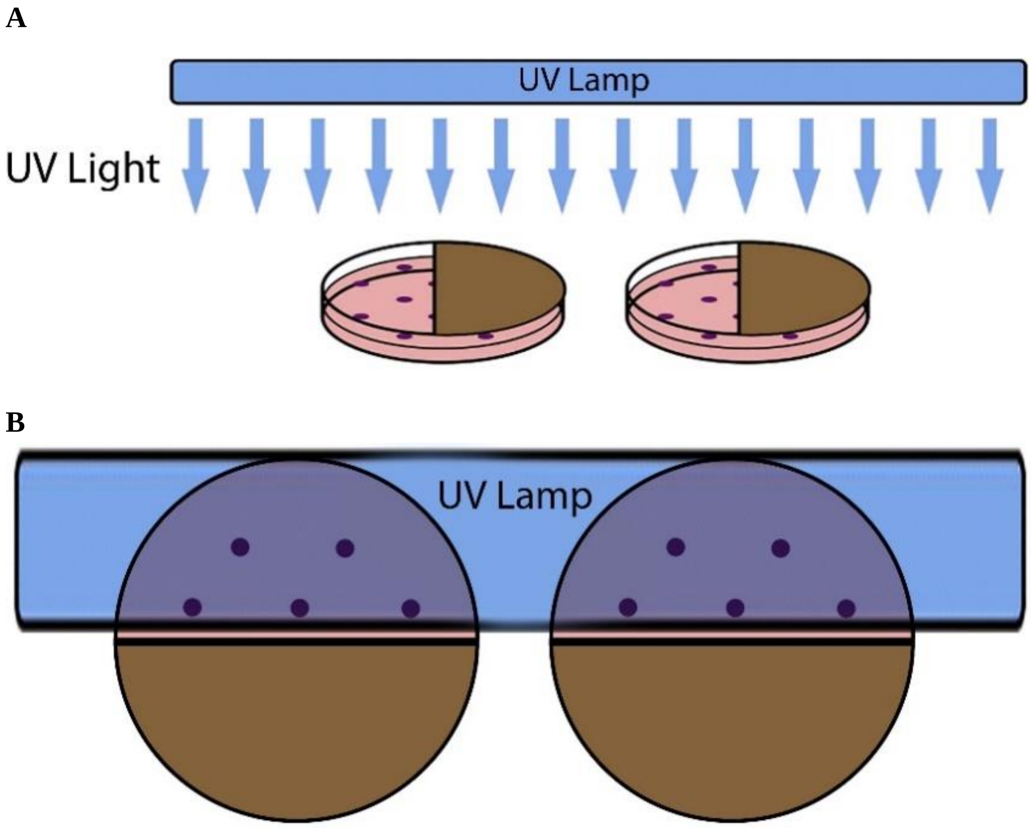


Figure 4.3a: Diagram of modified mirror drop plate method. Organisms are arranged as a "mirror" in the same position on each side of the agar. Each dot represents a 3µL aliquot of culture. One half was covered with cardboard and the other was left exposed to UV **4.3b: Bird's-eye view of plates under the UV lamp during exposure (Parker 2018).**

Using the method illustrated in Figure 4.3a & b, isolates were categorised as UVC resistant if they were able to survive 15 W/m^2 of UVB or 5 W/m^2 of UVC for 10 mins. Isolates were considered sensitive to UV if they did not survive 5 mins of 1 or 5 W/m^2 of UVC or 10 or 15 W/m^2 of UVB. Isolates that did not survive UVA 365 nm at 30 W/m^2 were considered sensitive to UV radiation.

4.2.3 Identification of UVC resistant and UVC sensitive isolates

4.2.3.1 DNA extraction

UVC resistant and UVC sensitive organisms were selected for genus identification by Sanger sequencing. Organisms were grown in NB until stationary phase and 200 µL of culture was placed into sterile 1.5 mL microcentrifuge tubes. The culture was then boiled at 100°C for 15 minutes to burst the cells and release the DNA. This served as the DNA template for the PCR reaction as described below. *Escherichia coli* DNA was also extracted at the same time to serve as the positive control for the PCR.

4.2.3.2 PCR reaction

The 16S rRNA gene region was amplified using the bacterial forward primer 27F (5'AGAGTTTGATCMTGGCTCAG3') (Lane 1991) and the universal reverse primer 1492R (5'TACGGYTACCTTGTTACGACTT3'). PCR components were: 12.5 µL GoTaq Green Master Mix (Promega, Wisconsin, USA), 1 µL of 10 µM 27F, 1 µL of 10 µM 1492R primer, 2.5 µL DNA template, and 8 µL of nuclease-free water. Nuclease-free water (2.5 µL) was used in place of the DNA template for the no template negative control (NTC). PCR reactions were conducted on a TC-512 Gradient Thermal Cycler (Techne TM, Bibby Scientific TM, England) using the following conditions: initial denaturation at 95°C for 5 min; 30 cycles of denaturation (30 s at 95°C), annealing (30 s at 51°C), and extension (2 min at 72°C); and a final extension at 72°C for 10 min.

4.2.3.3 Agarose gel electrophoresis

PCR products were electrophoresed on a 1% agarose 1 X Tris-Borate EDTA (TBE) gel stained with 0.5 µL/mL ethidium bromide at 75 V for 45 mins using the PowerPac™ Basic Power Supply (Bio-Rad Laboratories., Auckland, New Zealand). A 0.1 µg/µL 100 bp DNA ladder (Solis BioDyne., Estonia) was used as a size marker. Each lane contained 5 µL of product or marker. Gel photos were taken using the AlphaImager® HP system (Alpha Innotech., California, USA).

4.2.3.4 Sanger sequencing

Sanger sequencing was conducted by Macrogen, Inc. (South Korea) using the 27F and 1492R primers. Sequences were curated and forward and reverse sequences aligned to create a consensus for each 16S rRNA sequence. Consensuses 16S rRNA sequences were used to search GenBank on NCBI (https://www.ncbi.nlm.nih.gov/) using BLASTn for preliminary isolate identification. Isolates were preliminarily identified using the top hit based on E-value.

4.2.4 Phylogenetic tree construction

Phylogenetic trees were created using 16S rRNA gene sequences of the UVC resistant and UVC sensitive isolates, with type strains for each genus and type species of other members of the same family. The 16S rRNA gene sequences of type strains were downloaded from the Ribosomal Database Project v 11.5 (https://rdp.cme.msu.edu/) (Cole et al. 2013). Sequences were aligned using MUSCLE alignment in Geneious Prime (https://www.geneious.com) version 2019.0. A maximum-likelihood tree was created using the FastTree 2.1.10 plugin (Price et al. 2010) with 1,000 bootstraps in Geneious Prime version 2019.0.

4.2.5 Death curve of resistant and sensitive bacteria

To determine isolate survivability following UVC exposure, further UVC survivability tests were carried out on resistant and sensitive organisms of the *Arthrobacter* and *Pseudarthrobacter* isolates B2, B4, E2 and E5. Organisms were grown in 5 mL NB by inoculating 200 µL from a single 1.5 mL 1:1 25% vol/vol glycerol stock until stationary phase; OD_{600} 0.5. Three further 5 mL NB were inoculated with 0.5 mL of this culture and grown until late log phase. The culture was then diluted in peptone water (Difco, USA) to achieve dilutions of 10^{-3} – 10^{-5}. Triplicate 9 mL dilutions were put into empty sterile Petri dishes and agitated using a magnetic stirrer set at speed 400 and 20°C (VWR® Professional Hot Plate, Avantor, PA, USA). For UVC exposure, the UVP 3UV™ Lamp (Analytik Jena, Upland, CA, USA) was used on the UVC setting (256 nm), 43 cm above Petri dish surface to achieve an irradiance of 1 W/m^2, and 16 cm above the Petri dish surface to achieve an irradiance of 5 W/m^2. The irradiance of the lamp was measured using the Solarmeter® (Solarlight Company, Glenside, PA, USA) Model 8.0. Each dilution (10^{-3} – 10^{-5}) was exposed to UVC and at each time point (1 minute, 5 minutes and 10 minutes), 100 µL of culture was removed and plated onto nutrient agar. Plates were then incubated for 96 hours at 20°C and colonies were counted. Plates that were not exposed were used as a control. The $\log_{10}$ CFU values were used to create a death curve.

4.3 Results

4.3.1 Cultured bacteria

A total of 309 presumed bacterial colonies (based on visual observation) were isolated on culture media from 11 sample sites. Of these, 24 isolates failed to recover from glycerol stock when inoculated into NB and were not used in the UV screening trials. Table 4.2 shows the number of isolates from each sample site for each temperature tested. Most organisms were isolated from the 20°C treatment (Table 4.2).

Table 4.2: Number of organisms isolated from each sample site.

Sample site	Location	Temperature				
		4°C	15°C	20°C	35°C	45°C
BP1	Dry Valleys	0	5	10	0	0
BP2	Dry Valleys	5	3	6	0	0
V1	Dry Valleys	10	10	13	1	0
V2	Dry Valleys	11	9	10	0	0
BW	Dry Valleys	8	0	14	3	0
Camp	Dry Valleys	7	8	9	1	0
Total		**41**	**35**	**62**	**5**	**0**
T2	Namib Desert	4	8	7	11	0
T4	Namib Desert	5	6	7	11	5
T6	Namib Desert	4	7	5	5	6
T8	Namib Desert	3	8	15	7	7
T10	Namib Desert	6	7	9	13	0
Total		**22**	**36**	**43**	**47**	**18**
Absolute total		**63**	**71**	**105**	**52**	**18**

Due to the large number of bacterial isolates recovered from soil in this study, only 25 isolates were selected for identification by Sanger sequencing using the 16S rRNA genes. This selection was based on UVC resistance and UV sensitivity. The identities of these organisms can be found in Table 4.4.

4.3.1 Rapid screening of UV resistant aerobic soil heterotrophs

A total of 309 aerobic heterotrophic bacteria were isolated onto agar from 11 sample sites (Table 4.2), but 24 were unable to be regrown from glycerol stocks. The remaining 285 isolates were cultivated, screened and compared phylogenetically during this Chapter. Isolates that were able to form visible colonies 72 hours (15°C, 20°C, 35°C and 45°C) or 3 weeks (4°C) after UV exposure were considered resistant to the radiation.

To identify UV resistant and UV sensitive isolates, a rapid screening method was developed, as detailed in Section 4.2.2. A total of 285 isolates were screened for their resistance to UVA (365nm), UVB (302nm) and UVC (254nm). Survival was classified as any observed colony growth following UV exposure. All 285 isolates survived 10 mins of UVA exposure at 30

W/m^2 ±0.07 (Table 4.3). This was expected, as the amount of UVA reaching the Earth's surface is between 40 and 66 W/m^2 (NIWA 2016; Pehnec et al. 2009). A total of 35 isolates survived 15 W/m^2 ±0.05 of UVB for 10 mins. A total of 15 isolates survived UVC at 5 W/m^2 ±0.05 for 10 mins; ten isolates from the Antarctic Dry Valleys, and six isolates from the Namib Desert (Table 4.3).

Table 4.3: Number of isolates that survived various UV wavelengths for 10 minutes. Total number of isolates screened n=285.

	UVA (365nm)	**UVB (302nm)**		**UVC (254nm)**	
	30 W/m^2	10 W/m^2	15 W/m^2	1 W/m^2	5 W/m^2
Antarctic Dry Valleys	135	46	26	53	10
Namib Desert	150	77	9	22	6
Total resistant isolates	**285**	**123**	**35**	**75**	**16**

This rapid screening method quickly identified organisms that were resistant to UV radiation at different intensities. Up to 12 isolates were assigned to a plate; isolates on each plate had been isolated from the same temperature and on the same agar. Figure 4.4 shows and example plate of 12 organisms exposed to 1 W/m^2 of UVC for 10 minutes. The full agar assignment spreadsheet can be found in Appendix 1.

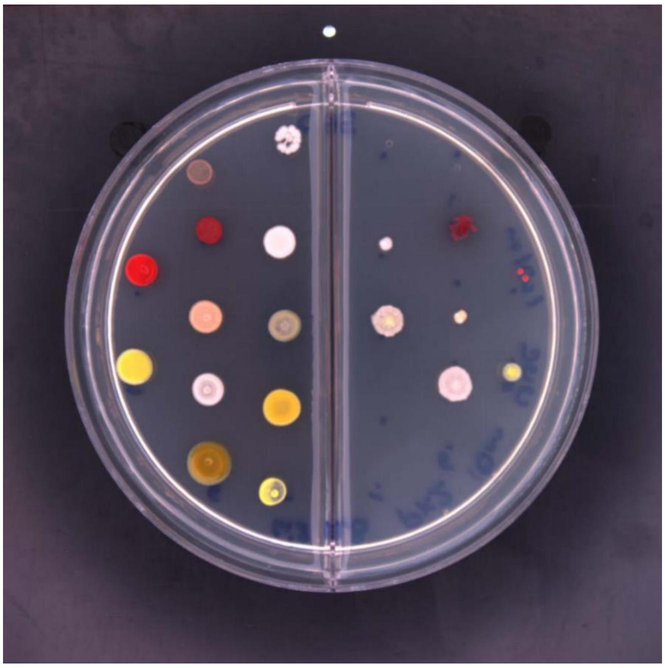


Figure 4.4: Example of mirror plate of UVC survivability at 1 W/m2 for 10 mins. The left side of the agar was covered with cardboard and served as the control. The pictured agar was LB agar with organisms grown at 15°C. See Appendix 1 for full photo files.

Due to the limited amount of UVC radiation that reaches Earth, the 16 isolates that were able to survive UVC radiation were of interest for further study. The primary goal of this thesis is to investigate genomic insights into how bacteria survive UV radiation. It is presumed that these bacteria have not co-evolved in the presence of UVC, therefore their ability to survive this highly DNA damaging radiation is of great interest. The remainder of this thesis will therefore focus on isolates that are resistant, and sensitive, to UVC radiation.

4.3.2 Taxonomic identification of UVC resistant and UVC sensitive organisms by PCR

Rapid screening identified 10 isolates that were UVC sensitive, and 16 that were UVC resistant. All 26 were identified to the genus level using the approximately 1,522 bp long 16S rRNA gene. *E. coli* was used as a positive control. Figure 4.5 shows an example of PCR product in agarose gel after electrophoresis.

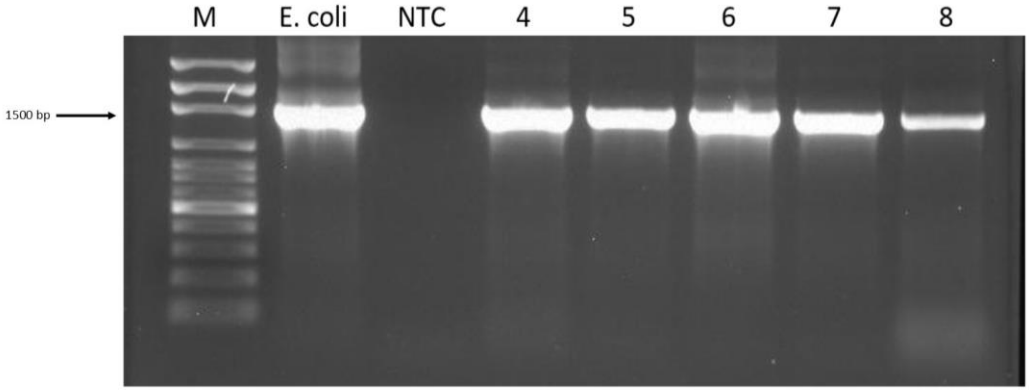


Figure 4.5: Analysis of PCR by agarose gel electrophoresis using the 16S rRNA bacterial forward primer 27F and universal reverse primer 1492R. Lane M: 100 bp ladder, *E. coli*, NTC, samples lane 4: B1, lane 5: B2, lane 6: B3, lane 7: B4, lane 8: B5.

A wide range of UVC resistant genera were identified through rapid screening followed by PCR and Sanger sequencing (Table 4.3). UVC sensitive organisms were also identified. Rapid screening identified organisms from 10 genera that appeared to be UVC resistant, while organisms from six genera appear to be UVC sensitive. Of these, there were four genera; *Arthrobacter*, *Pseudarthrobacter*, *Pseudomonas* and *Stenotrophomonas*, that had both resistant and sensitive organisms (Table 4.4).

As seen in Table 4.4, five of the six UVC resistant organisms from the Namib Desert were of the genus *Arthrobacter*. Notably, no UVC resistant bacteria were found at sample sites BP2 or T2.

Table 4.4: 16S rRNA identification of UVC resistant and sensitive isolates from the Dry Valleys and the Namib Desert.

Sample site	Location	Isolate code		UVC resistant or UVC sensitive	16S rRNA genus identification
BP1	Dry Valleys	B5		Resistant	*Brachybacterium* sp.
		C2		Resistant	*Pantoea* sp.
		C5	*	Resistant	*Stenotrophomonas* sp.
V1	Dry Valleys	C6		Resistant	*Brachybacterium* sp.
		C3	*	Sensitive	*Pseudomonas* sp.
		B3	*	Sensitive	*Pseudomonas* sp.
		D5		Sensitive	*Paenisporosarcina* sp.
		E4		Sensitive	*Plantibacter* sp.
V2	Dry Valleys	B1	*	Resistant	*Pseudomonas* sp.
BW	Dry Valleys	A1		Resistant	*Stenotrophomonas* sp.
		E3	*	Sensitive	*Stenotrophomonas* sp.
Camp	Dry Valleys	A5		Resistant	*Kluyvera* sp.
		B6		Resistant	*Enterobacter* sp.
		G1		Resistant	*Pseudomonas* sp.
T4	Namib Desert	B4	*	Resistant	*Pseudarthrobacter* sp.
		C1	*	Sensitive	*Arthrobacter* sp.
		E2	*	Sensitive	*Arthrobacter* sp.
T6	Namib Desert	E7	*	Resistant	*Arthrobacter* sp.
T8	Namib Desert	A2		Resistant	*Bacillus* sp.
		E5	*	Sensitive	*Pseudarthrobacter* sp.
T10	Namib Desert	A4	*	Resistant	*Arthrobacter* sp.
		B2	*	Resistant	*Arthrobacter* sp.
		D2	*	Resistant	*Arthrobacter* sp.
		E1	*	Sensitive	*Arthrobacter* sp.
		E6		Sensitive	*Massilia* sp.

*Isolates had UVC sensitive/UVC resistant counterparts and were selected for further phylogenetic analysis. Isolate codes in red did not have good quality Sanger sequencing reads and were discounted from further analysis.

The most common phyla observed in the Dry Valleys and the Namib Desert were Proteobacteria and Actinobacteria (Figure 3.4). This is reflected in Table 4.4, which shows that most of the observed UVC resistant genera belong to Proteobacteria (*Pantoea*, *Stenotrophomonas*, *Pseudomonas*, *Kluyvera* and *Enterobacter*) and Actinobacteria (*Arthrobacter*, *Pseudarthrobacter* and *Brachybacterium*). *Bacillus* belongs to the Firmicutes phylum, however, the isolate A2 which was identified as *Bacillus* did not have good quality Sanger sequence reads and was excluded from further analysis.

Table 4.4 shows that the most common genus that displayed UVC resistance based on the criteria outlined in Section 4.2.2 was *Arthrobacter*. Members of the *Pseudarthrobacter*, *Bacillus*, *Pseudomonas*, *Stenotrophomonas*, *Brachybacterium* and *Pantoea* also demonstrated UV resistance (Table 4.4). Although there were several isolates with poor quality Sanger sequencing reads (Table 4.4), the 16S rRNA sequence was not re-extracted for additional sequencing due to members of the *Arthrobacter*, *Pseudarthrobacter, Pseudomonas* and *Stenotrophomonas* genera having both UVC resistant and UVC sensitive isolates. This was the overall goal of 16S rRNA Sanger sequencing so as to observe phylogenetic relationships between UVC resistant and UVC sensitive isolates of the same genus. As such, these four genera were selected for 16S rRNA gene based phylogenetic analysis.

4.3.3 Phylogenetic analysis

As mentioned, four genera had both UVC resistant and UVC sensitive organisms identified by rapid screening. Species previously in the genus *Arthrobacter* has recently been reclassified into *Arthrobacter*, *Pseudarthrobacter*, *Glutamicibacter*, *Paeniglutamicibacter*, *Pseudoglutamicibacter* and *Paenarthrobacter* (Busse 2016). Maximum likelihood (ML) phylogenetic analysis showing the relationships between the UVC resistant and UVC sensitive organisms are presented in Figure 4.6 (*Arthrobacter* and *Pseudarthrobacter*), Figure 4.7 (*Pseudomonas*) and Figure 4.8 (*Stenotrophomonas*).

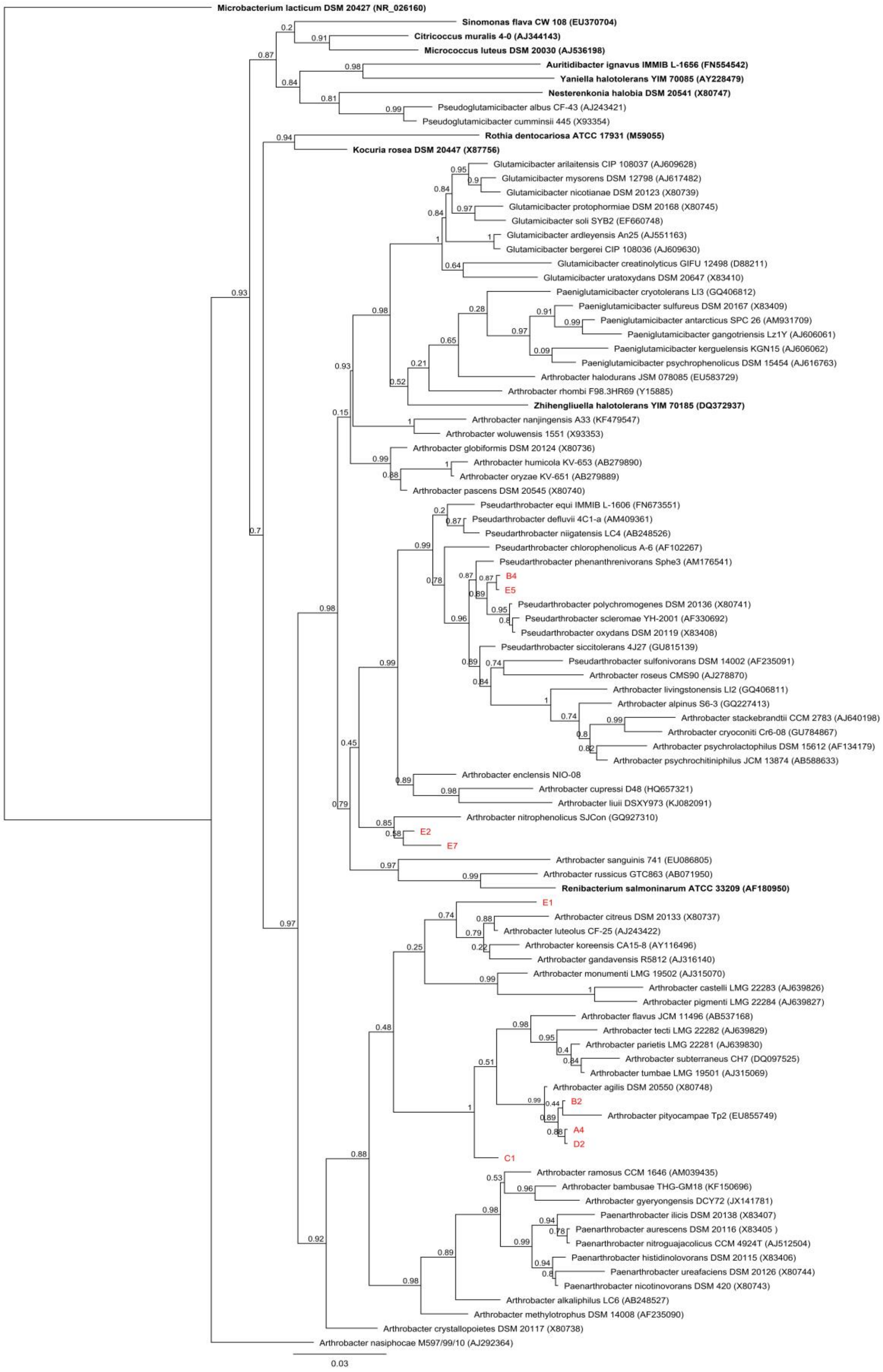

Microbacterium lacticum DSM 20427 (NR_026160)
Sinomonas flava CW 108 (EU370704)
Citricoccus muralis 4-0 (AJ344143)
Micrococcus luteus DSM 20030 (AJ536198)
Auritidibacter ignavus IMMIB L-1656 (FN554542)
Yaniella halotolerans YIM 70085 (AY228479)
Nesterenkonia halobia DSM 20541 (X80747)
Pseudoglutamicibacter albus CF-43 (AJ243421)
Pseudoglutamicibacter cumminsii 445 (X93354)
Rothia dentocariosa ATCC 17931 (M59055)
Kocuria rosea DSM 20447 (X87756)
Glutamicibacter arilaitensis CIP 108037 (AJ609628)
Glutamicibacter mysorens DSM 12798 (AJ617482)
Glutamicibacter nicotianae DSM 20123 (X80739)
Glutamicibacter protophormiae DSM 20168 (X80745)
Glutamicibacter soli SYB2 (EF660748)
Glutamicibacter ardleyensis An25 (AJ551163)
Glutamicibacter bergerei CIP 108036 (AJ609630)
Glutamicibacter creatinolyticus GIFU 12498 (D88211)
Glutamicibacter uratoxydans DSM 20647 (X83410)
Paeniglutamicibacter cryotolerans LI3 (GQ406812)
Paeniglutamicibacter sulfureus DSM 20167 (X83409)
Paeniglutamicibacter antarcticus SPC 26 (AM931709)
Paeniglutamicibacter gangotriensis Lz1Y (AJ606061)
Paeniglutamicibacter kerguelensis KGN15 (AJ606062)
Paeniglutamicibacter psychrophenolicus DSM 15454 (AJ616763)
Arthrobacter halodurans JSM 078085 (EU583729)
Arthrobacter rhombi F98.3HR69 (Y15885)
Zhihengliuella halotolerans YIM 70185 (DQ372937)
Arthrobacter nanjingensis A33 (KF479547)
Arthrobacter woluwensis 1551 (X93353)
Arthrobacter globiformis DSM 20124 (X80736)
Arthrobacter humicola KV-653 (AB279890)
Arthrobacter oryzae KV-651 (AB279889)
Arthrobacter pascens DSM 20545 (X80740)
Pseudarthrobacter equi IMMIB L-1606 (FN673551)
Pseudarthrobacter defluvii 4C1-a (AM409361)
Pseudarthrobacter niigatensis LC4 (AB248526)
Pseudarthrobacter chlorophenolicus A-6 (AF102267)
Pseudarthrobacter phenanthrenivorans Sphe3 (AM176541)
B4
E5
Pseudarthrobacter polychromogenes DSM 20136 (X80741)
Pseudarthrobacter scleromae YH-2001 (AF330692)
Pseudarthrobacter oxydans DSM 20119 (X83408)
Pseudarthrobacter siccitolerans 4J27 (GU815139)
Pseudarthrobacter sulfonivorans DSM 14002 (AF235091)
Arthrobacter roseus CMS90 (AJ278870)
Arthrobacter livingstonensis LI2 (GQ406811)
Arthrobacter alpinus S6-3 (GQ227413)
Arthrobacter stackebrandtii CCM 2783 (AJ640198)
Arthrobacter cryoconiti Cr6-08 (GU784867)
Arthrobacter psychrolactophilus DSM 15612 (AF134179)
Arthrobacter psychrochitiniphilus JCM 13874 (AB588633)
Arthrobacter enclensis NIO-08
Arthrobacter cupressi D48 (HQ657321)
Arthrobacter liuii DSXY973 (KJ082091)
Arthrobacter nitrophenolicus SJCon (GQ927310)
E2
E7
Arthrobacter sanguinis 741 (EU086805)
Arthrobacter russicus GTC863 (AB071950)
Renibacterium salmoninarum ATCC 33209 (AF180950)
E1
Arthrobacter citreus DSM 20133 (X80737)
Arthrobacter luteolus CF-25 (AJ243422)
Arthrobacter koreensis CA15-8 (AY116496)
Arthrobacter gandavensis R5812 (AJ316140)
Arthrobacter monumenti LMG 19502 (AJ315070)
Arthrobacter castelli LMG 22283 (AJ639826)
Arthrobacter pigmenti LMG 22284 (AJ639827)
Arthrobacter flavus JCM 11496 (AB537168)
Arthrobacter tecti LMG 22282 (AJ639829)
Arthrobacter parietis LMG 22281 (AJ639830)
Arthrobacter subterraneus CH7 (DQ097525)
Arthrobacter tumbae LMG 19501 (AJ315069)
Arthrobacter agilis DSM 20550 (X80748)
B2
Arthrobacter pityocampae Tp2 (EU855749)
A4
D2
C1
Arthrobacter ramosus CCM 1646 (AM039435)
Arthrobacter bambusae THG-GM18 (KF150696)
Arthrobacter gyeryongensis DCY72 (JX141781)
Paenarthrobacter ilicis DSM 20138 (X83407)
Paenarthrobacter aurescens DSM 20116 (X83405)
Paenarthrobacter nitroguajacolicus CCM 4924T (AJ512504)
Paenarthrobacter histidinolovorans DSM 20115 (X83406)
Paenarthrobacter ureafaciens DSM 20126 (X80744)
Paenarthrobacter nicotinovorans DSM 420 (X80743)
Arthrobacter alkaliphilus LC6 (AB248527)
Arthrobacter methylotrophus DSM 14008 (AF235090)
Arthrobacter crystallopoietes DSM 20117 (X80738)
Arthrobacter nasiphocae M597/99/10 (AJ292364)
0.03

Figure 4.6: ML phylogenetic tree of *Arthrobacter* spp. and *Pseudarthrobacter* spp. based on 16S rRNA gene sequences (max 1530 bp). The 16S rRNA gene sequence of *Microbacterium lacticum* DSM 20427 (NR_026160) was used as the outgroup. Bar, 0.03 substitutions per nucleotide position. Type species of genera are shown in bold face. Organisms identified in this study are in red. Values at nodes represent FastTree support values from 1,000 bootstraps, where 1 is equal to 100 percent. GenBank accession numbers are shown in parenthesis. Tree created using Geneious Prime version 2019.0 by Biomatters.

As seen in Figure 4.6, the *Pseudarthrobacter* clade shows that the two *Pseudarthrobacter* isolates B4 and E5 (Table 4.4) share a common ancestor. A BLASTn analysis identified *Pseudarthrobacter phenanthrenivorans* as the most similar sequence to isolates B4 (Resistant [Res]) and E5 (Sensitive [Sen]). The phylogenetic tree shows that *P. phenanthrenivorans*, B4 and E5 share a common ancestor, but B4 and E5 share a more recent ancestor with *Pseudarthrobacter polychromogenes*, *Pseudarthrobacter sclromae* and *Pseudarthrobacter oxydans* (Figure 4.6)

Isolates E2 (Sen) and E7 (Res) share a common ancestor and appear on a similar clade (*Arthrobacter sanguinis* clade) (Busse 2016), with isolate E2 appearing closer to the last common ancestor. Phylogeney shows that *Arthrobacter nitrophenolicus* is the closest 16S rRNA sequence to isolates E2 and E7. This relationship is supported (0.85 FastTree support value), however, the relationship between E7 and E2 does not have high support (0.58 FastTree support value).

Isolate E1 (Sen) is different from other isolates and is the only isolate from this study within the '*Arthrobacter citreus*' group (Busse 2016). This relationship is well supported (0.74 FastTree support value), however, the relationship between the '*Arthrobacter citreus*' clade and the '*Arthrobacter pigmenti*' group is not well supported (0.25 FastTree support value).

Isolates B2 (Res), D2 (Res) and A4 (Res) appear in the *Arthrobacter agilis* clade with high support (0.97 FastTree support value). Phylogeny shows these isolates are similar to *Arthrobacter pityocampae* and related to *A. agilis*. Finally, isolate C1 (Sen) shares a common ancestor with B2, D2 and A4 but appears closer in sequence to the ancestral strain.

Within the *Pseudomonas* phylogenetic tree isolates B3 (Sen), C3 (Sen) and B1 (Res) cluster on the same clade. The relationships between these isolates can be seen in Figure 4.7 below.

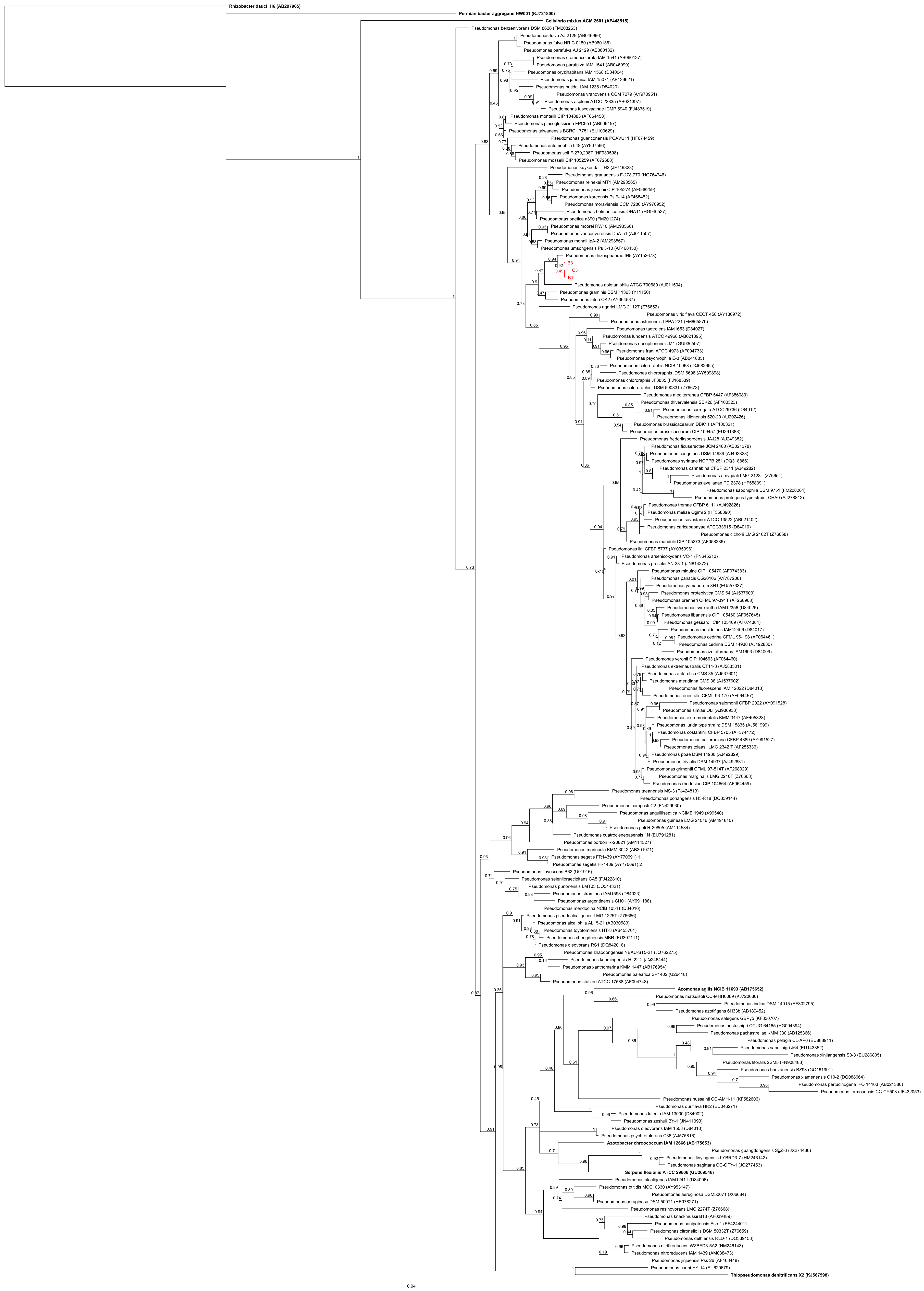

Rhizobacter dauci H6 (AB297965)
Permianibacter aggregans HW001 (KJ721800)
Cellvibrio mixtus ACM 2601 (AF448515)
Pseudomonas benzenivorans DSM 8628 (FM208263)
Pseudomonas fulva AJ 2129 (AB046996)
Pseudomonas fulva NRIC 0180 (AB060136)
Pseudomonas parafulva AJ 2129 (AB060132)
Pseudomonas cremoricolorata IAM 1541 (AB060137)
Pseudomonas parafulva IAM 1541 (AB046999)
Pseudomonas oryzihabitans IAM 1568 (D84004)
Pseudomonas japonica IAM 15071 (AB126621)
Pseudomonas putida IAM 1236 (D84020)
Pseudomonas vranovensis CCM 7279 (AY970951)
Pseudomonas asplenii ATCC 23835 (AB021397)
Pseudomonas fuscovaginae ICMP 5940 (FJ483519)
Pseudomonas monteilii CIP 104883 (AF064458)
Pseudomonas plecoglossicida FPC951 (AB009457)
Pseudomonas taiwanensis BCRC 17751 (EU103629)
Pseudomonas guariconensis PCAVU11 (HF674459)
Pseudomonas entomophila L48 (AY907566)
Pseudomonas soli F-279,208T (HF930598)
Pseudomonas mosselii CIP 105259 (AF072688)
Pseudomonas kuykendallii H2 (JF749828)
Pseudomonas granadensis F-278,770 (HG764746)
Pseudomonas reinekei MT1 (AM293565)
Pseudomonas jessenii CIP 105274 (AF068259)
Pseudomonas koreensis Ps 9-14 (AF468452)
Pseudomonas moraviensis CCM 7280 (AY970952)
Pseudomonas helmanticensis OHA11 (HG940537)
Pseudomonas baetica a390 (FM201274)
Pseudomonas moorei RW10 (AM293566)
Pseudomonas vancouverensis DhA-51 (AJ011507)
Pseudomonas mohnii IpA-2 (AM293567)
Pseudomonas umsongensis Ps 3-10 (AF468450)
Pseudomonas rhizosphaerae IH5 (AY152673)
B3
C3
B1
Pseudomonas abietaniphila ATCC 700689 (AJ011504)
Pseudomonas graminis DSM 11363 (Y11150)
Pseudomonas lutea OK2 (AY364537)
Pseudomonas agarici LMG 2112T (Z76652)
Pseudomonas viridiflava CECT 458 (AY180972)
Pseudomonas asturiensis LPPA 221 (FM865870)
Pseudomonas taetrolens IAM1653 (D84027)
Pseudomonas lundensis ATCC 49968 (AB021395)
Pseudomonas deceptionensis M1 (GU936597)
Pseudomonas fragi ATCC 4973 (AF094733)
Pseudomonas psychrophila E-3 (AB041885)
Pseudomonas chlororaphis NCIB 10068 (DQ682655)
Pseudomonas chlororaphis DSM 6698 (AY509898)
Pseudomonas chlororaphis JF3835 (FJ168539)
Pseudomonas chlororaphis DSM 50083T (Z76673)
Pseudomonas mediterranea CFBP 5447 (AF386080)
Pseudomonas thivervalensis SBK26 (AF100323)
Pseudomonas corrugata ATCC29736 (D84012)
Pseudomonas kilonensis 520-20 (AJ292426)
Pseudomonas brassicacearum DBK11 (AF100321)
Pseudomonas brassicacearum CIP 109457 (EU391388)
Pseudomonas frederiksbergensis JAJ28 (AJ249382)
Pseudomonas ficuserectae JCM 2400 (AB021378)
Pseudomonas congelans DSM 14939 (AJ492828)
Pseudomonas syringae NCPPB 281 (DQ318866)
Pseudomonas cannabina CFBP 2341 (AJ49282)
Pseudomonas amygdali LMG 2123T (Z76654)
Pseudomonas avellanae PD 2378 (HF558391)
Pseudomonas saponiphila DSM 9751 (FM208264)
Pseudomonas protegens type strain: CHA0 (AJ278812)
Pseudomonas tremae CFBP 6111 (AJ492826)
Pseudomonas meliae Ogimi 2 (HF558390)
Pseudomonas savastanoi ATCC 13522 (AB021402)
Pseudomonas caricapapayae ATCC33615 (D84010)
Pseudomonas cichorii LMG 2162T (Z76658)
Pseudomonas mandelii CIP 105273 (AF058286)
Pseudomonas lini CFBP 5737 (AY035996)
Pseudomonas arsenicoxydans VC-1 (FN645213)
Pseudomonas prosekii AN 28-1 (JN814372)
Pseudomonas migulae CIP 105470 (AF074383)
Pseudomonas panacis CG20106 (AY787208)
Pseudomonas yamanorum 8H1 (EU557337)
Pseudomonas proteolytica CMS 64 (AJ537603)
Pseudomonas brenneri CFML 97-391T (AF268968)
Pseudomonas synxantha IAM12356 (D84025)
Pseudomonas libanensis CIP 105460 (AF057645)
Pseudomonas gessardii CIP 105469 (AF074384)
Pseudomonas mucidolens IAM12406 (D84017)
Pseudomonas cedrina CFML 96-198 (AF064461)
Pseudomonas cedrina DSM 14938 (AJ492830)
Pseudomonas azotoformans IAM1603 (D84009)
Pseudomonas veronii CIP 104663 (AF064460)
Pseudomonas extremaustralis CT14-3 (AJ583501)
Pseudomonas antarctica CMS 35 (AJ537601)
Pseudomonas meridiana CMS 38 (AJ537602)
Pseudomonas fluorescens IAM 12022 (D84013)
Pseudomonas orientalis CFML 96-170 (AF064457)
Pseudomonas salomonii CFBP 2022 (AY091528)
Pseudomonas simiae OLi (AJ936933)
Pseudomonas extremorientalis KMM 3447 (AF405328)
Pseudomonas lurida type strain: DSM 15835 (AJ581999)
Pseudomonas costantinii CFBP 5705 (AF374472)
Pseudomonas palleroniana CFBP 4389 (AY091527)
Pseudomonas tolaasii LMG 2342 T (AF255336)
Pseudomonas poae DSM 14936 (AJ492829)
Pseudomonas trivialis DSM 14937 (AJ492831)
Pseudomonas grimontii CFML 97-514T (AF268029)
Pseudomonas marginalis LMG 2210T (Z76663)
Pseudomonas rhodesiae CIP 104664 (AF064459)
Pseudomonas taeanensis MS-3 (FJ424813)
Pseudomonas pohangensis H3-R18 (DQ339144)
Pseudomonas composti C2 (FN429930)
Pseudomonas anguilliseptica NCIMB 1949 (X99540)
Pseudomonas guineae LMG 24016 (AM491810)
Pseudomonas peli R-20805 (AM114534)
Pseudomonas cuatrocienegasensis 1N (EU791281)
Pseudomonas borbori R-20821 (AM114527)
Pseudomonas marincola KMM 3042 (AB301071)
Pseudomonas segetis FR1439 (AY770691) 1
Pseudomonas segetis FR1439 (AY770691) 2
Pseudomonas flavescens B62 (U01916)
Pseudomonas seleniipraecipitans CA5 (FJ422810)
Pseudomonas punonensis LMT03 (JQ344321)
Pseudomonas straminea IAM1598 (D84023)
Pseudomonas argentinensis CH01 (AY691188)
Pseudomonas mendocina NCIB 10541 (D84016)
Pseudomonas pseudoalcaligenes LMG 1225T (Z76666)
Pseudomonas alcaliphila AL15-21 (AB030583)
Pseudomonas toyotomiensis HT-3 (AB453701)
Pseudomonas chengduensis MBR (EU307111)
Pseudomonas oleovorans RS1 (DQ842018)
Pseudomonas zhaodongensis NEAU-ST5-21 (JQ762275)
Pseudomonas kunmingensis HL22-2 (JQ246444)
Pseudomonas xanthomarina KMM 1447 (AB176954)
Pseudomonas balearica SP1402 (U26418)
Pseudomonas stutzeri ATCC 17588 (AF094748)
Azomonas agilis NCIB 11693 (AB175652)
Pseudomonas matsuisoli CC-MHH0089 (KJ720680)
Pseudomonas indica DSM 14015 (AF302795)
Pseudomonas azotifigens 6H33b (AB189452)
Pseudomonas salegens GBPy5 (KF830707)
Pseudomonas aestusnigri CCUG 64165 (HG004394)
Pseudomonas pachastrellae KMM 330 (AB125366)
Pseudomonas pelagia CL-AP6 (EU888911)
Pseudomonas sabulinigri J64 (EU143352)
Pseudomonas xinjiangensis S3-3 (EU286805)
Pseudomonas litoralis 2SM5 (FN908483)
Pseudomonas bauzanensis BZ93 (GQ161991)
Pseudomonas xiamenensis C10-2 (DQ088664)
Pseudomonas pertucinogena IFO 14163 (AB021380)
Pseudomonas formosensis CC-CY503 (JF432053)
Pseudomonas hussainii CC-AMH-11 (KF582606)
Pseudomonas duriflava HR2 (EU046271)
Pseudomonas luteola IAM 13000 (D84002)
Pseudomonas zeshuii BY-1 (JN411093)
Pseudomonas oleovorans IAM 1508 (D84018)
Pseudomonas psychrotolerans C36 (AJ575816)
Azotobacter chroococcum IAM 12666 (AB175653)
Pseudomonas guangdongensis SgZ-6 (JX274436)
Pseudomonas linyingensis LYBRD3-7 (HM246142)
Pseudomonas sagittaria CC-OPY-1 (JQ277453)
Serpens flexibilis ATCC 29606 (GU269546)
Pseudomonas alcaligenes IAM12411 (D84006)
Pseudomonas otitidis MCC10330 (AY953147)
Pseudomonas aeruginosa DSM50071 (X06684)
Pseudomonas aeruginosa DSM 50071 (HE978271)
Pseudomonas resinovorans LMG 2274T (Z76668)
Pseudomonas knackmussii B13 (AF039489)
Pseudomonas panipatensis Esp-1 (EF424401)
Pseudomonas citronellolis DSM 50332T (Z76659)
Pseudomonas delhiensis RLD-1 (DQ339153)
Pseudomonas nitritireducens WZBFD3-5A2 (HM246143)
Pseudomonas nitroreducens IAM 1439 (AM088473)
Pseudomonas jinjuensis Pss 26 (AF468448)
Pseudomonas caeni HY-14 (EU620679)
Thiopseudomonas denitrificans X2 (KJ567598)
0.04

Figure 4.7: ML phylogenetic tree of *Pseudomonas* spp. based on 16S rRNA gene sequences (max 1545 bp). The 16S rRNA gene sequence of *Rhizobacter dauci* H6 (AB297965) was used as an outgroup. Bar, 0.04 substitutions per nucleotide position. Type species of genera are shown in bold face. Organisms identified in this study are in red. Values at nodes represent FastTree support values from 1,000 bootstraps, where 1 is equal to 100 percent. GenBank accession numbers are shown in parenthesis. Tree created using Geneious Prime version 2019.0 by Biomatters.

As seen in Figure 4.7, the isolates B1 (Res), B3 (Sen) and C3 (Sen) share a common ancestor. Phylogeny suggests that they share ancestry with *Pseudomonas rhizosphaerae* (FastTree support value 0.94).

Finally, isolates E3 (Sen) and C5 (Res) cluster on the same clade for the *Stentrophomonas* 16S rRNA tree, as seen below in Figure 4.8.

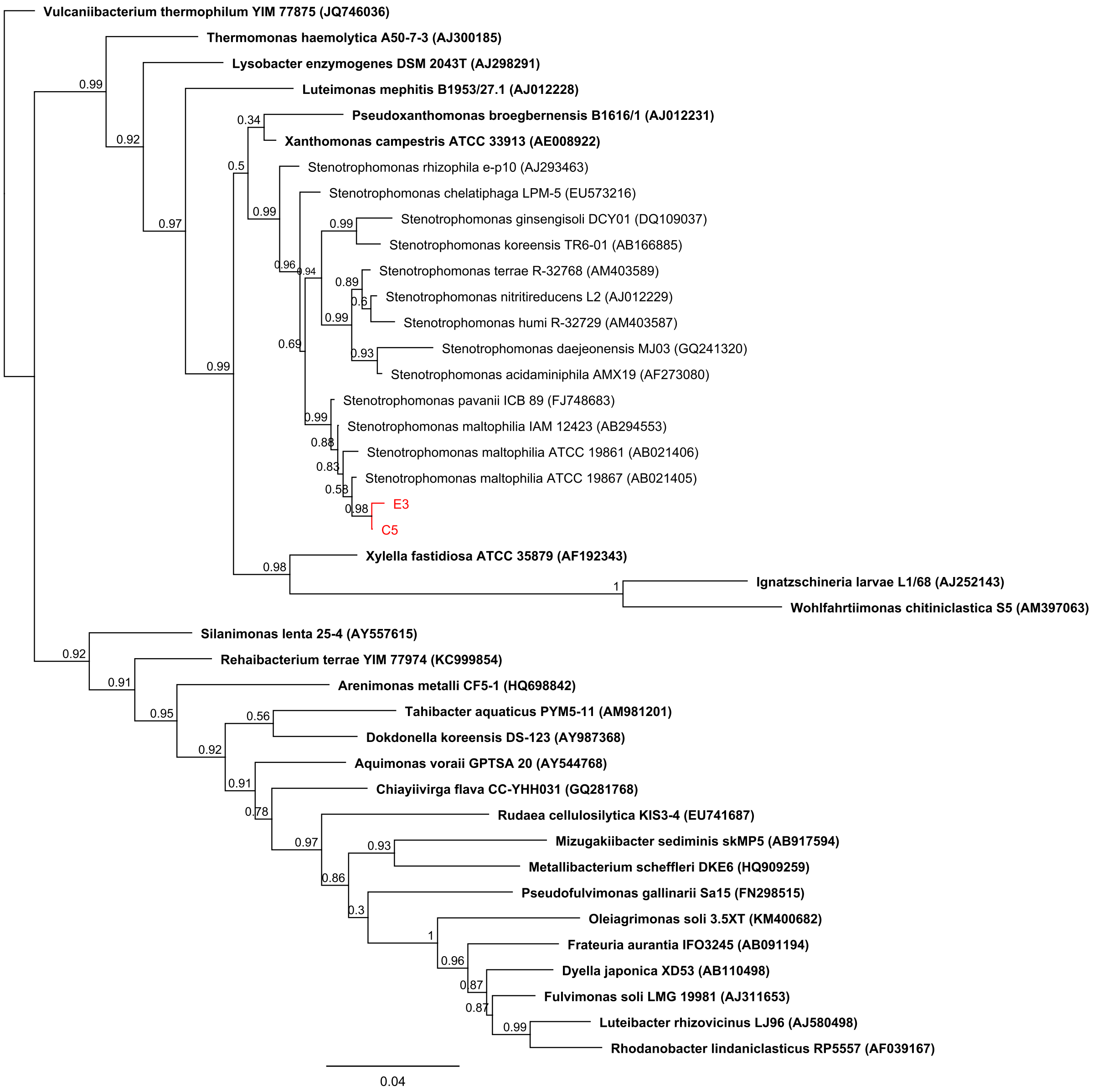

Vulcaniibacterium thermophilum YIM 77875 (JQ746036)
Thermomonas haemolytica A50-7-3 (AJ300185)
Lysobacter enzymogenes DSM 2043T (AJ298291)
Luteimonas mephitis B1953/27.1 (AJ012228)
Pseudoxanthomonas broegbernensis B1616/1 (AJ012231)
Xanthomonas campestris ATCC 33913 (AE008922)
Stenotrophomonas rhizophila e-p10 (AJ293463)
Stenotrophomonas chelatiphaga LPM-5 (EU573216)
Stenotrophomonas ginsengisoli DCY01 (DQ109037)
Stenotrophomonas koreensis TR6-01 (AB166885)
Stenotrophomonas terrae R-32768 (AM403589)
Stenotrophomonas nitritireducens L2 (AJ012229)
Stenotrophomonas humi R-32729 (AM403587)
Stenotrophomonas daejeonensis MJ03 (GQ241320)
Stenotrophomonas acidaminiphila AMX19 (AF273080)
Stenotrophomonas pavanii ICB 89 (FJ748683)
Stenotrophomonas maltophilia IAM 12423 (AB294553)
Stenotrophomonas maltophilia ATCC 19861 (AB021406)
Stenotrophomonas maltophilia ATCC 19867 (AB021405)
E3
C5
Xylella fastidiosa ATCC 35879 (AF192343)
Ignatzschineria larvae L1/68 (AJ252143)
Wohlfahrtiimonas chitiniclastica S5 (AM397063)
Silanimonas lenta 25-4 (AY557615)
Rehaibacterium terrae YIM 77974 (KC999854)
Arenimonas metalli CF5-1 (HQ698842)
Tahibacter aquaticus PYM5-11 (AM981201)
Dokdonella koreensis DS-123 (AY987368)
Aquimonas voraii GPTSA 20 (AY544768)
Chiayiivirga flava CC-YHH031 (GQ281768)
Rudaea cellulosilytica KIS3-4 (EU741687)
Mizugakiibacter sediminis skMP5 (AB917594)
Metallibacterium scheffleri DKE6 (HQ909259)
Pseudofulvimonas gallinarii Sa15 (FN298515)
Oleiagrimonas soli 3.5XT (KM400682)
Frateuria aurantia IFO3245 (AB091194)
Dyella japonica XD53 (AB110498)
Fulvimonas soli LMG 19981 (AJ311653)
Luteibacter rhizovicinus LJ96 (AJ580498)
Rhodanobacter lindaniclasticus RP5557 (AF039167)
0.99
0.92
0.97
0.34
0.5
0.99
0.99
0.96
0.94
0.89
0.6
0.99
0.69
0.93
0.99
0.99
0.88
0.83
0.58
0.98
0.98
1
0.92
0.91
0.95
0.56
0.92
0.91
0.78
0.97
0.93
0.86
0.3
1
0.96
0.87
0.87
0.99
0.04

Figure 4.8: ML phylogenetic tree of *Stentrophomonas* spp. based on 16S rRNA gene sequences (max 1548 bp). The 16S rRNA gene sequence of *Vucanlibacterium thermophilum* YIM 77575 (JQ746036) was used as an outgroup. Bar, 0.04 substitutions per nucleotide position. Type species of genera are shown in bold face. Organisms identified in this study are in red. Values at nodes represent FastTree support values from 1,000 bootstraps, where 1 is equal to 100 percent. GenBank accession numbers are shown in parenthesis. Tree created using Geneious Prime version 2019.0 by Biomatters.

Isolates C5 (Res) and E3 (Sen) appear to share a common ancestor (Figure 4.8). Isolate C5 appears to be an ancestral sequence to E3. In addition, C5 appears to share ancestry with *Stentrophomonas maltophilia* ATCC 19867.

As seen in Figures 4.6, 4.7 and 4.8, all phylogenetic trees have a UVC resistant and UVC sensitive ‘pair’ in the same clade. *Pseudomonas* sp. isolates B1 (Res), B3 (Sen) and C3 (Res) (Figure 4.7) and *Stentrophomonas* sp. isolates E3 (Sen) and C5 (Res) (Figure 4.8) were all isolated from the Dry Valleys. As seen in Figures 4.7 and 4.8, these isolates cluster together on their respective phylogenetic trees.

The phylogenetic tree of *Arthrobacter* and *Pseudarthrobacter* (Figure 4.6) shows more phylogenetic diversity amongst the isolates identified in this study. However, as noted by Nouioui et al. (2018), the *Arthrobacter* and *Pseudarthrobacter* genera form paraphyletic groups with low comprehensive tree support values, meaning that inferring accurate phylogenetic relationships is difficult. Despite these issues, isolates B4 (Res) and E5 (Sen) appear to be closely related based on their 16S rRNA gene.

The genera *Arthrobacter*, *Pseudarthrobacter*, *Pseudomonas* and *Stentrophomonas* contained closely related isolates (based on 16S rRNA gene sequences) which were identified as either UVC resistant or UVC sensitive. To the author’s knowledge, this is the first time that UV resistance has been reported in the *Pseudarthrobacter* genus. The primary aim of this thesis is to investigate genomic insights into how bacteria survive radiation. Comparing the whole genome of UV resistant and UV sensitive isolates of close relation may help in answering how UV resistance has developed, or been lost, in two isolates of close relation. For further analysis into the genomic differences between UVC resistant and UVC sensitive bacteria, isolates B4 and E5 were selected as representatives of the *Pseudarthrobacter* genus.

In addition, isolates B2 (Res) and E2 (Sen) were also selected for further genome analysis. E2 was selected as a UVC sensitive representative of the *A. sanguinis* clade. Members of the *A. sanguinis* clade have not previously been reported as UV resistant, making E2 a good candidate for comparative genomics against UV resistant isolates. Previous studies have found UVC resistant *Arthrobacter* that also share common ancestry with *A. agilis* (Ii et al. 2019). Since UVC resistance within this clade has previously been reported, B2 was selected for further genomic comparison with isolate B4 to determine if these two UVC resistant isolates share proteins that are not shared by the UVC sensitive isolates. The presence of DNA repair pathways in isolates B2 and B4 that are absent in isolates E2 and E5 may provide further insight into bacterial DNA repair following UV exposure.

While isolate E7 (Figure 4.6) was identified as a UVC resistant isolate, this isolate proved difficult to re-cultivate. It appears that isolate E2 is closer to the last common ancestor it shares with isolate E7 (Figure 4.6). Further investigations into the relationship of E2 and E7 may reveal further insights into UV resistance of these UVC resistant and UVC sensitive organisms. Isolates

E1 (Sen), E7 (Res) and C1 (Sen) (Figure 4.6) are of interest for further study as they belong to distinct clades in the 16S rRNA phylogenetic tree.

4.3.4 Comparison of UVC survivability between UVC resistant and UVC sensitive isolates

To determine isolate survivability following UVC exposure, further UVC survivability tests were carried out on resistant and sensitive organisms of the *Arthrobacter* and *Pseudarthrobacter* isolates B2, B4, E2 and E5. All four isolates were irradiated with UVC (256 nm) at 1 W/m^2 for 10 mins. The results can be seen below in Table 4.5. The data from Table 4.5 is illustrated in Figure 4.9.

Table 4.5: Mean log$_{10}$ CFU of the Namib isolates during UVC exposure at 1 W/m2 for up to 10 minutes.

	Exposure time at 1 W/m^2 UVC 256 nm			
	0 min	**1 min**	**5 mins**	**10 mins**
B2	8.695	8.640	7.597	5.588
B4	7.176	6.079	5.954	0
E2	8.056	7.499	2.889	0
E5	8.595	7.615	0	0

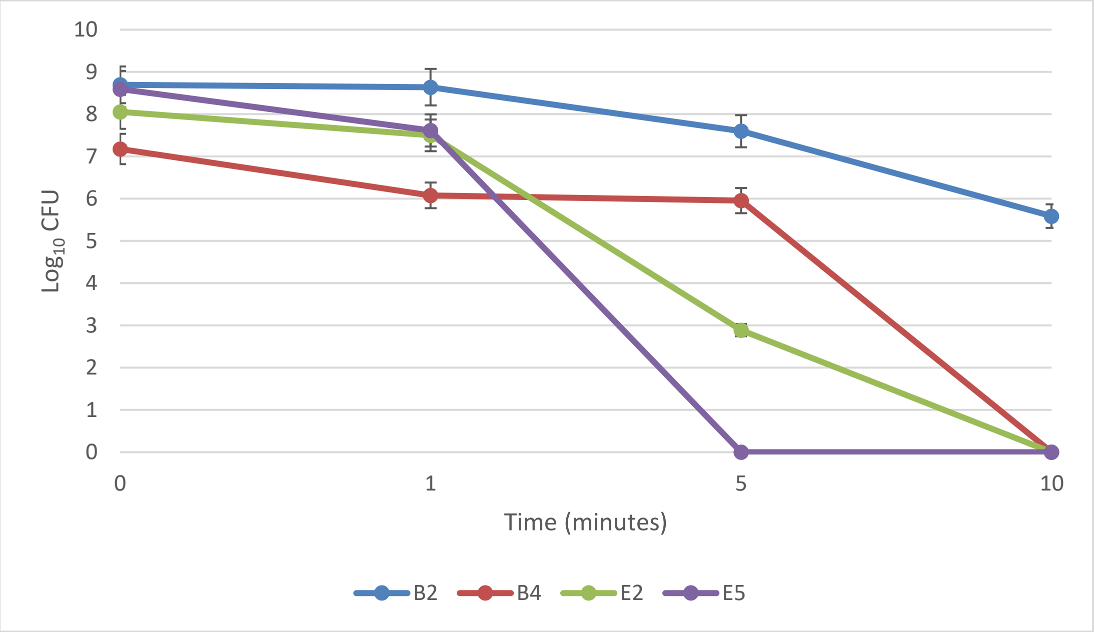


Figure 4.9: Log CFU of the Namib isolates during UVC exposure at 1 W/m2 for up to 10 minutes.

Isolate B2 was able to survive 1 W/m^2 of UVC at a detectable level for 10 minutes. This supports the observed results of the rapid screening test, which found that B2 was able to survive 10 minutes of UVC 256 nm at both 1 W/m^2 and 5 W/m^2. Isolate B2 reduced by 3.1 log$_{10}$ CFU following 10 minutes of 1 W/m^2 UVC exposure, showing that approximately 99.9% of live bacterial cells were killed by UVC. Isolate E5 was able to survive for 1 minute of 1 W/m^2 UVC before dropping to an undetectable level by 5 minutes of exposure. Isolate E2 was able to survive to 5 minutes of 1 W/m^2 UVC before dropping to an undetectable level by 10 minutes of exposure.

There was a 4.5 log drop for isolate E2 between 1 minute and 5 minutes of exposure. Both E2 and E5 appear to be sensitive to 1 W/m^2 of UVC. This is again consistent with the rapid screening results.

Unexpectedly, isolate B4, which was identified as UVC resistant during rapid screening, also showed sensitivity to UVC under these testing conditions by surviving 5 minutes of 1 W/m^2 of UVC before experiencing a 6 log drop between 5 and 10 minutes of exposure. Rapid screening previously found that isolate B4 was able to survive 10 mins of 1 W/m^2 UVC. Visual analysis of the B4 colonies showed that B4 grows as white colonies, rather than the red-pink colonies that were first observed during rapid screening. This loss of pigmentation may have contributed to the discrepancy in results between the two testing methods.

Further testing on the *Arthrobacter* spp. and *Pseudarthrobacter* spp. strains was conducted at 5 W/m^2 of UVC. All isolates were irradiated with UVC (256 nm) at 5 W/m^2 for 10 mins. The results can be seen below in Table 4.6. The data from Table 4.6 is illustrated in Figure 4.10.

Table 4.6: Mean log$_{10}$ CFU of the Namib isolates during UVC exposure at 5 W/m2 for up to 10 minutes.

		Exposure time at 5 W/m^2 UVC 256 nm		
	0 min	**1 min**	**5 mins**	**10 mins**
B2	8.695	7.574	5.748	2.588
B4	7.176	5.176	0	0
E2	8.056	6.366	0	0
E5	8.595	6.891	0	0

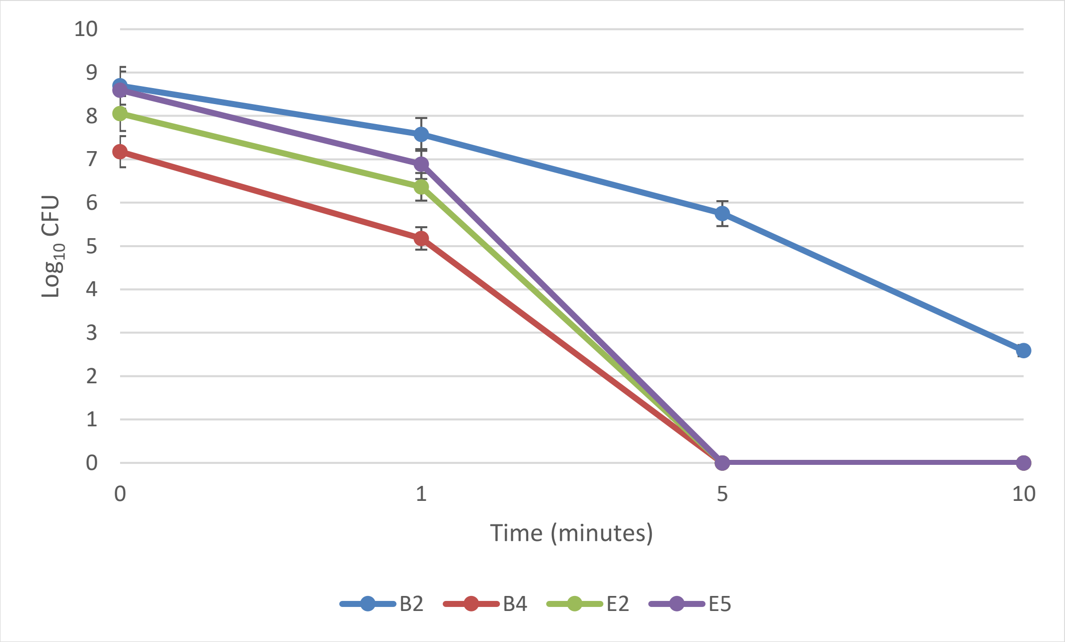


Figure 4.10: Figure 4.10: Log CFU of the Namib isolates during UVC exposure at 5 W/m2 for up to 10 minutes.

As seen in Figure 4.10, isolate B2 was able to survive 5 W/m^2 of UVC at a detectable level for 10 minutes. This is again consistent with the results observed through rapid screening and indicates that isolate B2 has retained UVC resistance after laboratory storage. Isolate B2 reduced by 6.1 $\log_{10}$ CFU following 10 minutes of 1 W/m^2 UVC exposure, showing that approximately 99.9999% of live bacterial cells were killed by UVC at 10 minutes. Isolates E2 and E5 were able to survive for 1 minute of 5 W/m^2 UVC before a severe drop to an undetectable level by 5 minutes of exposure. Both E2 and E5 appear to be sensitive to 5 W/m^2 of UVC, which is consistent with the rapid screening results. Isolate B4 was only able to survive 1 minute of 5 W/m^2 of UVC before dropping to an undetectable level by 5 minutes of exposure. During rapid screening, isolate B4 was able to survive 10 minutes of 1 W/m^2 and 10 minutes of 5 W/m^2 UVC (Appendix 1). These results are therefore inconsistent with the results obtained from rapid screening and may be due to the loss of pigmentation by B4.

4.4 Discussion

The aim of this Chapter was to identify and analyse UV resistant organisms isolated from the Dry Valleys and the Namib Desert. To achieve one part of this aim, a rapid screening method was developed to quickly identify UV resistant bacteria. Following rapid screening, UVC resistant and UVC sensitive isolates were identified to the genus level by Sanger sequencing of the 16S rRNA gene and phylogenetic relationships between these isolates were investigated. Finally, to determine the UV survivability of isolates B2, B4, E2 and E5, further survivability experiments were carried out to observe the $\log_{10}$ CFU reduction following UV exposure.

4.4.1 Isolated bacteria

Soil was inoculated onto culture media and grown at temperatures ranging from 4°C to 45°C (Table 4.2). Overall, most bacteria were isolated at 20°C followed by 15°C for both locations. Four presumed bacterial isolates were isolated at 35°C from the Dry Valleys. No bacteria were isolated at 45°C for the Dry Valleys. Bacteria were isolated from the Namib Desert at both 35°C and 45°C. More bacteria were isolated from the Dry Valleys at 4°C than the Namib Desert. This is most likely due to psychotropic bacteria being present in the Dry Valley soil (Cowan and Tow 2004).

Through the course of this study 14 to 40 bacterial isolates were cultivated from each sample site used in this study. This is an expected number of bacterial isolates from restrictive soil environments (Aislabie et al. 2006). The estimation of biomass within the Dry Valleys soil ranges from 10^6 – 10^8 cells per gram of soil (Cowan et al. 2002), while other studies have found between 10^3 – 10^4 cells per gram of soil (Goordial et al. 2016; Stomeo et al. 2012). The biomass for the sample sites used in this study was not investigated and as such the percentage of isolated bacteria cannot be determined.

A disadvantage of plate based-isolation methods is that visual observation of the cultures does not determine if bacteria isolated at different temperatures are the same organism. Thus, the actual number of bacteria isolated in this study may be far lower than indicated. One way to preliminary identify an organism would be to conduct Sanger sequencing of the 16S rRNA gene amplified from each isolate. However, this process is time consuming and expensive with the number of bacteria isolated, and there are doubts about the accuracy of using the 16S rRNA gene for bacterial identification (Janda and Abbott 2007; Woo et al. 2008; Zengler 2009). In addition, this study does not aim to be an exhaustive search of identifying every organism isolated onto agar. Bacteria were isolated to see if they are resistant to UV radiation, with the identity of UVC resistant and UVC sensitive isolates of further interest. Further studies may wish to identify each of the bacterial isolates via the 16S rRNA gene. This would provide a comprehensive analysis of the percentage of ASVs that were grown on agar vs the estimated number that were present in the community. However, again, this was not within the scope of this thesis.

4.4.2 UV resistance and genus identification

The development of a rapid screening method led to the preliminary identification of four genera with UVC resistant and UVC sensitive isolates. It was expected that some isolates from both the Dry Valleys and the Namib Desert would be UV resistant. The Dry Valleys are exposed to greater levels of UV radiation than most other global locations due to the ozone hole (Dib et al. 2008). During the Dry Valley soil collection for this study in January 2017, UVC irradiance was recorded as 1.07 W/m^2 at the soil surface (UVC radiometer, UVP Inc, Upland, CA). The Western Coast of Africa receives more UVB radiation than most other global locations (Beckmann et al. 2014; Lucas et al. 2016). As such, it was expected that some isolates would be able to survive high levels of UVB or UVC radiation. The *Pseudomonas* and *Stenotrophomonas* isolates identified in this study were isolated from the Dry Valleys, while the *Arthrobacter* and *Pseudarthrobacter* isolates were isolated from the Namib Desert.

Both *Pseudomonas* and *Stenotrophomonas* have previously been observed in the Antarctic soils (Lambrechts et al. 2019; Vazquez et al. 2005). Cowan et al. (2014) notes that Antarctica is a hostile environment that generally responds slowly to change. The high genetic similarity of the 16S rRNA gene of isolates from within this region is most likely due to the isolated nature of the Dry Valleys. Previous investigations into UV resistance of bacteria in the Dry Valleys have shown that *Deinococcus* sp. and *Halomonas* sp. are resistant to ionising radiation (Dartnell et al. 2010; Hirsch et al. 2004; Musilova et al. 2015). Dartnell et al. (2010) noted that *Pseudomonas* sp. isolated from the Dry Valleys displayed low ionising radiation tolerance. This was supported by Flores et al. (2009) who noted that *Pseudomonas* sp. isolated from Andean lakes were sensitive to UV radiation. However, as noted by Dib et al. (2008) and Ordoñez et al. (2009), some strains of *Pseudomonas* are resistant to UVB radiation. Additionally, *Pseudomonas* spp. such as *P. aeruginosa* have previously been reported to be resistant to UV radiation due to FP sex factors such as FP50 and FP58 (Krishnapillai 1975). Vazquez et al. (2005) has previously reported the presence of *Stenotrophomonas* spp. in the Dry Valleys. Flores et al. (2009) noted that *Stenotrophomonas* sp. isolated from Andean lakes were resistant to UV radiation. As such, the UV resistance observed in isolate C5 (Figure 4.8) and B1 (Figure 4.7) is expected based on previous literature. Isolate B1 may have attained UV resistance based on environmental pressures of the Dry Valleys, either through adaptation or the uptake of a plasmid.

Arthrobacter and *Pseudarthrobacter* have previously been reported in soil from the Namib Desert (Makhalanyane et al. 2015; Ronca et al. 2015). However, investigations into the UV resistance of bacteria from the Namib Desert has not previously been reported. Although this is the first report of UVC resistant *Arthrobacter* from the Namib Desert, previous studies have documented UV resistance of this genus (Ii et al. 2019; Kumar et al. 2016; Yoshinaka et al. 1973). The UVC resistant *Arthrobacter* and *Pseudarthrobacter* (Figure 4.6) were also found to be resistant to UVB radiation (Appendix 1). The Western Coast of Africa receives more UVB radiation than most other global locations, therefore the resistance of these isolates to UVB is expected. However, the Namib Desert does not receive UVC radiation at the soil surface (NIWA 2016), indicating that the survival mechanisms for UVB and UVC in these isolates may be initiated by similar pathways. The Namib Desert is the oldest desert in the world, and it could therefore be used as an important environment for providing information on the evolution of survival mechanisms for DNA repair.

Sequencing of the 16S rRNA region allowed for identification to at least the genus level, and phylogenetic analysis showed that each of the four genera had a clade with one resistant and one sensitive isolate (Figures 4.6, 4.7 and 4.8). While the 16S rRNA gene is only an indication of phylogenetic relationships (Janda and Abbott 2007), to the author's knowledge, this is the first observation of highly related UV resistant and UVC sensitive isolates that appear in the same genus clade. The four genera found to have UVC resistant and UVC sensitive isolates were isolated from the two most dominant phyla. *Arthrobacter* and *Pseudarthrobacter* are from the Actinobacteria phylum, while *Pseudomonas* and *Stenotrophomonas* are from the Proteobacteria phylum. This is expected, as both Actinobacteria and Proteobacteria are readily cultivable on culture media (Rasmussen et al. 2008) and were able to be tested for UV resistance and sensitivity by the testing conditions of this study.

4.4.3 Phylogenetic analysis

A wide range of organisms were identified as UVC resistant or UVC sensitive by the testing conditions of this thesis (Table 4.4). Isolates were identified from the genera *Arthrobacter*, *Pseudarthrobacter*, *Pseudomonas* and *Stenotrophomonas* have UVC resistant and UVC sensitive counterparts. These genera have previously been observed as displaying some degree of UV resistance (Dib et al. 2008; Flores et al. 2009; Kuhlman et al. 2005; Ordoñez et al. 2009; Osman et al. 2008), as well as UV sensitivity (Claußß 2006; Flores et al. 2009; Ordoñez et al. 2009). To date, there is no information regarding the UV resistance of *Pseudarthrobacter* spp. Isolates of the *Brachybacterium*, *Bacillus* and *Pantoea* genera were also UV resistant by the testing conditions of this study. UVC resistance has been observed previously in *Brachybacterium* sp. (Dib et al. 2008; Ordoñez et al. 2009; Osman et al. 2008), *Bacillus* sp. (Dib et al. 2008; Ordoñez et al. 2009; Osman et al. 2008) and *Pantoea* sp. (Mohammadi et al. 2012; Walterson and Stavrinides 2015). These isolates were not pursued for further analysis because there were no UVC sensitive isolates identified by Sanger sequencing for these genera.

Phylogenetic analysis showed that some resistant and sensitive isolates shared common ancestry (Figures 4.6, 4.7 and 4.8). Isolates B4 (Res) and E5 (Sen) share a common ancestor with *P. polychromogenes*, although the closest BLASTn sequence for both isolates was *P. phenanthrenivorans*. Isolates E2 (Sen) and E7 (Res) share a common ancestor and the closest 16S rRNA sequence to both isolates was *A. nitrophenolicus*. The isolates B3 (Sen), C3 (Sen) and B1 (Res) share a common ancestor with *P. rhizophaerae* and *P. abietaniphila*. Neither *P. rhizosphaerae* or *P. abietaniphila* have previously been reported as resistant to UV radiation, indicating that isolate B1 may have attained UV resistance as a survival mechanism within the Dry Valleys following divergence from isolate C3. Finally, isolates C5 (Res) and E3 (Sen) appear to share a common ancestor, with the C5 sequence having a shorter branch length than E3 in the rooted tree. The closest common ancestor to isolate C5 is *S. maltophilia*; an organism that has previously been reported as resistant to UV radiation (Flores et al. 2009).

The community profile (Figure 3.4, Section 3.3) indicated that members of the extremophile phyla Deinococcus-Thermus were present in both the Dry Valley and Namib soils. Organisms from this phylum may not have been isolated on agar during this study, or the 16S rRNA genes identified in the community profile could have been from a historical community (Cowan et al. 2014).

Isolates B4 (Res) and E5 (Sen) were selected for further genomic analysis as representatives of the *Pseudarthrobacter* genus. Investigations into the UV resistance of *Pseudarthrobacter* spp. is important as UV resistance has not previously been reported in this genus. Isolates B2 (Res) and E2 (Sen) were also selected for further genome analysis to examine if DNA repair pathways that were present in the UVC resistant isolates were absent in the UVC sensitive isolates.

Phylogenetic analysis of the *Arthrobacter* and *Pseudarthrobacter* 16S rRNA sequences showed low bootstrap support values for the phylogenetic relationship within these genera. This has previously been reported as a persistent issue within the two genera (Busse et al. 2012; 2015; Nouioui et al. 2018). Therefore, to assign accurate taxonomy to isolates from these genera, a whole genome approach is required. A detailed analysis of the phylogeny and taxonomy of the *Arthrobacter* and *Pseudarthrobacter* genera will be carried out in the following Chapter.

4.4.3 UVC survivability

Isolates B2, B4, E2 and E5 were selected for further UVC survivability testing and genomic analysis. Previous UV screening in this thesis indicated that isolates B2 and B4 were resistant to 5 W/m^2 of UVC for 10 minutes, while isolates E2 and E5 were sensitive to radiation after 1 minute of 5 W/m^2 UVC. Survivability analysis revealed that isolate B2 was still resistant to UVC radiation, but isolate B4 appeared to have lost some of its UVC resistance. This experiment demonstrated that isolate B4 appeared to have lost the ability to survive UVC radiation. In the time between these two experiments, isolate B4 appeared to lose its red colony pigmentation, which is theorised to have contributed towards the sensitivity of B4 observed in this experiment.

4.4.4 UV survival mechanisms

UV repair mechanisms are often multi-parametric in that there are many systems that act together to repair UV induced DNA damage (Ii et al. 2019). UV tolerance is speculated to be related to desiccation resistance; however, repair mechanisms of radiation damage are diverse amongst different organisms (Ii et al. 2019), and it is likely that most of the UVC resistant isolates have desiccation resistance genes.

Pavlopoulou et al. (2016) noted that induced radiation damage does not significantly vary depending on the radioresistant or radiosensitivity of an organism. That is, the accumulated damage in UVC resistant and UVC sensitive organisms is thought to be the same. UV resistant organisms appear to efficiently remove ROS, possibly due to antioxidants playing a regulative role in protecting DNA (Pavlopoulou et al. 2016). More recently, it has also been speculated that intracellular Mn/Fe ratios and the presence of Mn^{2+} complexes play key roles in defence against ROS (Ii et al. 2019; Paulino-Lima et al. 2016; Pavlopoulou et al. 2016). Interestingly, the availability of Mn extracellularly in the environment does not correlate with UVC resistance (Paulino-Lima et al. 2016). Further investigations into the ability to manipulate the UV survivability of B2, B4, E2 and E5 could therefore include increasing the Mn/Fe intracellular ratio.

Previous studies have noted UV resistance within the *Arthrobacter* genus. *A. alpinus* ERGS4:06, *Arthrobacter* sp. MN05-02 and *Rubrobacter radiotolerans* [previously *A. radiotolerans*] have all been observed to survive high levels of UV (Ii et al. 2019; Kumar et al. 2016; Yoshinaka et al. 1973). Although these *Arthrobacter* spp. displayed pigmentation, which the authors speculated helped the isolates survive radiation, other non-pigmented *Arthrobacter* have been known to survive UV also (Kuhlman et al. 2005). Both UVC resistant isolates B2 and B4 showed red pigmented colonies on agar, while the UVC sensitive isolates E2 and E5 produced white-cream colonies on agar.

Carotenoids have been known to help protect bacterial cells from ROS that cause DNA damage (Dieser et al. 2010; Sutthiwong et al. 2014). The production of a carotenoid in B2 and B4 may be one of the mechanisms by which these organisms survive UV ROS damage in a hostile desert environment. Colonies of E5 and E2 appeared white on agar at pH 7 indicating that they do not have active genes for pigment production. This may be a contributing factor in the UV sensitivity of E2 and E5.

Previous studies have noted that pigment production in bacteria leads to increased survivability when exposed to UV radiation (Dieser et al. 2010; Sutthiwong et al. 2014; Yoshinaka et al. 1973). Isolates B2 and B4 produce a red-pink pigment on agar, which is water insoluble. Phylogenetic analysis (Figure 4.6) showed that *A. agilis* and *A. pityocampae* are closely related to B2. Both *A. agilis* and *A. pityocampae* produce a red-pink pigment on agar (Busse et al. 2015; İnce et al. 2014). Dsouza et al. (2015) noted that genes for the synthesis of the red-pigment lycopene have been identified in *Arthrobacter* H41 and Br18, both of which are phylogenetically similar to *A. agilis* and *A. pityocampae*. However, Dieser et al. (2010) and Sutthiwong et al. (2014) note that the carotenoid bacterioruberin has been isolated from *A. agilis* MB8-13. Further investigation into the carotenoid biosynthesis genes of B2 may help to distinguish if the pigment is lycopene or bacterioruberin.

Phylogenetic analysis (Figure 4.6) also showed that *P. phenanthrenivorans* and *P. polychromogenes* are closely related to B4. *P. phenanthrenivorans* and *P. polychromogenes* have not been observed to produce red-pink colony pigmentation. *P. phenanthrenivorans* grows as beige colonies on agar (Kallimanis et al. 2009), while *P. polychromogenes* grows as blue colonies (Schippers-Lammertse et al. 1963). B4 may have adapted to produce a red-pink carotenoid under the restrictive Namib Desert conditions; however, during the course of laboratory experimentations, B4 lost its red-pink pigment and began to grow as white colonies. Loss of pigmentation can occur during bacterial subculturing (Lowe et al. 2010) and the favourable laboratory conditions compared to the desert environment may have caused B4 to lose its pigment production ability. Following the loss of this pigmentation, it was observed that B4 became more sensitive to UVC radiation, suggesting that the pigment produced by B4 was instrumental in prolonged UVC survival.

4.4.5 UV survivability and critique of method

The rapid screening method used in this thesis was developed as a way for quickly identifying the UV survivability of a large number of bacterial isolates. Several isolates were identified as UVC resistant through this method (Table 4.4).

The testing conditions of the rapid screening method proved to be a reliable indicator of UV resistance or sensitivity. The results achieved in rapid screening were able to be replicated using different testing conditions and media for exposure. The rapid screening method utilised a small amount of culture broth exposed on agar while subsequent UV resistance testing utilised culture diluted and exposed in peptone water. Both testing conditions indicated that B2 was resistant to UVC while isolates E2 and E5 were sensitive to UVC radiation. During rapid screening, isolate B4 was able to survive 10 minutes of 1 W/m^2 and 10 minutes of 5 W/m^2 UVC. However, during subsequent testing, B4 lost its pigmentation and was found to be resistant to only 5 minutes of 1 W/m^2 UVC and 1 minute of 5 W/m^2 UVC. This strengthens the observations of previous studies which have observed that pigment production in bacteria leads to increased survivability when exposed to UV radiation (Dieser et al. 2010; Sutthiwong et al. 2014; Yoshinaka et al. 1973).

Isolates B2, B4, E2 and E5 were selected for whole genome analysis based on their phylogenetic relationships, and the importance of their phylum in desert soil systems. *Arthrobacter* are readily cultured from desert soils but, currently, limited information is known about the physiological traits that allow these organisms to survive in restrictive desert systems.

From the 16S rRNA phylogenetic analysis of *Arthrobacter* and *Pseudarthrobacter* (Figure 4.6, Section 4.3), it was established that the isolates B4 and E5 shared a common ancestor. To the author's knowledge, literature has not reported on apparently highly related organisms in which one is sensitive to and the other is resistant to UV radiation. Isolates B2 (Res) and E2 (Sen) were also selected for further genome analysis alongside isolates B4 and E5. B2 and E2 serve as representatives of the *A. agilis* clade and the *A. sanguinis* clade, respectively. Using comparative genomics to compare genes potentially involved in UV resistance between the resistant and sensitive isolates is of great interest going forward.

Chapter 5: Whole genome analysis of *Arthrobacter* and *Pseudarthrobacter*

5.1 Introduction

In Chapter 4, two UVC resistant (B2 and B4) and two UVC sensitive organisms (E2 and E5) were selected for further analysis based on their 16S rRNA gene phylogeny. Based on near-full length 16S rRNA gene sequences generated by Sanger sequencing, B2 and E2 were identified as *Arthrobacter*, and B4 and E5 were identified as the recently established genus *Pseudarthrobacter* (Busse 2016). To better understand how these organisms have adapted to the harsh environmental conditions of the Namib Desert, the isolates B2, E2, B4 and E5 were metabolically and genomically characterised.

5.1.1 The *Arthrobacter* genus

The genus *Arthrobacter*, belonging to the family Micrococcaceae and class Actinobacteria, was officially established by Conn and Dimmick (1947) after the genus name was proposed and abandoned in 1895 (Busse 2016; Hu et al. 2016; Yu et al. 2015). Three type species were proposed; *Arthrobacter globiforme* [later renamed to *Arthrobacter globiformis* (Skerman et al. 1980)], *Arthrobacter tumescens* [later reclassified to *Terrabacter tumescens* (Busse et al. 2012)] and *Arthrobacter helvolum* [later reclassified to *Pseudoclavibacter helvolum* (Busse et al. 2012)].

Arthrobacter as a genus are metabolically and ecologically diverse, and can survive environmentally challenging conditions for lengthy periods of time (Mongodin et al. 2006). The *Arthrobacter* genus is generally regarded as obligate aerobes with an oxidative mode of metabolism (Eschbach et al. 2003). However, changes in oxygen tension often occur in the upper soil layers, leading some *Arthrobacter* species to develop oxygen-independent growth strategies such as nitrogen utilisation (Eschbach et al. 2003). Based on the EzBioCloud [originally EzTaxon] server (Chun et al. 2007) the complete assembly genome size for *Arthrobacter* ranges from 2.59-5.89 Mb, with a GC content of 59%-69.8%.

5.1.2 Morphology

The morphology of *Arthrobacter* spp. as described by Conn and Dimmick (1947) is varied, with the tendency of the genus to present as Gram-negative rods in young cultures, and Gram-positive coccoid forms in older cultures (Busse 2016; Yu et al. 2015). As such, *Arthrobacter* spp. are characterised as pleomorphic and Gram variable (Busse et al. 2015).

5.1.3 Cultural characteristics and physiology

Skerman et al. (1980) stated that *Arthrobacter* candidates should: i) grow on the surface of solid media as soft, smooth colonies; ii) growth in broth is usually slow but not overly abundant; iii) colonies on poured agar are typically small.

For physiology, *Arthrobacter* spp. can use either ammonium salts or nitrates as sole nitrogen sources. Glucose, and occasionally other sugars, can be utilised as carbon and energy sources, but without excess acid production. Gelatin can be liquefied, albeit slowly, and *Arthrobacter* spp. will usually cause blackening on Mueller's tellurite agar (Busse 2016; Skerman et al. 1980). *Arthrobacter* spp. are catalase positive, contains the quinones MK-9 (H2) or MK-8/MK-9, and the peptidoglycan A3α or A4α (Hu et al. 2016; Yu et al. 2015). These characteristics were the basis for classification of novel *Arthrobacter* species up until the genus reclassification (Busse 2016).

Arthrobacter spp. have previously been recovered from a wide variety of environments, including soil, sea water, fresh water, human skin, oil, tobacco leaves, air, sewage, clinical specimens and cyanobacterial mats (Busse et al. 2012; Fu et al. 2014). They are among the most commonly isolated aerobic bacteria from soils (Busse et al. 2012; Mongodin et al. 2006), and they have previously been isolated from deserts (Dsouza et al. 2015; Hu et al. 2016; Yu et al. 2015). Some species of *Arthrobacter* can utilise halogenated organic compounds such as 4-chlorophenol, or reduce environmentally damaging heavy metals such as hexavalent chromium Cr(VI) to the less toxic Cr(III), suggesting they may be useful for bioremediation (Camargo et al. 2004; Dsouza et al. 2015; Fu et al. 2014; Westerberg et al. 2000).

5.1.4 Reclassification of the *Arthrobacter* genus into six different genera

Busse (2016) noted that previous phylogenetic analyses utilising the Neighbour-Joining method have demonstrated that the *Arthrobacter* genus is non-monophyletic and does not have a stable internal structure. This is supported by the many branching nodes having <70% bootstrap values (Busse 2016; Busse et al. 2012), and that several species, such as *A. crystallopoietes* and *A. woluwensis*, occupy varying positions within different trees (Busse 2016).

The *Arthrobacter* genus was reclassified into six genera by Busse (2016) based on the 16S rRNA gene. Prior to this, Busse et al. (2012) identified 11 *Arthrobacter* 'groups' from phylogenetic clustering, high 16S rRNA gene sequence similarities, and group-specific chemotaxonomic traits such as peptidoglycan structures and quinone systems. The *Arthrobacter* genus was subsequently reclassified based on these traits and polar lipids. The specific polar lipids used for reclassification were monogalactosyldiacylglycerol (MGDG), digalactosyldiacylglycerol (DGDG), dimannosylglyceride (DMG), trimannosyldiacylglycerol (TMDG), phosphatidylinositol (PI) and unidentified phospholipids (PL1-5) (Busse 2016). The differences between the two genera can be seen in Table 5.1. The genera *Arthrobacter* and *Pseudarthrobacter* exhibit similar presence or absence of polar lipids (Table 5.1).

Table 5.1: Differences between the Arthrobacter and Pseudarthrobacter genus (Busse 2016). (+) indicates positive, (-) indicates negative, (v) indicates variable positive or negative.

Trait	***Arthrobacter***	***Pseudarthrobacter***
Quinone	MK-9 (H_2)	MK-9 (H_2)
Peptidoglycan	A3α (Lys-Ala_{2-3}) A11.5 or A11.6	A3α (Lys-Ser-Thr-Ala) A11.23
Polar lipids		
MGDG	+	+
DGDG	v	-
DMG	+	v
TMDG	+	+
PI	+	+
PL1-5	-	-

The *Arthrobacter* type species is *A. globiformis* (Busse 2016; Skerman et al. 1980). With the introduction of the *Pseudarthrobacter* genus, a new type species, *Pseudarthrobacter polychromogenes*, was established (Busse 2016).

One of the persistent issues with the *Arthrobacter* genus is that 16S rRNA gene phylogenetic trees do not have significant bootstrap values (>75%), making it difficult to determine phylogenetic relationships among *Arthrobacter* species using only the 16S rRNA gene. Nouioui et al. (2018) notes that problems remain after the reclassification of *Arthrobacter* and *Pseudarthrobacter*, due to the genera still forming paraphyletic groups and the low comprehensive tree support values for the revised genera. In addition to this, several *Micrococcaceae* genera are interspersed throughout the *Arthrobacter* clades, and share high 16S rRNA gene sequence identities with the type species *Arthrobacter globiformis*. This makes identification of novel *Arthrobacter* strains at the genus level based only on the 16S rRNA gene ambiguous (Busse et al. 2015). Nouioui et al. (2018) recommends using more genome sequences to help with the continued classification of the whole *Arthrobacter* genus. The reclassification of *Arthrobacter* has not resolved the issue of low phylogenetic tree support values within the *Arthrobacter* genus.

Due to the issues of relying on the *Arthrobacter* 16S rRNA gene for comprehensive taxonomic classification, a whole genome comparative approach is required for a more reliable identification of new isolates.

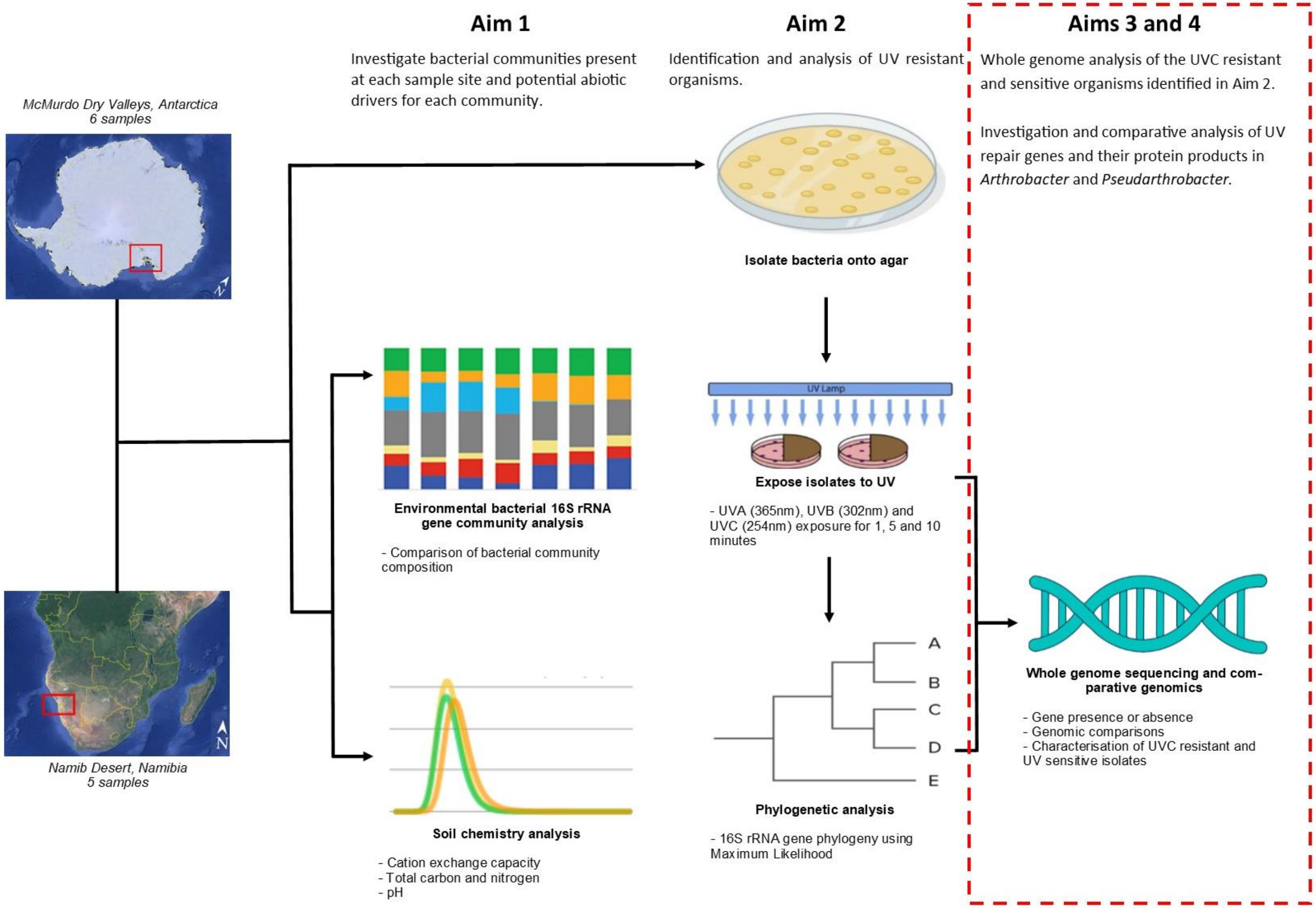
5.1.4 Comparative genomics

Comparative genomics is an important tool for understanding gene function. This often involves comparing genomic sequences to explore genetic similarity and diversity (Bhasin and Raghava 2006; Haubold and Wiehe 2004). Therefore, comparative genomic approaches will often start with a genomic alignment with a reference organism and a search for orthologous sequences among the genomes (Yao et al. 2015a).

Previous comparative genomic studies on *Arthrobacter* have been conducted (Chauhan et al. 2018; Dsouza et al. 2015; Yao et al. 2015b), however, these organisms were isolated from soils outside of the African continent. As the *Pseudarthrobacter* genus is more recent, comparative genomics studies into the genus have not been conducted yet, to the author's knowledge. This study is the first to report on UV resistant *Arthrobacter* and *Pseudarthrobacter* from the Namib Desert.

The focus of this Chapter is Aim 3: Whole genome analysis of the UVC resistant and sensitive organisms identified in Aim 2. To achieve this Aim, three objectives were outlined: (1) extract genomic DNA from the organisms selected in Objective 2.3 and sequence the whole genomes using the Illumina HiSeq™; (2) *de novo* assembly of the genomes and functional annotation using RAST and Prokka; (3) characterise the selected organisms using KEGG, COG and whole genome comparative analysis with other publicly available *Arthrobacter* and *Pseudarthrobacter* genomes.

In this chapter, the genomes of B2, E2, B4 and E5 were sequenced and compared with other publicly available genomes of the *Arthrobacter* and *Pseudarthrobacter* genera. The aim of this was to generate draft genome sequences of these isolates, to determine differences between the genomes of the newly sequenced Namib Desert strains and compare these isolates with other stains of *Arthrobacter* and *Pseudarthrobacter*.



Aim 1

Investigate bacterial communities present at each sample site and potential abiotic drivers for each community.

Aim 2

Identification and analysis of UV resistant organisms.

Aims 3 and 4

Whole genome analysis of the UVC resistant and sensitive organisms identified in Aim 2.

Investigation and comparative analysis of UV repair genes and their protein products in *Arthrobacter* and *Pseudarthrobacter*.

McMurdo Dry Valleys, Antarctica
6 samples

Namib Desert, Namibia
5 samples

Isolate bacteria onto agar

Expose isolates to UV

- UVA (365nm), UVB (302nm) and UVC (254nm) exposure for 1, 5 and 10 minutes

Phylogenetic analysis

- 16S rRNA gene phylogeny using Maximum Likelihood

Environmental bacterial 16S rRNA gene community analysis

- Comparison of bacterial community composition

Soil chemistry analysis

- Cation exchange capacity
- Total carbon and nitrogen
- pH

Whole genome sequencing and comparative genomics

- Gene presence or absence
- Genomic comparisons
- Characterisation of UVC resistant and UV sensitive isolates

Figure 5.1: Breakdown of schematic overview of thesis. Aim 3 will be covered in Chapter 5 and Aim 4 will be covered in Chapter 6.

5.2 Methods and materials

5.2.1 Genomic DNA extraction

Genomic DNA was extracted as previously described (Minas et al. 2011). DNA was extracted at early stationary phase from 50 mL of NB. The broth was spun at room temperature in 50 mL centrifuge tubes at 6,300 x *g* for 10 min using the 5810 Centrifuge (Eppendorf, Hamburg, Germany). The supernatant was discarded, and the remaining cell pellet was washed with 25 mL of pH 7.2 phosphate buffer (100 mM NaH_2PO_4) at 6,300 x *g* for 10 min at room temperature. The pellet was added to a tube containing 1 g of 0.1 mm silica-zirconia beads. To each tube, 270 µL of phosphate buffer (100 mM NaH2PO4) and 270 µL of SDS lysis buffer (100 mM NaCl, 500 mM Tris pH 8.0, 10% SDS) were added and samples were homogenised using the FastPrep®-24 (MP Biomedicals, Ohio, USA) at 4.0 for one 20 sec run. Samples were centrifuged at 18,506 x *g* for 3 min at 20°C and 180 µL of cetyltrimethylammonium bromide-polyvinylpyrrolidone (CTAB) extraction buffer (100 mM Tis-HCl, 1.4 M NaCl, 20 mM EDTA, 2% CTAB, 1% polyvinylpyrrolidone and 0.4% β-mecaptoethanol) was added. Samples were vortexed for 10 sec before incubation at 60°C and 300 rpm on a Incubated Shaker SI-300R (Acorn Scientific, New Zealand) for 30 min. Samples were centrifuged at 18,506 x *g* for 1 min and then 350 µL of chloroform:isoamyl alcohol (24:1) was added. Samples were then vortexed for 10 sec and centrifuged at 18,506 x *g* for 5 min. The upper aqueous layer was removed into a new microcentrifuge tube and a further 500 µL of chloroform:isoamyl alcohol (24:1) was added. Samples were vortexed for 15 sec and again centrifuged at 18,506 x *g* for 5 min. The upper aqueous phase was removed into a new microcentrifuge tube and 10 M ammonium acetate was added to samples to achieve a final concentration of 2.5 M. Samples were vortexed for 10 sec and centrifuged at 18,506 x *g* for 5 min. The aqueous layer was removed to a new tube and 0.54 volume of isopropanol was added and mixed by inversion. Samples were left for 24 hours at -20°C and then centrifuged at 4°C for 20 min at 18,506 x *g*. The supernatant was discarded, and the DNA pellet was washed with 1 mL 70% ethanol and centrifuged for 10 min at 13,200 rpm. Ethanol was removed via 4 min of speed vacuum using the Concentrator plus (Eppendorf, Hamburg, Germany) and DNA was re-suspended in 20 µL of sterile DNase free water (ThermoFisher Scientific, MA, USA). Samples were then quantified using the Qubit® 2.0 Fluorimeter (InVitrogen, Waltham, MA, USA). Genomic DNA was checked for contamination via electrophoresis on a 0.8% agarose / 1x TBE gel stained with 1 µL 0.5 µg/mL ethidium bromide at 75 V for 90 min using the PowerPac™ Basic Power Supply (Bio-Rad Laboratories., Auckland, New Zealand). Samples were stored at -20°C until required and then sent to Genewiz (Suzhou, China) for sequencing.

5.2.2 Genome sequencing and assembly

Genomic DNA was sequenced using Illumina sequencing on an Illumina HiSeq™ 2500 sequencer by Genewiz (Suzhou, China). The method used by Genewiz is as follows: sequencing library was constructed using the NEBNext® Ultra™DNA Library Prep Kit following the manufacturer's instructions (Illumina, San Diego, CA, USA). For each sample, 1 μg of genomic DNA was randomly fragmented to <500 bp by sonication using the S220 (Covaris, Massachusetts, USA). The fragments were treated with End Prep Enzyme Mix for end repairing, 5' phosphorylation and dA-tailing in one reaction, followed by a T-A ligation to add adaptors to both ends. Size selection of adaptor-ligated DNA was then performed using AxyPrep Mag PCR Clean-up (Axygen Scientific, Bath, UK). Following this, fragments of ~410 bp, with the approximate insert size of 350 bp, were recovered. Each sample was then amplified by PCR for 8 cycles using P5 and P7 primers, with both primers carrying sequences that can anneal to the flowcell to perform bridge PCR and P7 primer carrying a six-base index allowing for multiplexing. The PCR products were purified using AxyPrep Mag PCR Clean-up (Axygen Scientific, Bath, UK), validated using an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA), and quantified by Qubit2.0 Fluorimeter (InVitrogen, Waltham, MA, USA). Genome libraries with different indexes were pooled and loaded on an Illumina HiSeq™ 2500 instrument according to the manufacturer's instructions (Illumina, San Diego, CA, USA). Sequencing was carried out using a 2x150 paired-end configuration; image analysis and base calling were conducted by the HiSeq Control Software (HCS) + RTA 2.7 (Illumina, San Diego, CA, USA) on a HiSeq instrument.

Raw reads were binned according to their index sequences through the Illumina conversion software bcl2fastq 2.17 (Illumina, San Diego, CA, USA). Raw reads were discarded if they were: i) pair-ended reads with an adapter; ii) pair-end reads when the content of N bases was more than 10% in either read; iii) pair-end reads where the ratio of bases of low quality (Q<20) was more than 0.5 in either read. Further quality control was performed on the raw sequence data using fastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The reads were assembled into contigs using SPAdes v 3.12.0 (Bankevich et al. 2012) and Unicycler (Wick et al. 2017b). *De novo* assembly graphs were visualised with Bandage (Wick et al. 2015).

5.2.3 Creation of genome dataset

The genome sequences of 18 *Arthrobacter* and two *Pseudarthrobacter* were downloaded from the NCBI RefSeq and Whole Genome Shotgun (WGS) project databases for comparison with the four newly sequenced Namib Desert strains (Table 5.2). Representative genomes were used where the species type strain was not available. *Kocuria rosea* ATCC 186 and *Micrococcus luteus* NCTC 2665 were used as comparative reference organisms.

Table 5.2: Whole annotated bacterial genomes used in this study.

Organism name	Abbreviated name	Isolation source	Reference	RefSeq number
B2	-	Site T10, Namib Desert	This study	SZWI00000000.1
E2	-	Site T4, Namib Desert	This study	VAHM00000000.1
B4	-	Site T4, Namib Desert	This study	VAHN00000000.1
E5	-	Site T8, Namib Desert	This study	VAHO00000000.1
A. agilis 4041	Aag4041	Soil from Western France	Crovadore et al. (2018)	NZ_NFSC00000000.1
A. alpinus R3.8	AalR3.8	Rothera Point, Adelaide Island, Antactica	See-Too et al. (2017)	NZ_CP012677.1
A. castelli DSM 16402 ^{T}	Aca164	Biofilm on a mural paining, Saint-Catherine chapel, Herberstein, Austria	Heyrman et al. (2005)	NZ_AUMN01000001.1
A. crystallopoietes DSM 20117 ^{T}	Acr201	Soil of unknown origin	Ensign and Rittenberg (1963)	NZ_CP018863.1
A. cupressi CGMCC 1.10783 ^{T}	AcuCG1.1	Cypress tree in Mianyang, Sichuan Province, China	Zhang et al. (2012a)	NZ_FNEI01000001.1
A. enclensis NIO-1008 ^{T}	Aen1008	Chorao Island, India	Dastager et al. (2014)	NZ_LNQM00000000.1
A. globiformis NBRC 12137	Agl12137	Unknown	Unpublished	NZ_BAEG00000000.1
A. humicola LBUM149	Ahu149	Unknown	Unpublished	NZ_PYUI00000000.1
A. koreensis 5J12A	Ako12A	*Nerium oleander* rhizosphere soil, South Korea	Manzanera et al. (2015)	NZ_CECE01000006.1
A. livingstonensis LI2 ^{T}	AliLI2	Soil from Livingston Island, Antarctica	Ganzert et al. (2011)	NZ_QJVD00000000.1
A. luteolus NBRC 107841	AluNB107	Unknown	Unpublished	NZ_BCQM01000001.1
A. nitrophenolicus SJCon ^{T}	AniSJC	Soil from Punjab, India	Arora and Jain (2013)	NZ_AOFD00000000.1
A. oryzae DSM 25586 ^{T}	Aor25586	Soil from Saitama prefecture, Japan	Kageyama et al. (2008)	NZ_RBIR01000001.1
A. pityocampae Tp2 ^{T}	ApiTp2	*Thaumetopoea pityocampa* larvae, Middle Black Sea region, Turkey	İnce et al. (2014)	NZ_PRKW00000000.1
A. psychrolactophilus B7 ^{T}	ApsB7	Soil from Pennsylvania, USA	Loveland-Curtze et al. (1999)	NZ_QJVC00000000.1
A. rhombi B Ar 00.02	ArhBA02	Unknown	Unpublished	NZ_FUHW00000000
A. subterraneus NP_1H	AsuNP1H	Sediment from North Pond, Atlantic Ocean	Russell et al. (2016)	NZ_FNDT01000001
A. woluwensis DSM 10495 ^{T}	Awo104	Blood from HIV patient, Woluwe, Belgium	Funke et al. (1996)	NZ_FNSN00000000.1

P. chlorophenolicus A6 ^{T}	PchA6	Soil from Fort Collins, Colorado, USA	Westerberg et al. (2000)	NC_011886.1
P. phenanthrenivorans Sphe3 ^{T}	PphSphe3	Soil from Perivleptos, Greece	Kallimanis et al. (2009)	NC_015145.1
Kocuria rosea ATCC 186 ^{T}	Koc186	Soil of unknown origin	Stackebrandt et al. (1995)	NZ_QFBJ00000000.1
Micrococcus luteus NCTC 2665 ^{T}	Mle2665	Human nasal sample	Fleming (1922)	NC_012803.1

^{T} indicates that the whole genome is from the type strain of this species.

5.2.4 Functional annotation

To annotate the four Namib Desert strains, two methods were used: the RAST v2.0 pipeline (Aziz et al. 2008) and Prokka v1.12 (Seemann 2014). RAST annotation was achieved through uploading the genomes to the RAST server (http://rast.nmpdr.org/). Prokka annotation was integrated into an in-house script (Appendix 2), which generated a functional annotation for each draft genome using the tools described in Table 5.3.

Table 5.3: Tools used during Prokka annotation.

Tool (reference)	Features predicted
Prodigal (Hyatt et al. 2010)	Coding sequence
Barrnap (Seemann 2014)	Ribosomal RNA genes (rRNA)
Aragorn (Laslett and Canback 2004)	Transfer RNA genes (tRNA)
MinCED (Skennerton 2006)	CRISPR identification

General features such as GC content, number of contigs, genome length and N50 and L50 were determined via fastQC and RAST. Prokka associated tools Barrnap, Aragorn and MINCED were applied for rRNA, tRNA and CRISPR identification as described in Table 5.3 above. The output file from Prokka v1.12 and RAST annotation can be seen below in Figure 5.2.

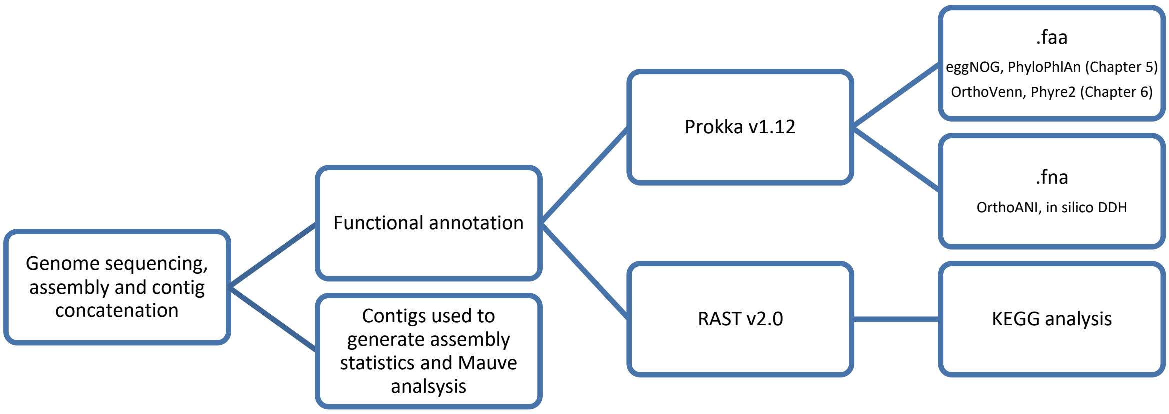

Genome sequencing, assembly and contig concatenation
Functional annotation
Contigs used to generate assembly statistics and Mauve analysis
Prokka v1.12
RAST v2.0
.faa
eggNOG, PhyloPhlAn (Chapter 5)
OrthoVenn, Phyre2 (Chapter 6)
.fna
OrthoANI, in silico DDH
KEGG analysis

Figure 5.2: Workflow of genome sequencing, assembly, contig concatenation and functional annotation in this thesis. The programmes each file was used for is indicated at the end of the process.

5.2.5 Metabolic potential of *Arthrobacter* and *Pseudarthrobacter* isolates

Potential metabolic characteristics of the newly sequenced Namib Desert isolates were compared with previously sequenced *Arthrobacter* and *Pseudarthrobacter* genomes to identify similarities and differences amongst the genomes. The Kyoto Encyclopaedia of Genes and Genomes (KEGG) (Kanehisa et al. 2016) database on RAST was used to identify general features of the isolates, such as functional tricarboxylic acid cycle, carbon fixation pathways, sugar utilisation and nitrogen metabolism. These features were used as an interactor of assembly and annotation quality, as well as to further characterise the isolates.

Biochemical assays were also conducted to confirm the KEGG results. Sugar utilisation tests (xylose, lactose, glucose, maltose, sucrose and fructose), carbon utilisation (citrate) and nitrate reduction tests were set up using standard procedures (Tortora et al. 2009). Biochemical tests were incubated at 25°C for 72 hours.

5.2.6 Assigning predicted proteins to Clusters of Orthologous Groups

The Clusters of Orthologous Groups (COG) database was established to assign all known proteins of completed microbial genomes to Orthologous Groups (OG) (Tatusov et al. 2000); this helps identify what is common and different between the taxa analysed. COGs are generated by comparing the protein sequence of complete genomes. EggNOG v4.5 (http://eggnogdb.embl.de) is a publicly available database of biological information hosted by the European Molecular Biology Laboratory (EMBL). EggNOG is able to identify and assign protein sequences to established COGs (Huerta-Cepas et al. 2016). Functional classifications of COGs were achieved by uploading the Prokka .faa file to eggNOG for OG identification under the Actinobacteria taxonomic scope. The annotation file was downloaded and viewed using a Microsoft Excel v1902 spreadsheet and used to create a bar chart of normalised COG distribution.

The COG distribution was also visualised using a heatmap. *Arthrobacter* COG distribution was counted and normalised using the heatmap.2 function in the R gplots package v. 3.5.2 (Warnes et al. 2015) with R software environment (R Core Team 2017) and visualised in RStudio (v 1.1.463).

5.2.7 Whole genome analysis

As previously mentioned, there are issues with relying on the *Arthrobacter* 16S rRNA gene for complete phylogenetic analysis. To help resolve these issues whole genome analysis was conducted using three methods; Mauve genome alignment (Darling et al. 2004), orthologous average nucleotide identity (OrthoANI) (Yoon et al. 2017) and the PhyloPhlAn software (Segata et al. 2013).

5.2.7.1 Mauve

Mauve is a multiple genome alignment tool for research into comparative genomics (Darling et al. 2004). Mauve can identify large-scale evolutionary events such as gene rearrangement and inversion. Genomes were aligned using the Mauve Contig Mover (MCM) algorithm (Rissman et al. 2009) in the Mauve plugin v1.1.1 for Geneious™ v2019.0 with the NCBI genome as the reference and the Namib isolate as the query sequence. Locally collinear blocks (LCBs) were not manipulated and were left as the assigned default for each alignment.

5.2.7.2 OrthoANI

OrthoANI (https://www.ezbiocloud.net/tools/ani) is an online software for whole genome analysis based on the ANI algorithm (Lee et al. 2016; Yoon et al. 2017). OrthoANI can replace wet-lab DNA-DNA hybridisation (DDH), as long as the DNA sequences are known (Lee et al. 2016). Under DDH guidelines, a new species is indicated by <70% similarity. The new species cut-off value for OrthoANI is <95% (Lee et al. 2016; Yoon et al. 2017). OrthoANI values were obtained by uploading the FASTA sequences of each isolate and reference genome to the OrthoANI website. OrthoANI values were then recorded and visualised using the heatmap.2 function in the gplots library of R (v. 3.5.2) and visualised in Rstudio (v 1.1.463).

5.2.7.3 *In silico* DNA-DNA Hybridisation

Genome-to-Genome Distance Calculator 2.1 (GGDC) (http://ggdc.dsmz.de/ggdc.php) is an online software tool for *in silico* DNA-DNA hybridisation (DDH). Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ) is a recognised culture repository and developed the GGDC as a recommended method. Using a computational programme removes the issues with DDH wet-lab procedures. *In silico* DDH values were obtained by uploading the FASTA sequences of each isolate and closest reference genome to the GGDC website. DDH values were then downloaded as Excel spreadsheets and the DDH prediction from Formula 2 was recorded, as recommended (DSMZ 2019).

5.2.7.4 PhyloPhlAn

PhyloPhlAn is a computational pipeline that creates highly accurate and resolved phylogenetic trees based on whole-genome sequence information (Segata et al. 2013). To generate a phylogenetic tree through this programme, 400 conserved proteins are identified and used for phylogenetic signalling (Segata et al. 2013). This results in a high topological accuracy phylogenetic tree for taxonomic curation and estimation (Segata et al. 2013).

In addition to the reference organisms in Table 5.2, the genome sequences of an additional 48 uncharacterised *Arthrobacter*, two *Glutamicibacter*, two *Paenarthrobacter* and one *Paeniglutamicibacter* were also downloaded from the NCBI website for use in the PhyloPhlAn programme. Isolates with greater than 97% similarity were selected for PhyloPhlAn analysis. *M. luteus* NCTC 2665 and *K. rosea* ATCC 186 were used as outgroup organisms. The concatenated aligned amino acid positions were reconstructed into a tree using amino acid evolution model JTT + CAT in RAxML v8.2.9, which was then edited in Geneious version 2019.0 and Inkscape v0.92.4 to produce the final tree.

The combination of OrthoANI, Mauve alignment, DNA-DNA hybridisation and PhyloPhlAn analysis helped to comprehensively categorise the Namib Desert isolates and indicated closest relatives for each isolate for further genomic analysis.

5.3 Results

Two *Arthrobacter*; B2 and E2, and two *Pseudarthrobacter* isolates; B4 and E5, were selected for genome sequencing based on the results from Chapter 3. They were isolated from the Namib Desert at sites T4 (B4 and E2), T8 (E5) and T10 (B2), as described in Section 3.2 (refer to Table 3.1 for sampling location details).

5.3.1 Next generation sequencing and assembly outputs

High throughput sequencing was conducted on the four Namib isolates. The coverage of each genome was normalised to approximately 135x (Table 5.4) to improve assembly quality. Good quality reads and low error rates were obtained for the Namib isolates, as shown in Appendix 3.

Table 5.4: Coverage for each genome sequenced.

Isolate code	Number of raw reads	Average nucleotide coverage
B2	51,352,518	137x
E2	56,223,629	136x
B4	34,774,522	134x
E5	35,472,273	142x

The assembled genomes produced reasonable numbers of contigs, with B2 having the least number of contigs at 22 and E2 having the largest number of contigs at 81 (Table 5.5). This suggested that the genomes were not complete since they were still in a number of fragments. K-mer values checked for assembly were 27, 47, 63, 77, 89, 99, 107, 115, 121 and 127 for all genomes, with an optimum Kmer used for assembly (Table 5.5).

Table 5.5: Assembly statistics for Namib Desert isolates.

Assembly statistics	B2	E2	B4	E5*
Kmer value used for assembly	121	127	115	89
Number of contigs	23	81	69	77
Number of contigs >500 bp	22	52	62	42
N50	290,490	235,942	200,647	227,930
L50	3	6	7	7
Longest contig (bp)	795,432	614,230	333,962	425,160

*Following genome assembly, it was found that sample E5 contained a single contig that was separated from the genome. The single large circular contig (size 227,930 bp) suggests that it is a plasmid. This contig was included in the total assembly statistics.

5.3.1.1 Genomic features

The size of the four strains ranged from 3.26 Mb (B2) to 4.03 Mb (E2) (Table 5.6). Prokka annotation found a single copy of the 5S rRNA, 16S rRNA and 23S rRNA genes in the assemblies of isolates B2, B4 and E5. Prokka annotation found two copies of the 5S rRNA, and one copy of the 16S rRNA and 23S rRNA genes in isolate E2 (Table 5.6). The number of tRNA genes for the Namib isolates ranged from 48 (B2) to 55 (B4). No CRISPR repeats were found in any of the Namib isolates. Other general features are listed in Table 5.6. The newly sequenced *Pseudarthrobacter* strains have a similar number of coding sequences, with 79.8% of these allocated to a COG (Table 5.6).

Table 5.6: Draft genome features for the sequenced Arthrobacter strains B2 and E2, and Pseudarthrobacter strains B4, E5 and E5 plasmid as annotated by the RAST database.

Features	**B2**	**E2**	**B4**	**E5**	**E5 plasmid**
Genome size (bp)	3,262,671	4,032,940	3,819,717	3,831,203	227,930
GC content (%)	67.5%	65.3%	64.9%	64.9%	60.2%
Number of contigs	23	81	69	76	1
Total genes	3,116	4,029	3,793	3,825	287
Protein encoding genes (%)	3,051 (97.91%)	3,929 (97.52%)	3,728 (98.28%)	3,741 (97.80%)	287 (100%)
Hypothetical proteins (% of CDS)	983 (32.22%)	1,261 (32.09%)	1,282 (34.39%)	1,289 (34.46%)	199 (69.34%)
Genes assigned to COGs (% of CDS) *****	2,492 (81.68%)	3,071 (78.16%)	2,897 (77.71%)	2,888 (77.20%)	145 (50.52%)
Total RNA genes	51	54	58	57	0
rRNA genes (5S, 16S, 23S rRNA)	3 (1, 1, 1)	4 (2, 1, 1)	3 (1, 1, 1)	3 (1, 1, 1)	0
tRNA genes **	48	50	55	54	0
CRISPR repeats	0	0	0	0	0

***** This includes proteins assigned to the S (Function unknown) COG classification.

** Determined by Aragorn (Laslett and Canback 2004).

5.3.2 Gene prediction and annotation

Both the RAST and Prokka pipelines were used for gene prediction and annotation. RAST is a web-based programme that is simple to use, but takes between 12-24 hours to complete an annotation (Aziz et al. 2008). On the other hand, Prokka is a command line-based software, making it more challenging to utilise, but provides a complete annotation in approximately 10 minutes with multiple file types for analysis (Seemann 2014).

As seen in Table 5.7, both Prokka and RAST were able to call predicted proteins, however, some proteins were called differently. Table 5.7 shows a collection of example genes not specific to DNA repair. The fact that each programme may call the hits differently justified the use of both RAST and Prokka in conjunction for continued analysis to ensure as comprehensive annotations as possible.

Table 5.7: Comparison of annotation descriptions between RAST and Prokka for selected DNA repair genes in isolate B2.

Prokka locus tag	RAST description	Prokka description	BLASTX result		
			Top hit	E value	Accession number
KFNIHMNI_00323	ATP-dependent DNA helicase RecQ	DEAD-box ATP-dependent RNA helicase CshA	ATP-dependent DNA helicase RecQ	0.0	WP_104117520.1
KFNIHMNI_00269	DNA recombination protein RmuC	Hypothetical protein	DNA recombination protein RmuC	0.0	WP_104117174.1
KFNIHMNI_02797	Endonuclease III	Ultraviolet N-glycosylase/AP lyase	Endonuclease III	0.0	WP_104116483.1
KFNIHMNI_02794	DNA polymerase I	Hypothetical protein	Bifunctional 3'-5' exonuclease/DNA polymerase	0.0	WP_104116367.1
KFNIHMNI_01772 KFNIHMNI_02578 KFNIHMNI_02729	Excinuclease ABC subunit A paralog of unknown function	UvrABC system protein A	Excinuclease ABC subunit UvrA	0.0	WP_104117321.1 WP_104118747.1 WP_104116498.1
KFNIHMNI_02012	ADP-ribose pyrophosphatase	Hypothetical protein	NUDIX hydrolase	2e-162	WP_104117554.1
KFNIHMNI_00994	Diadenosine hexaphosphate (Ap6A) hydrolase; Bis(5'-nucleosyl)-tetraphosphatase (Ap4A) (Asymmetrical)	Putative mutator protein MutT4	DNA mismatch repair protein MutT	9e-124	AUZ89154.1

In addition, both RAST and tRNAcane-SE (Lowe and Chan 2016) were only able to detect 46 (B2) and 50 (B4, E2 and E5) tRNAs, while Prokka was able to detect an additional 2 – 8 tRNAs (Table 5.7).

5.3.3 Metabolic potential of Namib Desert isolates

The KEGG database on RAST was used to assess general features such as the presence of a tricarboxylic acid cycle (TCA) to determine the quality of the assembly and annotation of the Namib isolates. Important soil environmental functions such as carbon and nitrogen cycling, sugar utilisation and sulphur metabolism were also investigated to provide an understanding of possible functional roles of these isolates in their natural environment.

RAST employs the Enzyme Commission (EC) number system for numbering and identifying enzymes associated with specific metabolic processes. If an enzyme is detected, the EC number is highlighted. However, the identification of an enzyme does not indicate that the pathway is active. Similarly, the absence of enzyme identification does not mean the enzyme is definitely absent. To confirm some of the predicted metabolic functions of the Namib isolates, biochemical testing was conducted. The results of the biochemical tests can be seen below in Table 5.8.

Table 5.8: Biochemical testing results for the Namib isolates. Positive (+) and negative (-) results are indicated. Y = yellow, G= green, MR = Methyl Red reaction, K/K = alkaline slope and alkaline butt.

Biochemical test	**B2**	**E2**	**B4**	**E5**
IMViC				
Indole	-	-	-	-
MRVP	MR	MR	MR	MR
Citrate	-	-	-	-
Sugar utilisation				
Xylose	+	+	+ (weak)	+ (weak)
Lactose	-	-	-	-
Glucose	+	+	-	-
Maltose	+	+	+	+
Sucrose	-	-	-	-
ONPG	+	+	-	+
TSI	K/K	K/K	K/K	K/K
Nitrate test	+	+	+	+
Hugh-Leifson OF	Y/G	Y/G	G/G	G/G
Motility	-	-	-	-
Urease	-	-	-	-

As seen in Table 5.8 above the Namib isolates were metabolically similar for the indole, methyl red, Voges-Proskauer and citrate (IMViC) tests. None of the Namib isolates were able to utilise indole or citrate, and all were mixed acid fermenters, as indicated by the positive MR result. For sugar utilisation, all Namib isolates were able to utilise maltose and xylose (B4 and E5 weak utilisation). None of the Namib isolates were able to utilise lactose or sucrose; however, B2, E2 and E5 were able to utilise ONPG, indicating that B2, E2 and E5 have the β-galactosidase enzyme. Further, B2 and E2 were able to utilise glucose, but only under aerobic conditions according to the Hugh-Leifson OF test while B4 and E5 could not utilise glucose under any conditions tested. TSI tests revealed that none of the Namib isolates were able to use the triple sugar medium.

All Namib isolates were able to reduce nitrate to nitrite, were not able to utilise urease and were non-motile. These biochemical results were used to further analyse the KEGG annotations for the Namib isolates.

5.3.3.1 Primary metabolism

According to the KEGG annotation on RAST, approximately half of the predicted proteins belong to carbohydrate, amino acid and lipid metabolism for isolates E2 (48%), B4 (47%) and E5 (47%). This is in line with annotations for *A. agilis* UMCV2, *A. globiformis* NCRC 12137 and *P. phenanthrenivorans* Sphe3 as predicted by the RAST server. For B2 and the E5 plasmid, metabolism made up 42% and 28% of predicted proteins, respectively. Table 5.9 shows the number of metabolic features for each Namib isolate within the carbohydrate, amino acid or lipid metabolisms.

Table 5.9: Metabolic features of the Namib isolates as annotated by KEGG through RAST.

KEGG metabolism	B2	E2	B4	E5	E5 plasmid
Carbohydrate	213	398	346	327	3
Amino Acid	283	381	348	364	1
Lipid	66	82	77	75	0

The genes responsible for glycolysis/gluconeogenesis and the TCA cycle were found in all draft genomes, along with the enzymes required to carry out oxidative phosphorylation. This indicates that these genomes belong to aerobic heterotrophs. Biochemical testing confirmed that these isolates are obligate aerobic heterotrophs, as determined by the Hugh-Leifson OF test (Table 5.8).

5.3.3.1.1 Tricarboxylic acid cycle

The TCA cycle is a key process for aerobic organisms that utilises chemical reactions to release energy for cell use. All isolates appeared to encode the same enzymes for a complete TCA cycle, as seen in Figure 5.3 below. The E5 plasmid did not appear to have genes for any enzymes of the TCA cycle.

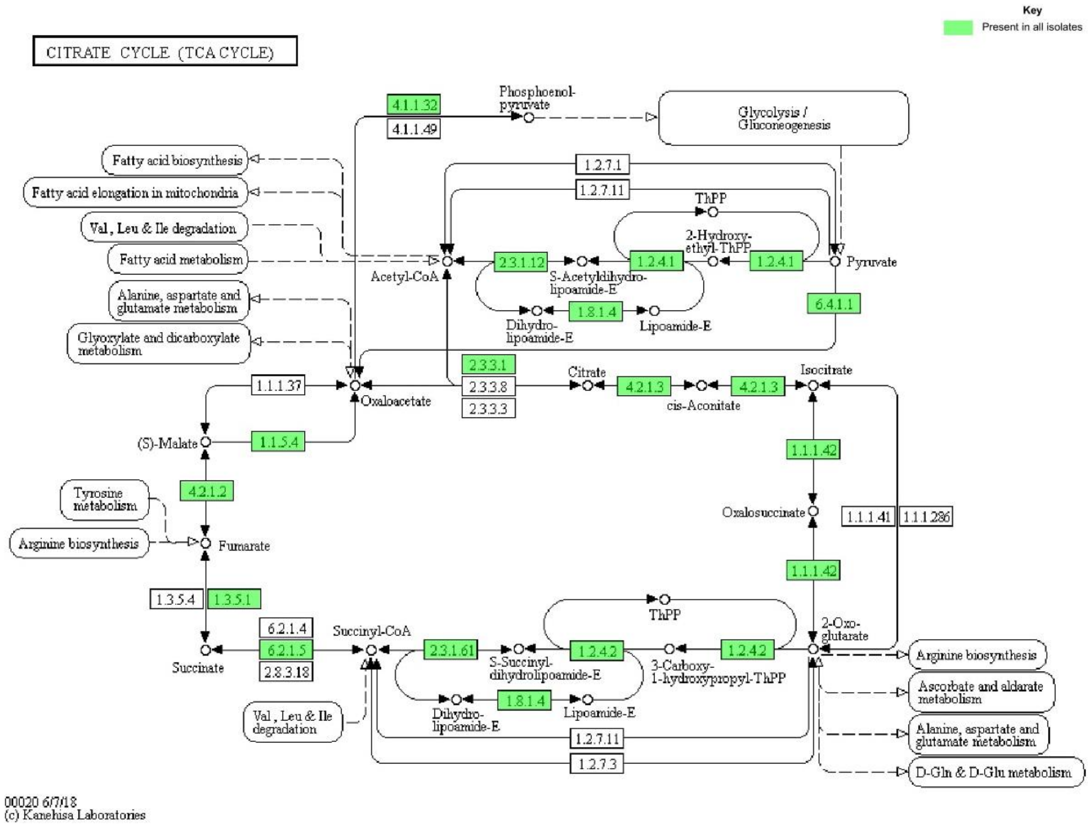
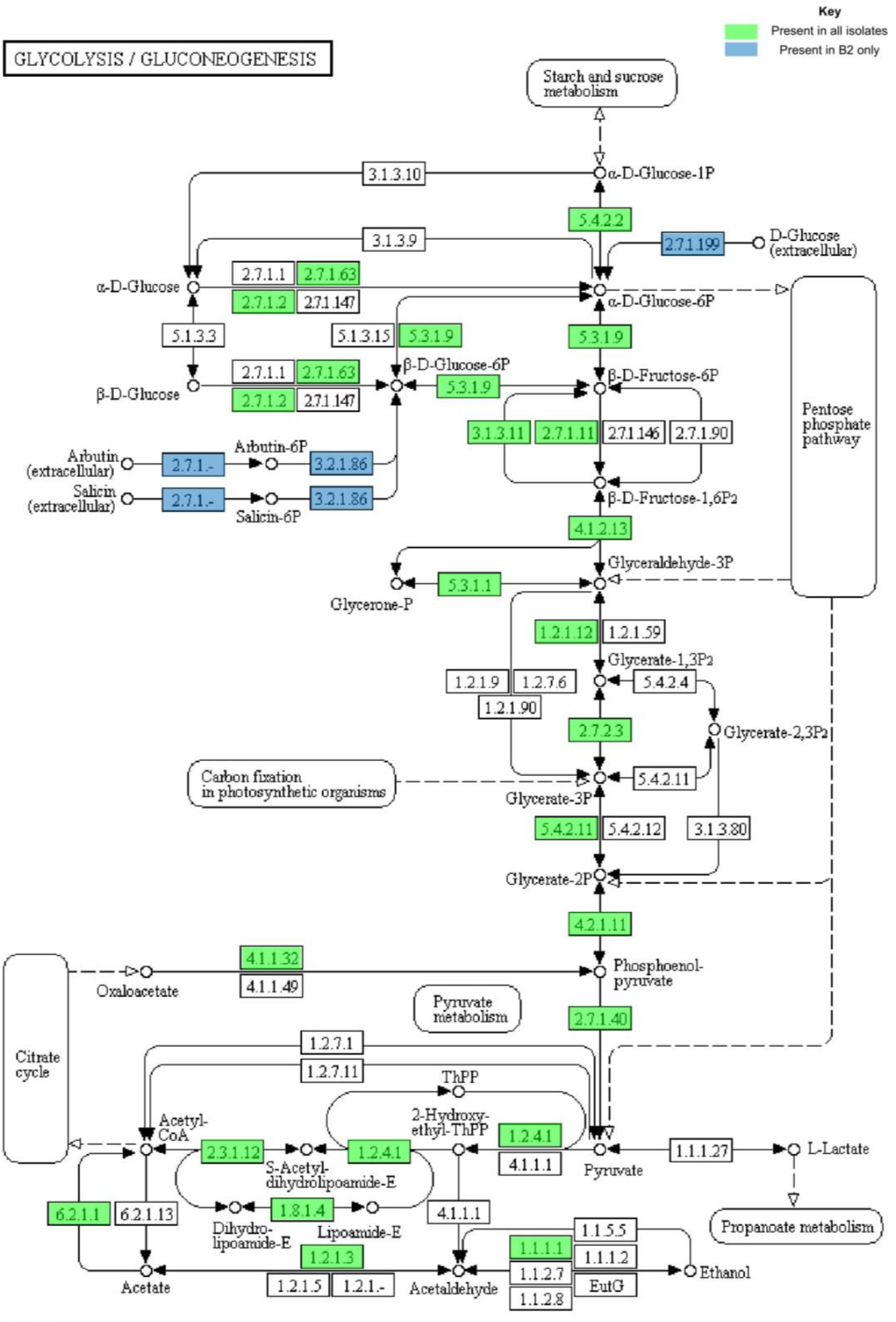


Figure 5.3: Citrate Cycle (TCA Cycle) showing the ECs present in the Namib isolates. Green colour indicates the EC number is found in all Namib isolates. White colour indicates EC number is not present in any isolate. Image from KEGG Pathway Database (2018a).

5.3.3.1.2 General carbohydrate metabolism

Genes for the metabolism of glycogen and starch were identified in these draft genomes. Genes related to the uptake of extracellular glucose were only observed in isolate B2. Isolates E2, B4 and E5 do not appear to be able to transport extracellular D-glucose or metabolise maltose or xylose. However, during biochemical testing it was found that E2 was able to utilise glucose (Table 5.8). This indicates that there are genes missing from the assembled genome of isolate E2 and further results for each draft genome should be analysed with caution. Isolates B2, E2 and B4 had the genes for enzymes required for the uptake of extracellular sucrose (EC 2.7.1.69), however, biochemical assays determined that none of the Namib isolates were able to utilise this sugar. E5 has the genes for extracellular fructose (EC 2.7.1.202) and mannitol uptake (EC 2.7.1.197), as well as genes for the metabolism of internal sucrose and glucose even though it could not metabolise either by biochemical testing.

In addition, isolate B2 has the genes required for the extracellular uptake of arbutin and salicin (EC 2.7.1.-). Salicin and arbutin are plant-derived aromatic β-glucosides that can be converted into β-D-Glucose-6P (Sonowal et al. 2013). The presence of these enzymes suggests that B2 may have adapted to an environment that contains plants or plant-derived aromatic compounds for energy utilisation. These differences are indicated below in Figure 5.4.



00010 4/12/18
(c) Kanehisa Laboratories

Figure 5.4: Glycolysis pathway showing the ECs present in the Namib isolates. Green colour indicates the EC number is found in all Namib isolates. Blue colour indicates the EC is present in isolate B2 only. White colour indicates EC number is not present in any isolate. Image from KEGG Pathway Database (2018b).

5.3.3.1.3 Carbon fixation

Carbon fixation is a metabolic process whereby CO_2 is converted into an organic carbon form. While the Namib isolates are classified as heterotrophic organisms, KEGG detected enzymes related to cyclic carbon fixation in all isolates. While the KEGG pathways for carbon fixation are incomplete in the Namib isolates, further culture-based experiments are required to confirm if these isolates are able to fix carbon. From the reductive citric acid cycle, seen on the left-hand side of the KEGG pathway, it was observed that all isolates can encode the same enzymes (Figure 5.3). Isolate B2 did not seem to have the gene for EC 2.7.9.2 for pyruvate to phosphoenolpyruvate, but instead had the gene for EC 6.4.1.1, which is able to perform the same process (KEGG Pathway Database 1961).

The E5 plasmid did not encode any enzymes any ECs for the 'carbon fixation in photosynthetic organisms' KEGG pathway. All isolates also possessed the genes for the conversion of atmospheric CO_2 to oxaloacetate (EC 4.1.1.31) as seen in Figure 5.5.

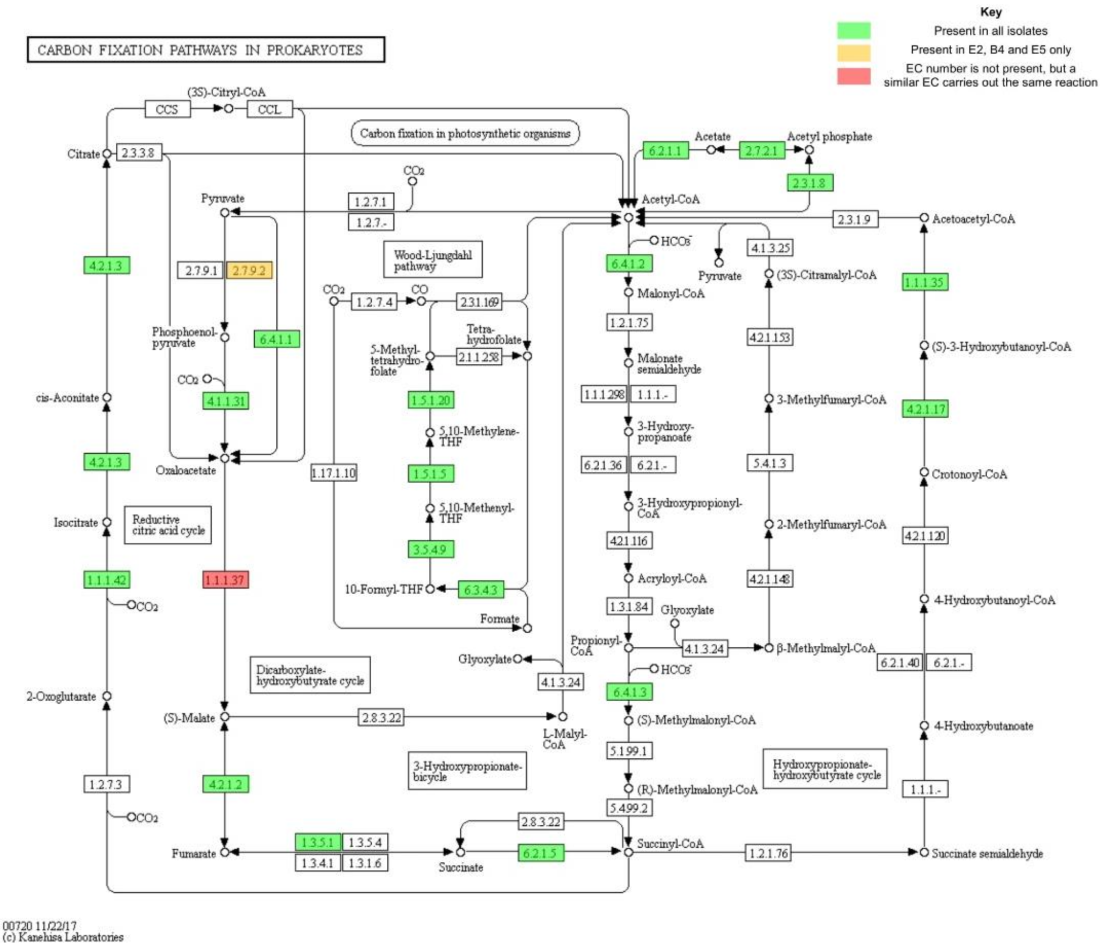


Figure 5.5: Carbon fixation pathway in prokaryotes showing the EC annotations present in the Namib isolates. Green colour indicates the EC number is found in all Namib isolates. Orange colour indicates the EC is present in isolates E2, B4 and E5 only. Red colour indicates the specific EC for the pathway is absent, but a similar EC is able to carry out the same process. White colour indicates EC number is not present in any isolate. Image from KEGG Pathway Database (2017).

Figure 5.5 above shows that the gene encoding EC number 1.1.1.37 is absent in all organisms. Instead however, all isolates possess the gene for EC 1.1.5.4, an enzyme which is able to reduce (S)-malate to oxaloacetate. The conversion of (S)-malate to oxaloacetate is important for the production of succinate and succinyl-CoA: an important enzyme for the TCA cycle.

5.3.3.1.4 Nitrogen metabolism

The nitrogen cycle is a biogeochemical cycle in which nitrogen is interconverted into several different chemical forms. During this cycle, nitrogen will circulate between atmospheric, marine and terrestrial ecosystems (Gruber and Galloway 2008). Complete nitrogen fixation is the conversion of N_2 to NH_3 with a series of intermediaries in between.

All isolates contain the gene for enzymes required to convert CO_2 to H_2CO_3 (EC 4.2.1.1). The genes required for the uptake of extracellular nitrate were present, along with genes for enzymes involved in the reduction of nitrate to nitrite and further to ammonia. This was supported by the observed biochemical results (Table 5.8) which found that all four Namib isolates were able to reduce nitrate to nitrite.

In addition, all isolates appear to be able to encode the enzymes required to convert ammonia to L-glutamine (EC 6.3.1.2). The E5 plasmid did not appear to encode any enzymes for nitrogen metabolism. Isolate B2 is the only isolate that appears have the genes to be able to convert nitroalkane to nitrate (EC 1.13.12.16). As the Namib isolates are aerobic heterotrophs, they are unlikely to be nitrogen fixers. However, as mentioned, the Namib isolates were able to reduce nitrate to nitrite (Table 5.8), indicating that they are able to conduct denitrification. The Namib isolates do not have the ECs required for complete nitrogen fixation as shown below in Figure 5.6, however, further tests would be required to confirm this.

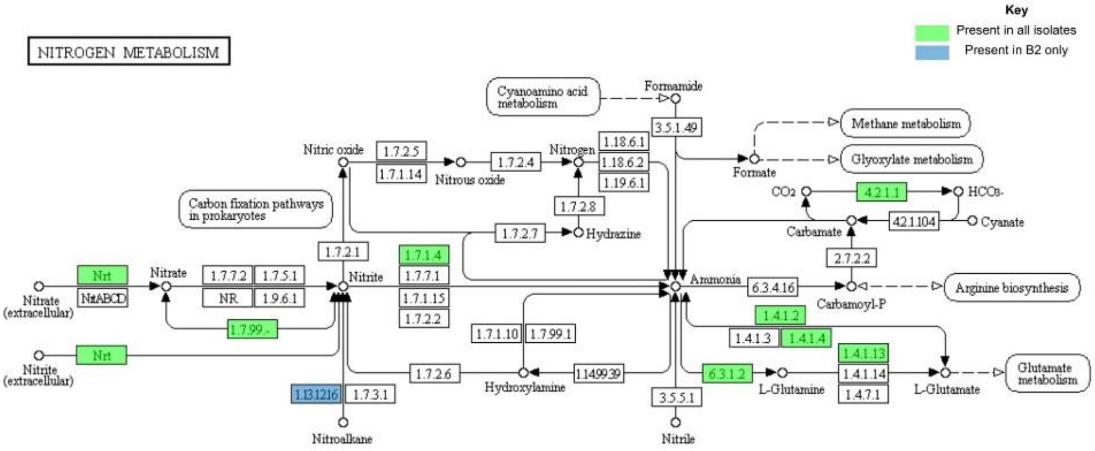


Figure 5.6: Nitrogen metabolism pathway showing the ECs present in all Namib isolates. Green colour indicates the EC number is found in all Namib isolates. Blue colour indicates the EC is present in isolate B2 only. White colour indicates EC number is not present in any isolate. Image from KEGG Pathway Database (2018c).

As seen above, all isolates appeared to be able to reduce nitrate to nitrite (EC 1.7.99.-). Biochemical testing confirmed that E2 could reduce nitrate to nitrite as per the nitrate reduction test (Table 5.8). After the addition of reagents sufanilic acid and α-naphthylamine, E2 turned red, indicating a positive nitrate reduction test. Isolates B4 and E5 did not turn red after the addition of nitrate A and B and did not turn red after the addition of zinc. This also indicates a positive nitrate reduction test. Nitrite reductase genes were also detected in all four Namib isolates by RAST.

5.3.3.1.5 Pentose phosphate pathway

The pentose phosphate pathway is a metabolic pathway that runs parallel to glycolysis. During this pathway, ribose 5-phosphate, NADPH and pentones are produced. Ribose 5-phosphate is a precursor for the synthesis of nucleotides. All isolates contained a wide quantity of EC numbers for the pentose phosphate pathway as seen in Figure 5.7. All isolates were predicted to have enzymes to convert D-ribulose-5P to D-ribose, and β-D-glucose-6P to α-D-glucose-6P. All isolates also appear to have enzymes for the conversion of D-xylulose-5P to β-D-fructose-6P, suggesting that these isolates may utilise a range of sugar sources for metabolic processes.

Isolates B2 and E2 appear to be missing one enzyme (EC 3.1.1.17) that was present in both isolates B4 and E5. This enzyme facilitates processing between D-glucono-1,5-lactone and D-gluconate. Isolate B2 was also missing genes for ECs 4.2.1.39, 2.7.1.45 and 4.1.2.14, which the other isolates had. These enzymes assist the cell in processing between D-gluconate and 2-dehydro-3-deoxy-D-gluconate, 2-dehydro-3-deoxy-D-gluconate and 2-dehydro-3-deoxy-D-gluconate-6P and 2-dehydro-3-deoxy-D-gluconate-6P and D-glyceraldehyde-3P respectively.

The E5 plasmid possessed genes for just one enzyme for the pentose phosphate pathway; EC 2.7.1.15 for D-ribose-5P and D-ribose and for 2-deoxy-D-ribose and 2-deoxy-D-ribose-5P conversion. Genes for these enzymes were also present on the E5 chromosome.

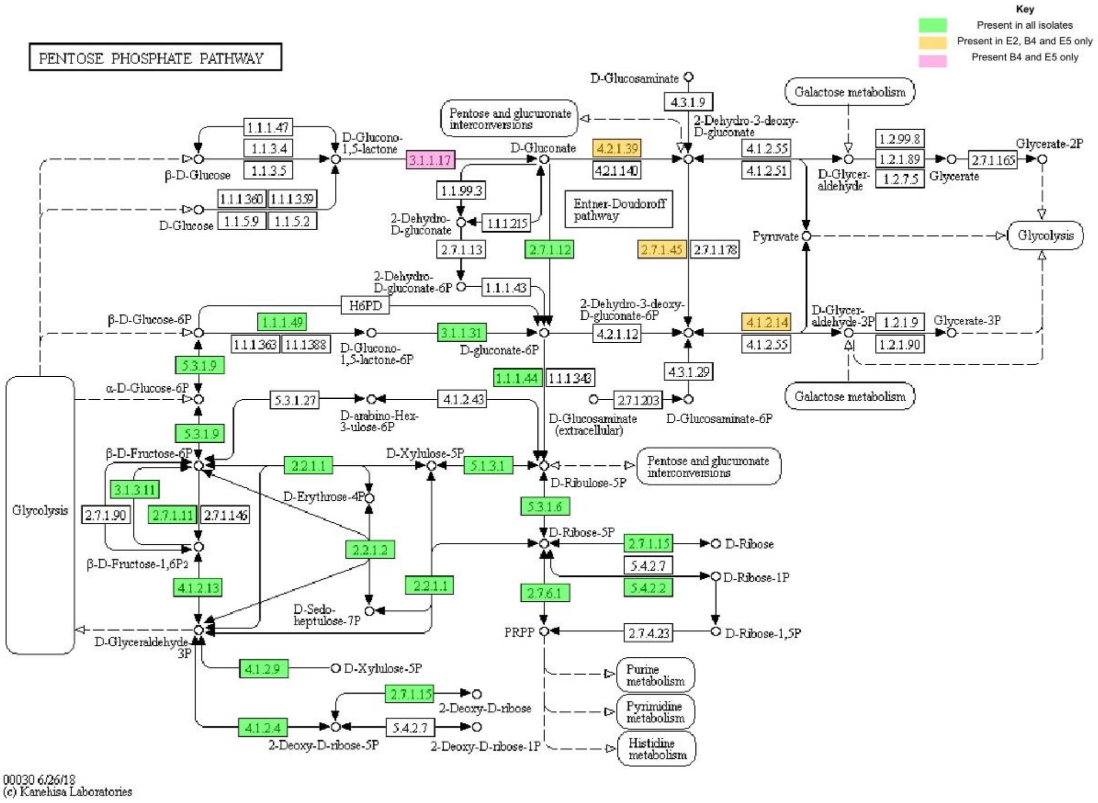


Figure 5.7: Pentose Phosphate Pathway showing the ECs present in all Namib isolates. Green colour indicates the EC number is found in all Namib isolates. Orange colour indicates the EC is present in isolates E2, B4 and E5 only. Pink colour indicates the EC is present in isolates B4 and E5 only. White colour indicates EC number is not present in any isolate. Image from KEGG Pathway Database (2018d).

5.3.3.1.6 Sulphur metabolism

All isolates were assigned the same ECs for the sulphur metabolism KEGG pathway, as seen in Figure 5.8. The isolates had the genes that can encode enzymes for the uptake of extracellular sulphate. As discussed in Chapter 3 (Table 3.2 & 3.3, Section 3.3), the amount of sulphur in the soil from the Namib sample sites was an average of 167.8 µg/g, making this activity likely.

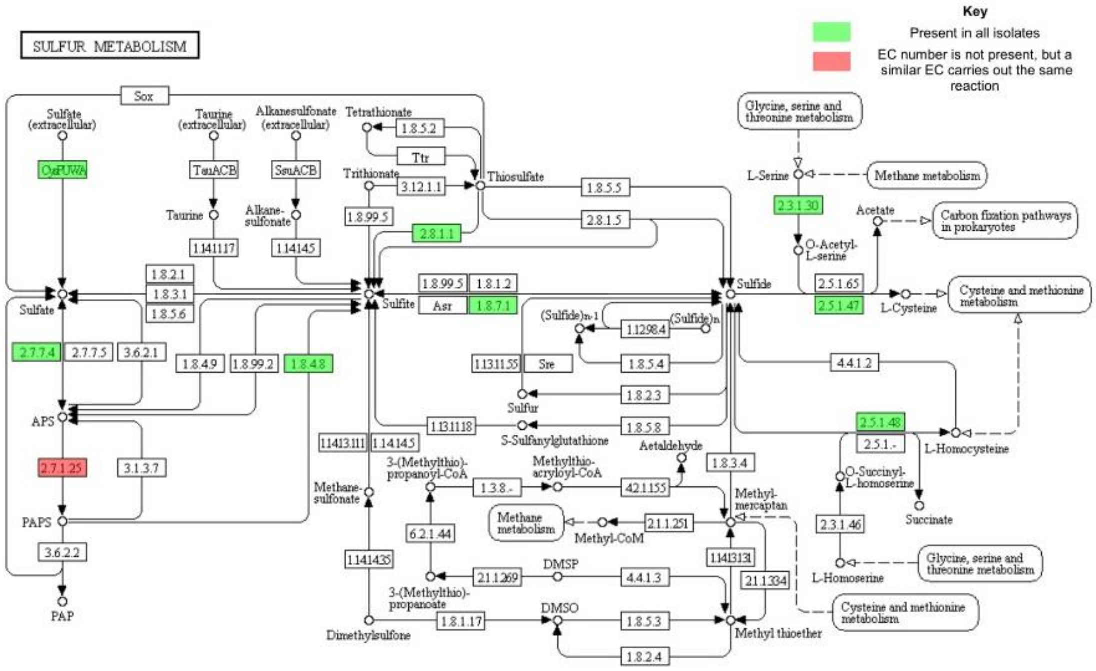


Figure 5.8: Sulphur metabolism pathway showing the ECs present in all Namib isolates. Green colour indicates the EC number is found in all Namib isolates. Red colour indicates the specific EC for the pathway is absent, but a similar EC is able to carry out the same process. White colour indicates EC number is not present in any isolate. Image from KEGG Pathway Database (2019).

Figure 5.8 above shows that the gene for enzyme EC 2.7.1.25 is absent in all organisms. EC 2.7.1.25 is responsible for the catalytic reaction of adenosine 5'-phosphosulfate (APS) to phosphoadenosine 5'-phosphosulfate. This process is important for the cycling of sulphate to sulphite for further metabolic usage. However, EC 2.7.1.25 has so far only been reported in humans (Venkatachalam et al. 1998). In bacteria, fungi and plants two enzymes are able to perform this process: EC 2.7.7.4 and EC 2.7.1.25. EC 2.7.7.4 was found in all the Namib isolates, indicating that the Namib isolates have the potential to reduce sulphate to sulphite.

5.3.3.2 Pigment production

As mentioned in Section 4.4, isolate B2 produces a red-pink colour on agar at pH 7. It was previously proposed that this pigment could be either lycopene or bacterioruberin; two carotenoids that are produced by *Arthrobacter* sp. H41 and Br18, and *A. agilis* MB8-13 respectively. Lycopene and bacterioruberin are produced along the same carotenoid pathway (Figure 5.9).

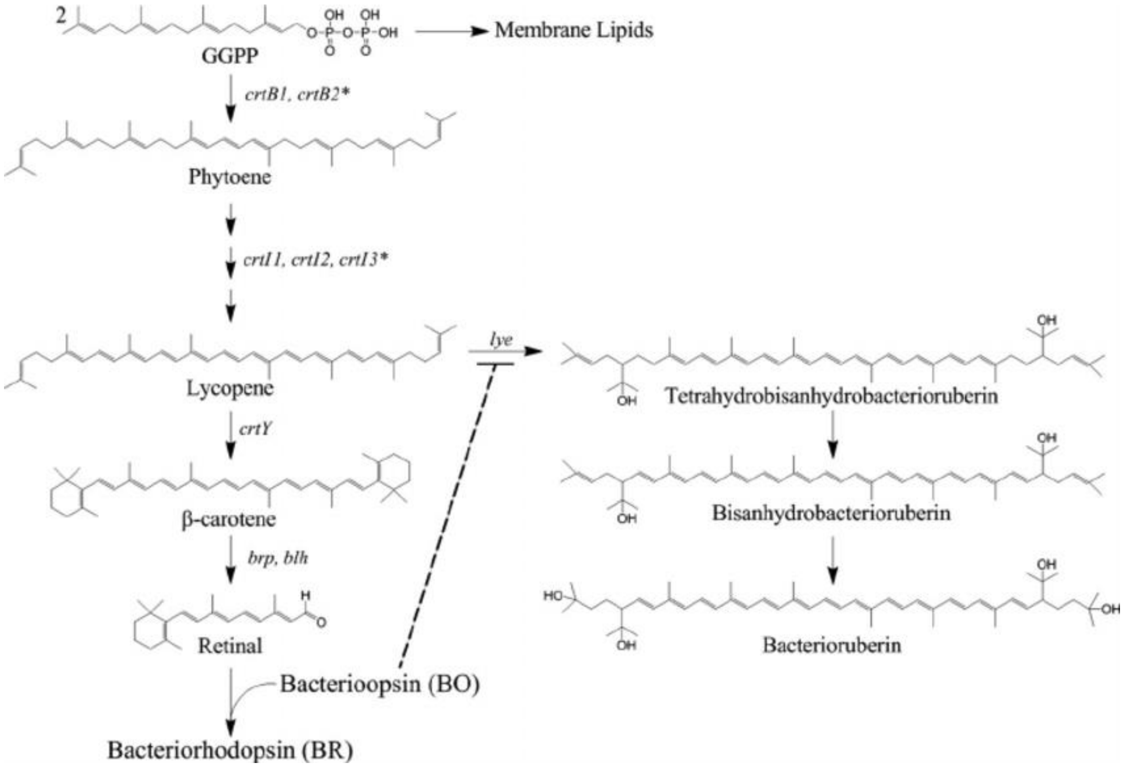


Figure 5.9: Production of bacterioruberin from lycopene The enzyme *lye* facilitates the transition from lycopene to tetrahydrobisanhydrobacterioruberin and can be inhibited by the apoprotein bacterioopsin (Dummer et al. 2011).

Analysis of RAST annotations showed that the genes for the enzymes EC 2.5.1.32 and EC 1.3.99.31 (renamed from the obsolete enzyme EC 1.14.99.-) for lycopene production were detected in isolate B2. Lycopene elongase (EC 2.5.1.150) for the conversion of lycopene to bacterioruberin was also detected in B2. The apoprotein bacterioopsin which can inhibit lycopene elongase (Dummer et al. 2011) was not detected in B2. The pigment produced by B2 is therefore most likely bacterioruberin. Isolate E5 appeared to be able to convert lycopene to bacterioruberin (EC 2.5.1.150), but the enzymes for converting the carotenoid precursor phytoene into lycopene (EC 2.5.1.32 and EC 1.3.99.31) were classified as 'inactive' in E5 by RAST. This may be due to a damaged promoter region or a mutated gene pathway.

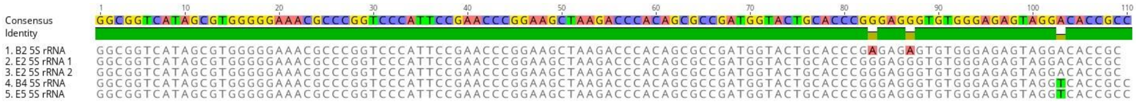
It was also mentioned in Section 4.4 that isolate B4 produced a red-pink pigment on agar at pH 7. Following UV testing and subsequent regrowth of B4, this red-pink pigmentation was lost, and the colonies began to grow white. Loss of pigmentation genes can occur during bacterial subculturing (Lowe et al. 2010) and the favourable laboratory conditions may have caused B4 to lose its pigment production requirement or ability. Analysis of the RAST KEGG pathway showed that B4 could encode one enzyme for phytoene dehydrogenase (EC 1.3.99.31) but was missing phytoene synthase (EC 2.5.1.32) for the full conversion of phytoene to lycopene. Subsequent testing will be required to determine if B4 is still resistant to UV radiation. Isolate E2 did encode enzymes relating to carotenoid pathways (EC 1.3.99.28), but the enzymes for converting phytoene into lycopene (EC 2.5.1.32 and EC 1.3.99.31) were not detected.

5.3.4 Analysis of the rRNA genes

Each genome was annotated in Prokka to determine if rRNA genes were present. The 5S rRNA gene was present as a single-copy gene in all isolates, except isolate E2, which had two identical copies. The Prokka identification of the 5S gene differs from RAST. RAST correctly identified only one 5S gene of 123 bp in length. The second 5S gene was not detected by RAST as a 5S rRNA, but rather as a repeat region. The sequence of the 5S gene for E2 from Prokka matched 100% the RAST annotation where the sequences overlapped. All isolates contained a single copy of both the 16S and 23S rRNA genes. This may be because some regions of rRNA genes are variable and can cause disruption to the *de novo* assembly (Wick et al. 2015).

5.3.4.1 5S rRNA gene

The 5S rRNA genes were located using Prokka and were 109 bp (B2 and E2) and 110 bp (B4 and E5) in length. As seen in Figure 5.9 B2 and E2 have 107 identical sites, with 98.2% similarity, while B4 and E5 have 110 identical sites and 100% similarity. All four isolates have the same sequence, except for an A at positions 83 and 87 in B2 and a T in position 103 in B4 and E5. The alignment of the 5S rRNA gene of the Namib isolates can be seen below in Figure 5.10.



```
                 1        10        20        30        40        50        60        70        80        90        100       110
Consensus        GGCGGTCATAGCGTGGGGGAAACGCCCGGTCCCATTCCGAACCCGGAAGCTAAGACCCACAGCGCCGATGGTACTGCACCCGGGAGGGTGTGGGAGAGTAGGACACCGCC
Identity
1. B2 5S rRNA    GGCGGTCATAGCGTGGGGGAAACGCCCGGTCCCATTCCGAACCCGGAAGCTAAGACCCACAGCGCCGATGGTACTGCACCCGAGAGAGTGTGGGAGAGTAGGACACCGC
2. E2 5S rRNA 1  GGCGGTCATAGCGTGGGGGAAACGCCCGGTCCCATTCCGAACCCGGAAGCTAAGACCCACAGCGCCGATGGTACTGCACCCGGGAGGGTGTGGGAGAGTAGGACACCGC
3. E2 5S rRNA 2  GGCGGTCATAGCGTGGGGGAAACGCCCGGTCCCATTCCGAACCCGGAAGCTAAGACCCACAGCGCCGATGGTACTGCACCCGGGAGGGTGTGGGAGAGTAGGACACCGC
4. B4 5S rRNA    GGCGGTCATAGCGTGGGGGAAACGCCCGGTCCCATTCCGAACCCGGAAGCTAAGACCCACAGCGCCGATGGTACTGCACCCGGGAGGGTGTGGGAGAGTAGGTCACCGCC
5. E5 5S rRNA    GGCGGTCATAGCGTGGGGGAAACGCCCGGTCCCATTCCGAACCCGGAAGCTAAGACCCACAGCGCCGATGGTACTGCACCCGGGAGGGTGTGGGAGAGTAGGTCACCGCC
```

Figure 5.10: Alignment of the 5S rRNA copies of B2 E2, B4 and E5. Created using Geneious version 2019.0.4.

5.3.4.2 16S rRNA gene

The identified Prokka 16S rRNA genes were compared with the 16S consensus obtained from Sanger sequencing in Chapter 4. The 16S gene ranged from between 1,332 (E2) and 1,348 (B2) bp. The Sanger consensus of isolates B2, B4 and E5 matched the Illumina 16S rRNA gene to 100% identity in the overlapping sequences. The Sanger sequence of isolate E2 matched the Illumina sequence 99.9% with a single base pair difference. This base pair (A in Illumina, G in Sanger) had a low call confidence in the Sanger sequence. Both Illumina and Sanger are sequence-by-synthesis methods, however, the Illumina reads had a base coverage of 136x for E2. Therefore, it is assumed that the correct base for this position is an A.

The 16S rRNA gene for each isolate, as identified by Prokka annotation, was searched using BLASTn. B2 shared a 99.7% identity with the 16S rRNA gene of *A. agilis* DSM 20550 (X80748). Isolates B4 and E5 shared 98.9% and 99% identity with the 16S rRNA gene of *P. phenanthrenivorans* Spe3 (AM176541) respectively. Isolate E2 shared 98.65% identity with the 16S rRNA gene of *A. enclensis* NIO-1008 (JF421614). The alignments of each can be seen below in Figures 5.11a, b, c and d.

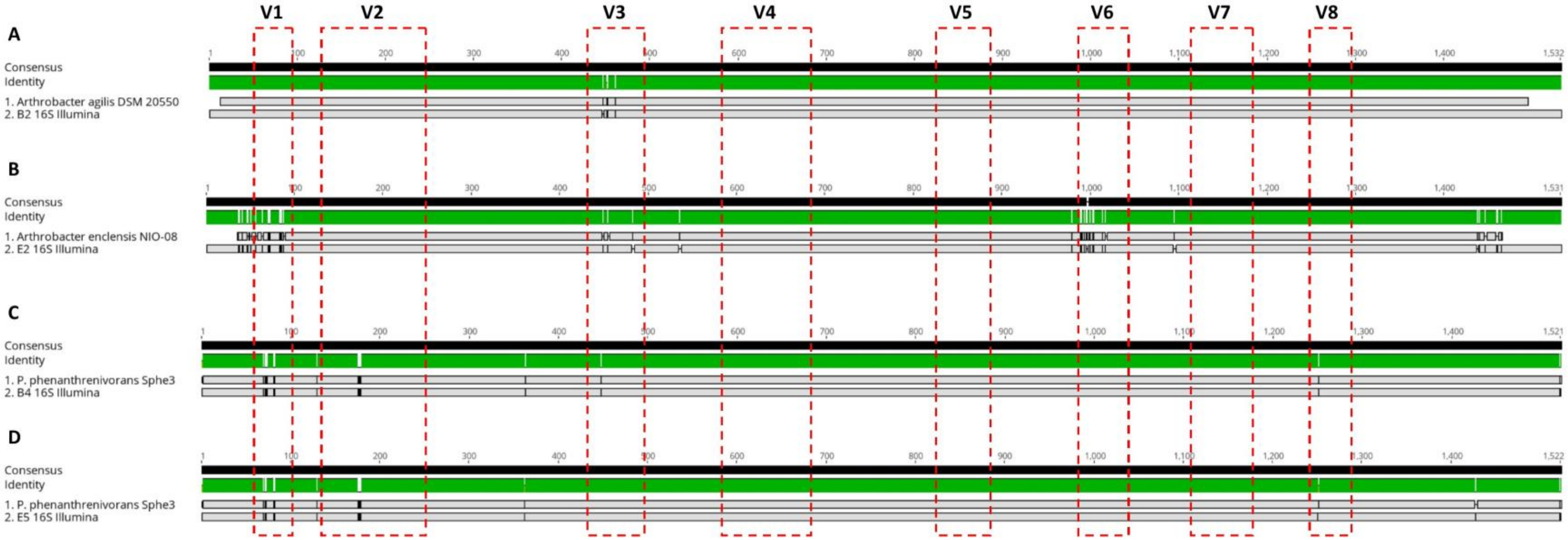


Figure 5.11a: Alignment of 16S rRNA gene sequences from A. agilis DSM 20550 (top) compared with B2 identified by Prokka via Illumina sequencing (bottom). 5.11b: Alignment of 16S rRNA gene sequences from the *A. enclensis* NIO-1008 (top) compared with E2 identified by Prokka via Illumina sequencing (bottom). 5.11c: Alignment of 16S rRNA gene sequences from *P. phenanthrenivorans* Sphe3 (top) compared with B4 identified by Prokka via Illumina sequencing (bottom). 5.11d: Alignment of 16S rRNA gene sequences from *P. phenanthrenivorans* Sphe3 (top) compared with E5 identified by Prokka via Illumina sequencing (bottom). Hypervariable ("V") regions are indicated by the red boxes. Created using Geneious version 2019.0.4 by Biomatters.

The 16S rRNA gene of B2 and *A. agilis* DSM 20550 showed variations within the V3 region only (Figure 5.11a). The 16S rRNA gene of E2 and *A. enclensis* NIO-1008 showed variations within the V1, V3 and V6 regions (Figure 5.11b). Finally, the 16S rRNA gene of B4 and E5 and *P. phenanthrenivorans* Sphe3 showed variations in the V1, V2, V3 and V8 regions (Figure 5.11c) and V1, V2 and V8 regions (Figure 5.11d), respectively.

As previously mentioned, E2 was selected as a UV-sensitive representative of the *A. sanguinis* clade. However, the E2 16S rRNA gene identified using Prokka does not share close similarity with *A. sanguinis* or *A. russicus* as was previously established (Figure 4.6, Section 4.3). Subsequently, a new 16S rRNA phylogenetic tree was created using the 16S rRNA as identified by Prokka for isolates B2, E2, B4 and E5. As previously mentioned, the *Arthrobacter* spp. clade does not have a stable internal structure, with several species occupying various positions within trees (Busse 2016), which can lead to difficulty in identifying a common relative. Figure 5.12 shows that isolates B2, B4 and E5 have the same positions as the original phylogenetic tree (Figure 4.6, Section 4.3), while E2 now occupies a different position within the tree.

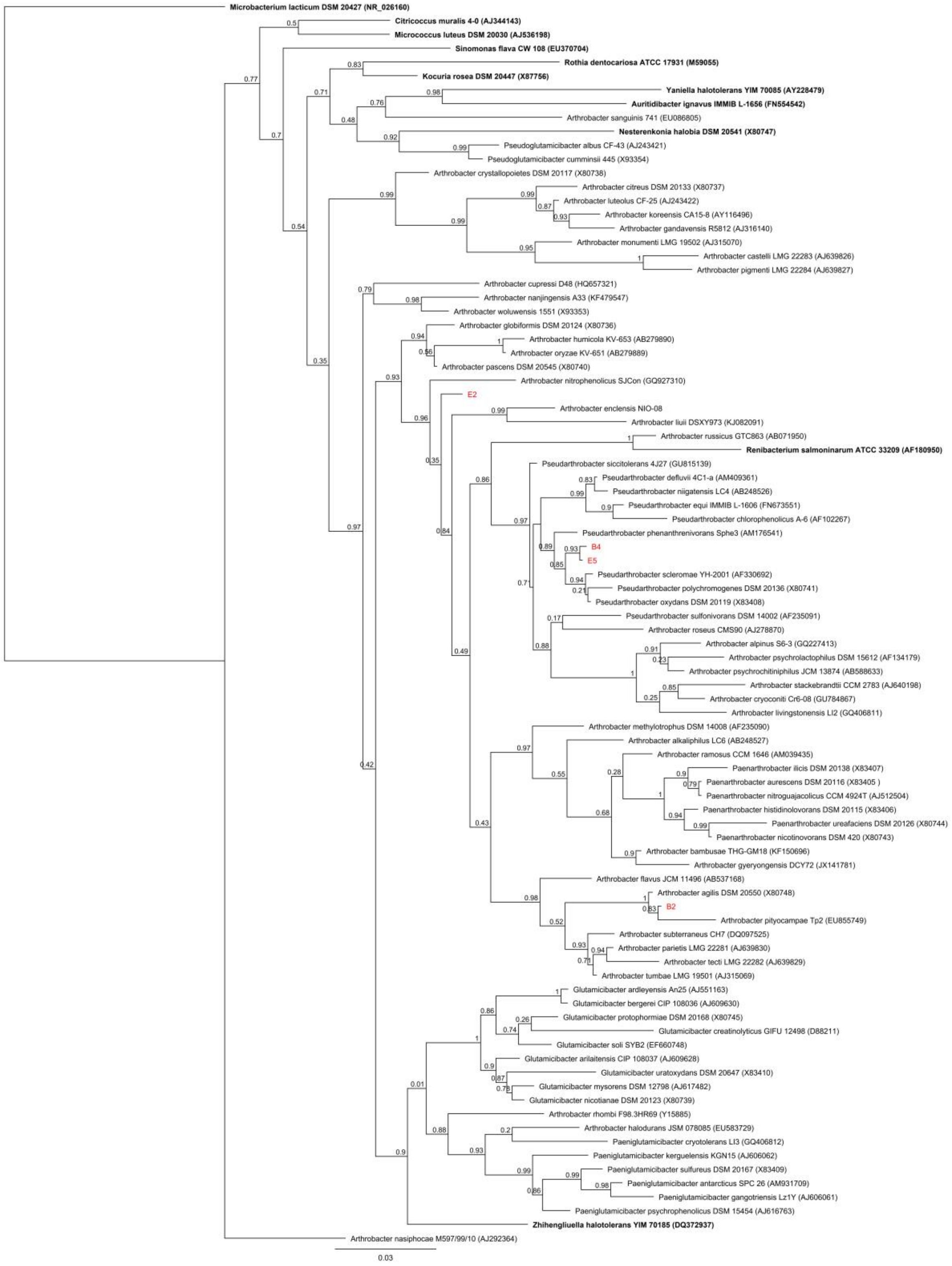

Microbacterium lacticum DSM 20427 (NR_026160)
Citricoccus muralis 4-0 (AJ344143)
Micrococcus luteus DSM 20030 (AJ536198)
Sinomonas flava CW 108 (EU370704)
Rothia dentocariosa ATCC 17931 (M59055)
Kocuria rosea DSM 20447 (X87756)
Yaniella halotolerans YIM 70085 (AY228479)
Auritidibacter ignavus IMMIB L-1656 (FN554542)
Arthrobacter sanguinis 741 (EU086805)
Nesterenkonia halobia DSM 20541 (X80747)
Pseudoglutamicibacter albus CF-43 (AJ243421)
Pseudoglutamicibacter cumminsii 445 (X93354)
Arthrobacter crystallopoietes DSM 20117 (X80738)
Arthrobacter citreus DSM 20133 (X80737)
Arthrobacter luteolus CF-25 (AJ243422)
Arthrobacter koreensis CA15-8 (AY116496)
Arthrobacter gandavensis R5812 (AJ316140)
Arthrobacter monumenti LMG 19502 (AJ315070)
Arthrobacter castelli LMG 22283 (AJ639826)
Arthrobacter pigmenti LMG 22284 (AJ639827)
Arthrobacter cupressi D48 (HQ657321)
Arthrobacter nanjingensis A33 (KF479547)
Arthrobacter woluwensis 1551 (X93353)
Arthrobacter globiformis DSM 20124 (X80736)
Arthrobacter humicola KV-653 (AB279890)
Arthrobacter oryzae KV-651 (AB279889)
Arthrobacter pascens DSM 20545 (X80740)
Arthrobacter nitrophenolicus SJCon (GQ927310)
E2
Arthrobacter enclensis NIO-08
Arthrobacter liuii DSXY973 (KJ082091)
Arthrobacter russicus GTC863 (AB071950)
Renibacterium salmoninarum ATCC 33209 (AF180950)
Pseudarthrobacter siccitolerans 4J27 (GU815139)
Pseudarthrobacter defluvii 4C1-a (AM409361)
Pseudarthrobacter niigatensis LC4 (AB248526)
Pseudarthrobacter equi IMMIB L-1606 (FN673551)
Pseudarthrobacter chlorophenolicus A-6 (AF102267)
Pseudarthrobacter phenanthrenivorans Sphe3 (AM176541)
B4
E5
Pseudarthrobacter scleromae YH-2001 (AF330692)
Pseudarthrobacter polychromogenes DSM 20136 (X80741)
Pseudarthrobacter oxydans DSM 20119 (X83408)
Pseudarthrobacter sulfonivorans DSM 14002 (AF235091)
Arthrobacter roseus CMS90 (AJ278870)
Arthrobacter alpinus S6-3 (GQ227413)
Arthrobacter psychrolactophilus DSM 15612 (AF134179)
Arthrobacter psychrochitiniphilus JCM 13874 (AB588633)
Arthrobacter stackebrandtii CCM 2783 (AJ640198)
Arthrobacter cryoconiti Cr6-08 (GU784867)
Arthrobacter livingstonensis LI2 (GQ406811)
Arthrobacter methylotrophus DSM 14008 (AF235090)
Arthrobacter alkaliphilus LC6 (AB248527)
Arthrobacter ramosus CCM 1646 (AM039435)
Paenarthrobacter ilicis DSM 20138 (X83407)
Paenarthrobacter aurescens DSM 20116 (X83405)
Paenarthrobacter nitroguajacolicus CCM 4924T (AJ512504)
Paenarthrobacter histidinolovorans DSM 20115 (X83406)
Paenarthrobacter ureafaciens DSM 20126 (X80744)
Paenarthrobacter nicotinovorans DSM 420 (X80743)
Arthrobacter bambusae THG-GM18 (KF150696)
Arthrobacter gyeryongensis DCY72 (JX141781)
Arthrobacter flavus JCM 11496 (AB537168)
Arthrobacter agilis DSM 20550 (X80748)
B2
Arthrobacter pityocampae Tp2 (EU855749)
Arthrobacter subterraneus CH7 (DQ097525)
Arthrobacter parietis LMG 22281 (AJ639830)
Arthrobacter tecti LMG 22282 (AJ639829)
Arthrobacter tumbae LMG 19501 (AJ315069)
Glutamicibacter ardleyensis An25 (AJ551163)
Glutamicibacter bergerei CIP 108036 (AJ609630)
Glutamicibacter protophormiae DSM 20168 (X80745)
Glutamicibacter creatinolyticus GIFU 12498 (D88211)
Glutamicibacter soli SYB2 (EF660748)
Glutamicibacter arilaitensis CIP 108037 (AJ609628)
Glutamicibacter uratoxydans DSM 20647 (X83410)
Glutamicibacter mysorens DSM 12798 (AJ617482)
Glutamicibacter nicotianae DSM 20123 (X80739)
Arthrobacter rhombi F98.3HR69 (Y15885)
Arthrobacter halodurans JSM 078085 (EU583729)
Paeniglutamicibacter cryotolerans LI3 (GQ406812)
Paeniglutamicibacter kerguelensis KGN15 (AJ606062)
Paeniglutamicibacter sulfureus DSM 20167 (X83409)
Paeniglutamicibacter antarcticus SPC 26 (AM931709)
Paeniglutamicibacter gangotriensis Lz1Y (AJ606061)
Paeniglutamicibacter psychrophenolicus DSM 15454 (AJ616763)
Zhihengliuella halotolerans YIM 70185 (DQ372937)
Arthrobacter nasiphocae M597/99/10 (AJ292364)
0.03

Figure 5.12: ML phylogenetic tree of *Arthrobacter* and *Pseudarthrobacter* based on 16S rRNA gene sequences (max 1530 bp). The 16S rRNA gene sequence of *Microbacterium lacticum* DSM 20427 (NR_026160) was used as an outgroup. Bar, 0.03 substitutions per nucleotide position. Type species of genera are shown in bold face. Organisms identified in this study are in red. Values at nodes represent FastTree support values from 1,000 bootstraps, where 1 is equal to 100 percent. Tree created using Geneious version 2019.0 by Biomatters.

As seen in Figure 5.12, isolate E2 is no longer within the same clade as *A. sanguinis* or *A. russicus*, but instead appears more closely to *A. enclensis* NIO-1008. BLASTn analysis of the complete E2 16S rRNA gene also indicates that E2 shares high 16S rRNA gene sequence similarity with the *P. chlorophenolicus* A6 and *P. phenanthrenivorans* Sphe3 (98.13% percentage identity, 100% query cover for both) 16S rRNA genes. This relationship is also supported by the phylogenetic analysis in Figure 5.12, although with some low support values.

5.3.4.3 23S rRNA gene

The 23S rRNA genes were located using Prokka and were between 3,119 (E2) and 3,149 (B2) bp in length. The 23S rRNA gene is not widely used for bacterial identification, but it is able to offer more thorough sequence analysis due to greater sequence variation in 23S rRNA than 16S rRNA (Pei et al. 2009). An alignment of B4 and E5 shows high similarity between the pair, with 99.6% identical sites. The 23S gene of E2 shares 93.6% pairwise identity with B2 and 96.7% identity with each of B4 and E5. A multiple alignment of the 23S rRNA gene for the Namib isolates can be seen in Figure 5.13 below.

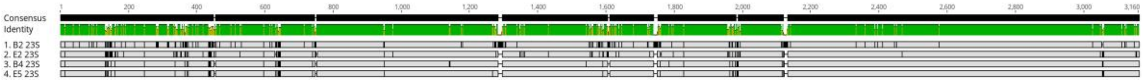


Figure 5.13: Multiple alignment of the 23S rRNA gene between B2, E2, B4 and E5. Created using Geneious version 2019.0.4 by Biomatters.

5.3.5 COG assigned protein function

The COG (clusters of orthologous groups) system was developed to assign the microbial proteins of complete genomes to OGs (orthologous groups) (Tatusov et al. 2000). Each OG is assigned a general category, of which many proteins of different functions can be assigned. These are known as COG functional categories. EggNOG uses non-supervised orthologous relationships to expand on the previous COG database (Huerta-Cepas et al. 2016). The eggNOG identified COGs for each genome was normalised to 100%. The proportion of genes assigned to a COG functional category can be seen in Figure 5.14. The COG functional category with the most variation across the genomes were categories G, E, and K, which are carbohydrate metabolism and transport, amino acid metabolism and transport, and transcription, respectively. However, as seen in Figure 5.14, these do not differ largely.

COG N, which is the functional category for cell motility, did not represent a large percentage of COG assignments in most of the reference *Arthrobacter* genomes. One N COG was present in isolates E2 and E5. Both E2 and E5 are non-motile organisms, as confirmed by the motility test during biochemical testing (Table 5.8). Isolates *A. alpinus* (23), *A. enclensis* (1), *A. crystallopoietes* (22), *A. nitrophenolicus* (1) and *P. phenanthrenivorans* (1) had N assigned COGs, however, these organisms have not previously been observed to be motile (Busse et al. 2012; Dastager et al. 2014). Other organisms with N COGs such as *A. luteolus* (23), *A. koreensis* (24), *A. oryzae* (23) and *P. chlorophenolicus* (27) have been reported as motile (Busse et al. 2012).

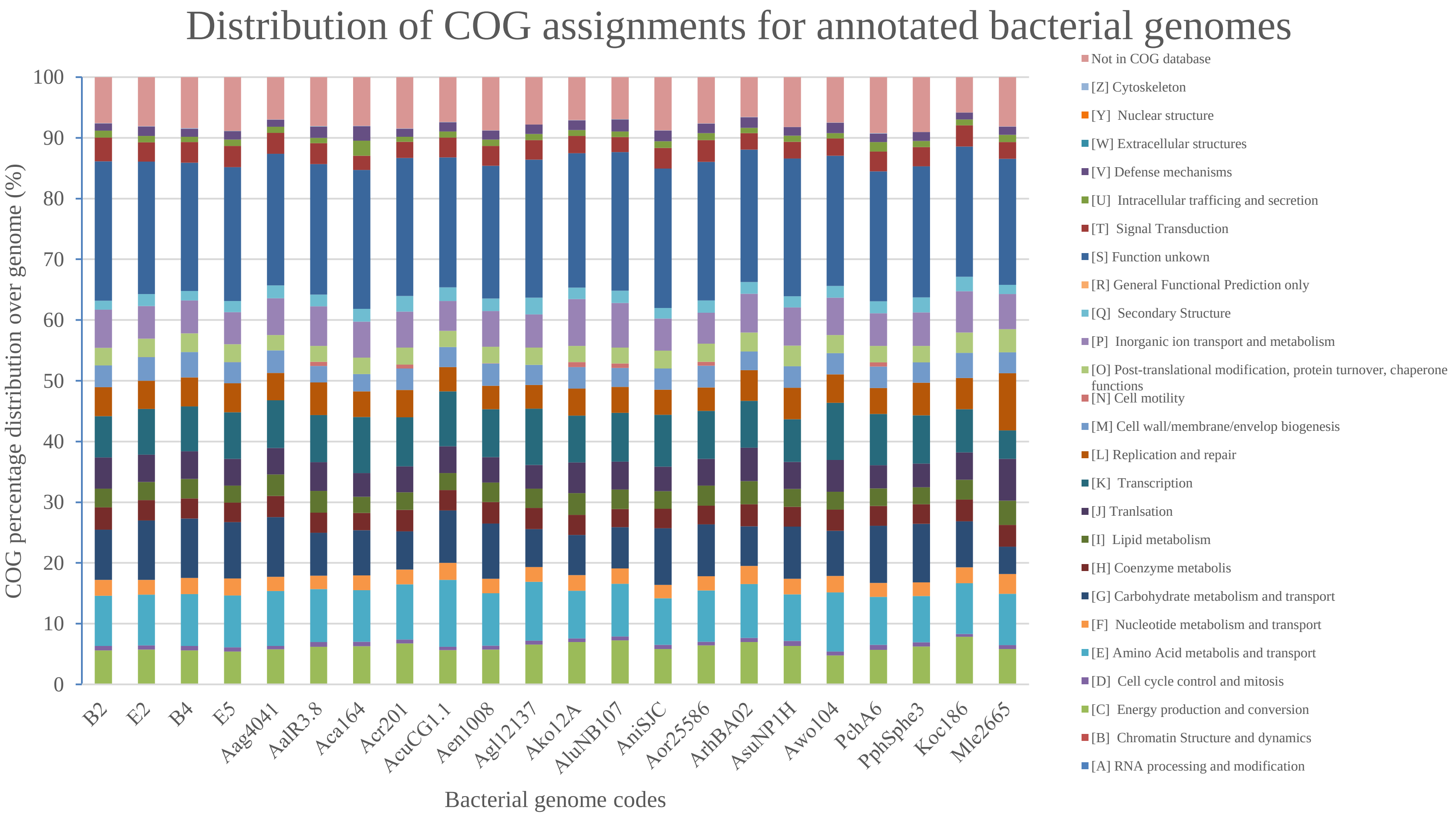


Figure 5.14: Functional annotations (COGs) of annotated bacterial genomes used in this study. COGs [A] RNA processing and modification, [B] Chromatin Structure and dynamics, [N] Cell motility, [W] Extracellular structures and [Z] Cytoskeleton were not included in this graph due to their overall low abundance in each isolate. The distribution for E5 includes the COGs found on the chromosome and plasmid. *Micrococcus luteus* NCTC 2665 (NC_012803.1) and *Kocuria rosea* ATCC 186 (NZ_QFBJ0000000.1) were used as comparative organisms.

As seen in Figure 5.14 above, the COG distribution across *Arthrobacter* and *Pseudarthrobacter* species is highly similar. To observe COG distribution in more detail, a heatmap of COGs was created, as seen below in Figure 5.15. Hierarchical clustering was ordered using the rows (COGs) as the basis.

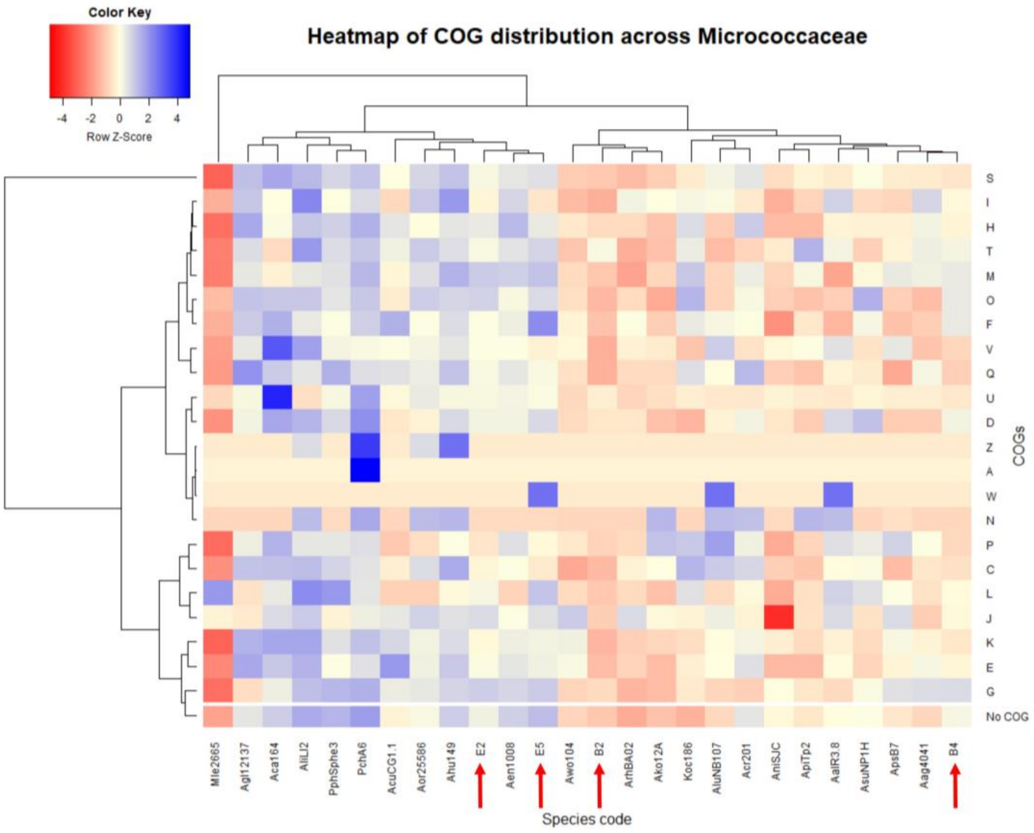


Figure 5.15: Heatmap of COG distribution across *Arthrobacter* and *Pseudarthrobacter*. Namib isolates are indicated by the red arrows. *M. luteus* NCTC 2665 was used as the outgroup organism. *K. rosea* ATCC 186 was also included in the heatmap and clustered with *A. luteolus* NBRC 107841 and *A. crystallopoietes* DSM 20117 on the dendrogram. Red colour indicates below average COG distribution, while blue represents above average COG distribution. Heatmap was created using heatmap.2 function of the R gplots package.

As seen in Figure 5.15 above, the COG distributions of *Arthrobacter* and *Pseudarthrobacter* do not vary largely across the genera; however, the heatmap is able to give a more detailed perspective. COGs G, E, and K were again shown to be the most variable COGs, while the distribution of A (RNA processing and modification), W (extracellular structure) and Z (cytoskeleton) COGs is largely consistent across *Arthrobacter* and *Pseudarthrobacter*, with a few exceptions. For example, all gene sequences used for eggNOG annotation returned only one A COG, except for *P. chlorophenolicus* A6, which had two, both of which were classified as 3'-to-5' exoribonuclease specific for small oligoribonucleotides. The presence of two of these genes may be due to gene duplication.

Figure 5.15 shows that *A. nitrophenolicus* SJCon has fewer genes assigned to the J (translation) COG than other *Arthrobacter*. Isolate B4 had a higher distribution of Z (cytoskeleton) COGs and E5 had a higher distribution W (extracellular structure) COGs than the other Namib isolates. The comparison of COGs in the Namib isolates with *M. luteus* NCTC 2665 as an outgroup can be seen in Figure 5.16 below.

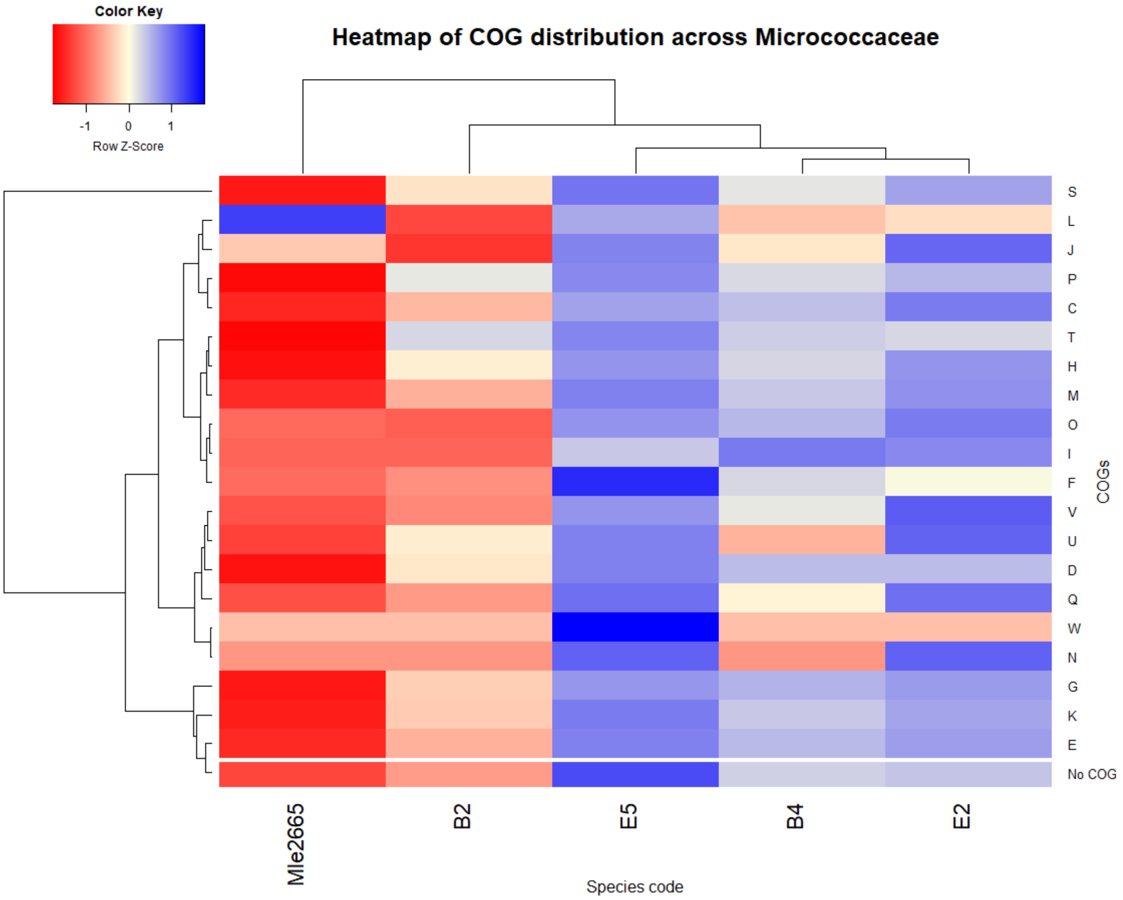


Figure 5.16: Heatmap of COG distribution across the newly sequenced Namib isolates. *M. luteus* NCTC 2665 was used as the outgroup organism. Red colour indicates below average COG distribution, while blue represents above average COG distribution. Heatmap created using the gplots package and heatmap.2 in R.

As seen from Figure 5.16 above, isolate B2 generally has lower COG distribution for each category when compared with the other Namib isolates. This is due to the much smaller genome size (3.2 Mb) of isolate B2 compared to the other isolates. Isolates E5 and E2 generally have a higher distribution of COGs due to their larger genome size (3.8 Mb and 4.03 Mb respectively). Isolate E5 has a higher distribution of W (extracellular structures) and F (nucleotide transport and metabolism) COGs compared to the other isolates. Proteins assigned to COG W are often pilus assembly proteins. Isolate E5 is the only Namib isolate to have a plasmid, indicating that the W COG observed in E5 may be used for pilus construction for plasmid transferral.

B4 has a higher distribution of I (lipid transport and metabolism) COGs compared to the other Namib isolates, while E2 has a higher distribution of V (defence mechanisms), J (translation, ribosomal structure and biogenesis) and U (intracellular trafficking and secretion) COGs.

The discrepancies in colour assignment of COG distribution in Figure 5.15 compared with Figure 5.16 is due to the smaller number of isolates being compared in Figure 5.16, resulting in an adjusted average distribution. COG categories A and Z were excluded from Figure 5.16, due to all isolates used only having one of each.

5.3.6 Whole genome comparisons

Due to difficulties with relying on the 16S rRNA gene for taxonomic classification of the *Arthrobacter* genus (Nouioui et al. 2018), whole genome comparative approaches were used to achieve a more reliable phylogenetic relationship between the Namib isolates and the reference genomes. Mauve alignments, OrthoANI, *in silico* DDH and PhyloPhlAn are the four methods for whole genome analysis selected for this study.

5.3.6.1 Mauve alignments

Mauve is a tool that helps to visualise large-scale evolutionary events through multiple genome alignment (Darling et al. 2004). Mauve was used in this study to compare the genome similarity of the Namib isolates to their presumed closest relative based on the 16S rRNA gene (Figure 5.12). From this analysis, genome rearrangement events between the closest relatives for each of the Namib isolates can be observed.

Mauve alignments were conducted with an 'all against all' approach using the NCBI downloaded sequences as references and the Namib isolates as the query sequence. The MCM algorithm was applied to reorder the Namib isolates' contigs into the most appropriate order for multiple sequence alignment, and the LCB weight was left unspecified for each alignment.

The Mauve analysis when comparing B2 with *A. agilis* 4041 and *A. pityocampae* Tp2 indicated that these two reference genomes are close relatives of B2. The phylogenetic tree of 16S rRNA gene sequences previously created (Figure 5.12) also has high support values for B2, *A. agilis* 4041 and *A. pityocampae* Tp2 being located in the same clade. Figure 5.17a below shows the Mauve alignment for B2 with *A. agilis* 4041, while 5.17b shows the Mauve alignment for B2 and *A. pityocampae* Tp2.

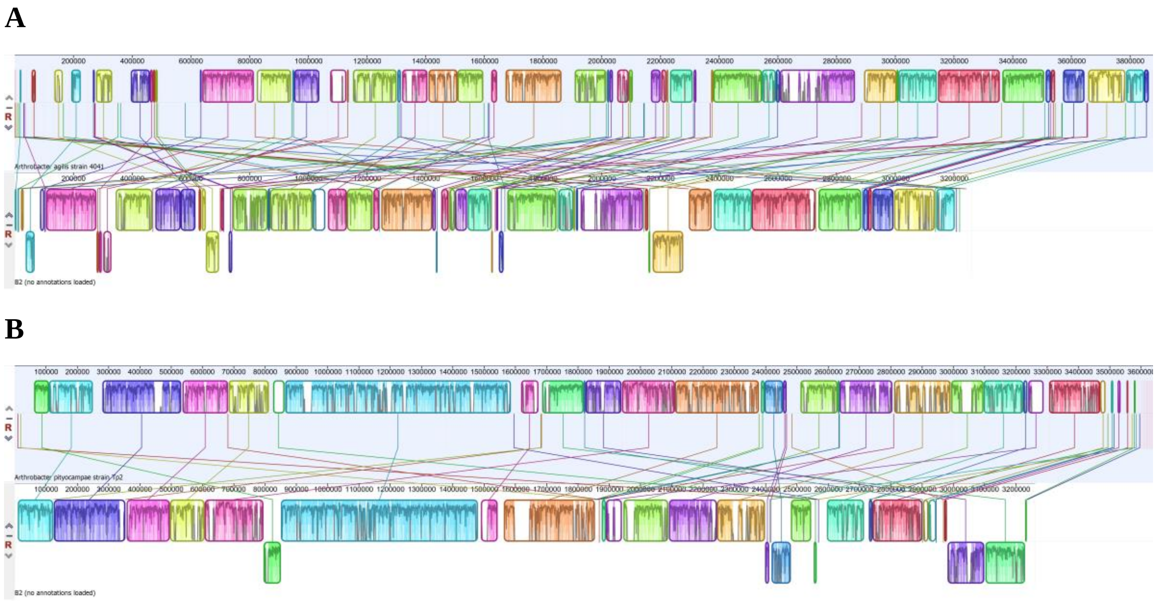


Figure 5.17a: Mauve alignment for *A. agilis* 4041 (top) with B2 (bottom). LCB weight is 98. **5.17b: Mauve alignment for *A. pityocampae* Tp2 (top) with B2 (bottom).** LCB weight is 358.

The Mauve analysis when comparing B4 with *P. phenanthrenivorans* Sphe3 and *P. chlorophenolicus* A6) indicated that these two reference genomes are close relatives of B4. This was also supported by the 16S rRNA phylogenetic tree. Figure 5.18a below shows the Mauve alignment for B4 with *P. phenanthrenivorans* Sphe3 while Figure 5.18b shows the Mauve alignment for B4 with *P. chlorophenolicus* A6.

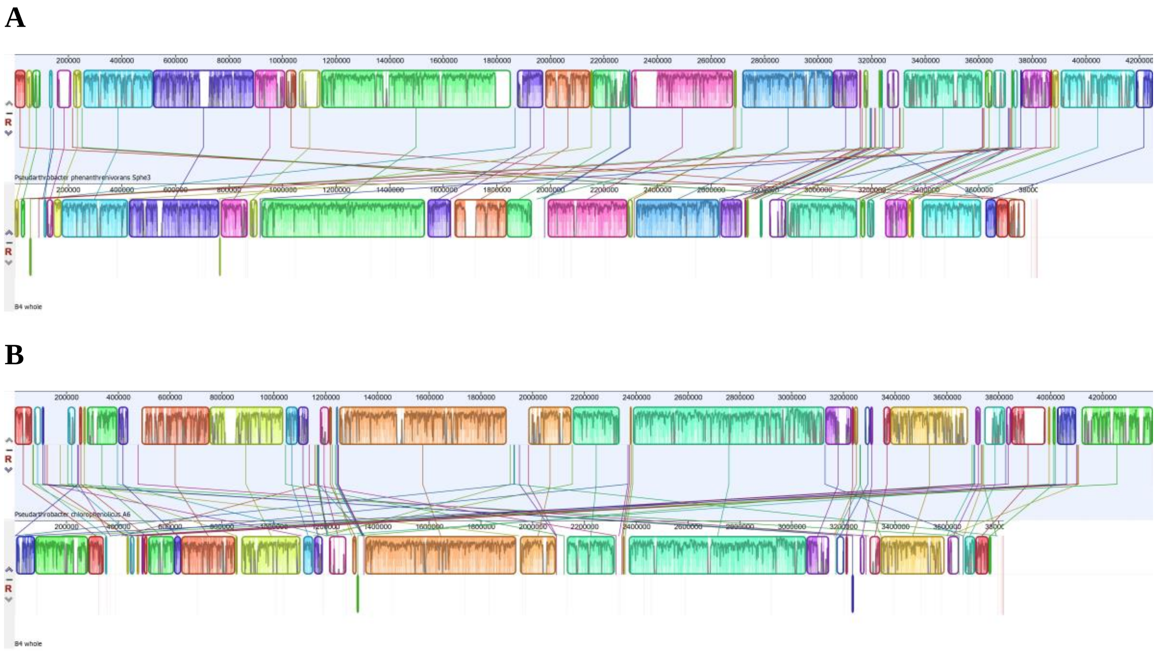


Figure 5.18a: Mauve alignment for *P. phenanthrenivorans* Sphe3 (top) with B4 (bottom). LCB weight is 129. **5.18b: Mauve alignment for *P. chlorophenolicus* A6 (top) with B4 (bottom).** LCB weight is 84.

The Mauve analysis when comparing E5 with *P. phenanthrenivorans* Sphe3 and *P. chlorophenolicus* A6 indicated that these two reference genomes are close relatives of E5. This was supported by the 16S rRNA phylogenetic tree. Figure 5.19a below shows the Mauve alignment for E5 with *P. phenanthrenivorans* Sphe3 while Figure 5.19b shows the Mauve alignment for E5 with *P. chlorophenolicus* A6.

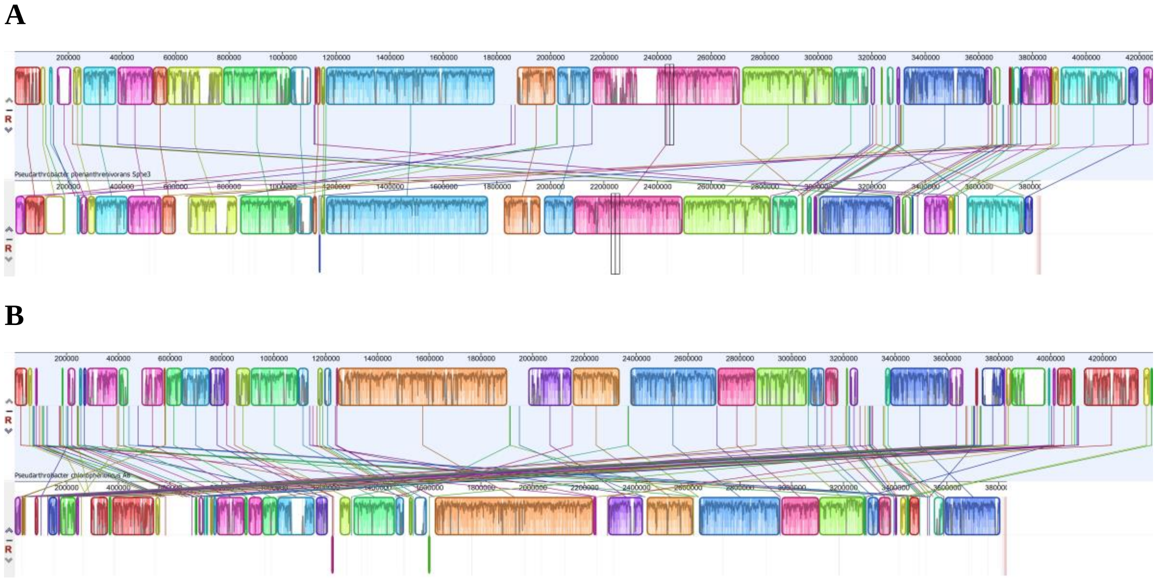


Figure 5.19a: Mauve alignment for *P. phenanthrenivorans* Sphe3 (top) with E5 (bottom). LCB weight is 66. **5.19b: Mauve alignment for *P. chlorophenolicus* A6 (top) with E5 (bottom).** LCB weight is 126.

The phylogenetic tree of 16S rRNA gene sequences previously created (Figure 5.12) determined that E2 appears to be closely related to *A. enclensis* NIO-1008 and *P. phenanthrenivorans* Sphe3. Further examination of the Mauve alignments also indicated that E2 may also be closely related to *P. chlorophenolicus* A6. Figure 5.20a below shows the Mauve alignment for E2 with *A. enclensis* NIO-1008, Figure 5.20b shows the Mauve alignment for E2 *P. phenanthrenivorans* Sphe3, and Figure 5.20c shows the Mauve alignment for E2 *P. chlorophenolicus* A6.

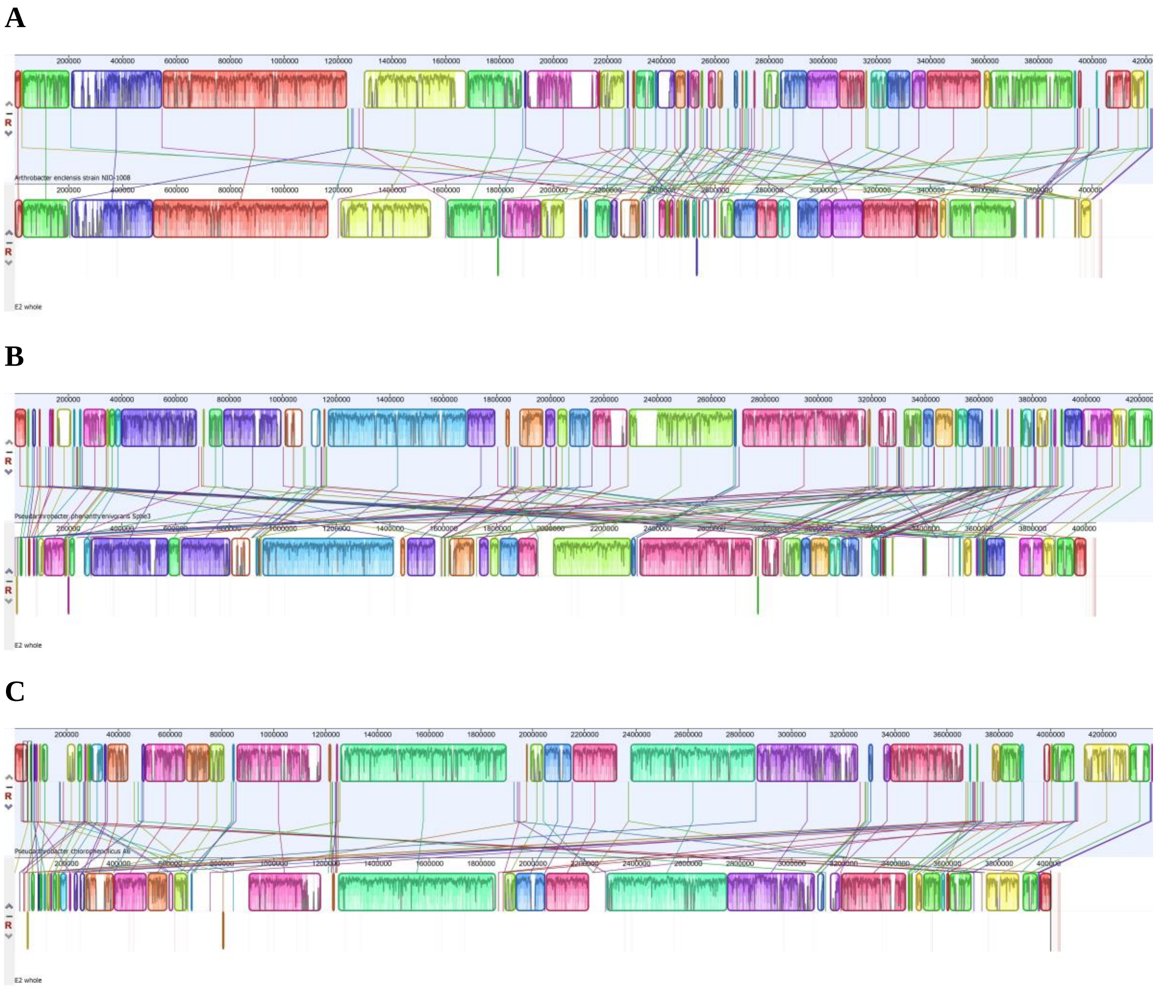

A
Arthrobacter enclensis strain NIO-1008
E2 whole
B
Pseudarthrobacter phenanthrenivorans Sphe3
E2 whole
C
Pseudarthrobacter chlorophenolicus A6
E2 whole

Figure 5.20a: Mauve alignment for *A. enclensis* NIO-1008 (top) with E2 (bottom). LCB weight is 203. **5.20b: Mauve alignment for *P. phenanthrenivorans* Sphe3 (top) with E2 (bottom).** LCB weight is 63. **5.20c: Mauve alignment for *P. chlorophenolicus* A6 (top) with E2 (bottom).** LCB weight is 82.

In addition to comparing the Namib isolates with other genomes of *Arthrobacter* and *Pseudarthrobacter*, the Namib isolates were also compared against each other. Isolates B4 and E5 appeared to share a large number of LCBs, indicating that there is low genome rearrangement between these two isolates, and that they are closely related. This is supported by phylogenetic trees with the B4 and E5 16S rRNA gene. Isolate E2 shared more Mauve similarity with isolates B4 and E5 than with B2 (Figure 5.21); however, the genome of isolate E2 appears to have undergone more rearrangement events than B4 and E5. The genome of isolate B2 is smaller than that of the other isolates (3.2 Mb) and this is reflected in the large sections of missing LCBs from the B2 Mauve alignment. Isolate B2 was used as the reference strain of the Mauve alignment in Figure 5.21 below.

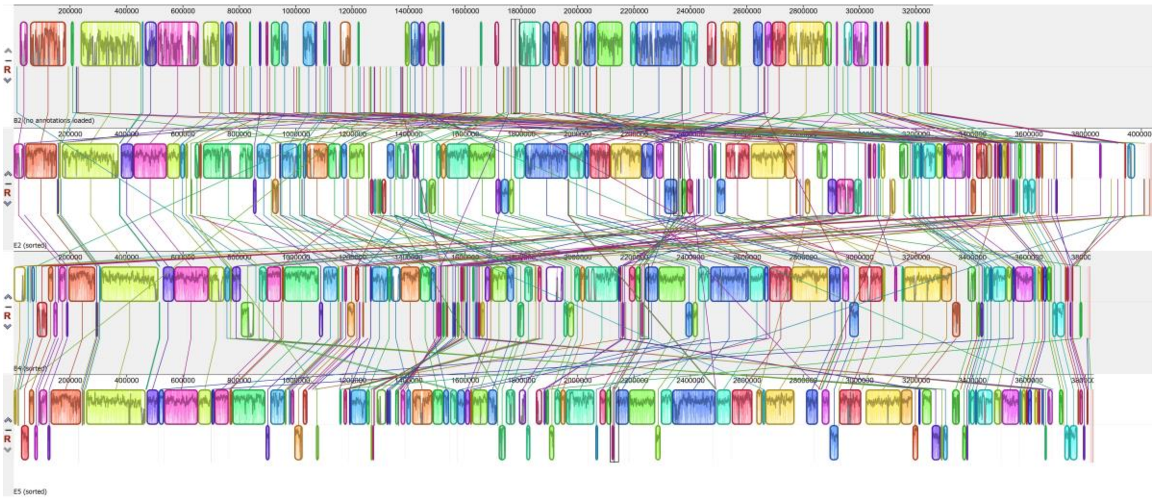


Figure 5.21: Mauve alignment for B2 (top), E2 (second from top), B4 (second from bottom) and E5 (bottom).

As mentioned, the whole genome of each Namib isolate was compared with each NCBI reference genome (Table 5.2). Only the most fitting Mauve alignments were presented in this section. The full Mauve alignments for all genomes can be found in Appendix 4.

5.3.6.2 OrthoANI

OrthoANI is used to compare the average nucleotide identity of two genomes. The average nucleotide identity indicates if two genomes are likely the same species (>95% similarity), or if it is likely that they are two different species (<95% similarity).

To further examine the overall taxonomic classification of the Namib isolates, OrthoANI percentages were calculated using an ‘all against all’ approach with the downloaded NCBI reference genomes. The mean OrthoANI similarity between species was 75%. OrthoANI percentages were visualised using a heatmap, as seen in Figure 5.22 below (refer to Table 5.2 for species abbreviation list). The new species cut-off value for OrthoANI is <95% (Lee et al. 2016; Yoon et al. 2017). The Namib isolates had OrthoANI values of <95% with all other available reference genomes, indicating that all four Namib isolates may be new species.

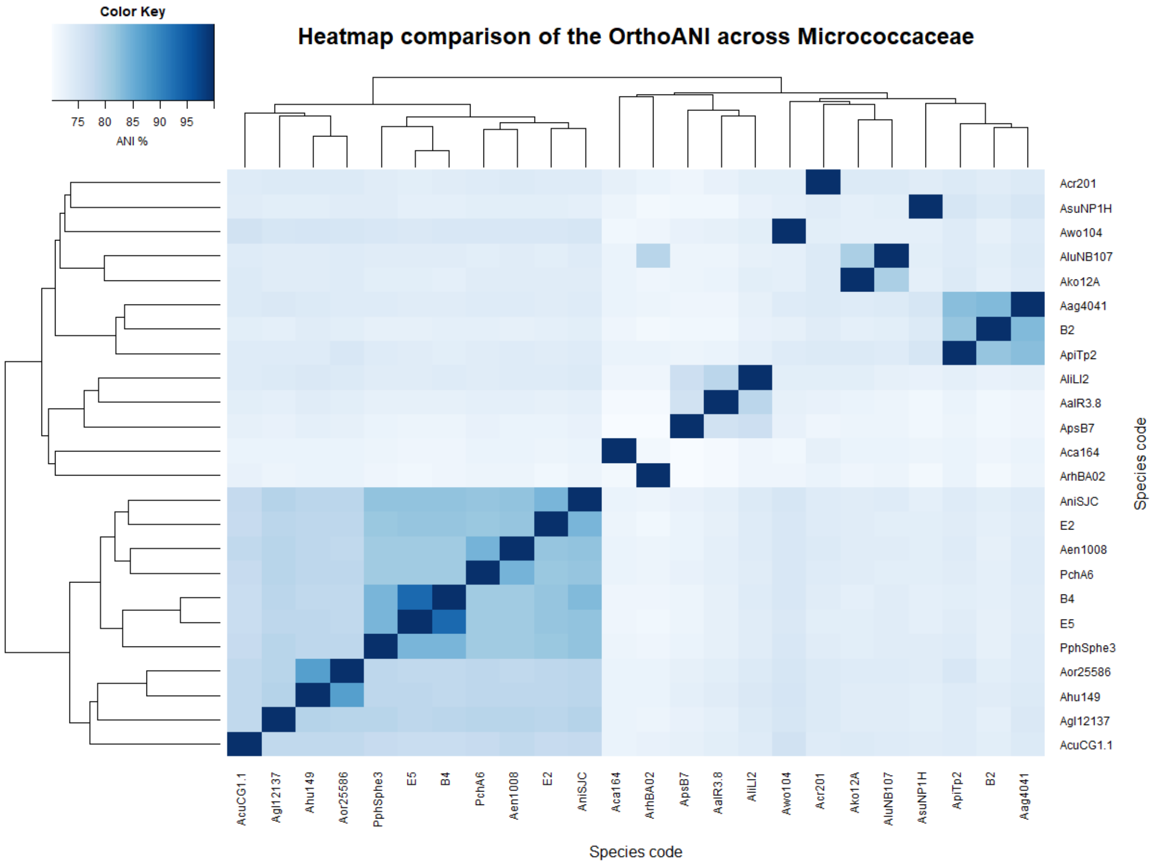


Figure 5.22: Heatmap of OrthoANI percentage across *Arthrobacter* and *Pseudarthrobacter* isolates. Colour key indicates the percentage of shared OrthoANI identity between genomes. Some species appear in a different order for the reciprocal pairwise comparison by the dendrogram option using the heatmap.2 function in R.

Figure 5.22 above shows that isolates B4 and E5 are highly related to one another, and that they share high OrthoANI values with *P. phenanthrenivorans* Sphe3, *P. chlorophenolicus* A6 and *A. enclensis* NIO-1008. Isolate E2 shares high OrthoANI values with *A. enclensis* NIO-1008 and *A. nitrophenolicus* SJCon. B2 shares high OrthoANI values with *A. agilis* 4041 and *A. pityocampae* Tp2. This is a similar result to the observed Mauve genome alignments. The

OrthoANI values for the downloaded NCBI isolates ranged from 70.16 to 82.78 percentage similarity. The OrthoANI percentages between the Namib isolates is shown in Table 5.10 below.

Table 5.10: OrthoANI values between the Namib isolates.

Genome sequence A	Genome sequence B	OrthoANI similarity (%)
B2	B4	72.92
B2	E2	73.01
B2	E5	72.75
B4	E2	81.98
B4	E5	93.21
E2	E5	81.85

As mentioned, the OrthoANI values with a percentage of <95% indicate a new species. Table 5.10 above shows that all Namib isolates have OrthoANI values of <95% when compared with each other, indicating that all four isolates may be distinct new species.

To further identify closest relatives using the OrthoANI algorithm, another heatmap was created with only the Namib isolates across the x-axis.

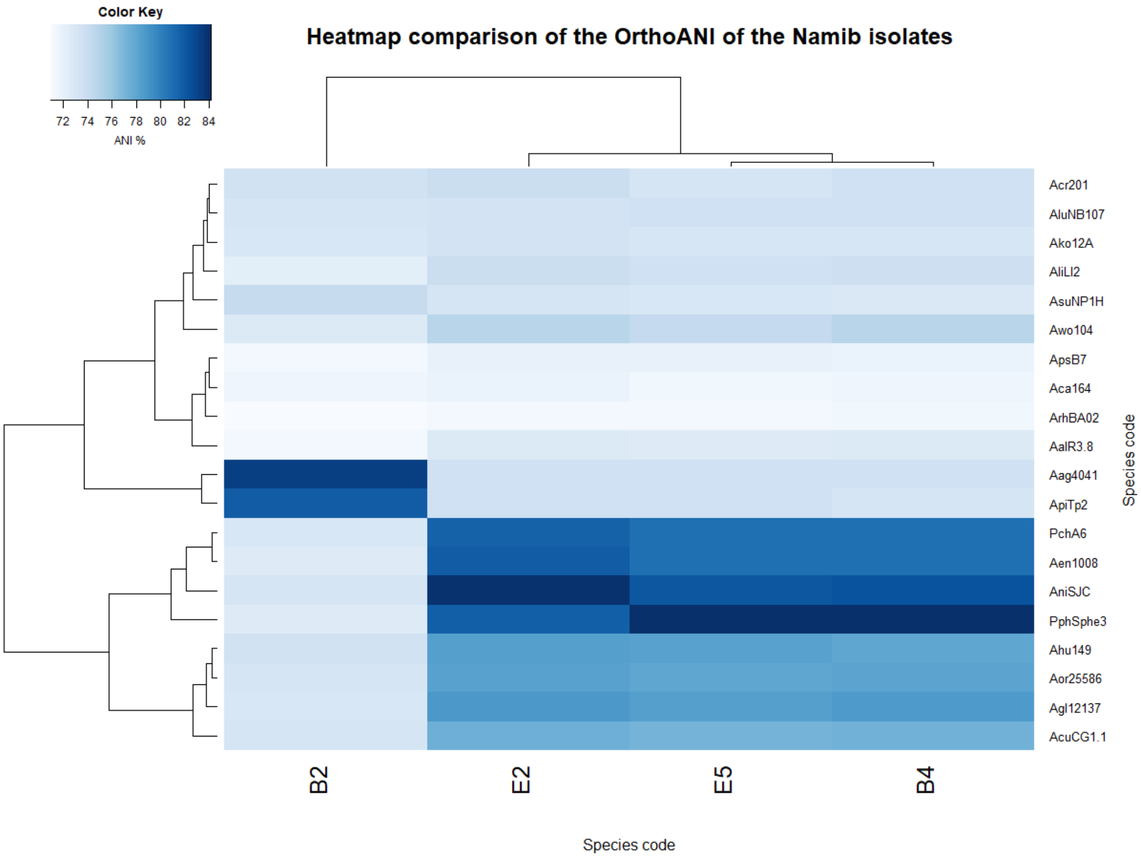


Figure 5.23: Heatmap of OrthoANI percentage across the Namib isolates. Colour key indicates the percentage of shared OrthoANI identity between genomes.

As seen in Figure 5.23 above, B2 shares a high OrthoANI value with *A. agilis* 4041 and *A. pityocampae* Tp2. This is supported by the Mauve alignments of the whole genome, which indicated a close relationship of B2 with both *A. agilis* 4041 and *A. pityocampae* Tp2 (Figure 5.17a & b). Isolates B4 and E5 share a high OrthoANI percentage similarity with *P.*

phenanthrenivorans Sphe3, which was also supported by the Mauve alignments for both genomes (Figures 5.18a & b and 5.19a & b). Isolates B4 and E5 also shared a high OrthoANI percentage similarity with *A. nitrophenolicus* SJCon, *A. enclensis* NIO-1008 and *P. chlorophenolicus* A6. For isolate E2 the closest OrthoANI percentage similarity is with *A. nitrophenolicus* SJCon. This was not determined by the 16S rRNA phylogenetic tree (Figure 4.6, Section 4.3) due to the curated 16S rRNA gene of *A. nitrophenolicus* being absent from the RDP database. Analysis of the Mauve alignment for E2 and *A. nitrophenolicus* SJCon, after using MCM reordering, showed that while the alignment had a relatively low LCB weight (53), there was a degree of reordering across the E2 genome, as shown in Figure 5.24 below.

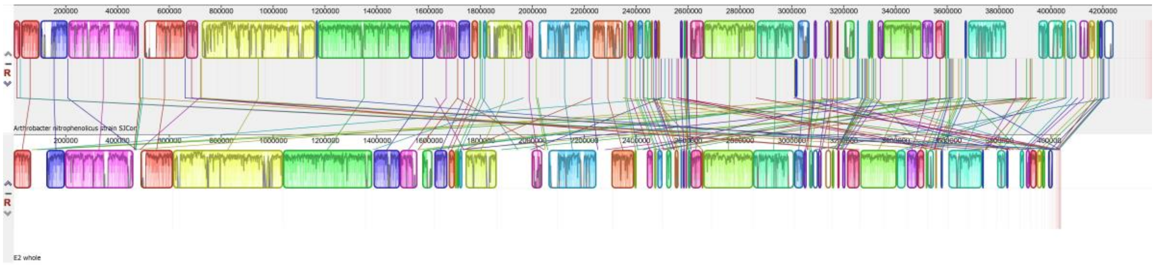


Figure 5.24: Mauve alignment of *A. nitrophenolicus* SJCon (top) and E2 (bottom). LCB weight is 53 which is the default for this alignment.

While there is some reordering along the E2 genome, there are large areas along both genomes which contain similar genes. The differences between the two genomes can be assumed to be due to rearrangement events.

5.3.6.3 *In silico* DNA-DNA Hybridisation

To establish a new species, DNA-DNA hybridisation (DDH) is still a required method when the 16S rRNA gene similarity is above 97% with closely related organisms. The Namib isolates all share above 98% similarity of the 16S rRNA gene with their closest relatives so DDH analysis is required.

As mentioned previously, the issues with wet-lab based DDH methods have resulted in *in silico* DDH programmes. The Genome Blast Distance Phylogeny approach (GBDP) (Meier-Kolthoff et al. 2013) is able to successfully emulate wet-lab DDH results with high accuracy. GBDP results were obtained by using each of the Namib isolates as the query sequence, and the NCBI reference genomes (Table 5.2) as the reference sequence. Only the two highest DDH percentage values between the Namib isolates and the reference sequences is presented here. *In silico* DDH percentages can be seen in Table 5.11 below and are reported from using Formula 2 of the comparison ('Greedy-with-trimming'), as recommended (Meier-Kolthoff et al. 2013).

Table 5.11: DDH estimation using Genome-to-Genome Distance Calculator.

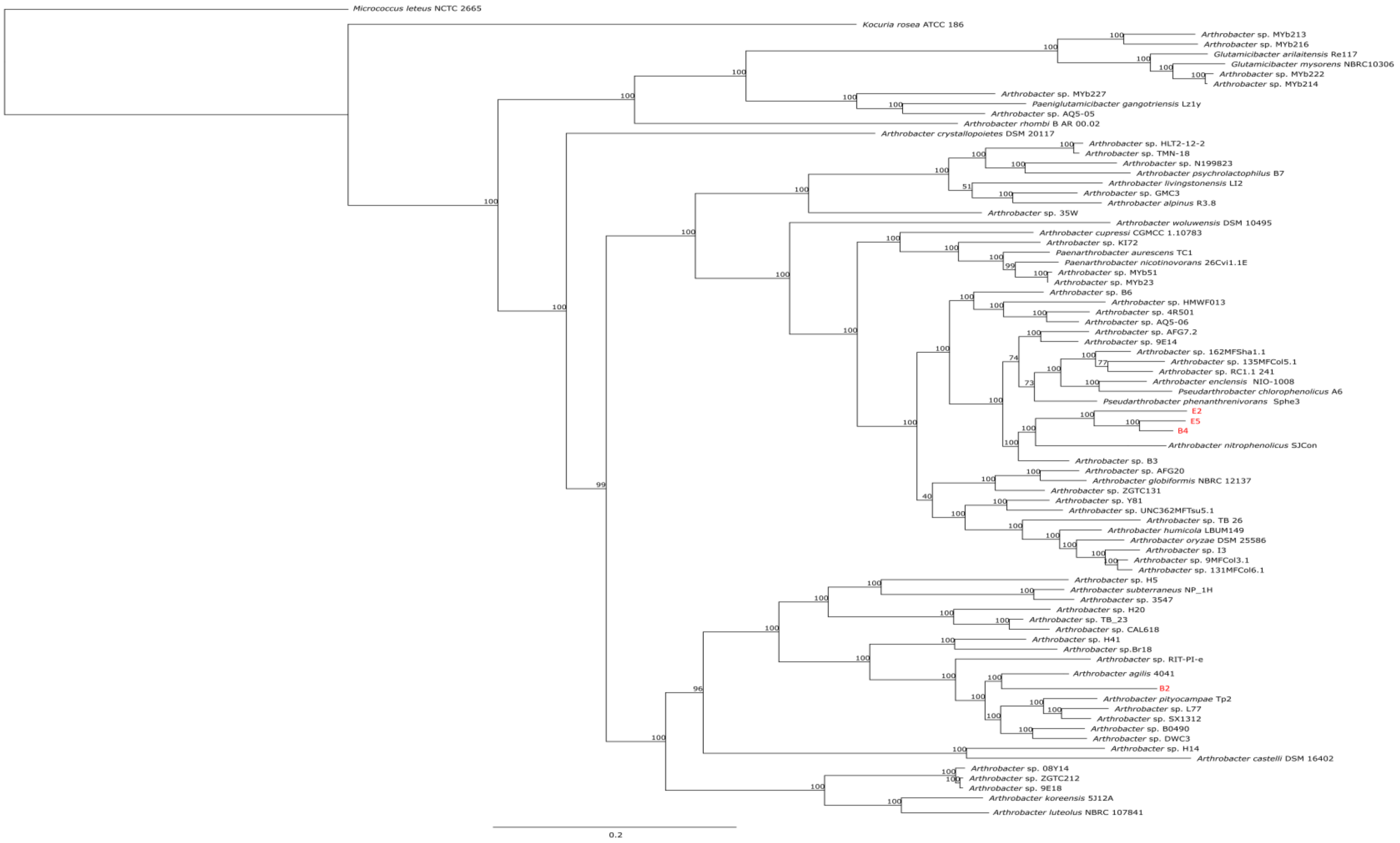
Reference sequence	Query sequence	DDH (%)
A. agilis 4041	B2	26.9
A. pityocampae Tp2	B2	24.8
A. nitrophenolicus SJCon	E2	27.4
P. phenanthrenivorans Sphe3	E2	24.3
P. phenanthrenivorans Sphe3	B4	27.7
A. nitrophenolicus SJCon	B4	25.5
P. phenanthrenivorans Sphe3	E5	27.5
A. nitrophenolicus SJCon	E5	25.1
B2	B4	19.9
B2	E2	20
B2	E5	20.4
B4	E2	24.9
B4	E5	51.8
E2	E5	24.8

As seen in Table 5.11 above, all four Namib isolates share <70% DDH similarity with their closest relatives. Furthermore, the Namib isolates shares <70% with each other. This supports the results observed through OrthoANI (Table 5.10 and Figure 5.23). By DDH specifications, this indicates that all four Namib isolates are distinct new species.

5.3.6.4 PhyloPhlAn

To further reconstruct the phylogenetic relationships between the *Arthrobacter* and *Pseudarthrobacter* genera, PhyloPhlAn was used as a multigenic approach. As previously discussed, some *Arthrobacter* occupy various positions within different phylogenetic trees (Busse 2016). As PhyloPhlAn considers over 400 conserved genomic proteins during tree construction (Segata et al. 2013), the resulting tree is considered highly accurate.

Isolate B2 was placed close to *A. agilis* 4041 and *A. pityocampae* Tp2 within the phylogenetic tree by PhyloPhlAn. This is consistent with the phylogenetic classification of both Mauve and OrthoANI. Isolates B4, E5 and E2 appear closely related based on the PhyloPhlAn phylogenetic tree. These isolates appear in the *Pseudarthrobacter* clade. Several *Arthrobacter* spp. are present in this clade also, however, these have been reclassified as *Pseudarthrobacter* spp. by EzBioCloud: a public analysis portal for bacterial taxonomy, genomics and metagenomics (Yoon et al. 2017). As seen in Figure 5.25 below, PhyloPhlAn gave taxonomic assignment for each of the Namib isolates consistent with the phylogenetic classification of both Mauve and OrthoANI.



Micrococcus leteus NCTC 2665
Kocuria rosea ATCC 186
Arthrobacter sp. MYb213
Arthrobacter sp. MYb216
Glutamicibacter arilaitensis Re117
Glutamicibacter mysorens NBRC10306
Arthrobacter sp. MYb222
Arthrobacter sp. MYb214
Arthrobacter sp. MYb227
Paeniglutamicibacter gangotriensis Lz1y
Arthrobacter sp. AQ5-05
Arthrobacter rhombi B AR 00.02
Arthrobacter crystallopoietes DSM 20117
Arthrobacter sp. HLT2-12-2
Arthrobacter sp. TMN-18
Arthrobacter sp. N199823
Arthrobacter psychrolactophilus B7
Arthrobacter livingstonensis LI2
Arthrobacter sp. GMC3
Arthrobacter alpinus R3.8
Arthrobacter sp. 35W
Arthrobacter woluwensis DSM 10495
Arthrobacter cupressi CGMCC 1.10783
Arthrobacter sp. KI72
Paenarthrobacter aurescens TC1
Paenarthrobacter nicotinovorans 26Cvi1.1E
Arthrobacter sp. MYb51
Arthrobacter sp. MYb23
Arthrobacter sp. B6
Arthrobacter sp. HMWF013
Arthrobacter sp. 4R501
Arthrobacter sp. AQ5-06
Arthrobacter sp. AFG7.2
Arthrobacter sp. 9E14
Arthrobacter sp. 162MFSha1.1
Arthrobacter sp. 135MFCol5.1
Arthrobacter sp. RC1.1 241
Arthrobacter enclensis NIO-1008
Pseudarthrobacter chlorophenolicus A6
Pseudarthrobacter phenanthrenivorans Sphe3
E2
E5
B4
Arthrobacter nitrophenolicus SJCon
Arthrobacter sp. B3
Arthrobacter sp. AFG20
Arthrobacter globiformis NBRC 12137
Arthrobacter sp. ZGTC131
Arthrobacter sp. Y81
Arthrobacter sp. UNC362MFTsu5.1
Arthrobacter sp. TB 26
Arthrobacter humicola LBUM149
Arthrobacter oryzae DSM 25586
Arthrobacter sp. I3
Arthrobacter sp. 9MFCol3.1
Arthrobacter sp. 131MFCol6.1
Arthrobacter sp. H5
Arthrobacter subterraneus NP_1H
Arthrobacter sp. 3547
Arthrobacter sp. H20
Arthrobacter sp. TB_23
Arthrobacter sp. CAL618
Arthrobacter sp. H41
Arthrobacter sp.Br18
Arthrobacter sp. RIT-PI-e
Arthrobacter agilis 4041
B2
Arthrobacter pityocampae Tp2
Arthrobacter sp. L77
Arthrobacter sp. SX1312
Arthrobacter sp. B0490
Arthrobacter sp. DWC3
Arthrobacter sp. H14
Arthrobacter castelli DSM 16402
Arthrobacter sp. 08Y14
Arthrobacter sp. ZGTC212
Arthrobacter sp. 9E18
Arthrobacter koreensis 5J12A
Arthrobacter luteolus NBRC 107841
0.2

Figure 5.25: Phylogenetic tree of the *Arthrobacter* and *Pseudarthrobacter* genera using PhyloPhlAn. The Namib isolate B2 is in the same clade as *A. agilis* 4041 and *A. pityocampae* Tp2, a relationship that is supported by both Mauve and OrthoANI. The Namib isolates B4, E5 and E2 share a clade with *A. nitrophenolicus* SJCon, *A. enclensis* NIO-1008, *P. phenanthrenivorans* Sphe3 and *P. chlorophenolicus* A6, relationships that are also supported by both Mauve and OrthoANI. The genome of *Micrococcus luteus* NCTC 2665 was used as an outgroup. The length of the scale bar indicates 0.2 substitutions per amino acid.

5.4 Discussion

Actinobacteria are the dominant phyla in desert systems and are important contributors to the functionality of these arid and extreme soils. *Arthrobacter* are ubiquitous organisms and their presence in soil is often reported (Chan et al. 2013; Makhalanyane et al. 2013). The Namib Desert is an environment that is very restrictive to the growth of microorganisms due to the lack of available water, temperature fluctuations and high surface UV incidence (Makhalanyane et al. 2015). It was established in Section 4.3 that isolates B2 and B4 demonstrated UVC resistance for 10 minutes, while E2 and E5 were UVC sensitive. This raises the question, how are isolates B2 and B4 able to survive this amount of radiation compared with other phylogenetically similar *Arthrobacter*? As a first step to answering this question, the genomes of one *Arthrobacter* (B2) and three *Pseudarthrobacter* (B4, E2 and E5), which were isolated from Namib Desert soil, were sequenced using next generation sequencing. As of November 2018 when this analysis was completed, 104 *Arthrobacter* and 28 *Pseudarthrobacter* had been sequenced (Yoon et al. 2017), with 18 *Arthrobacter* and two *Pseudarthrobacter* being assigned species names. This is the first study that has sequenced the genomes of *Arthrobacter* and *Pseudarthrobacter* originating from the Namib Desert due to their UV resistance or sensitivity.

5.4.1 Genome assembly

This study used the Illumina HiSeq system for next generation sequencing of the four Namib isolates. The genome assembly of the Namib isolates resulted in 23 – 81 contigs. This is a typical range of contigs for the *Arthrobacter* and *Pseudarthrobacter* genera, with publicly available assemblies from NCBI having between 1 and 118 contigs as of 2019. A presumed plasmid, represented by a single contig, was found in E5. The size of this plasmid is 227,930 bp, which is a typical size of plasmids in the *Arthrobacter* genus (Mongodin et al. 2006; Shintani et al. 2015).

The GC content of the sequenced Namib Desert isolates is within the range of other *Arthrobacter* (59.7 – 70.6%) and *Pseudarthrobacter* (64.3 – 67.8%) (Yoon et al. 2017). The estimated genome size of the Namib isolates is between 3,262,671 and 4,032,940 bp in length, which is also within the expected range for *Arthrobacter* (2,594,729 – 7,019,656 bp) and *Pseudarthrobacter* (3,779,248 – 5,056,151) (Yoon et al. 2017). Finally, the number of tRNA and rRNA genes is also within the expected range for *Arthrobacter* and *Pseudarthrobacter* (Yoon et al. 2017).

While the genome assembly features of the Namib isolates are in line with expected results, the presence of more than a single contig leaves the genomes in a draft form. Leaving these genomes in draft form may mean that important data is missed (Ricker et al. 2012). This is particularly important in understanding how these isolates have developed to become either resistant or sensitive to UV radiation. DNA repeat regions are often contributors to DNA fragmentation, leading to multiple contigs (Ricker et al. 2012). To resolve this, single molecule sequencing platforms such as PacBio SMRT or the ONT MinION would need to be used on the Namib isolates. Both technologies produce very long read lengths, often leading to a complete genome assembly when used as scaffolds for Illumina reads (Koren and Phillippy 2015; Wick et al. 2017a). The error rates for these long-read platforms is decreasing, making them more and more useful as a tool for understanding genome sequences and structures.

DNA repeat regions can have a range of functions, such as gene regulation for growth in specific environments. Ricker et al. (2012) found that repetitive DNA regions caused fragmentation, and that in *Cupriavidus metallidurans* some of the repeat regions contained genes for surviving heavy metal environmental contamination. In comparing their *de novo* assemblies against well-annotated finished reference assemblies, these 'genomic-islands' were fragmented, meaning that information regarding the adaptation of *C. metallidurans* to heavy metal environments was missed. With regard to the Namib isolates, it is possible that information regarding their adaptation to the restrictive Namib Desert environment could have been missed. This could be remediated through the use of either PacBio SMRT or the ONT MinION to establish a complete genome.

5.4.2 rRNA genes

Each of the Namib isolates contained one copy of the 16S and 23S rRNA genes. Isolates B2, B4 and E5 also contained only one copy of the 5S rRNA gene, while E2 contained two copies of the 5S rRNA gene. The two copies of the 5S rRNA gene in E2 were identical, indicating that either one of the 5S rRNA genes has been duplicated, or that one of these genes is an orphan gene, or due to assembly error.

The copy number of 16S rRNAs observed with the *Arthrobacter* genus ranges from 4 – 8, while the copy number for the *Pseudarthrobacter* genus ranges from 4 – 6 (Chun et al. 2007). This puts the copy number of the 16S rRNA gene of the Namib isolates lower than the expected range for the *Arthrobacter* and *Pseudarthrobacter* genera. Since rRNA genes are highly conserved, the SPAdes assembler may have had issues with putting these into contigs, or the genes may have been fragmented. Obtaining a complete genome of the Namib isolates may reveal further rRNA genes.

The 16S rRNA genes of the Namib isolates were found to share over 98.5% similarity via BLASTn with their closest identified relatives. The 16S rRNA gene of B2 shared 99.7% similarity with *A. agilis* DSM 20550, with variations only found within the V3 region (Figure 5.11a). The 16S rRNA gene of E2 shared 98.65% similarity with *A. enclensis* NIO-1008, with variations observed within the V1, V3 and V6 regions (Figure 5.11b). The 16S rRNA gene of B4 and E5 shared 98.9% and 99% similarity with *P. phenanthrenivorans* Sphe3. Variations between *P. phenanthrenivorans* Sphe3 and B4 occurred in the V1, V2, V3 and V8 regions (Figure 5.11c). Variations between *P. phenanthrenivorans* Sphe3 and E5 occurred in the V1, V2 and V8 regions (Figure 5.11d).

5.4.3 CRISPR arrays

MINCED (Skennerton 2006) software v0.2.0 did not detect CRISPR repeat regions for any of the four Namib isolates. This aligns with the findings of Kallimanis et al. (2011), who found that the complete genome of *P. phenanthrenivorans* Sphe3 (then *A. phenanthrenivorans* Sphe3) did not contain any CRISPR arrays. Similarly, both Russell and Hatfull (2016) and Singh et al. (2016) noted that they did not identify any CRISPR repeats in *A. agilis* L77 or *Arthrobacter* sp. ATCC 21022.

Conversely, both See-Too et al. (2017) and Xu et al. (2017) did identify 1 and 125 CRISPR within *A. alpinus* R3.8 and *Arthrobacter* sp. B6 respectively. However, the phylogenetic positions of isolates E2, B4 and E5 are much closer to *P. phenanthrenivorans* Sphe3, while B2 is positioned close to *Arthrobacter agilis* 4041 (Figure 5.25). The absence of CRISPR appears to be phylogenetically conserved and consistent with closely related strains as described by previous studies (Kallimanis et al. 2009; Russell and Hatfull 2016; Singh et al. 2016).

5.4.4 Metabolic potentials

The metabolic potentials of the Namib isolates were investigated by examining KEGG annotations by the RAST server and biochemical testing. Overall, the Namib isolates were indole negative, non-motile and unable to utilise urease, citrate or lactose.

It was established that the draft genomes for the Namib isolates had the potential for complete glycolysis/gluconeogenesis and TCA cycle, suggesting that these genomes belong to aerobic heterotrophs. This is consistent with previous findings (Busse 2016; Busse et al. 2015). The presence of 1 (B2) to 5 (B4, E2 and E5) anaerobic reductases, and 1 (B4, E2 and E5) to 2 (B2) Fur family regulators, indicates the potential for all isolates to carry out anaerobic respiration (Yang et al. 2013). However, biochemical testing revealed that isolates B2 and E2 were unable to utilise glucose under anaerobic conditions. While it is unknown how deep the Namib isolates might occur within the soil, the isolates were isolated from top soil, which may indicate the anaerobic genes are ancestral. Further biochemical testing under anaerobic conditions is required to confirm this hypothesis. Alternatively, the Namib isolates may possess anaerobic genes because they are in higher abundance in deeper soil layers. Further community testing within greater soil depth is required to investigate this further.

Isolate B2 has the required genes for the uptake of extracellular glucose for glycolysis and gluconeogenesis. This was confirmed via biochemical testing (Table 5.8). Isolate E2 was also able to utilise glucose, but the required enzymes (EC 2.7.1.199) for the uptake of glucose were not observed in the genome of E2. *A. nitrophenolicus* SJCon, the closest relative to E2, can produce acid from glucose (Arora and Jain 2013), indicating that some genes may have been missed during the annotation of E2. B4 and E5 did not have the required enzymes for external glucose uptake and this was confirmed by biochemical testing. These results are supported by literature of the closest relative genomes to each isolate except for glucose utilisation in B2 (Busse et al. 2015; Kallimanis et al. 2009). Neither *A. agilis* nor *A. pityocampae* have been observed to ferment glucose (Busse et al. 2015; İnce et al. 2014). This further indicates that B2 is genetically distinct from *A. agilis* 4041 and *A. pityocampae* Tp2.

Isolates B2, B4 and E2 have the required genes for the uptake of extracellular sucrose. However, biochemical testing showed that none of the Namib isolates were able to utilise sucrose, indicating that the sucrose utilisation genes may be inactive or damaged. *P. phenanthrenivorans* Sphe3 and *A. agilis* 4041 have been reported as unable to utilise sucrose (Busse et al. 2015; Kallimanis et al. 2009). *A. nitrophenolicus* SJCon has been reported to ferment sucrose (Arora and Jain 2013). However, the Namib Desert does not have plant or grass life regularly, meaning that E2 may have lost the ability to ferment sucrose due to its low availability in a desert system. All four Namib isolates were able to utilise xylose and maltose.

The Namib isolates were able to reduce nitrate to nitrite. For isolates B4 and E5, this agrees with previous findings which noted that *P. phenathrenivorans* can reduce nitrate to nitrite (Busse et al. 2015). Busse et al. (2015) also noted that *A. agilis*, the closest relative of B2, is unable to reduce nitrate to nitrite. However, *A. pityocampae* Tp2, another close relative of B2, is able to reduce nitrate to nitrite (İnce et al. 2014). The ability of B2 to reduce nitrate to nitrite is therefore in line with closely related species. Finally, *A. nitrophenolicus* SJCon, the closest relative of E2, is also unable to reduce nitrate to nitrite (Arora and Jain 2013). The difference in nitrate reduction abilities between E2 and *A. nitrophenolicus* SJCon reinforces that these two isolates are different species. Both E2 and *A. nitrophenolicus* SJCon were isolated from soil, indicating that the ability of E2 to reduce nitrate to nitrite is due to environmental selection pressure instead of due to phylogenetic relationships.

The Namib isolates contain the enzymes responsible for reducing sulphate to sulphite; however, the functionality of this pathway was not investigated during biochemical testing. Sulphur can be utilised in microbial respiration and other cellular processes (Madigan et al. 2006). The Namib isolates also have the enzymes for the conversion of sulphite to sulphide, and then further to L-cysteine for use in the carbon fixation and cysteine and methionine metabolism (Figure 5.8). Other *Arthrobacter* have been reported to grow aerobically on methylated sulphur compounds, however, no *Arthrobacter* have previously been reported to grow autotrophically on inorganic sulphur compounds (Busse et al. 2015). This indicates that it is unlikely for the Namib isolates to be able to utilise sulphur as an energy source.

The E5 plasmid contains the enzymes ribokinase (RbsK) and ribose 5-phosphate isomerase B (RpiB), two enzymes that are involved in the pentose phosphate pathway. These enzymes facilitate the breakdown of D-ribose to D-ribose-5-phosphate using ATP, with Mg^{2+} or Mn^{2+} as a cofactor (Chuvikovsky et al. 2006) As shown in Table 3.3 (Section 3.3), soil from site T8 contained 33.6 mg/kg of magnesium, indicating that there is enough magnesium in the environment to support the use of this pathway. It was previously mentioned that the E5 genome does not appear to have the required genes for extracellular D-glucose. E5 is instead able to uptake extracellular fructose and mannitol, both of which are produced by other microorganisms and plants (Gunina and Kuzyakov 2015; Yu et al. 2016).

As mentioned, the enzymes for converting phytoene into the carotenoid lycopene, and for the conversion of lycopene to bacterioruberin, were found in B2. The pigment produced by B2 is most likely bacterioruberin; however, lycopene also produces a red-pink pigment (Dsouza et al. 2015). The chemical structures of lycopene and bacterioruberin are different. The identity of the pigment could therefore be confirmed by using gas chromatography mass spectrometry or high pressure liquid chromatography (Yang et al. 2015). Isolates E2, B4 and E5 did have some carotenoid biosynthesis enzymes; however, the pathways were incomplete, reflected in the phenotype of the white colonies on agar. Carotenoid production has been identified as a mechanism for UV survival in microorganisms (Dieser et al. 2010; Sutthiwong et al. 2014). Further investigations regarding carotenoid production against UV radiation may involve transforming a complete carotenoid pathway into isolates E2 and E5. Both isolates have been deemed UVC sensitive by the testing conditions of this thesis. Introducing carotenoid biosynthesis pathways into these organisms may cause a change in their UV sensitivity.

5.4.5 Genomic comparisons

To investigate phylogenetic and taxonomic relationships between the Namib isolates and the NCBI reference genomes, several comparative genomics approaches were utilised. COG distribution revealed that the Namib isolates have a similar COG distribution as other *Arthrobacter* and *Pseudarthrobacter*. Taxonomic classification of the Namib isolates revealed that all four Namib isolates are genetically distinct from the *Arthrobacter* and *Pseudarthrobacter* reference genomes used in this study.

5.4.5.1 COG distribution

The COG system assigns all known proteins into orthologous groups. The distribution of COGs of the Namib isolates was compared to other *Arthrobacter* and *Pseudarthrobacter* with Micrococcaceae representative species *K. rosea* and *M. luteus* used as outgroup organisms. Figure 5.14 showed that there was a general trend of COG distribution that was consistent across all the genomes used in this study. Further analysis of COG distribution using a heatmap (Figure 5.15) showed that COG categories G (carbohydrate metabolism and transport), E (amino acid metabolism and transport) and K (transcription) were the most variable across all Micrococcaceae genomes used in this study. Analysis of COG distribution only the Namib isolates (Figure 5.16) showed that isolates E5 and E2 generally had higher COG distribution when compared with B2 and B4, while B2 generally had lower COG distribution compared to the other Namib isolates due to the smaller genome size of B2. Each of the Namib isolates has one Z COG (cytoskeleton). The bacterial cytoskeleton is responsible for elements of cell division and chromosome segregation, as well as the maintenance of cell polarity and cell shape (Graumann 2007).

Overall, the Namib isolates had a higher distribution of G COGs (carbohydrate metabolism and support) compared to other *Arthrobacter* such as *A. crystallopoietes* DSM 20117, *A. globiformis* NBRC 12137, *A. koreensis* 5J12A, *A. luteolus* NBRC 107841 and *A. rhombi* B Ar 00.02. The reference sequences of *P. chlorophenolicus* A6 and *P. phenanthrenivorans* Sphe3, *A. psychrolactophilus* B7, *A. nitrophenolicus* SJCon, *A. enclensis* NIO-1008and *A. agilis* 4041 had a similar distribution of G COGs to the Namib isolates. All of these isolates originate from soil, indicating that the distribution of G COGs may be more due to environmental selection pressure instead of due to phylogeny (Figure 5.25). Furthermore, the number of genes categorised as COG C, I, V and N in the Namib genomes were lower than other *Arthrobacter* and *Pseudarthrobacter*. This indicates that survival in desert soil may require fewer energy production and conversion, lipid metabolism, defence mechanisms and cell motility functions. Both E2 and E5 are phylogenetically similar to *A. nitrophenolicus* and *P. phenanthrenivorans,* both of which only had one N COG, and have previously been reported as non-motile. One gene associated with chromatin structure and dynamics (COG B) was found in each Namib isolates and reference genomes.

One COG W (extracellular structure) was found for *A. luteolus* NBRC 107841, *A. alpinus* R3.8 and E5. Proteins assigned to COG W are often pilus assembly proteins. Isolate E5 is the only Namib isolate to have a plasmid. It could therefore be theorised that the E5 plasmid was obtained by conjugation with another organism, meaning that E5 now requires pilus formation genes to transfer the plasmid to other organisms.

5.4.5.2 Taxonomic classification of Namib isolates

Four methods; Mauve alignments, OrthoANI, *in silico* DDH and PhyloPhlAn, were used to examine the taxonomic classification of the Namib isolates. Analysis of the Mauve alignments showed that all Namib isolates shared a large amount of similarity in gene arrangement with their closest 16S rRNA gene relatives (Figures 5.17, 5.18, 5.19 and 5.20). However, while the Mauve alignments shows that there is a general gene arrangement trend across all Namib isolates and their closest reference genomes, there were still several rearrangements in each alignment. Rearrangement events between different species are expected (Darling et al. 2008; Prabha et al. 2014). During evolution, organisms experience mutational events such as gene rearrangement, lateral transfer, gene loss or duplication. Genome rearrangement, therefore, has a large impact on the organisms' phenotype and can influence gene expression (Darling et al. 2008; Prabha et al. 2014). This is of particular interest in this thesis due to the targeted expressed phenotype of UV resistance. While Mauve alignments do not show specific gene mutations, observing alignments can be used to characterise the shared amount of sequence found in each species (Darling et al. 2010). This in turn can indicate phylogenetic relationships. However, as the Namib isolates are in contig form and not a complete genome, this may introduce bias into the progressiveMauve alignments, even after re-ordering the contigs (Shaik et al. 2016). While progressiveMauve contains algorithms to reduce anchored bias (Darling et al. 2010), utilising a complete genome for Mauve will help to completely remove any bias.

Further taxonomic analysis of the Namib isolates using OrthoANI and *in silico* DDH indicates that the Namib isolates are genetically distinct from the available reference genomes. OrthoANI shows that isolates B4 and E5 share the most similarity of the Namib isolates. B4 and E5 are also more closely related to each other than any of the available reference genomes.

Our understanding of molecular based microbiology is constantly expanding. The novelty of two UVC resistant genomes from the Namib Desert further highlights how important it is to continue improving our investigations into microbial desert communities. Further investigation into the distribution of B2, B4, E2 and E5 within the Namib Desert and other non-desert soils would help to improve our current understanding of bacterial distribution and the importance of these isolates in community assemblies. Performing a 16S rRNA sequence from shotgun metagenomics study on the 16S rRNA gene of these isolates would probably provide inaccurate results however, due to the high similarity of the *Arthrobacter* and *Pseudarthrobacter* 16S rRNA genes within the genera (>99.5% similarity). Targeting other proteins, such as the heat shock protein 70, elongation factor Tu or RecA (Venter et al. 2004), or a combination of these, may be more appropriate to determine the presence of B2, B4, E2 and E5 in other soils.

Analysis of the PhyloPhlAn tree produces a similar result to OrthoANI and *in silico* DDH. From Figure 5.25, it can be seen that isolate B2 is on the same clade as *A. agilis* 4041, while *A. pityocampae* Tp2 is in a different clade. This supports the observed OrthoANI and *in silico* DDH results for B2, which determined that B2 shared more similarity with *A. agilis* 4041 than *A. pityocampae* Tp2. Similar results are observed for isolates B4 and E5. OrthoANI and *in silico* DDH showed that isolates B4 and E5 share more similarity with each other than with the reference genomes. This is reflected by B4 and E5 appearing on the same clade (Figure 5.25). In Figure 5.24, both B4 and E5 are close to both *P. phenanthrenivorans* Sphe3 and *A. nitrophenolicus* SJCon. OrthoANI and *in silico* DDH results for B4 and E5 determined that both isolates shared more similarity with *P. phenanthrenivorans* Sphe3 than with *A. nitrophenolicus* SJCon. This inference is difficult to determine from the PhyloPhlAn tree alone, due to *A. nitrophenolicus* SJCon, B4, E5 and E2 forming a clade that is separate from the *Pseudarthrobacter* clade. Isolate E2 is close to *A. nitrophenolicus* SJCon in the PhyloPhlAn tree. This is supported by the OrthoANI and *in silico* DDH results which determined that E2 shares the most similarity with *A. nitrophenolicus* SJCon.

As previously discussed, the *Arthrobacter* genus was recently reclassified into six different genera. The PhyloPhlAn tree presented here (Figure 5.25) includes representatives of five of these genera, as well as other uncategorised *Arthrobacter* spp. available from the NCBI database. *Arthrobacter* sp. MYb213, MYb216, MYb222 and MYb214 form a monophyletic clade with *Glutamicibacter* (Figure 5.25) and have been renamed as *Glutamicibacter* sp. on the EzBioCloud server (Chun et al. 2007). Similarly, *Arthrobacter* sp. MYb227 and AQ5-05 form a monophyletic clade with *Paeniglutamicibacter*, and *Arthrobacter* sp. KI72, MYb51 and MYb23 form a monophyletic clade with *Paenarthrobacter*. These isolates have also been renamed to their new respective genera on the EzBioCloud server.

However, as noted by Nouioui et al. (2018), the revised *Arthrobacter* genera do not have high support values under the new constrained comprehensive trees, particularly in the *Paenarthrobacter* clade. This is evident in the case of *Arthrobacter cupressi*. Figure 5.25 shows that *A. cupressi* CGMCC 1.10783 clusters in the same clade as the new *Paenarthrobacter* genus. The PhyloPhlAn tree shows high support values for this relationship, yet *A. cupressi* CGMCC 1.10783 does not meet the *Paenarthrobacter* genus peptidoglycan criteria of A3α (Lys–Ala–Thr–Ala) A11.17 (Busse 2016). Instead, *A. cupressi* CGMCC 1.10783 meets the *Arthrobacter* genus criteria of A3α (Lys–Ala$_{2–3}$) A11.5 or A11.6 (Zhang et al. 2012a). This highlights that further classification of the *Paenarthrobacter* genus is required to create a more comprehensive cluster.

Arthrobacter sp. RC1.1 241, 135MFCol5.1, 9E14, AGF7.2, AQ5-06, 4R501 and HMWF013 have been reclassified as *Pseudarthrobacter* on the EzBioCloud server; a public analysis portal for bacterial taxonomy, genomics and metagenomics (Yoon et al. 2017). This is supported by the produced PhyloPhlAn tree in this thesis. *Arthrobacter* sp. B3 and B6 [not related to this study] have not been renamed to *Pseudarthrobacter;* however, despite appearing on the *Pseudarthrobacter* clade. As with *Paenarthrobacter*, several issues remain with the classification of isolates in the *Pseudarthrobacter* genus. One notable issue is *Arthrobacter enclensis* NIO-

1008. The PhyloPhlAn tree (Figure 5.25) shows that *A. enclensis* NIO-1008 is in the middle of the *Pseudarthrobacter* clade. The peptidoglycan for *Pseudarthrobacter* is A3α (Lys–Ser–Thr–Ala) A11.23 (Busse 2016), which is the peptidoglycan that *A. enclensis* NIO-1008 exhibits (Dastager et al. 2014). Therefore, this thesis recommends that *A. enclensis* NIO-1008 be renamed to *Pseudarthrobacter enclensis* NIO-1008 based on phylogenetic analysis and peptidoglycan structure.

Following the completion of this analysis and discussion in March 2019, *A. enclensis* was renamed to *P. enclensis* (Busse and Schumann 2019).

A second issue with the *Pseudarthrobacter* clade is *A. nitrophenolicus* SJCon. *A. nitrophenolicus* SJCon appears in the *Pseudarthrobacter* clade but meets the peptidoglycan criteria for the *Arthrobacter* classification (Arora and Jain 2013; Busse 2016). This raises further issues with the proposed identity of B4, E2 and E5. All three Namib isolates share a PhyloPhlAn clade (Figure 5.25) which is also shared with *A. nitrophenolicus* SJCon. Analysis of the peptidoglycan of B4, E2 and E5 is therefore required to complete their taxonomic classification. The major peptidoglycan structures will help to classify which genera, *Arthrobacter* or *Pseudarthrobacter*, B4, E2 and E5 truly belong to.

The issues discussed here with the *Pseudarthrobacter* genus classification reflect a need for further genome sequences to be added to public databases. The availability of additional sequences will help to further analyse and classify these phylogenetically complex genera.

Based on the combination of phenotypic (biochemical tests) and phylogenetic analysis (Mauve alignments, OrthoANI, *in silico* DDH and PhyloPhlAn) performed in this Chapter, it appears that all four Namib isolates are genetically distinct from the available reference genomes. Further analyses including chemotaxonomic classification of peptidoglycan, polar lipids and quinones will need to be performed before the Namib isolates can officially be confirmed as new species.

This Chapter has demonstrated that the draft genomes of the Namib isolates are within the expected parameters as set by other members of the *Arthrobacter* and *Pseudarthrobacter* genera. The metabolic pathways identified within these isolates are expected amongst organisms isolated from soil. While these taxonomic techniques have demonstrated a comprehensive analysis of the *Arthrobacter* and *Pseudarthrobacter* genera, they do not provide information regarding whether the differences in observed phenotypes (UV resistance) could be attributed to genomic sequence differences. Further discussion in Chapter 6 will address the need for a transcriptomics study to identify which genes are being upregulated in response to UV exposure.

Chapter 6: Genomic insights into DNA repair mechanisms in *Arthrobacter* and *Pseudarthrobacter*

In desert systems, bacteria face many stressors, including temperature fluctuations, low water availability and UV radiation exposure. Despite this, bacteria can colonise desert soil using a variety of survival mechanisms. DNA repair mechanisms such as base excision repair (BER), nucleotide excision repair (NER), recombinational repair and oxidative stress response can help these organisms survive in the harsh desert environment. Proteins and their derivatives that are involved in osmoprotection, such as glycine betaine and trimethylglycine, sigma factors and oxidative stress proteins have been identified in the Namib isolates. As the focus of this thesis is on general DNA repair and UV-specific repair genes, genes involved in osmoprotection, sigma factors and oxidative stress responses are not reported in this section. Continuing the investigations into how the metabolically diverse *Arthrobacter* genus has adapted to survive UV, or has become sensitive to UV, will help to improve our understanding how an increased level of UV radiation reaching Earth's surface due to climate change can have an impact on soil microbiota.

Arthrobacter spp. have previously been isolated from other stressful environments, such as subsurface waters, arctic ice, radioactive environments and chemically contaminated sites (Dsouza et al. 2015; Fong et al. 2001; Fredrickson et al. 2004; Hirsch 1986; Mongodin et al. 2006). The molecular mechanisms by which *Arthrobacter* spp. survive stressful conditions such as UV has only been investigated on a limited scale; however, more recently this has become a research focus for several studies (Ii et al. 2019; Kumar et al. 2016; Rasuk et al. 2017).

There is a need to determine what factors influence the UV resistance and UV sensitivity of closely related species. Factors that influence protein function may include protein structure and internal gene regulation. Differences in protein structure, particularly in the active sites of enzymes, can influence protein binding activity. The *uvrA*, *uvrB* and *uvrC* genes belong to the NER pathway and mutations in these genes often result in a UVC sensitive phenotype (Crowley et al. 2006; Goosen and Moolenaar 2008; Hanada et al. 2000; Stracy et al. 2016; Van Houten et al. 2005). The protein structure of the *uvrA*, *uvrB* and *uvrC* genes may play an important role in the UV sensitivity of E2 and E5. To investigate the possibility of point mutations in these NER genes, the secondary and tertiary protein structure of the UvrA, UvrB and UvrC proteins in the UVC resistant B2 and B4 will be compared with the UVC sensitive E2 and E5.

This chapter focuses on Aim 4: Investigation and comparative genomics of UV repair genes in *Arthrobacter* and *Pseudarthrobacter*. To achieve this Aim, two objectives were outlined: (1) identify gene elements associate with UV resistance; (2) comparative analysis of predicted tertiary protein structures encoded by elements identified in Objective 4.1.

This chapter describes the comparison of the genomes of B2, E2, B4 and E5 against each other and publicly available genomes of the *Arthrobacter* and *Pseudarthrobacter* genera, with specific interest in their general DNA repair and UV-specific repair genes. One objective was to identify orthologous clusters shared by the UVC resistant isolates B2 and B4. Protein clusters shared between B2 and B4 would be candidates for genomic regions encoding UVC resistance. The presence or absence of DNA and UV repair genes in the Namib isolates was investigated and compared with the presence or absence of these genes within other stains of *Arthrobacter* and *Pseudarthrobacter*. The secondary and tertiary structure of the UvrA, UvrB and UvrC proteins was compared between the Namib isolates B2, B4, E2 and E5.

6.2 Methods

6.2.1 Comparative genome analysis using OrthoVenn

OrthoVenn was used to investigate homologous proteins shared by and unique to the Namib isolates. Prokka generated .faa files of the Namib isolates (Section 5.2.4, Chapter 5) were converted to FASTA format and uploaded to the OrthoVenn website (http://www.bioinfogenome.net/OrthoVenn/), along with two reference genomes; *A. agilis* 4041 (NZ_NFSC00000000.1) and *P. phenanthrenivorans* Sphe3 (NC_015145.1) (Table 5.2, Section 5.2.3). The E-value setting was the default of 1e-2 and the inflation value was 1.5.

Investigation of gene neighbourhoods for genes of interest was carried out using the Mauve plugin v1.1.1 for Geneious™ v2019.0 for Geneious version 2019.0.4 (https://www.geneious.com/). Images of gene regions were created in the R software environment (R Core Team 2017) using package gggenes (Wilkins 2019) and visualised in RStudio (v 1.1.463).

6.2.2 Presence of DNA repair mechanisms on the Namib isolates and reference genomes

To determine the presence of DNA repair mechanisms on each of the Namib isolates, the generated Prokka and RAST files from Chapter 5 were searched for known DNA repair genes (Appendix 5). The presence of DNA repair genes within the reference genomes (Table 5.2, Section 5.2.3) was determined by downloading the RefSeq files for each genome and searching the files. The Prokka genome files were searched using annotated gene calls and BLASTP for specific genes related to DNA repair, which were compiled from literature and included BER, NER and recombinational repair systems (Kowalczykowski et al. 1994; Krwawicz et al. 2007; Kurth et al. 2015; Martins-Pinheiro et al. 2007). DNA repair genes searched in this thesis, and the locus tags of the query genes can be found in Appendix 5.

Each gene was grouped into a repair category: BER, NER, recombinational repair and 'other' repair systems. 'Other' repair systems included genes involved in photoreactivation, as these genes do not belong to any of the other repair categories.

Protein alignment was conducted using the MUSCLE software (Madeira et al. 2019) and Maximum-likelihood phylogenetic trees were created using the FastTree plugin with 1,000 bootstraps using Geneious version 2019.0.4 (https://www.geneious.com/). EggNOG annotation files obtained in Chapter 5 were used for protein searches.

The presence and absence of DNA repair genes was visualised using a heatmap. The gene copy number of each gene was counted. A heatmap was generated by using the heatmap.2 function in the R gplots package v. 3.5.2 (Warnes et al. 2015) with R software environment (R Core Team 2017) and visualised in RStudio (v 1.1.463).

6.2.3 Primary and secondary protein structure analysis

Primary protein structures were obtained from the generated Prokka files (Section 5.2.2) for the Namib isolates and the RefSeq files for the reference genomes. Secondary protein structures were predicted using Geneious version 2019.0.4 by Biomatters. Sequence alignments of the Namib isolates were analysed to identify secondary structures that were unique between the isolates.

6.2.4 Tertiary structure prediction and analysis

To visualise any noticeable differences between the UVC resistant B2 and B4 isolates and the UVC sensitive E2 and E5 protein structures, the primary protein sequence to the Phyre2 V2.0 website (http://www.sbg.bio.ic.ac.uk/) (Kelley et al. 2015) for tertiary protein structure prediction. The 3D structures were then downloaded and imported into Geneious version 2019.0.4 for visual analysis and to colour-code the known active sites for each of the UvrA, UvrB and UvrC proteins. The protein active sites were estimated from the literature based on previous studies with *Thermotoga maritima*, *Bacillus* sp. and *D. radiodurans* (Goosen and Moolenaar 2008; Jaciuk et al. 2011; Van Houten et al. 2005).

6.3 Results and Discussion

6.3.1 Comparison of *Arthrobacter* and *Pseudarthrobacter* isolates

OrthoVenn was used to identify genomic differences between the newly sequenced Namib desert isolates and other *Arthrobacter*, *Pseudarthrobacter* spp. OrthoVenn focuses on comparative genomics and allows the user to identify orthologous protein clusters that are unique to the individual organisms, as well the clusters shared by all query searches.

OrthoVenn was used to identify regions within B2 and B4 that might provide insight regarding the mechanisms of UV resistance observed in these genomes. OrthoVenn illustrates this relationship by assigning numbers to the overlapping regions where two genomes share an orthologous gene cluster. An assumption of this analysis is that isolates B2 and B4 are repairing UV induced DNA damage using the same repair pathway. If this assumption is accurate, the isolates B2 and B4 should share an orthologous protein cluster that is absent in isolates E2 and E5. Reference genomes were also included in this analysis in an attempt to narrow down the pool of shared isolates between B2 and B4. The reference genomes chosen were *A. agilis* 4041 and *P. phenanthrenivorans* Sphe3. As found in Chapter 5, E2 shares has a high OrthoANI percentage similarity with *A. nitrophenolicus* SJCon. However, due to the limitation of OthroVenn accepting only six query sequences for each analysis, *P. phenanthrenivorans* Sphe3 was selected as one of the references due to the high OrthoANI percentage similarity with B4, E5 and E2. The UV resistance of *A. agilis* 4041 and *P. phenanthrenivorans* Sphe3 are unknown, however, neither isolate has been reported to be UV resistant or sensitive. For this analysis, the assumption was that the reference strains were UV sensitive, and that isolates B2 and B4 have gained UV resistance, as opposed to the reference strains and E2 and E5 losing this phenotype. Under this assumption, isolates B2 and B4 would not have possessed the genes for UV resistance through common ancestry, but rather by another evolutionary event such as horizontal gene transfer.

It is important to note that evolutionary events such as gene deletion, insertion and duplication are difficult to observe through OrthoVenn analysis. However, this analysis provides an initial insight into the shared orthologous protein clusters of B2 and B4 that may confer UV resistance that are absent in isolates E2 and E5. Figure 6.1 shows the comparison of B2 (pink), E2 (brown), B4 (yellow) and E5 (orange) against the reference genomes of *A. agilis* 4041 (green) and *P. phenanthrenivorans* Sphe3 (blue).

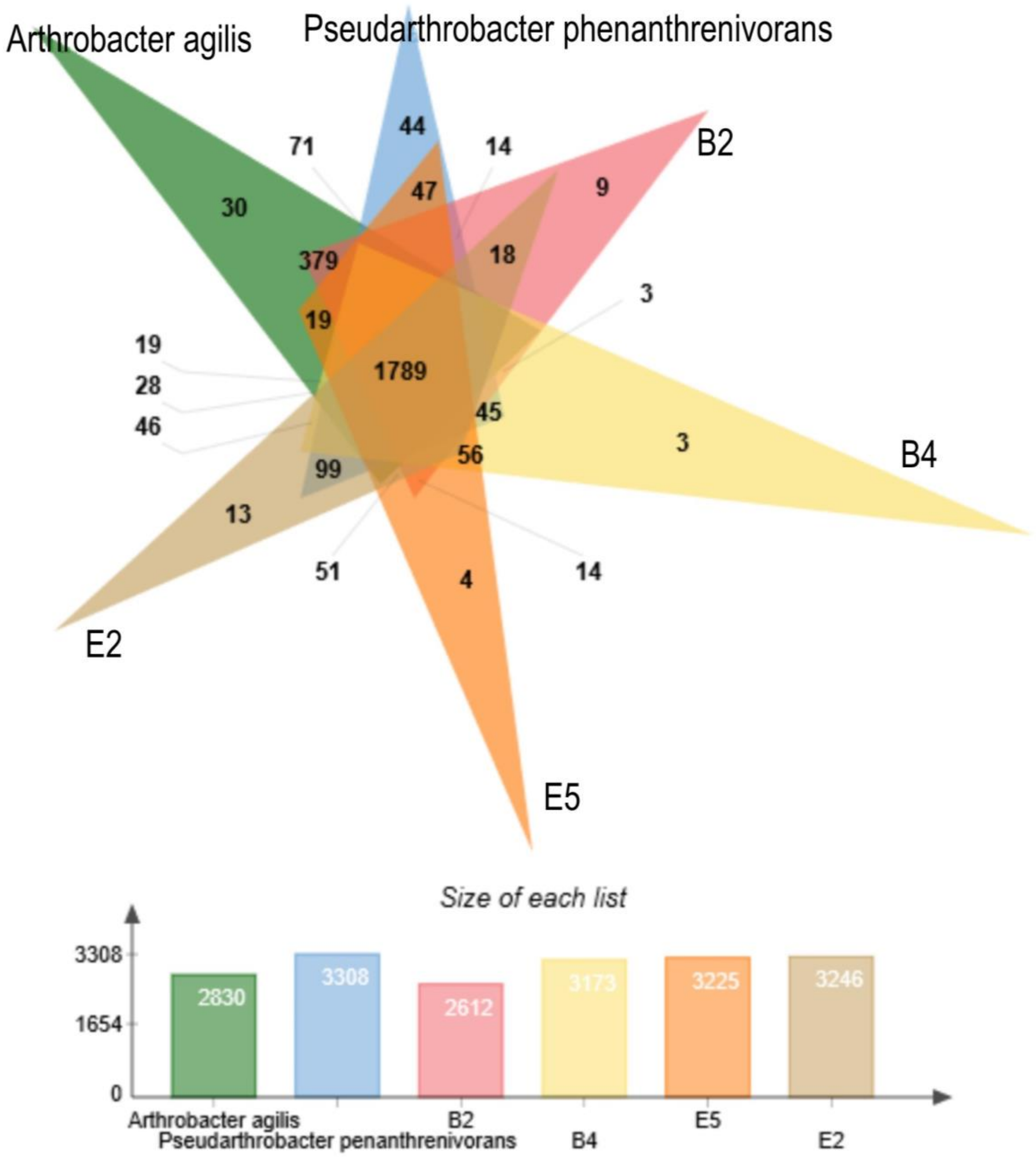


Figure 6.1: Generated OrthoVenn diagram comparing isolates B2, E2, B4 and E5 compared to each other, and to *A. agilis* DSM 4041 *and P. phenanthrenivorans* Sphe3 as reference genomes. Numbers in overlapping Venn regions indicate shared homologous protein clusters between the query sequences. The size of each list is the number of protein clusters within each uploaded genome.

As seen in Figure 6.1, all query genomes share 1,789 orthologous protein clusters. There are seven clusters that are shared by all four Namib isolates that are not shared by either of the reference genomes of *A. agilis* 4041 or *P. phenanthrenivorans* Sphe3. The UVC resistant isolates B2 and B4 have nine and three unique orthologous clusters, while the UVC sensitive isolates, E2 and E5 have 13 and four unique clusters, respectively. There are three clusters that are shared by the UVC resistant isolates B2 and B4 that are not shared by either of the reference genomes. This has been highlighted below in Figure 6.2.

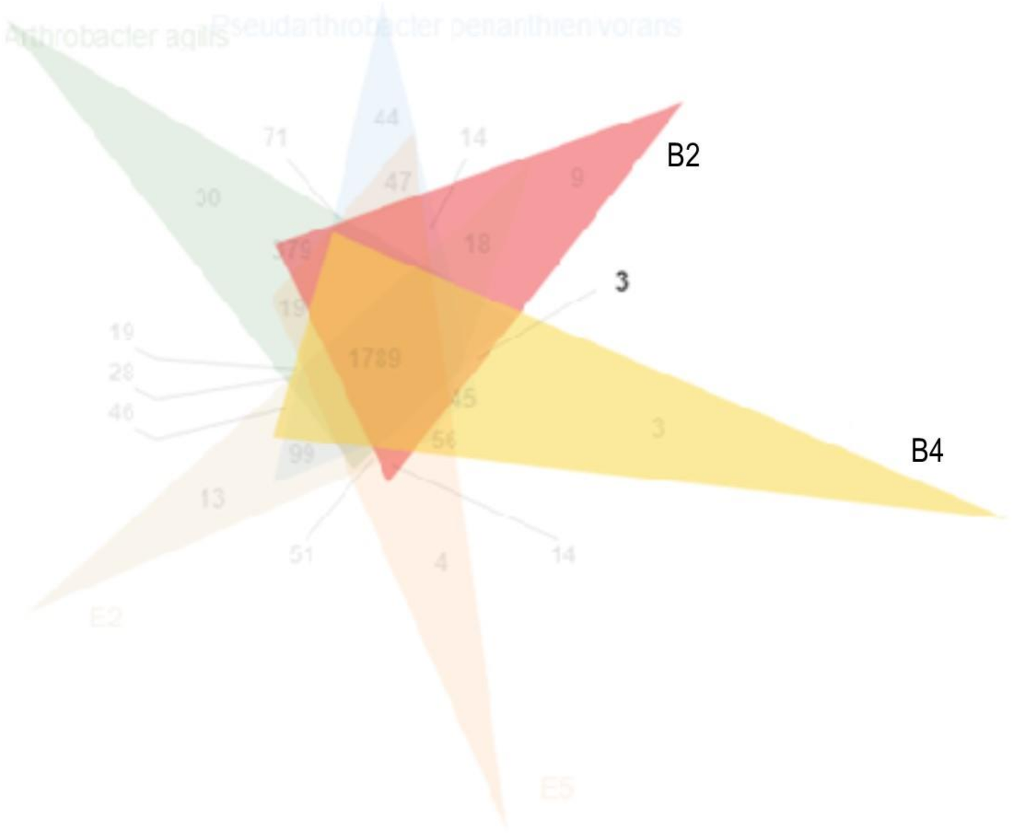


Figure 6.2: Generated OrthoVenn diagram highlighting the cluster of orthologous proteins shared by isolates B2 and B4.

The gene clusters identified in Figure 6.1 that are shared to isolates B2 and B4 were of interest due to the UVC resistant nature of these two isolates. The orthologous proteins shared between these two isolates was investigated for specific UV related repair genes that may have been absent in isolates E2 and E5. When compared to each of the reference genes, OrthoVenn identified 6 gene clusters shared between *A. agilis* 4041, B2 and B4, 8 gene clusters shared between *A. agilis* 4041, *P. phenanthrenivorans* Sphe3, B2 and B4, and 0 gene clusters shared between *P. phenanthrenivorans* Sphe3, B2 and B4 (Figure 6.1). The gene clusters shared only by UVC resistant isolates B2 and B4 can be seen below in Table 6.1. The identification of these proteins was based on their Prokka annotation files.

Table 6.1: Identification of gene clusters found only in UVC resistant isolates B2 and B4. The KFNIHMNI tag is for B2 and the HLFAFCBI tag is for B4. BLASTP was carried out using the B2 protein sequence as the query.

Prokka annotation			**Predicted role/ function**	**InterPro**		**NCBI BLASTP**					
Query locus tag	Gene	Gene description		**InterPro number**	**InterPro description**	**Best homologue:**	**Description**	**% ID**	**% query cover**	**E-value**	**Subject accession number**
KFNIHMNI_01031 HLFAFCBI_00939		Hypothetical protein		IPR022121	Peptidase M73 camelysin. Detailed signature match only	*Arthrobacter* sp. B1805	Hypothetical protein	100%	88%	1e-143	WP_104116766.1
KFNIHMNI_00274 HLFAFCBI_00316		Hypothetical protein		IPR017517 IPR024344	Conserved hypothetical protein, actinobacterial-type and Mycothiol-dependent maleylpyruvate isomerase	*Arthrobacter* sp. B1805	Maleylpyruvate isomerase family mycothiol-dependent enzyme	94%	100%	2e-137	WP_104117293.1
KFNIHMNI_00004 HLFAFCBI_01049	ALDH7B4	Aldehyde dehydrogenase family 2-member B4, mitochondrial	Arginine biosynthesis and glycerolipid metabolism	IPR015590	Aldehyde dehydrogenase domain	*Arthrobacter pityocampae* Tp2	Aldehyde dehydrogenase family protein	100%	88%	2e-46	WP_104121451.1

The three shared orthologous protein clusters between B2 and B4 were selected for further investigation to see if the shared proteins may indicate specific UV related repair genes that may have been absent in isolates E2 and E5. As the genomes for the Namib isolates are in draft form, it is unknown if additional shared orthologues are missed by this analysis.

When using different reference genomes, OrthoVenn detected that the gene cluster containing KFNIHMNI_00274 and HLFAFCBI_00316, a hypothetical protein, was also found to be orthologous with *A. alpinus, A. castilli, A. koreensis, A. psychrolactophilus, A. luteolus, Paeniglutamicibacter gangotriensis, Paenarthrobacter nicotinovorans, Paenarthrobacter aurescens, Glutamicibacter mysorens* and *Glutamicibacter arilaitensis*. Additionally, the orthologous cluster KFNIHMNI_01031 and HLFAFCBI_00939, which also contained a hypothetical protein, clustered with *Paenarthrobacter aurescens, A. oryzae*, *A. pityocampae* and *P. chlorophenolicus.*

Isolates B2 and B4 did not share ‘Aldehyde dehydrogenase family 2-member B4, mitochondrial’ from cluster KFNIHMNI_00004 and HLFAFCBI_01049 with any other *Arthrobacter* reference genome used in this study (data not shown). However, the predicted function of this gene is arginine biosynthesis and glycerolipid metabolism, making it unlikely that this gene is involved in UV resistance.

It does not appear that any of the orthologous protein clusters shared by B2 and B4 can infer UV resistance. It was previously assumed that B2 and B4 would not have possessed the genes for UV resistance through common ancestry. However, the absence of any UV specific repair genes on B2 or B4 that is not also shared by E2, E5 or the reference genomes implies that this assumption could be incorrect. UV resistance has not previously been reported in either *A. agilis* 4041 or *P. phenanthrenivorans* Sphe3. However, the purpose of using OrthoVenn was to identify genes that may be present in the UVC resistant isolates that are absent in the UVC sensitive isolates. To further investigate the presence of UV related repair genes within the UVC resistant isolates that were absent in the UVC sensitive isolates, gene clusters of B2 and B4 were compared with *A. agilis* 4041 and *P. phenanthrenivorans* Sphe3. The gene clusters shared by *A. agilis* 4041, *P. phenanthrenivorans* Sphe3, B2 and B4 can be seen below in Table 6.2.

Table 6.2: Gene clusters shared by *A. agilis* 4041, *P. phenanthrenivorans* Sphe3, B2 and B4. The KFNIHMNI tag is for B2 and the HLFAFCBI tag is for B4. *A. agilis* 4041 and *P. phenanthrenivorans* Sphe3 have been identified alongside their RefSeq locus tag. BLASTP was carried out using the B2 protein sequence as the query.

Prokka annotation			Predicted role/ function	InterPro		BLASTP					
Locus tag	**Gene**	**Gene description**		**InterPro number**	**InterPro description**	**Best homologue:**	**Description**	**% ID**	**% query cover**	**E-value**	**Accession number**
KFNIHMNI_01456 HLFAFCBI_01913 WP_013600217.1 (PphSphe3) WP_087075002.1 (Aag4041)	*ywrD*	Glutathione hydrolase-like YwrD proenzyme	Glutathione metabolism	IPR029055	Nucleophile aminohydrolases, N-terminal	*Arthrobacter* sp. B1805	Transferase	99%	100%	0	WP_104118633.1
KFNIHMNI_01136 HLFAFCBI_03374 WP_013600183.1 (PphSphe3) WP_087071749.1 (Aag4041)	*lsrC*	Autoinducer 2 import system permease protein LsrC	Translocation of substrates across the membrane	IPR001851	ABC transporter permease	*Arthrobacter* sp. B1805	ABC transporter permease	97%	100%	0	WP_104116094.1
KFNIHMNI_02248 HLFAFCBI_00892 WP_013599632.1 (PphSphe3) WP_087072669.1 (Aag4041)		Quinone oxidoreductase 2	Utilised as an electron donor	IPR036291	NAD (P)- binding domain family	*Arthrobacter* sp. B1805	SDR family oxidoreductase	97%	100%	0	WP_104117558.1
KFNIHMNI_02336 HLFAFCBI_03337 WP_013602617.1 (PphSphe3) WP_087072817.1 (Aag4041)		Hypothetical protein		IPR009597	Domain of unknown function DUF1206	*Arthrobacter* sp. B1805	DUF1206 domain-containing protein	99%	98%	0	WP_104117751.1
KFNIHMNI_02341 HLFAFCBI_02847 WP_013599289.1	*ywfD*	Dihydroanticapsin 7-dehydrogenase	Biosynthesis of antibiotics	IPR036291	NAD (P)- binding domain family	*Arthrobacter* sp. B1805	SDR family oxidoreductase	99%	97%	9e-175	WP_104117752.1

(PphSphe3) WP_087073949.1 (Aag4041)				IPR020904	Short-chain dehydrogenase reductase						
KFNIHMNI_02276 HLFAFCBI_03453 WP_013599656.1 (PphSphe3) WP_087071548.1 (Aag4041)		Hypothetical protein		IPR000064	Endopeptidase, NLPC/P60 domain	*Arthrobacter* sp. B1805	NIpC/P60 family protein	97%	61%	3e-168	WP_104117577.1
KFNIHMNI_02365 HLFAFCBI_01702 WP_013599769.1 (PphSphe3) WP_087072861.1 (Aag4041)		Hypothetical protein		IPR039419	TetR transcription regulator	*Arthrobacter* sp. B1805	TetR/AcrR family transcriptional regulator	99%	100%	1e-156	WP_104117658.1
KFNIHMNI_02774 HLFAFCBI_02800 WP_003792170.1 (PphSphe3) WP_003792170.1 (Aag4041)		Hypothetical protein		IPR013177	Domain of unknown function DUF1206	*Mobilicoccus pelagius*	Hypothetical protein	100%	100%	8e-11	GAB48542.1
KFNIHMNI_00718 HLFAFCBI_02249 WP_087075879.1 (Aag4041)		Hypothetical protein		IPR001128	Cytochrome P450	*Arthrobacter* sp. B1805	Cytochrome P450	98%	99%	0	WP_104118549.1
KFNIHMNI_00045 HLFAFCBI_02849 WP_087077007.1 (Aag4041)	*hutI*	Imidazolonepropionase	L-histidine degradation to N-formimidoyl-L-glutamate	IPR006680	Amidohydrolase-related domain	*Arthrobacter* sp. B1805	Amidohydrolase	99%	100%	0	WP_104118912.1
KFNIHMNI_01856 HLFAFCBI_01874 WP_087073461.1 (Aag4041)	*thiM*	Hydroxyethylthiazole kinase	Transferase	IPR000417	Hydroxyethylthiazole kinase	*Arthrobacter* sp. B1805	Hydroxyethylthiazole kinase	96%	96%	2e-166	WP_104117533.1

KFNIHMNI_01855 HLFAFCBI_01875 WP_087073459.1 (Aag4041)	*thiE*	Thiamine-phosphate synthase	Thiamine metabolism	IPR034291	Thiamine phosphate synthase	*Arthrobacter* sp. B1805	Thiamine phosphate synthase	97%	100%	6e-142	WP_104117386.1
KFNIHMNI_02921 HLFAFCBI_02147 WP_087072221.1 (Aag4041)	*hin*	DNA-invertase hin	Recombinase that initiates DNA cleavage and recombination	IPR036162 IPR006118	Resolvase-like N-terminal catalytic domain superfamily Recombinase	*Arthrobacter agilis*	Recombinase family protein	90%	100%	3e-120	WP_087072221.1
KFNIHMNI_01369 HLFAFCBI_02024 WP_087075529.1 (Aag4041)	*rppH*	RNA pyrophosphohydrolase	Regulates the 5'-end-dependent mRNA decay	IPR015797	NUDIX hydrolase-like domain superfamily	*Arthrobacter* sp. B1805	NUDIX domain-containing protein	98%	100%	1e-104	WP_104117883.1

Analysis of the orthologous proteins shared by B2, B4, *A. agilis* 4041 and *P. phenanthrenivorans* Sphe3 revealed two clusters of interest: cytochrome P450, found in B2, B4 and *A. agilis* 4041, and a recombinase family protein, found in B2, B4 and *A. agilis* 4041. These proteins are highlighted in red in Table 6.2 above. As mentioned, it is important to observe the orthologous protein clusters shared by B2 and B4, as well as by these isolates and other reference genomes, to observe any DNA or UV specific repair mechanisms that are present in these genomes that are absent from the UVC sensitive E2 and E5 genomes.

6.3.1.1 Recombinase family protein

The recombinase family protein *hin* was detected as an orthologous cluster between B2, B4 and *A. agilis* 4041. *Hin* is a site-specific DNA inversion recombinase, that has low frequency DNA rearrangements, but can be caused by a change in environmental conditions (Johnson 2015). *Hin* is found in genera such as *Salmonella* and it is responsible for inverting specific DNA regions, which results in a more efficient flagella gene expression in *Salmonella* (Dhar et al. 2004). Further investigation into the expression of this protein might also be of future research interest, as the *hin* protein may help B2 and B4 to more efficiently express genes for genome repair, when compared with E2 and E5. To the authors knowledge, *hin* has not previously been reported as a protein relating to UV resistance.

6.3.1.2 Cytochrome P450

OrthoVenn analysis of the UVC resistant isolates B2 and B4 showed that both B2, B4 and *A. agilis* 4041 shared the cytochrome P450 gene. Upon further investigation of the Genbank files for *P. phenanthrenivorans* Sphe3, a cytochrome P450 was found, but it shared low similarity with B2 and B4, and was therefore not identified as a shared cluster by the OrthoVenn software.

Cytochrome P450 (CYP) is a protein that contains heme as a cofactor. CYP proteins have a wide range of known functions, including acting as terminal acceptors in the electron transport chain, biotransformation of xenobiotic compounds in bacteria and vitamin D metabolism in humans (Groves 2005; Guengerich 2005). There are multiple family types of CYP; however, analysis using BLAST and InterPro was unable to determine the family type of the CYP of B2, B4 or *A. agilis* 4041. In addition, many CYPs still have unknown functions. In bacteria, CYP has been linked to the biotransformation of xenobiotic compounds, but in humans CYPs have been noted as proteins that assist in the removal of UV induced ROS (Luecke et al. 2010). Sang et al. (2012) notes that CYP genes were up regulated during UVA exposure in beetles. As UV induces ROS damage, this protein is of interest within the scope of this study. Further, the best homologue of the B2 CPY as per BLASTP analysis is the CYP gene of isolate *Arthrobacter* sp. B1805 (Table 6.2). This organism was isolated in the Tibetan Plateau, which has been noted as a restrictive environment due to the high levels of UV radiation that reach this location (Shen et al. 2014). The best homologue to the B4 CPY as per BLASTP analysis is *Arthrobacter* sp. IMM11 (WP_142028888.1, query cover 98%, percentage identity 70.54%), which was isolated from soil in Vilassar de Dalt, Barcelona (Caliz et al. 2011). Neither *Arthrobacter* sp. B1805 nor *Arthrobacter* sp. IMM11 have previously been reported as being UV resistant, indicating that further screening of the UV resistance of these isolates may help to further investigate the hypothesis that CPYs are involved in radiation resistance.

Further analysis of the CYP regions of B2, B4 and *A. agilis* 4041 was conducted to observe similarities and differences in surrounding gene arrangement between the genomes. The comparison of the gene regions of B2, B4 and *A. agilis* 4041 can be seen in Figure 6.3.

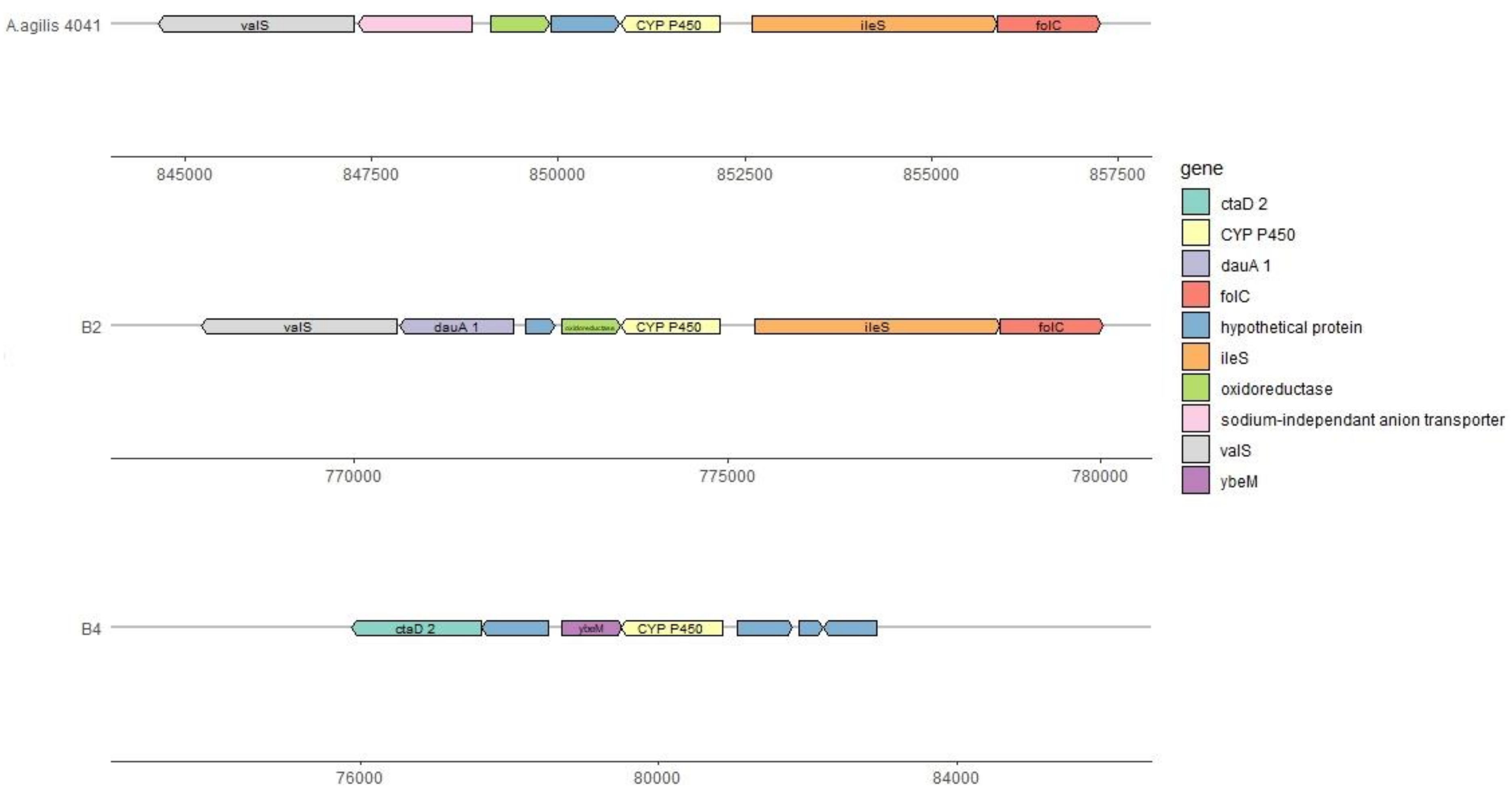


Figure 6.3: Organisation of cytochrome P450 region and surrounding genes of *A. agilis* 4041, B2 and B4. Arrows indicate coding sequences and direction of transcription. Images created in the R software environment (R Core Team 2017) package gggenes (Wilkins 2019) and visualised in RStudio (v 1.1.463).

Figure 6.3 shows that the surrounding genes in the CYP regions of *A. agilis* 4041 and B4 are different. The surrounding genes in the CYP regions of *A. agilis* 4041 and B2 are similar, with some slight differences in the gene neighbourhood between the two genomes. The CYP genes of *A. agilis* 4041, B2 and B4 were aligned with MUSCLE to observe protein similarity. The MUSCLE alignment of the shared CYP proteins showed 78.5% pairwise identity between *A. agilis* 4041 and B2, and 56.2% pairwise identity between *A. agilis* 4041 and B4.

A BLASTP analysis was carried out to identify the most similar sequences to the B2 and B4 CYP proteins against predicted proteins in the reference genomes. The most similar reference genome sequence to the B2 CYP was *A. pityocampae* Tp2 with 80.77% sequence identity with 99% query cover followed by *A. agilis* 4041 with 78.5% sequence identity with 99% query cover. The most similar reference genome sequence to the B4 CYP protein was *A. enclensis* NIO-1008 with 68.14% similarity with 98% query cover. The genome arrangement of the CYP and surrounding genes in isolates B2 and B4 against their similar reference sequence genomes was visualised using the R package gggenes (Wilkins 2019). The gene organisation of the CYP region for *A. agilis, A. pityocampae* and B2 can be seen in Figure 6.4. The organisation of the CYP gene regions for *A. enclensis* NIO-1008 and B4 can be seen in Figure 6.5.

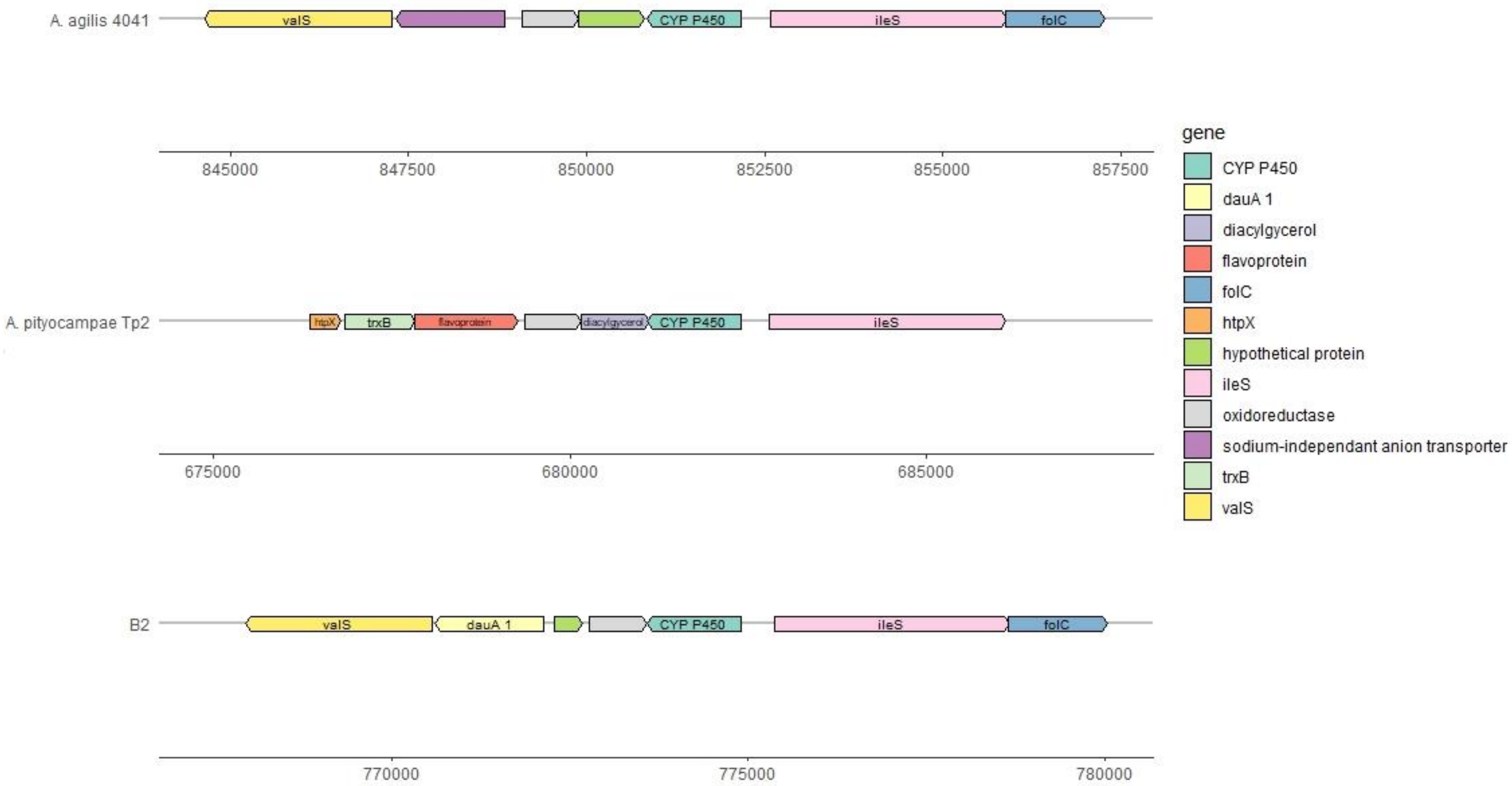


Figure 6.4: Organisation of cytochrome P450 region and surrounding genes of *A. agilis* 4041, *A. pityocampae* Tp2 and B2. Arrows indicate coding sequences and direction of transcription. Images created in the R software environment (R Core Team 2017) package gggenes (Wilkins 2019) and visualised in RStudio (v 1.1.463).

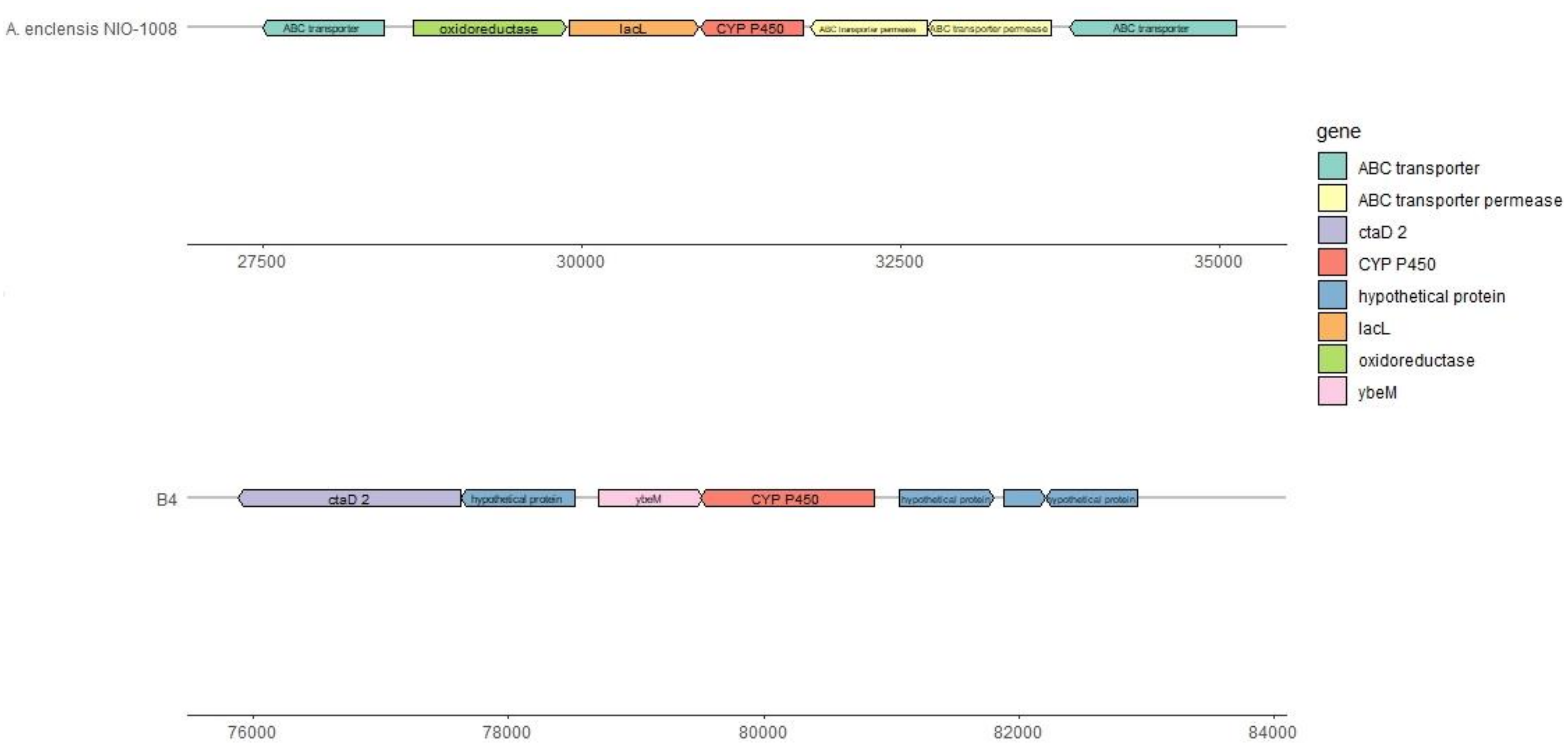


Figure 6.5: Organisation of cytochrome P450 region and surrounding genes of *A. enclensis* NIO-1008 and B4. Arrows indicate coding sequences and direction of transcription. Images created in the R software environment (R Core Team 2017) package gggenes (Wilkins 2019) and visualised in RStudio (v 1.1.463).

Figure 6.4 shows that the organisation of the gene neighbourhood around the CYP gene of B2 is more similar to *A. agilis* 4041 than to *A. pityocampae* Tp2. *A. pityocampae* has surrounding genes such as *htpX*, *trxB* and flavoprotein that are absent from this region in the B2 and *A. agilis* 4041 genomes. This indicates that some genome rearrangement has occurred in *A. agilis* 4041, and subsequently isolate B2 after divergence from *A. pityocampae* Tp2.

Figure 6.5 shows that the organisation of the gene neighbourhood around the CYP gene of B4 and *A. enclensis* NIO-1008 is very different between the two genomes. The CYP gene of *A. enclensis* NIO-1008 is surrounded by ABC transporter proteins, while the CYP gene of B4 is surrounded by several hypothetical proteins and *ybeM*. The differing gene neighbourhood between these two isolates indicates that there has been genome rearrangement in this region, possibly due to transposable elements. Alternatively, as the surrounding CYP gene regions of *A. enclensis* NIO-1008 and B4 differ, but the sequence similarity of these two genes is high, this may indicate the CYP gene of B4 was obtained via evolutionary events such as horizontal gene transfer. PhyloPhlAn analysis showed that *A. enclensis* NIO-1008 and B4 share a common ancestor, although B4 appears to be more closely related to *A. nitrophenolicus* SJCon based on conserved protein sequence (Figure 5.25). This would support the idea that isolate B4 has obtained or maintained this gene following divergence from the *A. enclensis* NIO-1008 common ancestor and has undergone genome rearrangements following this divergence.

As previously discussed, (Table 6.1), the genomes of E2 nor that of E5 appeared to have a cytochrome P450 gene. However, the above analysis assumes that the CYP gene of E2 and E5 would be in the same region, and that the gene would be called the same during annotation by RAST and Prokka. To confirm the presence or absence of CYP within the genome sequence of E2 and E5, a BLASTn analysis was done. When using a BLASTn search of the B2 and B4 CYP genes against the E2 and E5 genomes, no hits were returned. To further investigate if the E2 and E5 CYP gene was truly absent, the two CYP regions from B2 and B4 were aligned with the E2 and E5 genomes using Mauve. The Mauve analysis can be seen in Figure 6.6 for the comparison of B2, E2 and E5 and in Figure 6.8 for the comparison of B4, E2 and E5. The red arrow indicates the CYP region of B2 and B4. Figure 6.7 shows that the genomes of E2 and E5 are missing a gene in the B2 CYP region. Figure 6.8 shows that both E2 and E5 have a hypothetical protein in the B2 CYP region. When searched on NCBI BLASTP, the hypothetical protein of E2 and E5 did not return as a CYP protein. Figure 6.9 shows that the genomes of E2 and E5 are missing a gene in the B4 CYP region.

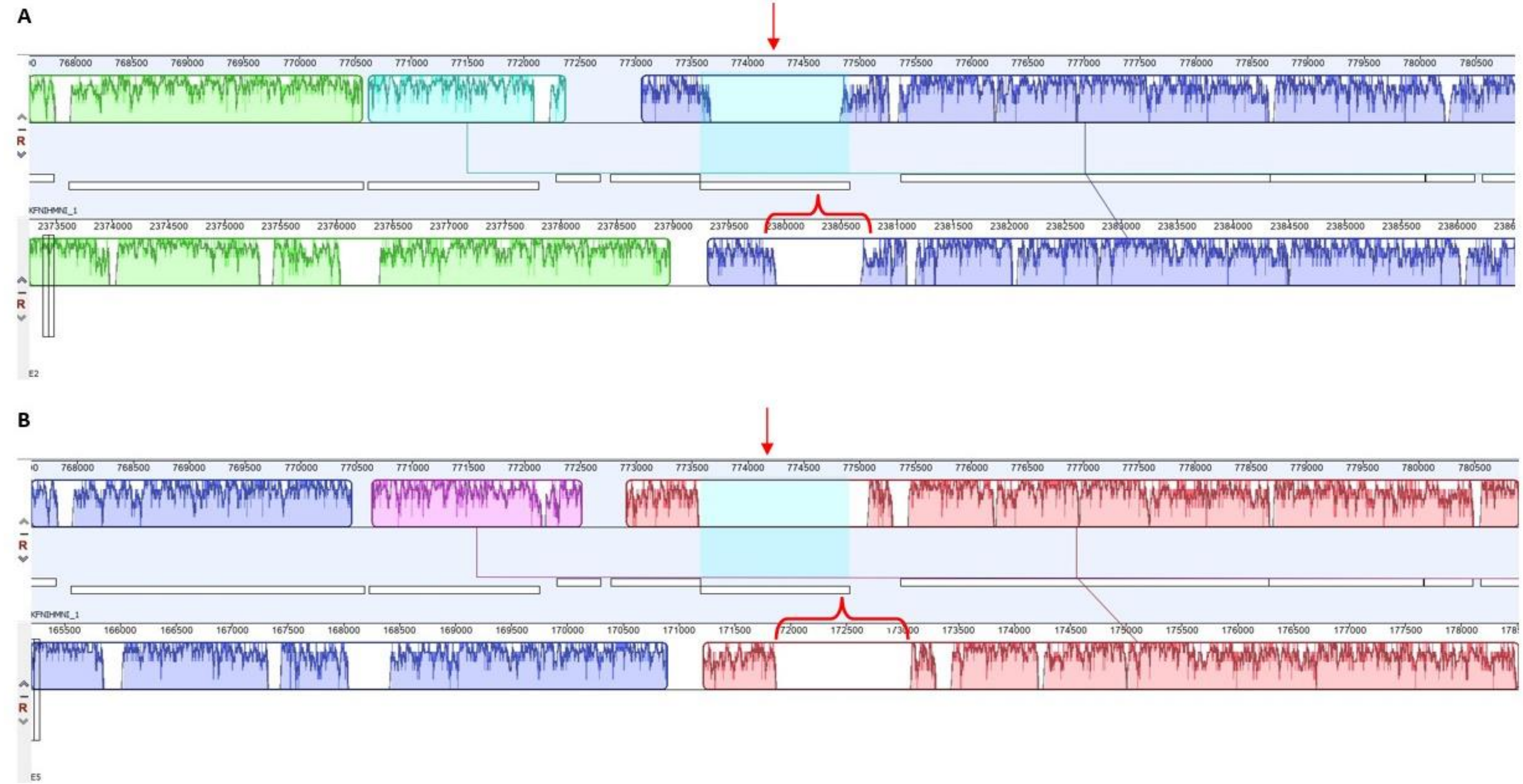


Figure 6.6a: Mauve alignment of B2 (top) and E2 (bottom) with the B2 cytochrome P450 location indicated by a red arrow (locus tag KFNIHMNI_00718). As seen in the alignment, the E2 region that corresponds to the B2 cytochrome P450 region is missing a gene in the reciprocal area. An absence of a colour block within a Mauve alignment indicates that the genes in that region are dissimilar, while genes surrounding the other areas are similar. **6.6b: Mauve alignment of B2 (top) and E5 (bottom) with the B2 cytochrome P450 location indicated by a red arrow. As seen in the alignment, the E5 region that corresponds to the B2 cytochrome P450 region is missing a reciprocal gene in the area.**

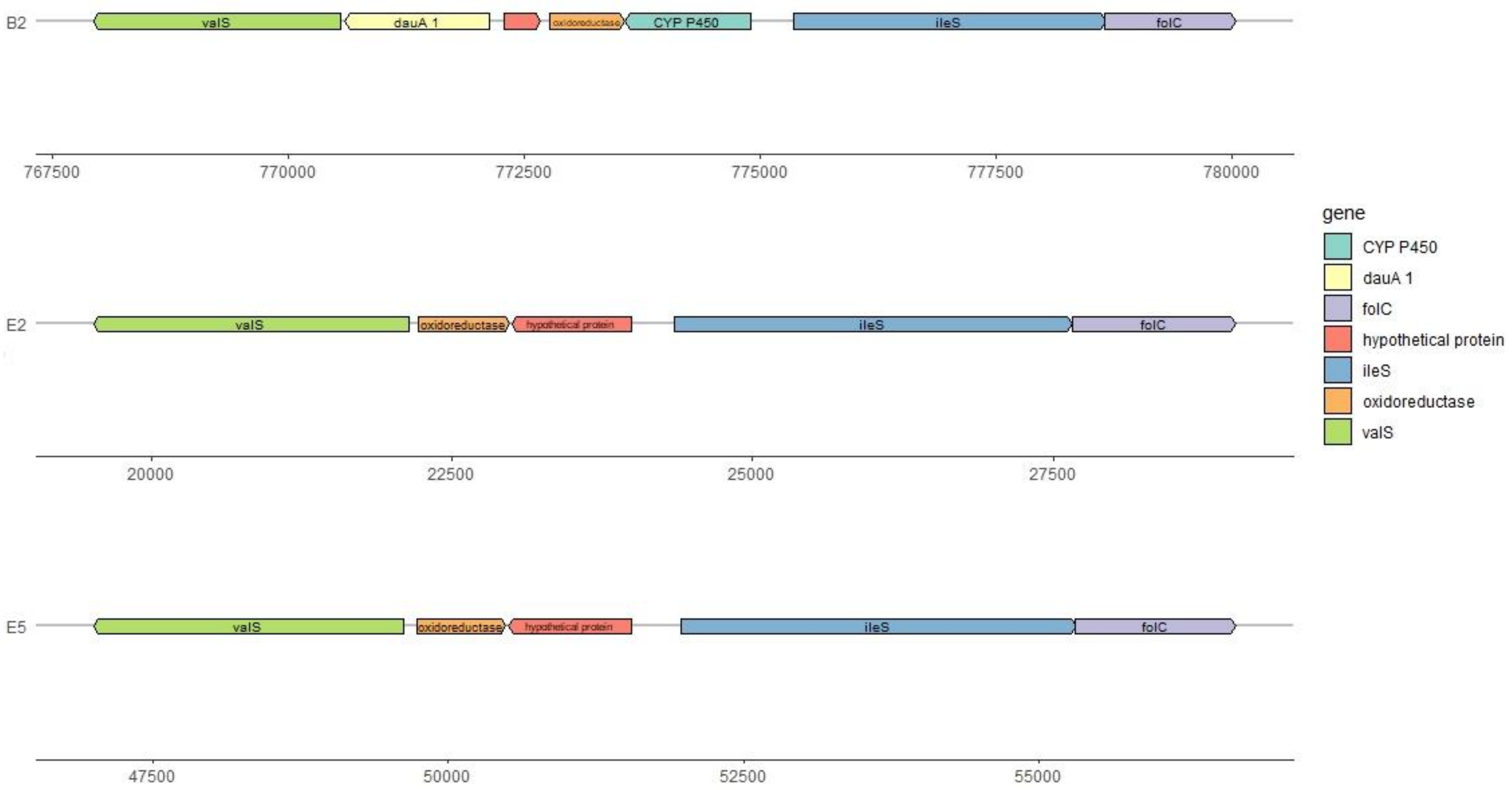


Figure 6.7: Gene region surrounding cytochrome P450 on isolate B2 (locus tag KFNIHMNI_00718, 6.8a) with the similar identified gene regions of E2 (locus tag FICEKGFG_00024, 6.8b) and E5 (locus tag GHIKLHBA_00463, 6.8c). A BLASTP analysis of the hypothetical proteins in the red boxes in 6.9b and 6.9c found that they are endonucleases and the best homologue for both was *A. nitrophenolicus* (WP_009372872.1) with a 73.49% similarity and 90% query cover for E2, and a 79.80% similarity with 88% query cover for E5. Arrows indicate coding sequences and direction of transcription. Images created in the R software environment (R Core Team 2017) package gggenes (Wilkins 2019) and visualised in RStudio (v 1.1.463).

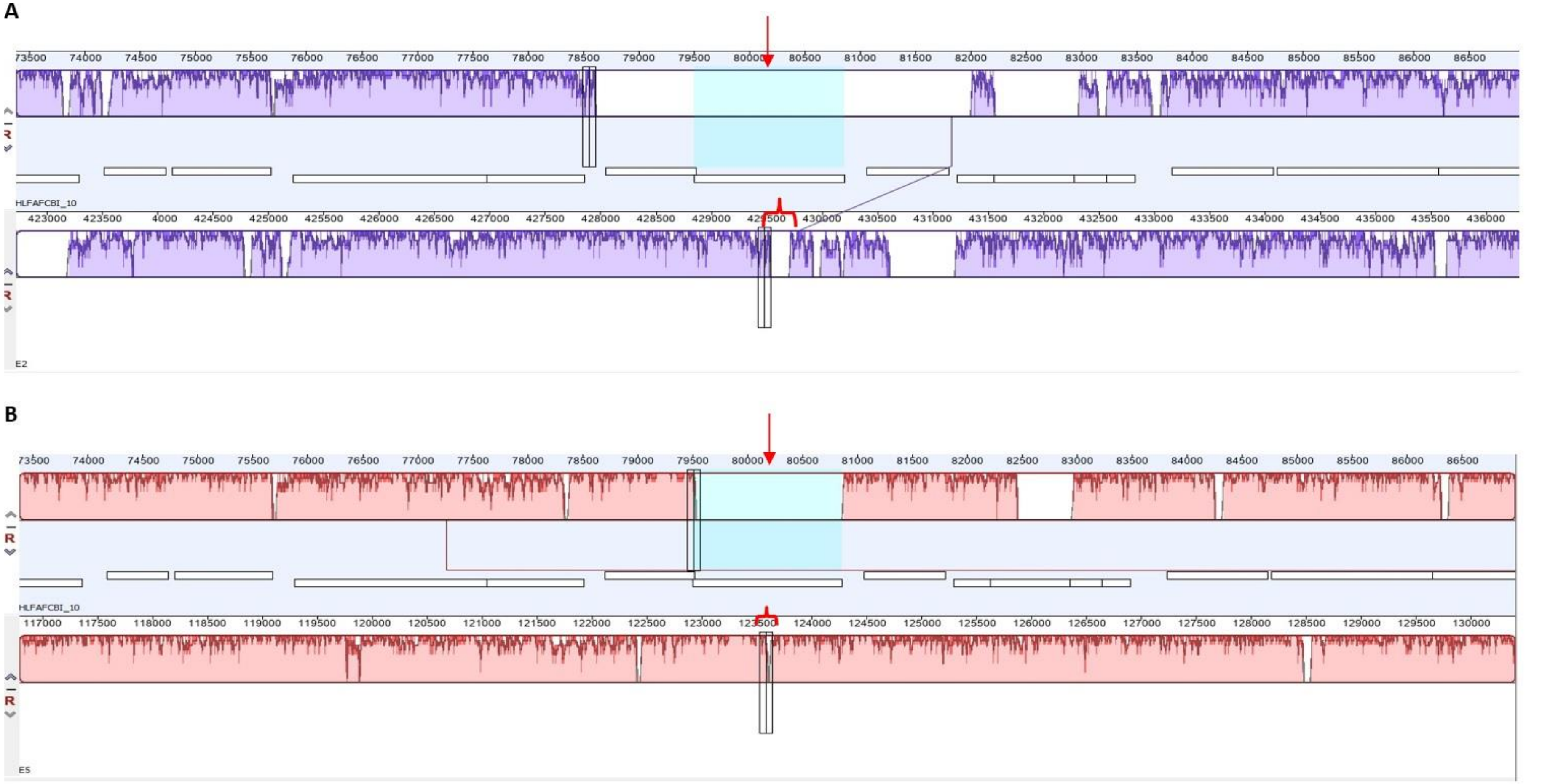


Figure 6.8a: Mauve alignment of B4 (top) and E2 (bottom) with the B4 cytochrome P450 location indicated by a red arrow (locus tag HLFAFCBI_02249). As seen in the alignment, the E2 region that corresponds to the B4 cytochrome P450 region is missing a corresponding gene in the alignment area. **6.8b: Mauve alignment of B4 (top) and E5 (bottom) with the B4 cytochrome P450 location indicated by a red arrow.** As seen in the alignment, the E5 region that corresponds to the B4 cytochrome P450 region has several similar genes, but there is a very small region in E5 that is missing the cytochrome P450 type gene in that area.

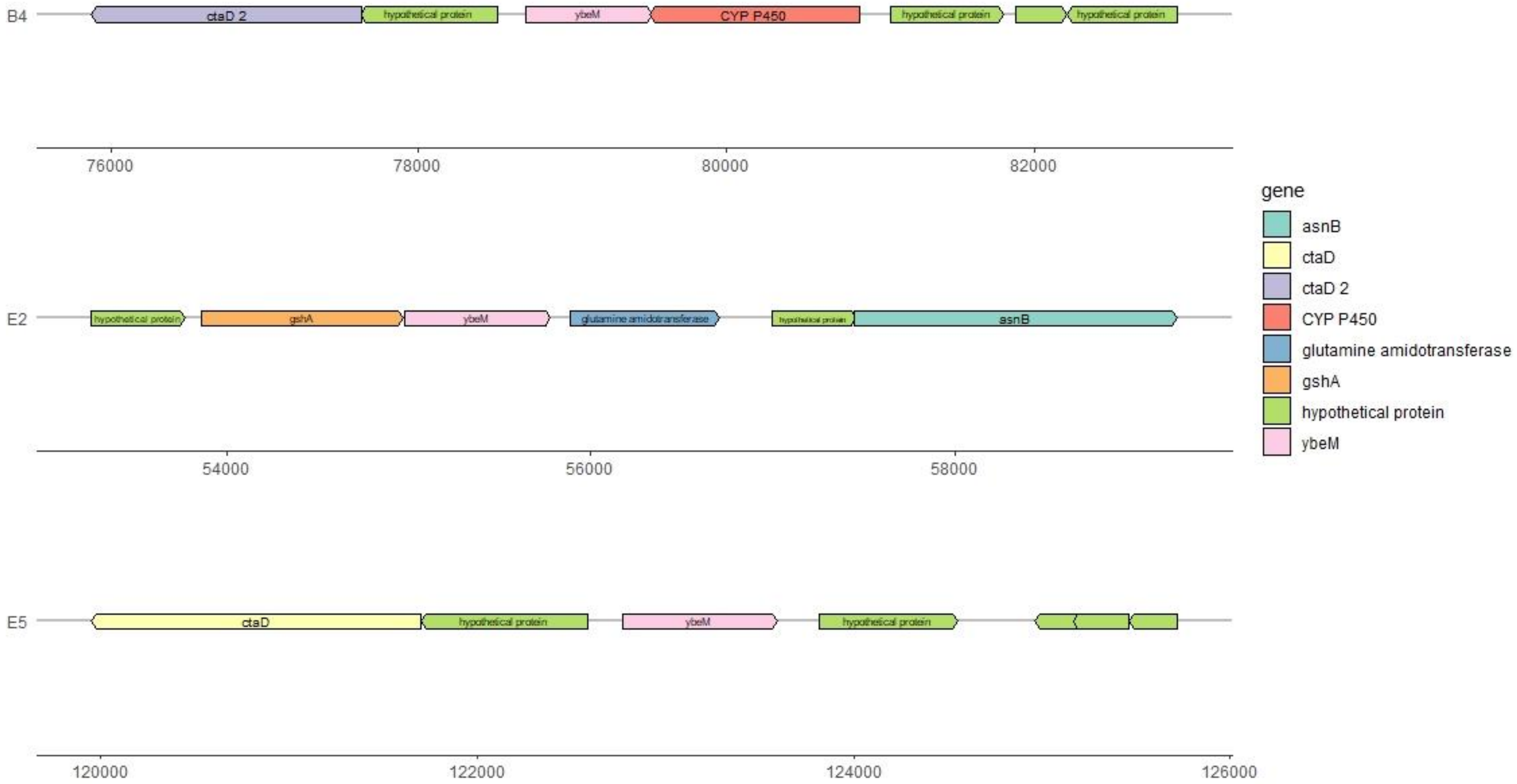


Figure 6.9: Gene region surrounding cytochrome P450 on isolate B4 (locus tag HLFAFCBI_02249, 6.10a) with the similar identified gene regions of E5 (locus tag for *ybeM* GHIKLHBA_02277, 6.10b) and E2 (locus tag for *ybeM* FICEKGFG_03043, 6.10c). As seen in the Figures above, the E5 gene organisation and region shared a large degree of gene organisation with B4, except for the absence of cytochrome P450. The gene region of E2 however, does not share a large degree of similarity with the B4 region. Arrows indicate coding sequences and direction of transcription. Images created in the R software environment (R Core Team 2017) package gggenes (Wilkins 2019) and visualised in RStudio (v 1.1.463).

As seen from Figure 6.7, both UVC sensitive isolates E2 and E5 have a hypothetical protein in the B2 CYP region. BLASTP analysis suggested that these are endonucleases, with a sequence from *A. nitrophenolicus* as the closes match. Endonucleases are enzymes that cleave a phosphodiester bond and are important in cleaving DNA for DNA repair, indicating that the hypothetical proteins in E2 and E5 may be used for DNA repair mechanisms following UV exposure.

Finally, the eggNOG annotation files of E2 and E5 were examined for CYP. The CYP protein was searched for by name and specific COG function. The eggNOG annotation files can be found in Appendix 6. E2 did not have any eggNOG annotation specific to cytochrome P450. However, E5 had two hits specific to cytochrome P450. These were located on locus tags GHIKLHBA_00506 (57 aa long) and GHIKLHBA_00507 (70 aa long). A BLASTP search with each of these proteins revealed that they both belong to the cytochrome P450 superfamily but share very low protein similarity with the CYPs found in B2 and B4 (3.9 – 6.8%). This suggests that the CYP proteins are unique to the UVC resistant B2 and B4. Therefore, if the homologous CYP observed in B2 and B4 is responsible for UV specific ROS repair, this may provide B2 and B4 an added advantage in surviving UV radiation. Further transcriptomic studies are required to confirm the up or down regulation of cytochrome P450 in B2 and B4 under UV exposure conditions. The alignment of the two E5 CYPs and the CYP of B2 and B4 can be found in Appendix 7.

Several UV resistant *Arthrobacter* such as *A. alpinus* ERGS4:06, *Arthrobacter* sp. MN05-02 and *Rubrobacter radiotolerans* [previously *A. radiotolerans*] all have a cytochrome P450 that has between 19.1 – 86.7% protein similarity with B2 and 15.4 – 54.8% protein similarity with B4. Some of these isolates have multiple copies of CYP genes, indicating that these genes may have different functions. The CYP family is very diverse, which is reflected in the protein alignment of the CYP in B2, B4, *A. alpinus* ERGS4:06 [two genes], *Arthrobacter* sp. MN05-02, *R. radiotolerans* [two genes] and *D. radiodurans* [three genes] seen in Figure 6.10. As seen in Figure 6.10 below, there are very few conserved regions of CYP proteins from known UV resistant organisms. CYPs were also identified in some of the reference *Arthrobacter* genomes used in this study (Figure 6.11).

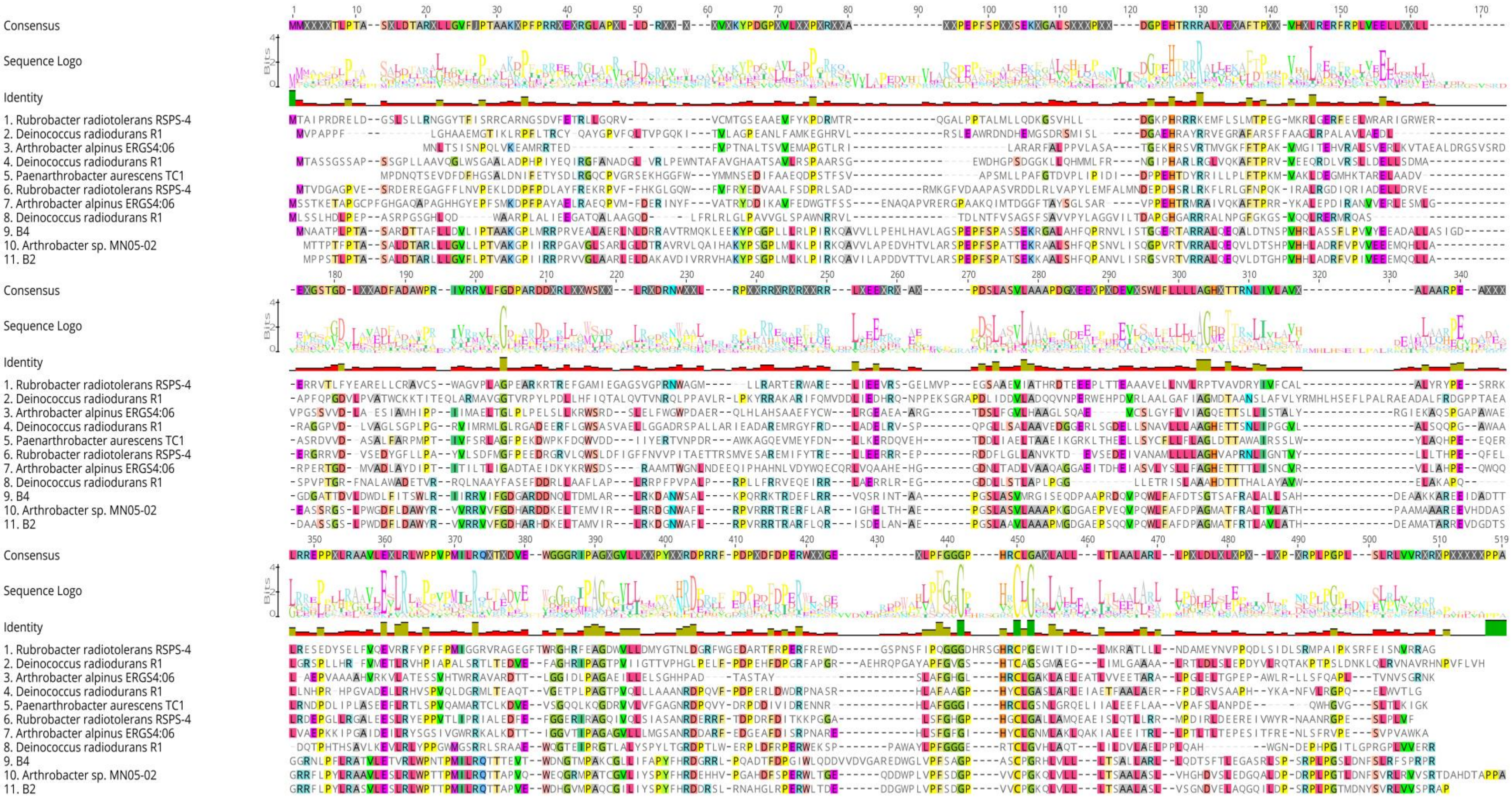

Consensus
Sequence Logo
Identity
Bits
1. Rubrobacter radiotolerans RSPS-4
2. Deinococcus radiodurans R1
3. Arthrobacter alpinus ERGS4:06
4. Deinococcus radiodurans R1
5. Paenarthrobacter aurescens TC1
6. Rubrobacter radiotolerans RSPS-4
7. Arthrobacter alpinus ERGS4:06
8. Deinococcus radiodurans R1
9. B4
10. Arthrobacter sp. MN05-02
11. B2

Figure 6.10: Protein alignment of cytochrome P450 from UV resistant isolates. The coded amino acids are shown above the alignment and the consensus sequence is shown above the amino acid sequence logo.

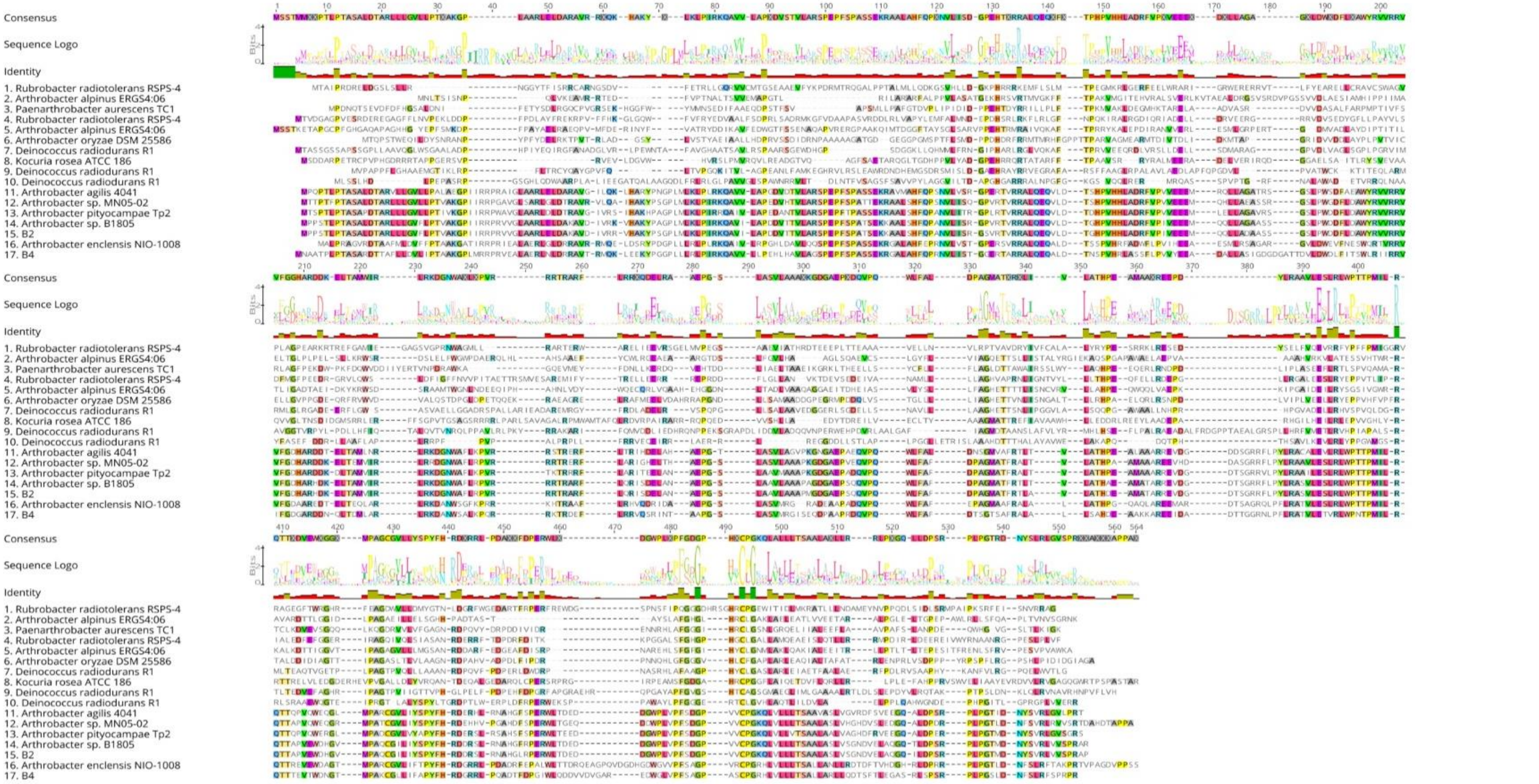

Consensus
Sequence Logo
Identity
1. Rubrobacter radiotolerans RSPS-4
2. Arthrobacter alpinus ERGS4:06
3. Paenarthrobacter aurescens TC1
4. Rubrobacter radiotolerans RSPS-4
5. Arthrobacter alpinus ERGS4:06
6. Arthrobacter oryzae DSM 25586
7. Deinococcus radiodurans R1
8. Kocuria rosea ATCC 186
9. Deinococcus radiodurans R1
10. Deinococcus radiodurans R1
11. Arthrobacter agilis 4041
12. Arthrobacter sp. MN05-02
13. Arthrobacter pityocampae Tp2
14. Arthrobacter sp. B1805
15. B2
16. Arthrobacter enclensis NIO-1008
17. B4
Bits
564

Figure 6.11: Protein alignment of CYP from UV resistant isolates and Arthrobacter reference genomes used in this study. The coded amino acids are shown above the alignment and the consensus sequence is shown above the amino acid sequence logo.

Figure 6.11 shows that there are only a few conserved regions in the CYPs identified in the reference *Arthrobacter* genomes and the known UV resistant isolates. As mentioned, CYPs have multiple roles in both animals and microorganisms (Groves 2005; Guengerich 2005). While there are very few conserved regions overall, the CYP structure of *A. agilis* 4041, *Arthrobacter* sp. MN05-02, *A. pityocampae* Tp2, *Arthrobacter* sp. B1805, *A. enclensis* NIO-1008, B2 and B4 appears to be highly conserved.

To confirm the amino acid (aa) sequence similarity seen in Figure 6.11 is due to common ancestry, a Maximum Likelihood (ML) analysis was carried out with the aa sequences. Figure 6.12 shows that the CYP proteins of B2 and B4 appear to have a common ancestor, with a well-supported relationship. The CYP proteins of *D. radiodurans* R1 (protein sequences 2 and 3) are ancestral to the B2 and B4 CYP proteins (Figure 6.12).

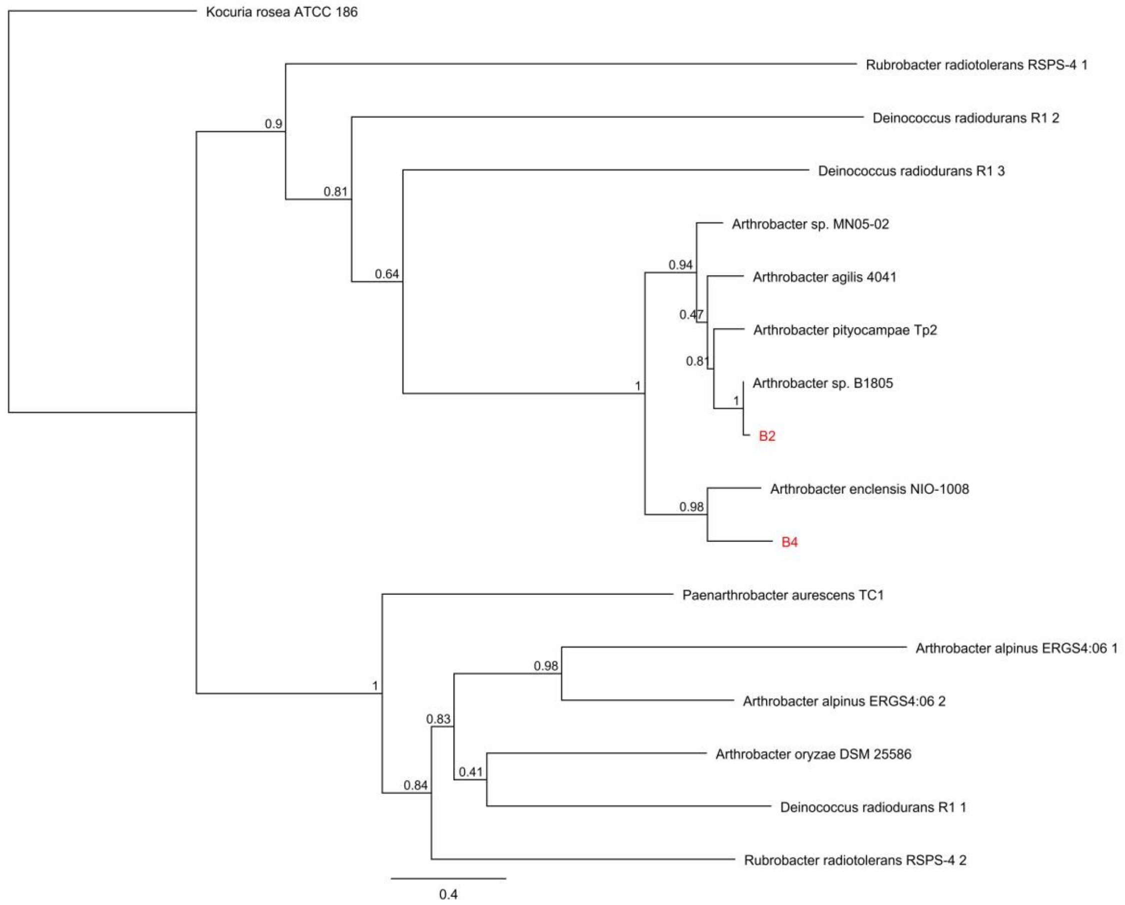

Kocuria rosea ATCC 186
Rubrobacter radiotolerans RSPS-4 1
0.9
Deinococcus radiodurans R1 2
0.81
Deinococcus radiodurans R1 3
Arthrobacter sp. MN05-02
0.94
0.64
Arthrobacter agilis 4041
0.47
Arthrobacter pityocampae Tp2
0.81
Arthrobacter sp. B1805
1
1
B2
Arthrobacter enclensis NIO-1008
0.98
B4
Paenarthrobacter aurescens TC1
Arthrobacter alpinus ERGS4:06 1
0.98
1
Arthrobacter alpinus ERGS4:06 2
0.83
Arthrobacter oryzae DSM 25586
0.41
0.84
Deinococcus radiodurans R1 1
Rubrobacter radiotolerans RSPS-4 2
0.4

Figure 6.12: ML phylogenetic tree of the CYP protein sequence alignment from UV resistant isolates and Arthrobacter reference genomes used in this study. *D. radiodurans* R1, *A. alpinus* ERGS4:06 and *R. radiotolerans* RSPS-4 have multiple CYPs and have been designated a number (1 through 3) to identify each as a separate sequence. *K. rosea* ATCC186 is used as the outgroup. Organisms identified in this study are in red. The scale indicates 0.4 substitutions per amino acid position. Values at each node represent FastTree support values from 1,000 bootstraps, where 1 is equal to 100 percent. The tree was created using Geneious version 2019.0.4 by Biomatters.

As seen in Figure 6.12 above, *Arthrobacter* sp. MN05-02, *A. agilis* 4041, *A. pityocampae* Tp2, *Arthrobacter* sp. B1805 and B2 form a monophyletic clade. B2 is part of a monophyletic clade with *Arthrobacter* sp. B1805, *A. pityocampae* Tp2, *A. agilis* 4041 and *Arthrobacter* sp. MN05-02. B4 appears in the same clade with *A. enclensis* NIO-1008. These clades show common ancestry, indicating the CYP protein of B2 and B4 are derived from a common ancestor, that is well supported. Although there is a low consensus among the selected CYPs, the presence of the CYP protein in known UV resistant organisms warrants further investigation into the activity of CYP to determine if this protein plays a part in UV resistance.

6.3.2 DNA repair genes

The Namib isolates were analysed for a range of DNA repair genes covering base excision repair, nucleotide excision repair to ssDNA damage, recombinational repair to dsDNA damage and other repair mechanisms.

As discussed in Section 2.4.3, base excision repair removes oxidative and radiation induced DNA damage via the repair enzyme DNA glycosylase (García-Ortiz et al. 2001; Kurth et al. 2015). DNA glycosylase splits the N-glycosidic bond between the deoxyribose and the target base (García-Ortiz et al. 2001; Kurth et al. 2015). Nucleotide excision repair removes a wide range of DNA damage caused by UV radiation and chemical mutagens (Crowley et al. 2006; Sinha and Häder 2002). Nucleotide excision repair recruits the UvrABC enzymes for DNA repair (Deaconescu et al. 2012; Kiskeer et al. 2013). Recombination repair (Section 2.4.5) can be used for the reparation of single or double strand breaks on the DNA strand (Ikehata and Ono 2011; Morimatsu and Kowalczykowski 2003).

Other UV-induced DNA damage repair mechanisms include photoreactivation and DNA ligase. Photoreactivation genes were found on each of the Namib isolates. ATP-dependent DNA ligase *ligC* was also found on each Namib genome. Error prone SOS-repair mechanisms such as translesion bypass using *umuD* and *umuC* were not found in the Namib isolates or the reference genomes used in this study when searched using BLAST. However, lexA was found in each of the Namib and reference genomes, indicating that SOS responses using recA may exist in these genera.

The repair system, genes for DNA repair, and their presence or absence on the Namib isolates is detailed in Table 6.3 below.

Table 6.3: DNA repair genes found in the Namib isolates. A green tick indicates that the gene was found on the genome. The Prokka locus tags are provided below the green tick for each isolate. Red crosses indicate that the gene was not found on the genome.

Repair system	Gene	Function	B2	B4	E2	E5
Excision repair						
Base excision repair	*ung*	Uracil-DNA glycosylase	✓ KFNIHMNI_01736	✓ HLFAFCBI_02324	✓ FICEKGFG_03179	✓ GHIKLHBA_02602
	ndk	AP endonuclease	✓ KFNIHMNI_00722	✓ HLFAFCBI_02923	✓ FICEKGFG_00028	✓ GHIKLHBA_00467
	mutM	Formamidopyrimidine-DNA glycosylase	✓ KFNIHMNI_00635	✓ HLFAFCBI_02631	✓ FICEKGFG_02655	✓ GHIKLHBA_02042
	mutY	A/G-specific adenine glycosylase	✓ KFNIHMNI_00867	✓ HLFAFCBI_00491	✓ FICEKGFG_02325	✓ GHIKLHBA_00125
	nth	Ultraviolet N-glycosylase/AP lyase/endonuclease III	✓ KFNIHMNI_02797	✓ HLFAFCBI_02734	✓ FICEKGFG_01297	✓ GHIKLHBA_03357
	exoIII	Exodeoxyribonuclease	✓ KFNIHMNI_01418	✓ HLFAFCBI_01973	✓ FICEKGFG_03500	✓ GHIKLHBA_01525
	alkA	3-methyladenine DNA glycosylase	✓ KFNIHMNI_02242	✓ HLFAFCBI_01472	✓ FICEKGFG_01548	✓ GHIKLHBA_03178
Nucleotide excision repair	*uvrA*	DNA binding	✓ KFNIHMNI_01772 KFNIHMNI_02578 KFNIHMNI_02729	✓ HLFAFCBI_02430 HLFAFCBI_03460 HLFAFCBI_02046	✓ FICEKGFG_00393 FICEKGFG_03363	✓ GHIKLHBA_00380 GHIKLHBA_00800 GHIKLHBA_02506
	uvrB	Helicase - 3' incision endonuclease	✓ KFNIHMNI_01780	✓ HLFAFCBI_02036	✓ FICEKGFG_00403	✓ GHIKLHBA_00811 GHIKLHBA_01404

	uvrC	5' incision endonuclease	✓ KFNIHMNI_01769	✓ HLFAFCBI_02049	✓ FICEKGFG_00390	✓ GHIKLHBA_00797
	uvrD	Excision helicase	✓ KFNIHMNI_01701 KFNIHMNI_00345	✓ HLFAFCBI_02364 HLFAFCBI_00134	✓ FICEKGFG_03216 FICEKGFG_00865	✓ GHIKLHBA_02562 GHIKLHBA_01234
	mfd	Transcription-repair coupling factor	✓ KFNIHMNI_02040	✓ HLFAFCBI_01403	✓ FICEKGFG_01048	✓ GHIKLHBA_01859
Recombinational repair						
Non-homologous end joining	*ligC*	ATP-dependent DNA ligase	✓ KFNIHMNI_00058	✓ HLFAFCBI_01562	✓ FICEKGFG_01645 FICEKGFG_01893 FICEKGFG_01896	✓ GHIKLHBA_01051 GHIKLHBA_02371
	ligD	Multifunctional non-homologous end joining protein LigD	✓ KFNIHMNI_01796 KFNIHMNI_00929	✓ HLFAFCBI_03587 HLFAFCBI_00589	✓ FICEKGFG_02181	✓ GHIKLHBA_03260 GHIKLHBA_00029
Homologous recombination						
Initiation – RecBCD pathway	*recA*	Recombinase	✓ KFNIHMNI_02198	✓ HLFAFCBI_01234	✓ FICEKGFG_01218	✓ GHIKLHBA_03312
	*recB**	Exo V helicase*	✓ KFNIHMNI_01701	✓ HLFAFCBI_02364	✓ FICEKGFG_03216	✓ GHIKLHBA_02562
Initiation – RecFOR pathway	*recF*	Assist RecA filamentation	✓ KFNIHMNI_00979	✓ HLFAFCBI_01487	✓ FICEKGFG_01563	✓ GHIKLHBA_03165
	recO	Binds ssDNA and assists RecF	✓ KFNIHMNI_02607	✓ HLFAFCBI_01727	✓ FICEKGFG_00236	✓ GHIKLHBA_00641
	recR	ATP-binding and assists RecF	✓ KFNIHMNI_02351	✓ HLFAFCBI_00775	✓ FICEKGFG_02809	✓ GHIKLHBA_00975

	recN	ATP-binding	✓ KFNIHMNI_02014	✓ HLFAFCBI_02986	✓ FICEKGFG_03097	✓ GHIKLHBA_03050
	recX	Regulatory protein	✓ KFNIHMNI_02200	✓ HLFAFCBI_01233	✓ FICEKGFG_01219	✓ GHIKLHBA_03311
	recQ	ATP-dependent DNA helicase	✓ KFNIHMNI_00323 KFNIHMNI_01782	✓ FICEKGFG_00833	✓ HLFAFCBI_00104	✓ GHIKLHBA_01263
Initiation – SbcBCD pathway	*sbcB*	Degrades ssDNA	✓ KFNIHMNI_02292 KFNIHMNI_00265 KFNIHMNI_00266	✓ HLFAFCBI_00841 HLFAFCBI_00062 HLFAFCBI_00063	✓ FICEKGFG_02754 FICEKGFG_00794 FICEKGFG_00795	✓ GHIKLHBA_00870 GHIKLHBA_01304 GHIKLHBA_01305
	sbcC	Cleaves DNA hairpin structures that inhibit DNA replication	✓ KFNIHMNI_00529	✓ HLFAFCBI_02694	✓ FICEKGFG_02593	✓ GHIKLHBA_02105
	sbcD	ATP-dependent dsDNA exonuclease	✓ KFNIHMNI_00528	✓ HLFAFCBI_02695	✓ FICEKGFG_02592	✓ GHIKLHBA_02106
	ssb	Single strand binding protein	✓ KFNIHMNI_01034 KFNIHMNI_00694	✓ HLFAFCBI_02898 HLFAFCBI_01531	✓ FICEKGFG_00003 FICEKGFG_01602 FICEKGFG_01881	✓ GHIKLHBA_02336 GHIKLHBA_00441
Branch migration	*ruvA*	5'-3' junction helicase	✓ KFNIHMNI_01829	✓ HLFAFCBI_01802	✓ FICEKGFG_00166	✓ GHIKLHBA_00565
	ruvB	5'-3' junction helicase	✓ KFNIHMNI_01828	✓ HLFAFCBI_01801	✓ FICEKGFG_00167	✓ GHIKLHBA_00566
Resolvases	*ruvC*	Junction endonuclease	✓ KFNIHMNI_01830	✓ HLFAFCBI_01803	✓ FICEKGFG_00165	✓ GHIKLHBA_00564

	recG	Resolvase, 3'-5' junction helicase	✓ KFNIHMNI_00618	✓ HLFAFCBI_02639	✓ FICEKGFG_02647	✓ GHIKLHBA_02050
Other systems						
SOS and error-prone repair	*lexA*	SOS activator	✓ KFNIHMNI_02207	✓ HLFAFCBI_01226	✓ FICEKGFG_03323 FICEKGFG_01226	✓ GHIKLHBA_03304
	dinB	Error prone DNA polymerase IV	✓ KFNIHMNI_01976 KFNIHMNI_00953	✓ HLFAFCBI_03028 HLFAFCBI_02199	✓ FICEKGFG_01389 FICEKGFG_03138	✓ GHIKLHBA_02226 GHIKLHBA_03092
	dinG	ATP-dependent DNA helicase	✓ KFNIHMNI_02206	✓ HLFAFCBI_01227	✓ FICEKGFG_01225	✓ GHIKLHBA_03305
Alkytransferases	*ogt*	Alkyltransferase with regulatory motif	✓ KFNIHMNI_00662	✓ HLFAFCBI_01194	✓ FICEKGFG_01262	✓ GHIKLHBA_03272
Photolyase	*phr*	Deoxyribodipyrimide photolyase	✓ KFNIHMNI_00351	✓ HLFAFCBI_00142	✓ FICEKGFG_00873	✓ GHIKLHBA_01226
	phlK	(6-4) photolyase	✓ KFNIHMNI_01102	✗	✗	✗
	splB	Radical SAM enzyme for DNA repair	✓ KFNIHMNI_01171	✓ HLFAFCBI_02878	✗	✓ GHIKLHBA_02732

* A BLAST search of this gene returned the same protein locus tag as for DNA ligase.

Table 6.3 shows that a range of repair mechanisms were found in both the UVC resistant and UVC sensitive Namib genomes. Genes for the BER, NER and recombinational repair systems were observed on all genomes. The *phr* gene for Deoxyribodipyrimide photolyase repair was found on all isolates, but the *phlK* gene for specific 6-4 photolyase repair was only observed in isolate B2. The *phlK* gene was not observed in any of the reference genomes used in this study. A BLASTP of this protein returned a sequence similarity with other *Arthrobacter*, with the closest similarity to *Arthrobacter* sp. B1805 (96.26% sequence identity, 96% query cover). This protein returned as cryptochrome/photolyase family protein; a DNA photo repair gene responsible for DNA repair using blue light photons (Kavakli et al. 2017).

Finally, Table 6.3 shows that spore photoproduct lyase (*splB*), a radical SAM enzyme that repairs UV induces lesions in bacterial DNA, is absent in isolate E2. The absence of this gene in E2 may contribute to its sensitivity to UV.

The presence, absence and copy number of the above-mentioned genes (Table 6.3) for the Namib isolates was compared with the reference genomes for *Arthrobacter* and *Pseudarthrobacter*. DNA repair mechanisms encoded in each of the reference genomes and the Namib isolates, and the copy number of these genes, is shown in the heatmap below (Figure 6.13).

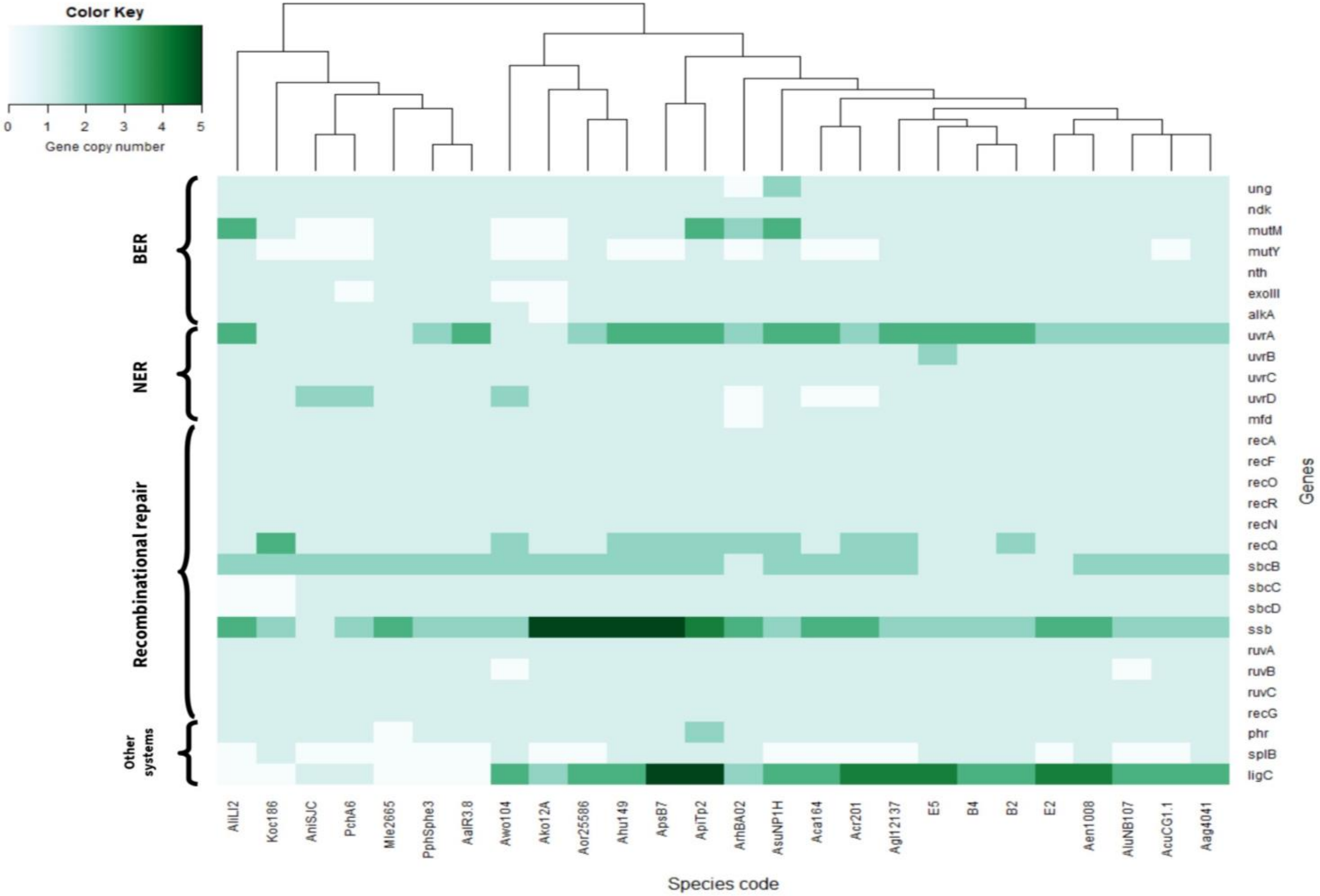


Figure 6.13: Heatmap of the presence and absence of DNA repair genes across Micrococcaceae. A white colour indicates a complete absence of the gene within that genome. A darker green colour indicates a higher copy number of the gene, while a lighter green colour indicates a lower copy number of the gene. Repair mechanisms are as indicated: BER, NER, recombinational repair and other DNA repair systems.

Figure 6.13 shows that the Namib isolates have a similar copy number of UV repair genes compared to other members of the *Arthrobacter* and *Pseudarthrobacter* genera.

All genes for the RecFOR recombinational repair pathway were found in the Namib isolates and the reference genomes. Single copies of the RecFOR pathway genes were observed in all genomes used in this study, except for multiple copies of RecQ observed in several of the genomes. The Namib isolates B4, E2 and E5 had single copies of RecQ, while isolate B2 contained two non-identical copies. RecB, of the RecBCD recombinational repair pathway, was observed in each of the Namib genomes and the reference *Arthrobacter* and *Pseudarthrobacter* genomes, but both RecC and RecD were absent. As previously discussed (Section 2.3.6), the RecBCD primarily initiates recombinational repair for double strand breaks (Singleton et al. 2004; Smith 2012; Spies and Kowalczykowski 2005) . This indicates that in the absence of RecBCD, the RecFOR pathway is responsible for double-strand break repair in *Arthrobacter* and *Pseudarthrobacter*. Interestingly, the Namib isolates and the reference genomes also contained a complete SbcBCD pathway. As noted (Section 2.3.6), the presence of SbcB inhibits the action of RecA and, in turn, the RecFOR pathway (Rocha et al. 2005). However, the presence of genes does not indicate that they are active within the genome; transcriptomics studies will indicate if there is inhibitory action of the SbcB nuclease on the action of RecFOR. Additionally, since the Namib isolate genomes are in draft form, additional gene copies may be missing from this analysis.

In general, the genes with the most variable copy number were *ligC, ssb* and *uvrA. LigC* is an ATP-dependant DNA ligase and *ssb* is a single strand binding gene in the SbcBCD recombinational repair pathway. The UvrA enzyme probes for DNA abnormalities before recruiting UvrB, UvrC and UvrD to assist in NER repair. E5 had two copies of the *uvrB* gene while the other Namib genomes had one copy. One of the E5 genes does not cluster with other *uvr* repair genes, as is usually expected (Goosen and Moolenaar 2008), suggesting that the second *uvrB* gene was acquired independently of the additional *uvrA* and *uvrC* genes. Alternatively, the second *uvrB* gene on isolate E5 could be a result of gene duplication (Goosen and Moolenaar 2008), which could result in the second *uvrB* gene being separate from the other *uvrA* and *uvrC* genes.

The *recJ* gene could not be identified by name search, nor by BLASTP in any of the Namib isolates or the *Arthrobacter* or *Pseudarthrobacter* reference genomes. RecJ is an important protein for ssDNA exonuclease incisions in the RecFOR pathway (Martins-Pinheiro et al. 2007). Bentchikou et al. (2010) notes that absence of the *recJ* gene in *D. radiodurans* results in cell death, indicating that *recJ* is vital for genome maintenance and cell viability of the species. The possible absence of this protein in *Arthrobacter* and *Pseudarthrobacter* indicates that another protein must perform the role of *recJ* in these genera. Rocha et al. (2005) has previously noted the absence of *recJ* in Actinobacteria and goes on to note that the helicase II *uvrD* can act in the

RecFOR pathway. It is therefore assumed that in isolates B2, E2, B4 and E5 the *uvrD* gene acts in both the UvrABC [NER repair] and RecFOR [recombinational repair] pathways.

The absence of *uvrD* in *A. rhombi* B Ar 00.02, *A. castelli* DSM 16402 and *A. crystallopoietes* DSM 20117 indicates that these organisms must have an alternative to UvrD to perform excision repair. Goosen and Moolenaar (2008) notes that UvrD is less well conserved than the other UvrABC subunits, hence may have been missed during annotation. It is suspected that transfer of the *uvrD* gene is independent of the *uvrABC* genes.

To achieve Aim 4, two objectives were outlined (Section 4.1). The first objective was to identify gene elements associated with UV resistance. This has now been achieved. The second objective was to compare the predicted tertiary protein structures encoded for by gene elements associated with UV resistance. To achieve this second objective, the focus for the final section of this thesis will be on the UvrABC proteins of the NER repair system. As noted in Figure 6.14, both the UVC resistant and UVC sensitive isolates have copies of these genes. However, the presence and absence of genes does not indicate the functionality of the corresponding protein. Analysing protein structures can provide insights into the likely functionality of proteins. Therefore, the predicted structures for the UvrABC proteins was investigated. This system was selected due to the importance of NER in UV-induced DNA repair, and that mutations in these proteins can lead to a UV sensitive phenotype (Crowley et al. 2006; Demple and Harrison 1994; Hanada et al. 2000; Van Houten et al. 2005).

6.3.3 Protein alignment and predicted secondary and tertiary structure of the UvrABC proteins

The bacterial NER (UvrABC system) is an important system in the repair of short patches of UV damaged DNA. The NER repair system has been shown to remove a range of lesions, including cyclobutane pyrimidine dimers (CPDs), chemical adducts and inter- and intrastrand crosslinking (Goosen and Moolenaar 2008). Previous studies have found that mutations in the UvrABC proteins results in a UVC sensitive phenotype (Crowley et al. 2006; Goosen and Moolenaar 2008; Hanada et al. 2000; Stracy et al. 2016; Van Houten et al. 2005). As such, the UvrABC proteins in the Namib isolates were selected for protein alignment and structure analysis to discern differences between the UVC resistant and UVC sensitive phenotypes.

6.3.3.1 UvrA protein structure

It has previously been demonstrated that mutations at the N-terminal zinc finger of UvrA do not impact on the ability of UvrA to attach to DNA (Deaconescu et al. 2012). However, mutations that disrupt the C-terminal zinc finger of UvrA have been shown to cause the protein to become ineffective and render the bacterium sensitive to UV radiation (Deaconescu et al. 2012; Goosen and Moolenaar 2008; Van Houten et al. 2005). An absent zinc finger at the C-terminal region of the UvrA protein in the isolates E2 and E5 may provide insight into their UVC sensitive phenotypes.

While there are several different subclasses of the UvrA protein, the general UvrA structure contains an ATP-binding signature (I, II, III and IV), a UvrB attachment region and a DNA binding region (Marszałkowska et al. 2014). The role of UvrA is to recognise and bind to a damaged DNA site through a UvrAB complex. Following detection and binding to damaged DNA, UvrA disassociates from the UvrAB complex and is replaced by the UvrC protein Marszałkowska et al. (2014) notes that there are four classes and five subclasses of UvrA proteins. A phylogenetic analysis of the UvrA subunit proteins was carried out to determine the relationship between the different UvrA proteins of each genome, and to help analyse which UvrA paralogue each protein belonged to. Figure 6.14 shows three major clades of the UvrA proteins analysed. This supports previous findings of multiple UvrA proteins within a single genome (Goosen and Moolenaar 2008). As seen in Figure 6.14, several species of *Arthrobacter* and *Pseudarthrobacter* have multiple copies of the UvrA protein. This is the first analysis of the UvrA subunits within the *Arthrobacter* and *Pseudarthrobacter* genera.

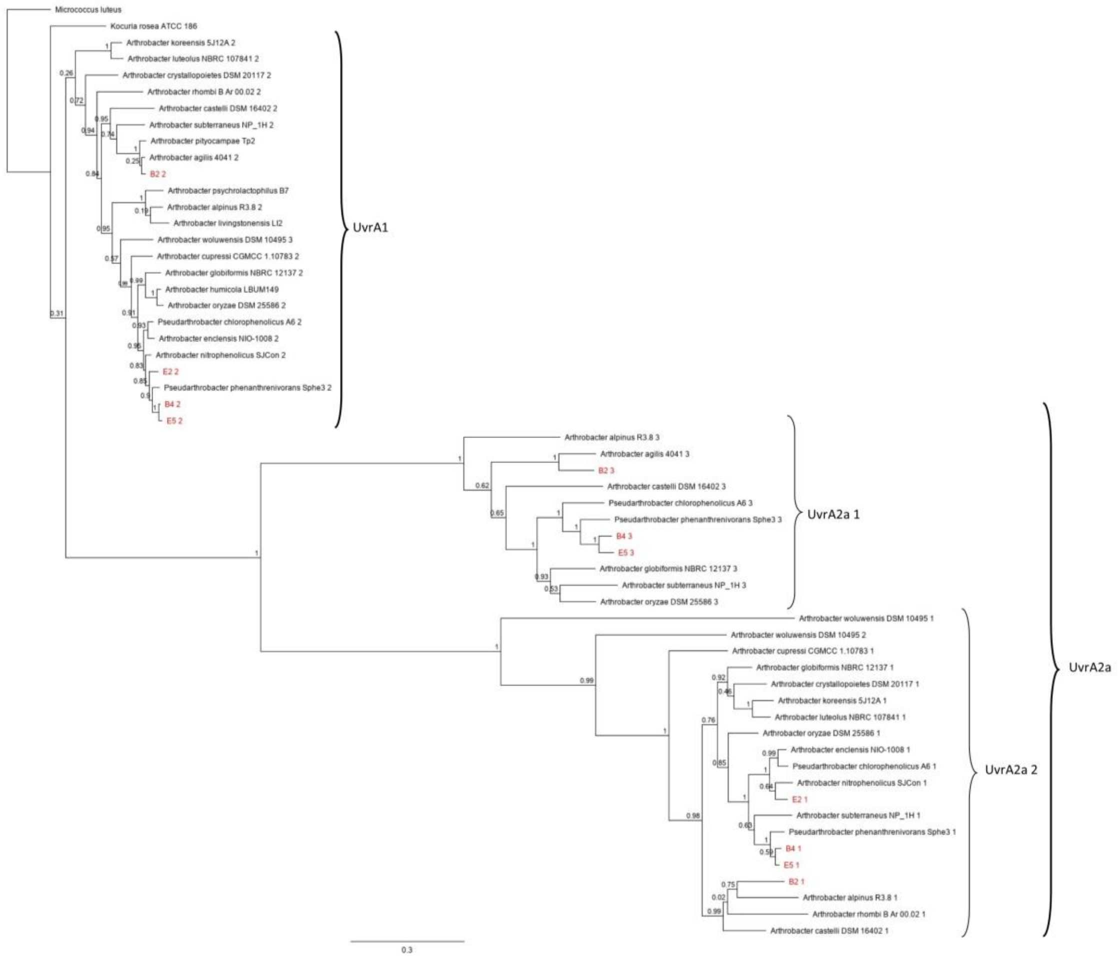

Micrococcus luteus
Kocuria rosea ATCC 186
1
Arthrobacter koreensis 5J12A 2
Arthrobacter luteolus NBRC 107841 2
0.26
Arthrobacter crystallopoietes DSM 20117 2
0.72
Arthrobacter rhombi B Ar 00.02 2
Arthrobacter castelli DSM 16402 2
0.95
0.94
0.74
Arthrobacter subterraneus NP_1H 2
Arthrobacter pityocampae Tp2
1
Arthrobacter agilis 4041 2
0.25
0.84
B2 2
Arthrobacter psychrolactophilus B7
1
Arthrobacter alpinus R3.8 2
0.19
Arthrobacter livingstonensis LI2
0.95
Arthrobacter woluwensis DSM 10495 3
0.57
Arthrobacter cupressi CGMCC 1.10783 2
0.99
Arthrobacter globiformis NBRC 12137 2
1
Arthrobacter humicola LBUM149
Arthrobacter oryzae DSM 25586 2
0.31
0.91
Pseudarthrobacter chlorophenolicus A6 2
0.93
Arthrobacter enclensis NIO-1008 2
0.95
Arthrobacter nitrophenolicus SJCon 2
0.83
E2 2
0.85
Pseudarthrobacter phenanthrenivorans Sphe3 2
0.9
1
B4 2
E5 2
UvrA1
Arthrobacter alpinus R3.8 3
1
1
Arthrobacter agilis 4041 3
B2 3
0.62
Arthrobacter castelli DSM 16402 3
Pseudarthrobacter chlorophenolicus A6 3
0.65
1
Pseudarthrobacter phenanthrenivorans Sphe3 3
1
1
B4 3
1
E5 3
Arthrobacter globiformis NBRC 12137 3
0.93
Arthrobacter subterraneus NP_1H 3
0.53
Arthrobacter oryzae DSM 25586 3
UvrA2a 1
Arthrobacter woluwensis DSM 10495 1
Arthrobacter woluwensis DSM 10495 2
1
Arthrobacter cupressi CGMCC 1.10783 1
Arthrobacter globiformis NBRC 12137 1
0.99
0.92
Arthrobacter crystallopoietes DSM 20117 1
0.46
1
Arthrobacter koreensis 5J12A 1
Arthrobacter luteolus NBRC 107841 1
0.76
1
Arthrobacter oryzae DSM 25586 1
Arthrobacter enclensis NIO-1008 1
0.99
1
Pseudarthrobacter chlorophenolicus A6 1
0.85
0.64
Arthrobacter nitrophenolicus SJCon 1
1
E2 1
0.98
Arthrobacter subterraneus NP_1H 1
0.63
Pseudarthrobacter phenanthrenivorans Sphe3 1
1
0.59
B4 1
E5 1
0.75
B2 1
0.02
Arthrobacter alpinus R3.8 1
0.99
Arthrobacter rhombi B Ar 00.02 1
Arthrobacter castelli DSM 16402 1
UvrA2a 2
UvrA2a
0.3

Figure 6.14: ML phylogenetic tree of the UvrA protein alignment from the Namib isolates and reference genomes used in this study. Two UvrA proteins were identified: UvrA1 and UvrA2a. Of UvrA2a, two subgroups were identified: UvrA2a 1 and UvrA2a 2. The UvrA1 gene of *M. luteus* NCTC 2665 was used as an outgroup. Organisms identified in this study are in red. Bar, 0.3 substitutions per amino acid position. Values at nodes represent FastTree support values from 1,000 bootstraps, where 1 is equal to 100 percent. Tree created using Geneious version 2019.0.4 by Biomatters.

Figure 6.14 shows three major clades of UvrA proteins: A1, A2a1 and A2a2 as named by Marszałkowska et al. (2014). UvrA1 is closer to the last common ancestor. UvrA2a 1 and UvrA2a 2 show greater divergence from the last common ancestor, and share an ancestor, although this relationship is not well supported (FastTree support value = 0.31). Greater sequence variability in the UvrA2a clades than UvrA1 suggest that there are greater constraints on UvrA1, leading to more conservation of the UvrA1 protein. The UvrA1, Uvr2A1 and Uvr2A2 shared between 23.6% (B2) and 33.8% (E2) identical sites amongst the Namib isolates.

Analysis of the length of the Namib isolate UvrA proteins showed that all isolates have one class I protein (900 – 1050 aa), one class IIa protein (740 – 880 aa), and B2, B4 and E5 have an additional class IIa protein (Figure 6.15). Marszałkowska et al. (2014) noted that UvrA class IIa proteins are missing a UvrB binding site between 118 – 256 aa. Despite the lack of a UvrB binding region, there is evidence of upregulation of class IIa UvrA proteins following UV exposure (Liu et al. 2003; Timmins et al. 2009). However, Tanaka et al. (2005) noted that inactivation of UvrA2a proteins does not lead to mutagenesis, indicating that the class IIa proteins have a minor role in UV resistance. This may help the UV resistant isolates B2 and B4 to survive extended UV exposure, in combination with other UV repair genes. Goosen and Moolenaar (2008) speculated that the UvrA2a class of proteins may be more involved in antibiotic resistance than UV repair.

6.3.3.1.1 UvrA1

UvrA1 proteins have four domains that are important for the function of the protein: ATP-binding site I, UvrB binding site, DNA binding site and ATP-binding site II. As the domains of the UvrA protein in Actinobacteria are unknown, the domains for this protein are estimated based on the size of the protein and on the known domains of *Thermotoga maritima* (Jaciuk et al. 2011). The binding regions of UvrA1 in *T. maritima* are: 1 – 87 and 503 – 588 aa (ATP-binding site I), 118 – 256 aa (UvrB binding site), 287 – 379 aa (DNA binding site) and 589 – 916 aa (ATP-binding site II) (Jaciuk et al. 2011). There are also four ‘zinc fingers’ within UvrA1 proteins. These are indicated by a Cys-X-X-Cys sequence. UvrA proteins also have two signature motifs which will not be investigated or highlighted on the following alignments. This is because the signature motifs are only indicators of the genus and do not play a role in the overall function of the protein (Jaciuk et al. 2011). The protein alignment and secondary structure of UvrA1 protein of the Namib isolates can be seen in Figure 6.15 below.

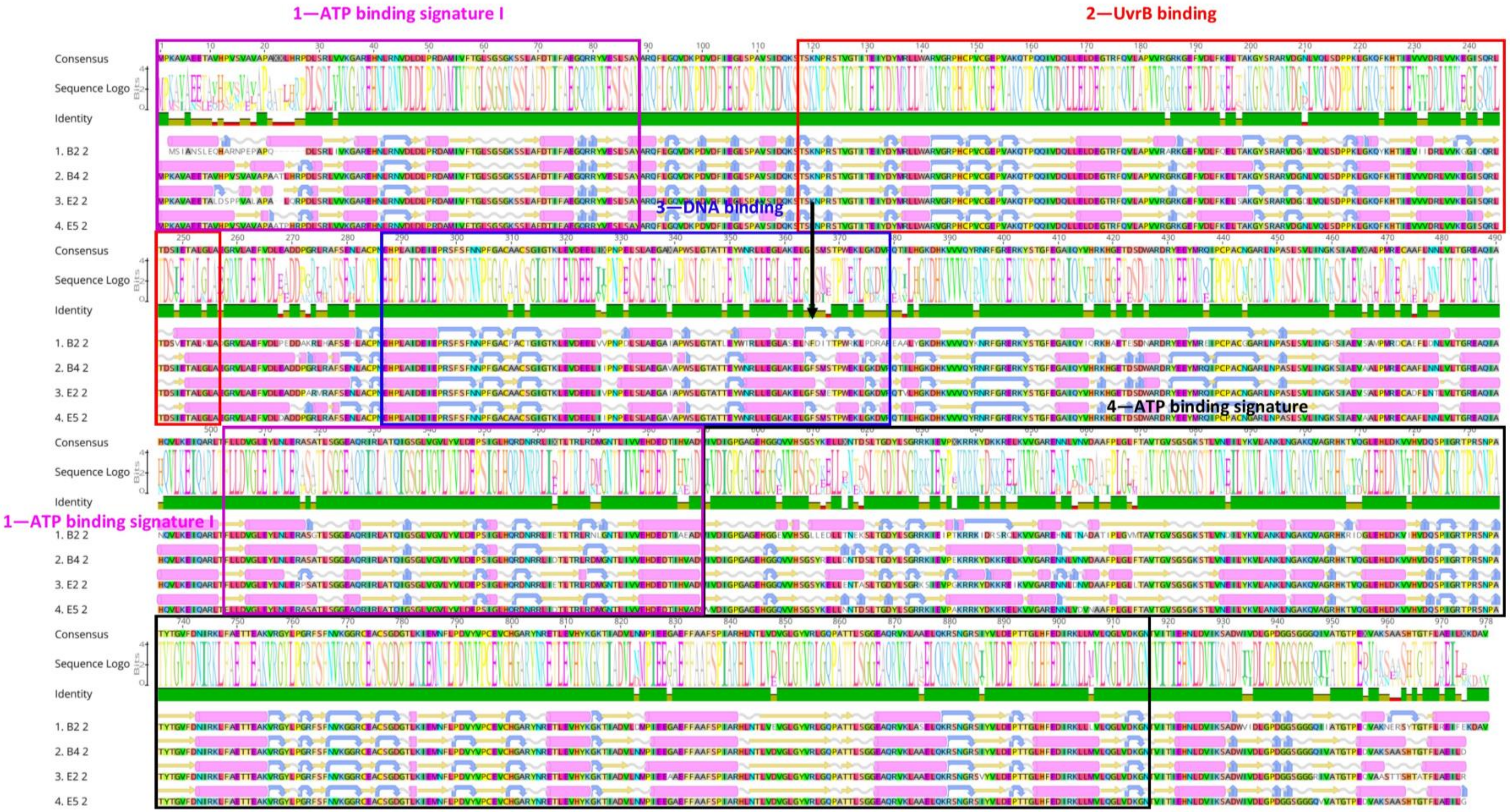


Figure 6.15: Primary and secondary structure alignment of the UvrA1 protein of the Namib isolates. Signature 1: ATP-binding site I (1 – 87, 503 – 590 aa) is shown in pink, signature 2: UvrB binding region (118 – 256 aa) is shown in red, signature 3: DNA binding region (287 – 379 aa) is shown in blue and signature 4: ATP binding signature II (667 – 822 aa) is shown in black. The coded amino acids are shown above the alignment and the consensus sequence is shown above the amino acid sequence logo. Black arrow indicates a motif that is different between the UVC resistant and UVC sensitive isolates. Coils are shown as grey lines, turns as blue U-turn arrows, α-helices are pink cylinders and β-sheets are as yellow block arrows. Secondary structures created using Geneious version 2019.0.4 by Biomatters.

Overall, the UvrA1 primary protein structure appears to be highly conserved amongst the Namib isolates. The secondary structure of the UvrA1 protein also appears highly conserved. There are regions that show some variability; however, these are outside of the estimated binding regions for this protein. However, as indicated in Figure 6.15, isolates B2, B4 and E2 demonstrate a predicted U-turn at aa 365, while isolate E2 does not have this U-tun, but an α-helix instead.

Phyre2 (Kelley et al. 2015) provided a confidence rating of 100% for the UvrA1 proteins using 'normal modelling' of each of the UvrA1 proteins in the Namib isolates. The percentage identity for each protein ranged from 66% (B2 and E2) to 74% (B4 and E5). The quality metrics for the predicted structure can be found in Appendix 8. The predicted tertiary structures for the UvrA1 protein for the Namib isolates can be seen in Figure 6.16. The ATP binding signature I (magenta), UvrB binding signature (red), DNA binding signature (yellow) and ATP binding signature II (cyan) are indicated.

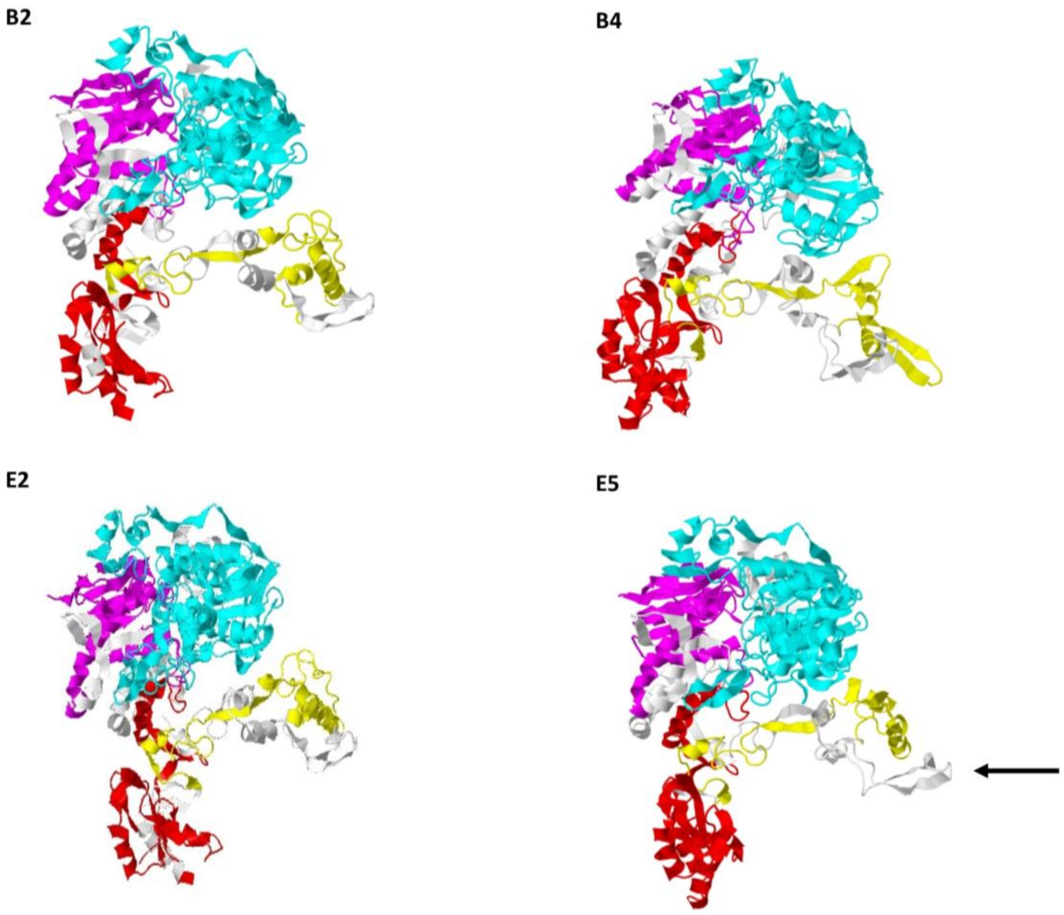


Figure 6.16: Predicted tertiary structure of the UvrA1 protein of the Namib isolates. Colours: ATP binding signature I (magenta), UvrB binding signature (red), DNA binding signature (yellow), ATP binding signature II (cyan). Black arrow indicates a protein tertiary structure position on E5 that looks different from that of B2, B4 and E2. Tertiary structures predicted using 'normal modelling' in Phyre2 V2.0 (Kelley et al. 2015) and coloured using Geneious version 2019.0.4 by Biomatters.

The predicted tertiary structures of the UvrA1 protein from the Namib isolates appears very similar between the isolates. However, the UvrA protein from isolate E5 appears to have an outward loop of beta sheets at an angle that is not seen in the other isolates. As this outward loop is not part of the known ATP, DNA or UvrB binding regions, it is assumed that this does not interfere with the functionality of the protein. The motif at aa 365 where isolates B2, B4 and E5 have a predicted U-turn and where E2 has an α-helix was investigated. However, the predicted tertiary structure of this region showed that UvrA1 from B2, B4 and E2 have a U-turn while from E5 it showed a coil, as shown in green in Figure 6.17 below.

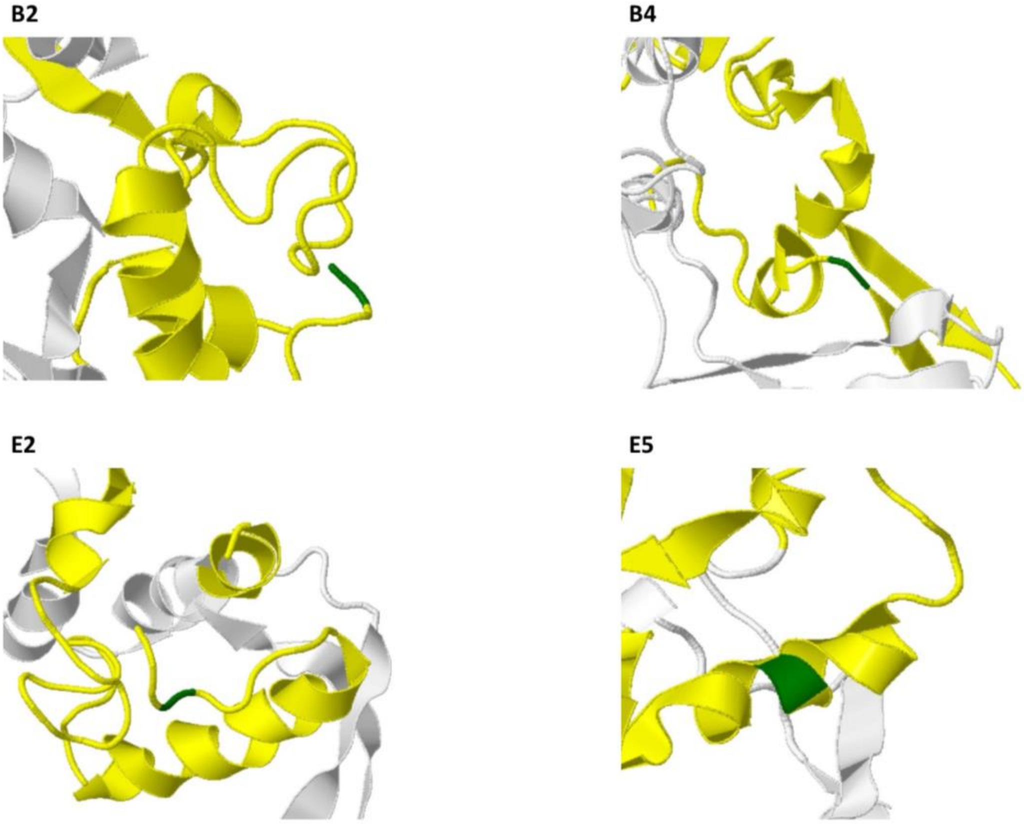


Figure 6.17: Close-up of the DNA binding region of the predicted UvrA1 protein tertiary structure of the Namib isolates. Colours: DNA binding signature (yellow), identified motif at aa 365 (green). Tertiary structures predicted using 'normal modelling' in Phyre2 V2.0 (Kelley et al. 2015) and coloured using Geneious version 2019.0.4 by Biomatters.

Figure 6.17 shows that the motif at aa 365 shows a U-turn for isolates B2, B4 and E2 and a coil for E5. This differs from the predicted secondary structure in Figure 6.15. This motif is outside of the known active sites for this protein and this discrepancy is unlikely to be the cause UV sensitivity in isolate E5. However, further investigations into proteomics may reveal that the UvrA1 protein of E5 has lower efficiency than B2, B4 and E2, although it is impossible to predict this based on current information.

As mentioned, the UvrA1 protein appears to be highly conserved. An amino acid change in the DNA binding site may impact on the ability of the protein to attach to damaged DNA sites. An inability to bind and begin nucleotide excision may cause the organism to express a UVC sensitive phenotype. Four zinc finger regions were found for each isolate, indicating the UV sensitivity in isolates E2 and E5 is not caused by the absence of zinc fingers in the UvrA1 protein.

6.3.3.1.2 UvrA2a 1

Two *uvrA* 2a genes were found in the genomes of Namib isolates B2, B4 and E5. Only one *uvrA* 2a gene was found in isolate E2. UvrA2a 1 and UvrA2a 2 (as identified in Figure 6.14) are described here separately. As mentioned, UvrA2a proteins are missing a UvrB binding site of between 120 and 140 aa. The UvrA2a 1 of the Namib isolates was approximately 120 aa shorter than UvrA1, while the UvrA2a 2 was approximately 140 aa shorter than UvrA1. The missing aa correspond to the UvrB binding region, as class UvrA2a proteins do not have this region (Marszałkowska et al. 2014). The UvrA2a 1 and Uvr2Aa 2 proteins shared 28.7% similarity amongst the Namib isolates. The protein alignment and secondary structure of the Namib UvrA2a 1 proteins can be seen in Figure 6.18 below.

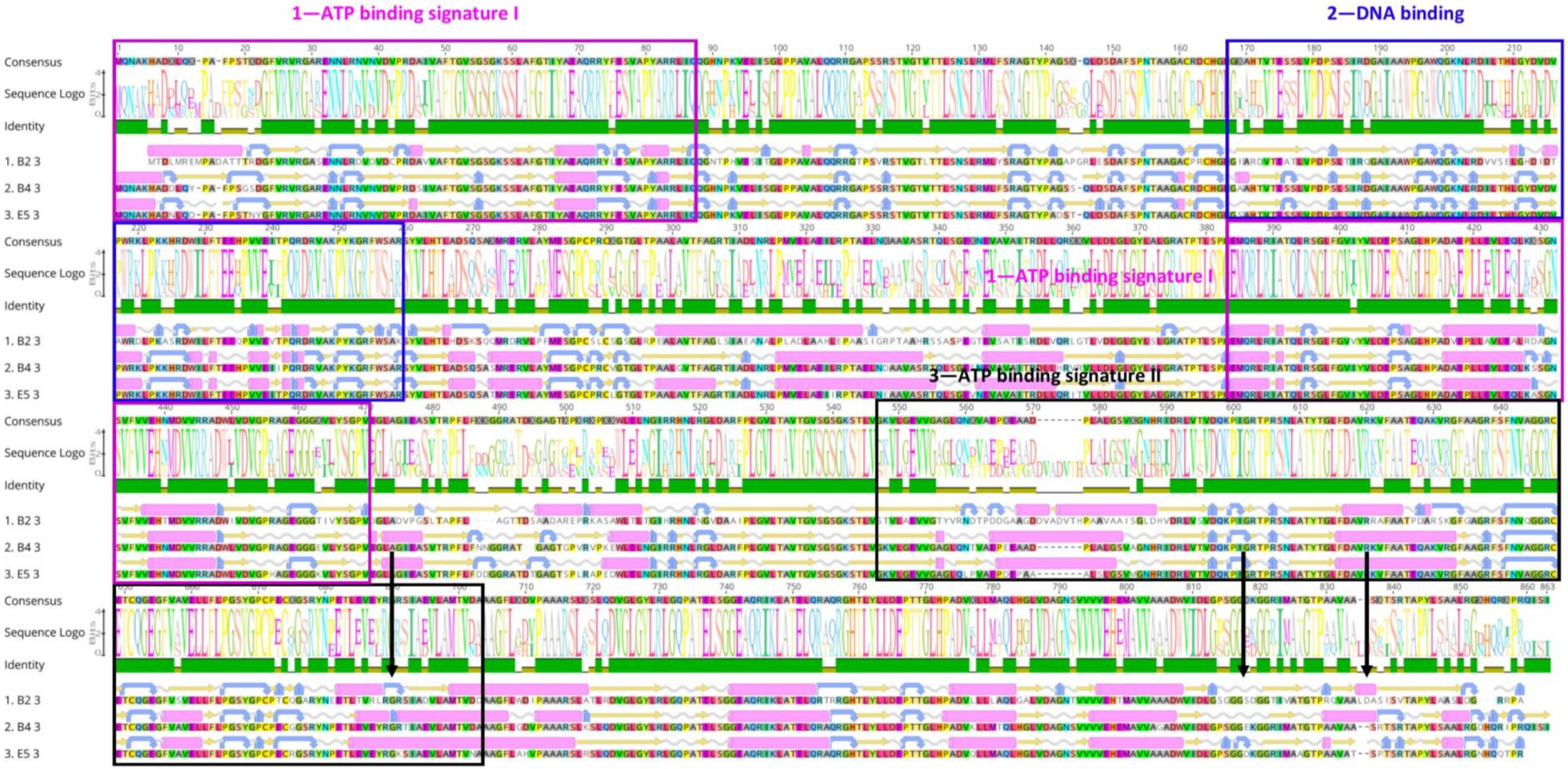

1—ATP binding signature I
2—DNA binding
1—ATP binding signature I
3—ATP binding signature II
Consensus
Sequence Logo
Identity
1. B2 3
2. B4 3
3. E5 3

Figure 6.18: Primary and secondary structure alignment of the UvrA2a 1 protein of the Namib isolates. Signature 1: ATP-binding site I (1 – 87, 383 – 470 aa) is shown in pink, signature 2: DNA binding region (167 – 259 aa) is shown in blue and signature 3: ATP binding signature II (547 – 702 aa) is shown in black. The amino acids are shown above the alignment and the consensus sequence is shown above the amino acid sequence logo. Black arrows indicate a motif that is different between the UVC resistant and UVC sensitive isolates. Coils are shown as grey lines, turns as blue U-turn arrows, α-helices are pink cylinders and β-sheets are as yellow block arrows. Secondary structures created using created using Geneious version 2019.0.4 by Biomatters.

Only isolates B2, B4 and E5 had the UvrA2a 1 protein, as seen in Figure 6.18. The UvrA2a 1 protein shows high sequence conservation in the DNA binding region. UvrA2a proteins do not have UvrB binding regions so are unable to signal UvrB to bind to damaged DNA regions during UV exposure (Marszałkowska et al. 2014). One motif was found on B2 and B4 that was not found in E5 in the ATP binding signature II. The motif is at position 690 where both B2 and B4 have a turn where E5 does not. Two additional motifs were found at position 818, where E5 has an additional turn, and position 836 where E5 is missing an α-helix that B2 and B4 have. The motif at position 690 in E5 may inhibit the ATP binding domain, which may impact on the proteins' ability to bind ATP. This may cause the protein to be dysfunctional or have lower ATP binding affinity. A mutation that renders this part of the protein dysfunctional may in turn lead to an overaccumulation of DNA damage that is unable to be repaired. As a result, the phenotype of the organism with this mutation may be sensitive to UV radiation.

The motifs at positions 818 and 836 are outside of the predicted active sites for UvrA2a and are unlikely to affect the function of the E5 protein, however, this cannot be confirmed using only predicted tertiary structure. The predicted tertiary structures for the UvrA2a 1 protein for the Namib isolates can be seen in Figure 6.19. The quality metrics for the predicted structure can be found in Appendix 8. The ATP binding signature I (magenta), DNA binding signature (yellow) and ATP binding signature II (cyan) are indicated.

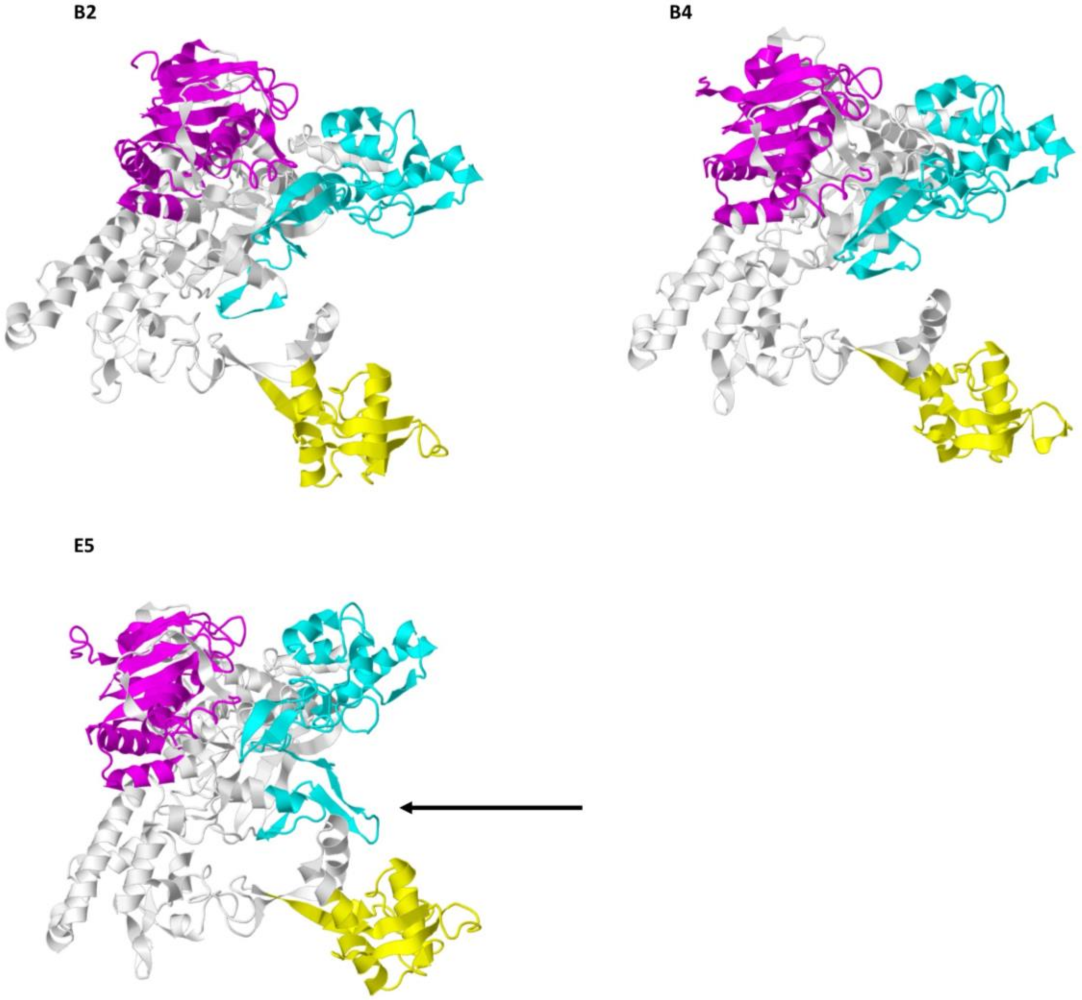


Figure 6.19: Predicted tertiary structure of the UvrA2a 1 protein of the Namib isolates. Colours: ATP binding signature I (magenta), DNA binding signature (yellow), ATP binding signature II (cyan). Black arrow indicates a protein tertiary structure position on E5 that looks different from B2 and B4. Tertiary structures predicted using 'normal modelling' in Phyre2 V2.0 (Kelley et al. 2015) and coloured using Geneious version 2019.0.4 by Biomatters.

As seen in Figure 6.19, there is a different predicted folding of the E5 UvrA2a 1 protein in the ATP binding signature II, compared to that of B2 and B4. Figure 6.18 showed that the E5 UvrA2a 1 protein was also missing a turn (aa 690) that B2 and B4 had. A close-up of the ATP binding signature II is shown in Figure 6.20 below.

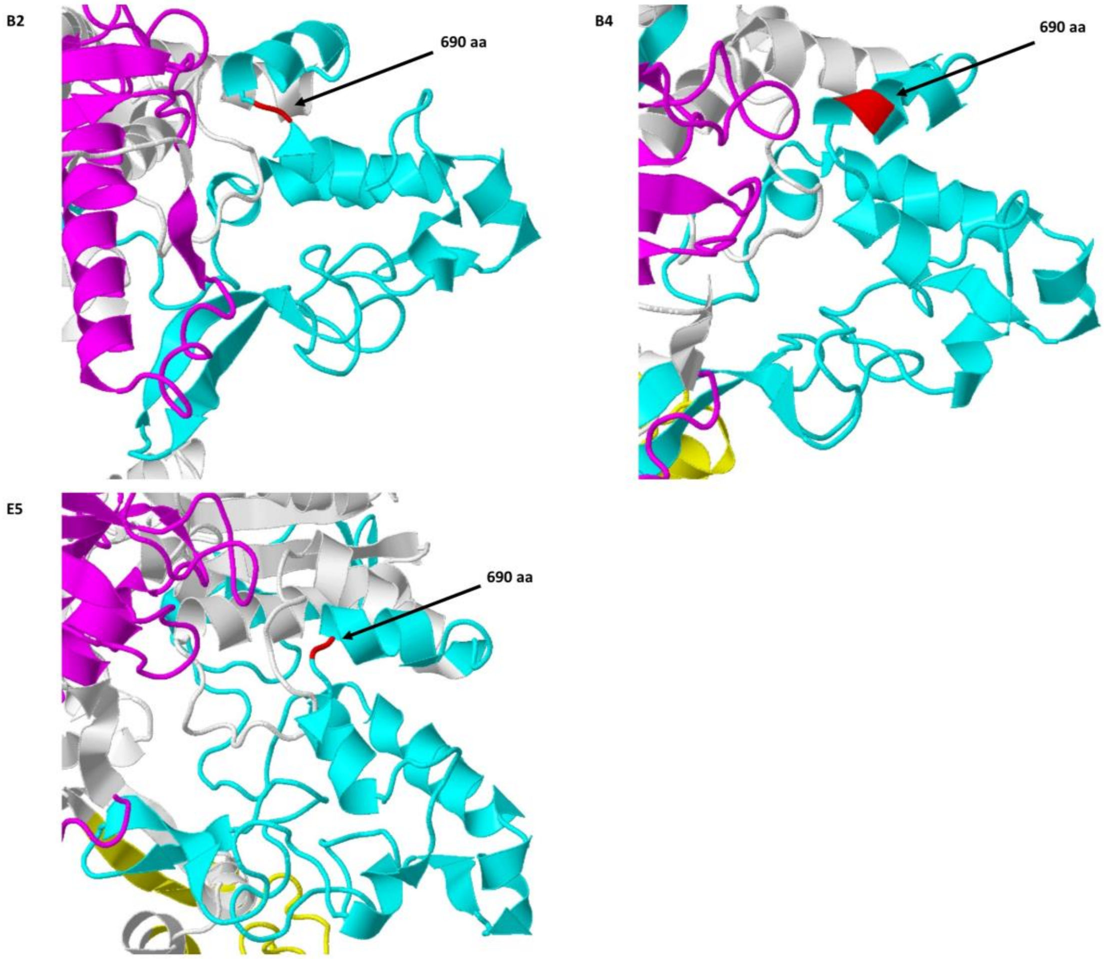


Figure 6.20: Close-up of the DNA binding region of the predicted UvrA2a 1 protein tertiary structure of the Namib isolates. Colours: ATP binding signature I (magenta), DNA binding signature (yellow), ATP binding signature II (cyan), identified motif at aa 690 (red). Tertiary structures predicted using 'normal modelling' in Phyre2 V2.0 (Kelley et al. 2015) and coloured using Geneious version 2019.04 by Biomatters.

Figure 6.20 shows that both B2 and E5 have a coil at aa 690 while B4 has an α-helix. This differs from the predicted secondary structure from Figure 6.17. The folding of the protein in the ATP binding signature II differs between the UVC resistant isolates and E5. As mentioned, ATP attaches to the UvrA protein and induces damage to the surrounding DNA regions, allowing for easier removal of the damaged nucleotide. If the ATP binding signature II is not folded in the correct way, this might cause the protein to become ineffective or have lower efficiency.

6.3.3.1.3 UvrA2a 2

The UvrA2a 2 was missing approximately 140 aa compared to UvrA1. The protein alignment and secondary structure of the UvrA2a 2 protein can be seen in Figure 6.21 below.

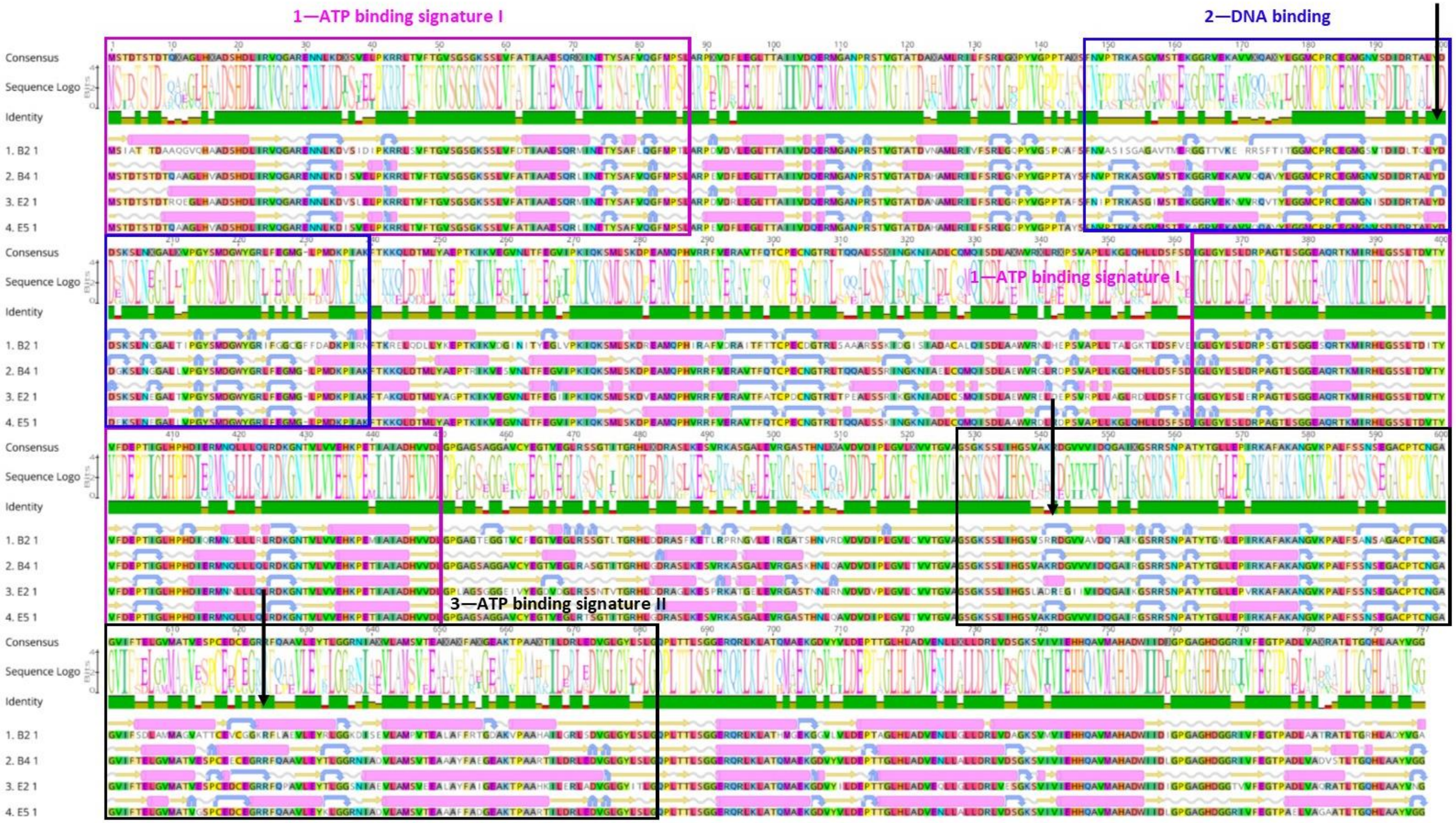

1—ATP binding signature I
2—DNA binding
1—ATP binding signature I
3—ATP binding signature II
Consensus
Sequence Logo
Identity
1. B2 1
2. B4 1
3. E2 1
4. E5 1

Figure 6.21: Primary and secondary structure alignment of the UvrA2a 2 protein of the Namib isolates. Signature 1: ATP-binding site I (1 – 87, 363 – 450 aa) is shown in pink, signature 2: DNA binding region (147 – 239 aa) is shown in blue and signature 3: ATP binding signature II (527 – 682 aa) is shown in black. The coded amino acids are shown above the alignment and the consensus sequence is shown above the amino acid sequence logo. Black arrows indicate a motif that is different between the UVC resistant and UVC sensitive isolates. Coils are shown as grey lines, turns as blue U-turn arrows, α-helices are pink cylinders and β-sheets are as yellow block arrows. Secondary structures created using created using Geneious version 2019.0.4 by Biomatters.

The UvrA2a 2 protein shows high conservation in the ATP binding signature I. The DNA binding region and the ATP binding signature II have a few areas of variability across the isolates. One motif in the DNA binding region (aa 199) shows that isolates B2, B4 and E2 have a U-turn, while E5 has an α-helix. Another two motifs were found in the ATP binding signature II region at 542 aa and 624 aa. In the 542 aa motif isolates B2, B4 and E5 have a U-turn, while E2 has an α-helix. In the 624 aa motif isolates B2, B4 and E5 have an α-helix, while E2 has a U-turn.

The predicted tertiary structures for the UvrA2a 2 protein for the Namib isolates can be seen in Figure 6.22 below. The quality metrics for the predicted structure can be found in Appendix 8. The ATP binding signature I (magenta), DNA binding signature (yellow) and ATP binding signature II (cyan) are indicated.

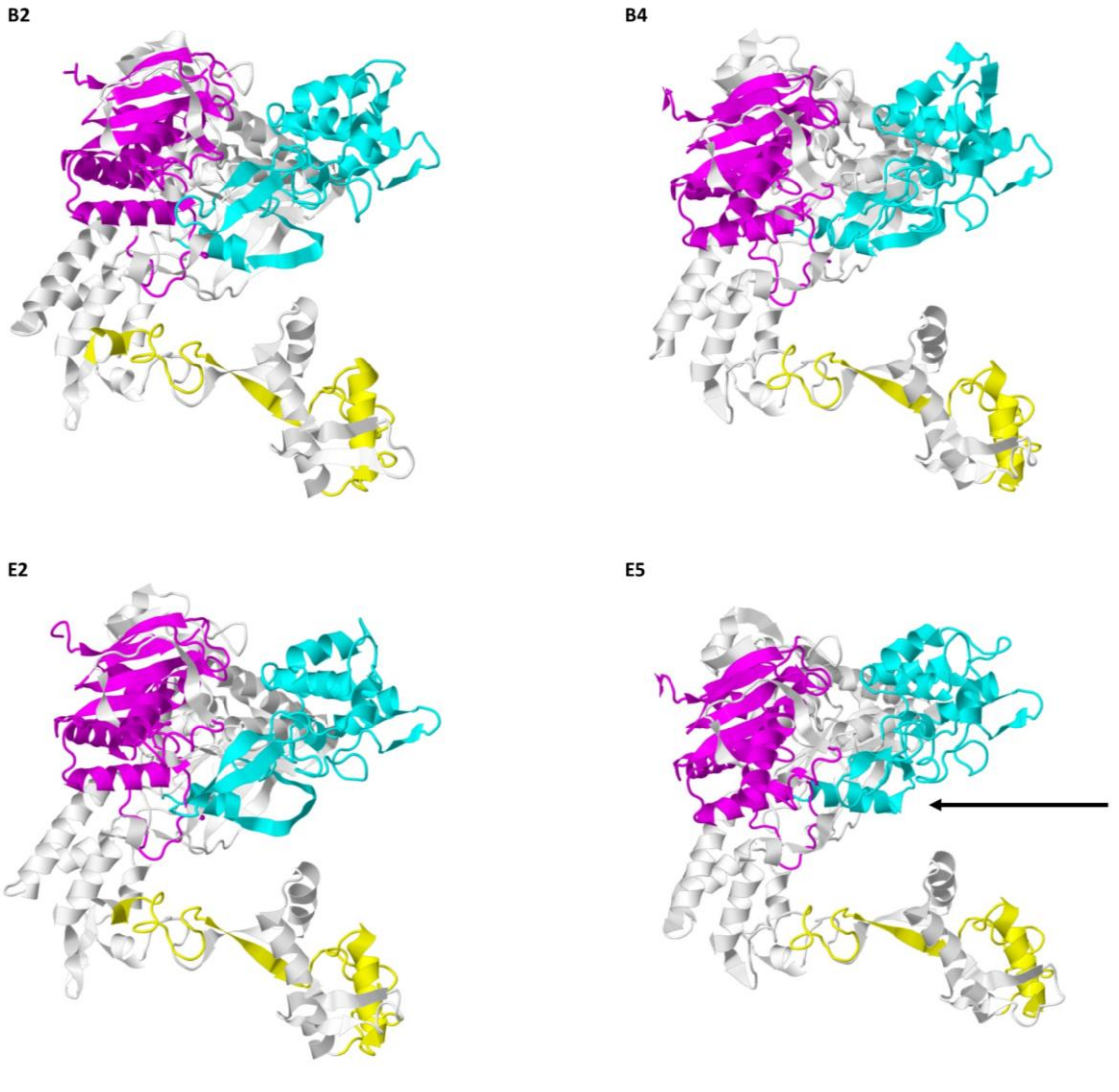


Figure 6.22: Predicted tertiary structure of the UvrA2a 2 protein of the Namib isolates. Colours: ATP binding signature I (magenta), DNA binding signature (yellow), ATP binding signature II (cyan). Black arrow indicates a protein tertiary structure position on E5 that looks different from B2, B4 and E2. Tertiary structures predicted using ‘normal modelling’ in Phyre2 V2.0 (Kelley et al. 2015) and coloured using Geneious version 2019.0.4 by Biomatters.

Figure 6.22 shows that the DNA binding region (yellow) of the UvrA2a 2 protein is very similar across the Namib isolates. However, one region of the ATP binding signature II in E5 appears different to the other isolates. This region directly proceeds the 542 aa motif seen in

Figure 6.21. In Figure 6.21, isolates B2, B4 and E2 show a coil – β-sheet – coil – U-turn – β-sheet secondary structure (aa 537 – 544) while E5 shows a coil – α-helix – β-sheet secondary structure (aa 537 – 544). The difference in the motif from 537 aa to 544 aa may cause a change in the binding ability of ATP.

A close-up of the 542 aa motif in the ATP binding signature II region (cyan) and the 199 aa motif in the DNA binding region (yellow) can be seen below in Figure 6.23 with the motifs shown in red.

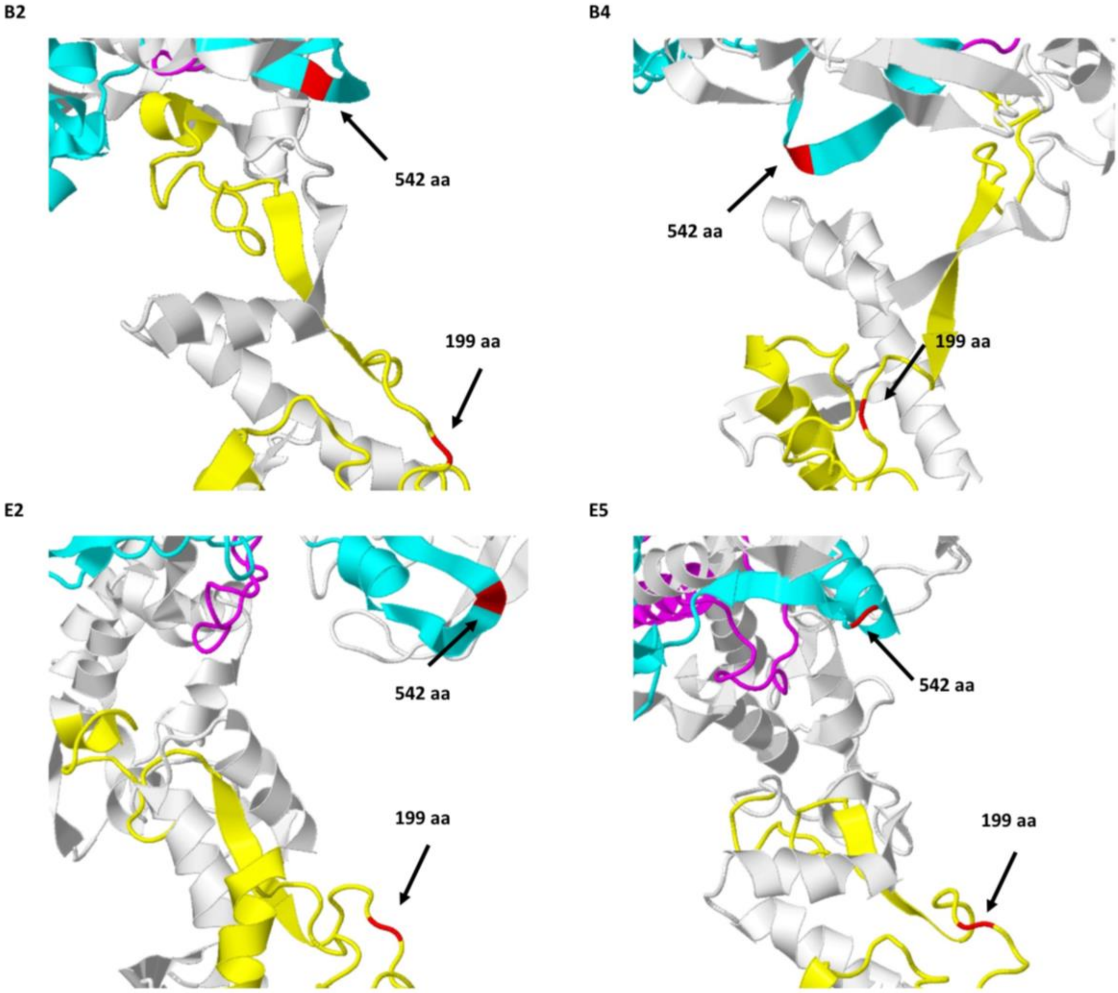


Figure 6.23: Close-up of the DNA binding region of the predicted UvrA2a 1 protein tertiary structure of the Namib isolates. Colours: ATP binding signature I (magenta), DNA binding signature (yellow), ATP binding signature II (cyan), identified motif at 199 and 542 aa (red). Tertiary structures predicted using 'normal modelling' in Phyre2 V2.0 (Kelley et al. 2015) and coloured using Geneious version 2019.0.4 by Biomatters.

As seen in Figure 6.23, the 199 aa motif in E5 does not appear to be different from the other three Namib isolates. However, the motif at 542 aa shows that where B2, B4 and E2 appear to have a U-turn in their β-sheet, isolate E5 has a coil before an α-helix. This may impact on the ability of this protein to utilise ATP. Analysis of the 624 aa motif showed that all four isolates had the same coil in the tertiary structure. This can be seen in Figure 6.24.

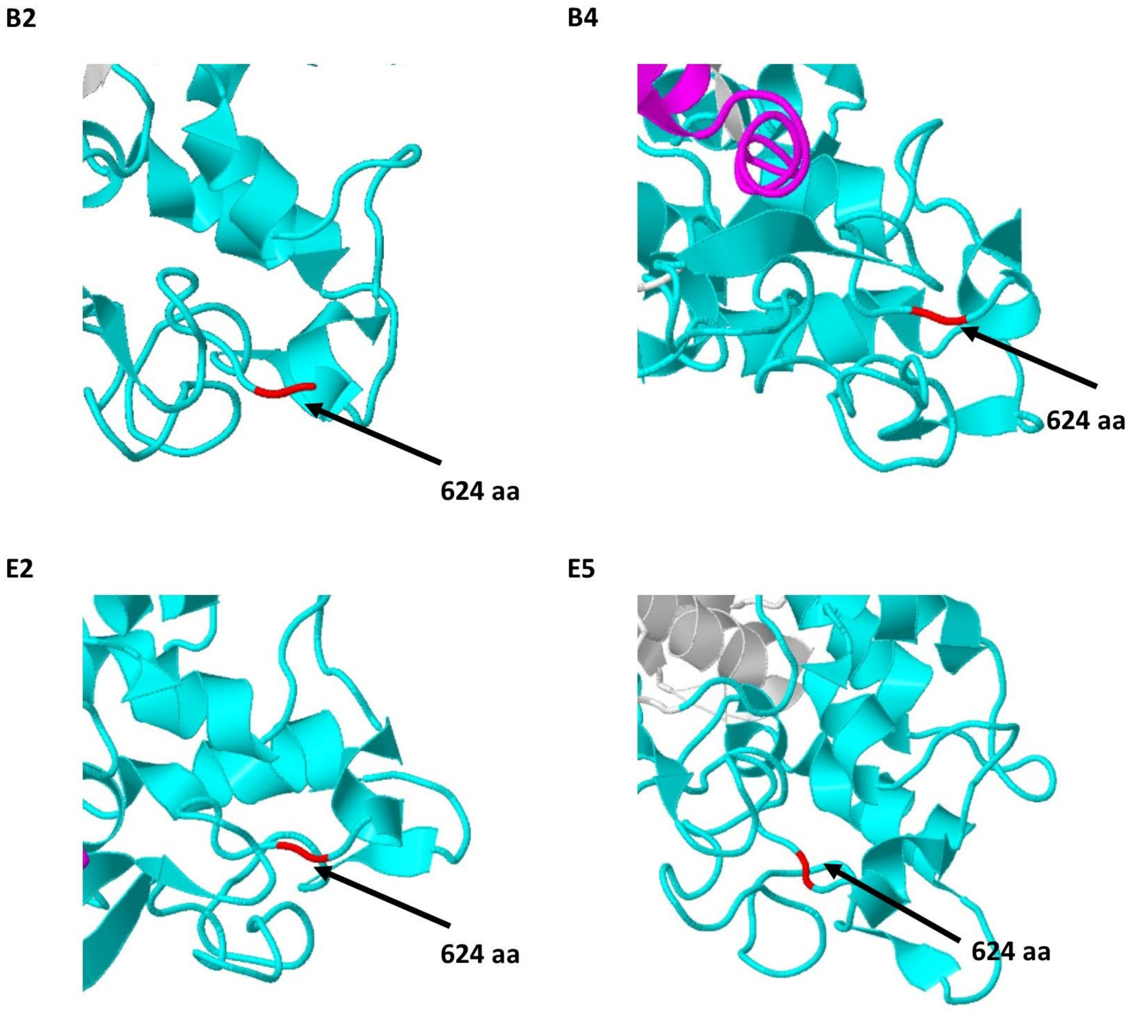


Figure 6.24: Close-up of the DNA binding region of the predicted UvrA2a 1 protein tertiary structure of the Namib isolates. Colours: ATP binding signature I (magenta), ATP binding signature II (cyan), identified motif at 624 and 542 aa (red). Tertiary structures predicted using 'normal modelling' in Phyre2 V2.0 (Kelley et al. 2015) and coloured using Geneious version 2019.0.4 by Biomatters.

Figure 6.24 shows that the identified motif at position 624 does not affect the predicted tertiary structure of E2. This indicates that this aa change is unlikely to cause E2 to display the observed UVC sensitive phenotype.

As mentioned, mutations that disrupt the C-terminal zinc finger of UvrA proteins can cause UV sensitivity (Deaconescu et al. 2012; Goosen and Moolenaar 2008; Van Houten et al. 2005). The two C-terminal zinc indicators were found for the UvrA2a 1 and UvrA2a 2 proteins. This indicates that the UV sensitivity observed in isolates E2 and E5 is most likely not caused by the absence of zinc fingers in the UvrA1 protein.

6.3.3.2 UvrB

The UvrB protein is recruited to assist in lesion removal by UvrA. As previously discussed, the UvrA2a proteins are missing their UvrB binding site. Once UvrA1 recruits UvrB, UvrB binds to the DNA and UvrC is able to associate with the UvrB-DNA complex and make an incision on the DNA.

The binding regions of UvrB are: UvrA binding (115 – 250 aa), DNA binding and bending (251 – 547 aa) and UvrA/UvrC binding (548 – 630 aa) (Hsu et al. 1995). Hsu et al. (1995) notes that both UvrA and UvrC can bind to the UvrB C-terminal binding region. Goosen and Moolenaar (2008) note that the UvrB protein is highly conserved in the β-hairpin region and UvrA binding (N-terminal) region. The UvrA/UvrC binding is the least conserved region of the protein (Goosen and Moolenaar 2008). This is supported by Deaconescu et al. (2012) who found that the residues 1 – 349 aa of the UvrB protein are essential for template strand repair without being required to be displaced by RNA polymerase.

As seen in Figure 6.13, E5 has two UvrB encoding genes. The first UvrB protein of E5 is 698 aa long which is in line with the other Namib isolates (B2 – 691 aa, B4 – 698 aa and E2 – 699 aa). The second UvrB protein of E5 is 1,051 aa long. This protein will be analysed separately. The protein alignment and secondary structure of UvrB protein of the Namib isolates can be seen in Figure 6.25 below.

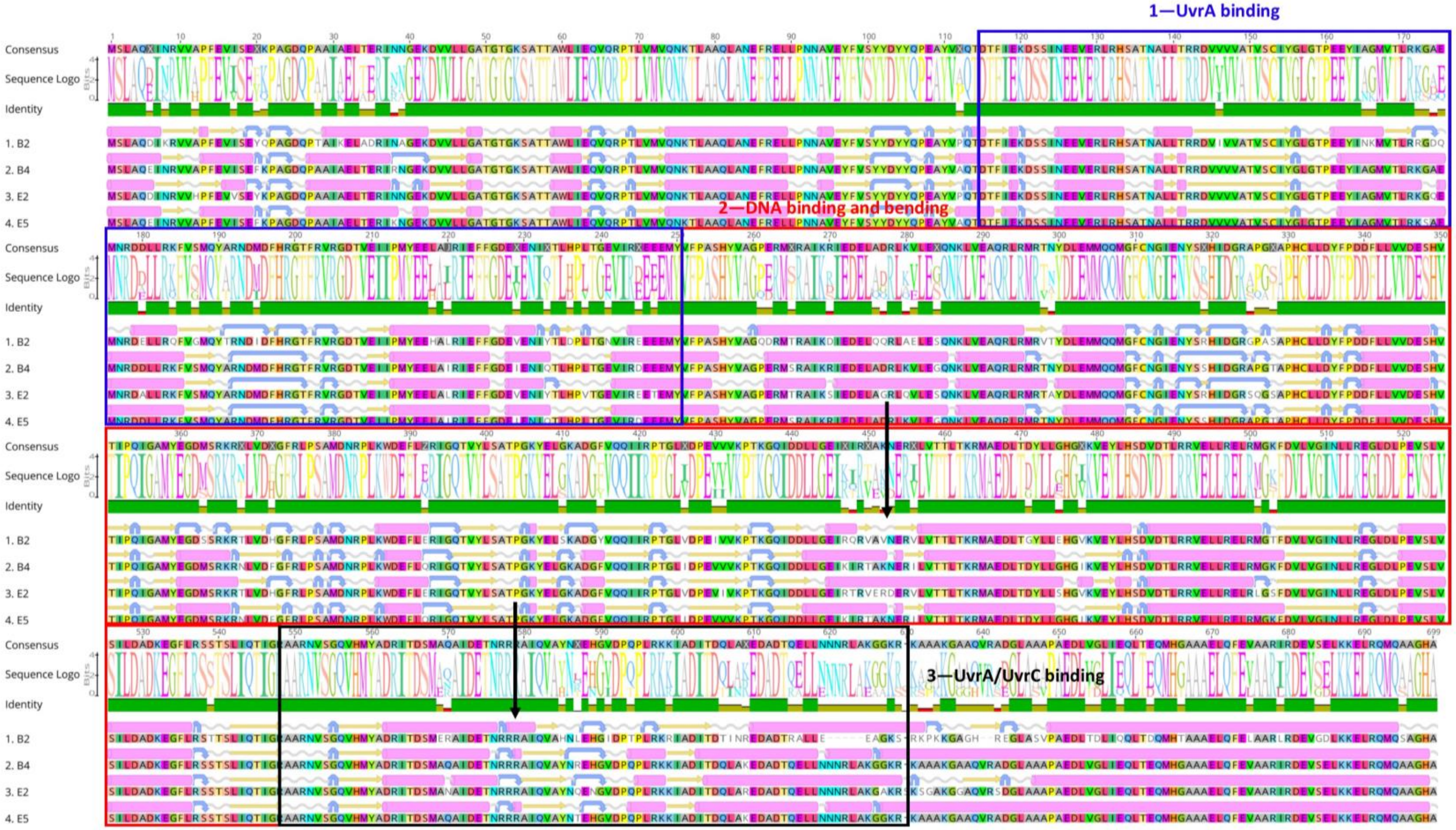


Figure 6.25: Primary and secondary structure alignment of the UvrB protein of the Namib isolates. Signature 1: UvrA binding site (115 – 250 aa) is shown in blue, signature 2: DNA binding and bending region (251 – 547 aa) is shown in red and signature 3: UvrA/UvrC binding site (548 – 630 aa) is shown in black. The coded amino acids are shown above the alignment and the consensus sequence is shown above the amino acid sequence logo. Black arrows indicate a motif that is different between the UVC resistant and UVC sensitive isolates. Coils are shown as grey lines, turns as blue U-turn arrows, α-helices are pink cylinders and β-sheets are as yellow block arrows. Secondary structures created using created using Geneious version 2019.0.4 by Biomatters.

As seen in Figure 6.25, the UvrB protein is highly conserved in the Namib isolates. There are two motifs in which isolate E2 appears to have a different predicted secondary structure compared to the other isolates. In the DNA binding and bending region, B2, B4 and E5 all have a coil at 453 aa while E2 has an α-helix. In the UvrA/UvrC binding region, E2 is missing an α-helix at 579 aa.

The predicted tertiary structures for the UvrB protein for the Namib isolates can be seen in Figure 6.26 below. The quality metrics for the predicted structure can be found in Appendix 8. The UvrA binding site (magenta), DNA binding and bending site (orange) and UvrA/UvrC binding site (cyan) are indicated.

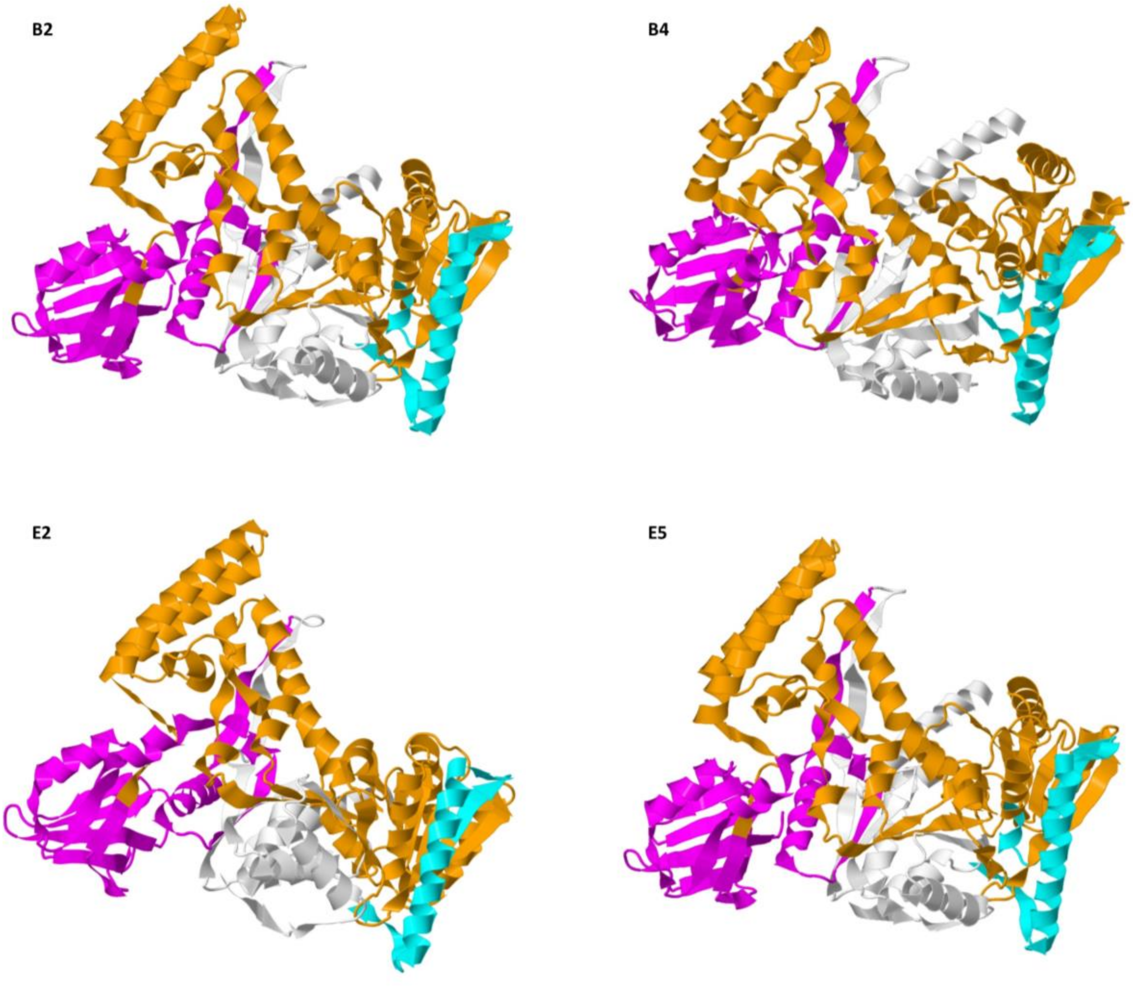


Figure 6.26: Predicted tertiary structure of the UvrB protein of the Namib isolates. Colours: UvrA binding site (magenta), DNA binding and bending site (orange), UvrA/UvrC binding site (cyan). Tertiary structures predicted using 'normal modelling' in Phyre2 V2.0 (Kelley et al. 2015) and coloured using Geneious version 2019.0.4 by Biomatters.

As seen in Figure 6.26, the UvrB protein is highly conserved amongst the Namib isolates. While there may be some slight visual variations, these are outside of the binding regions for this protein. Figure 6.27 below shows a close-up of the motifs highlighted in 6.25.

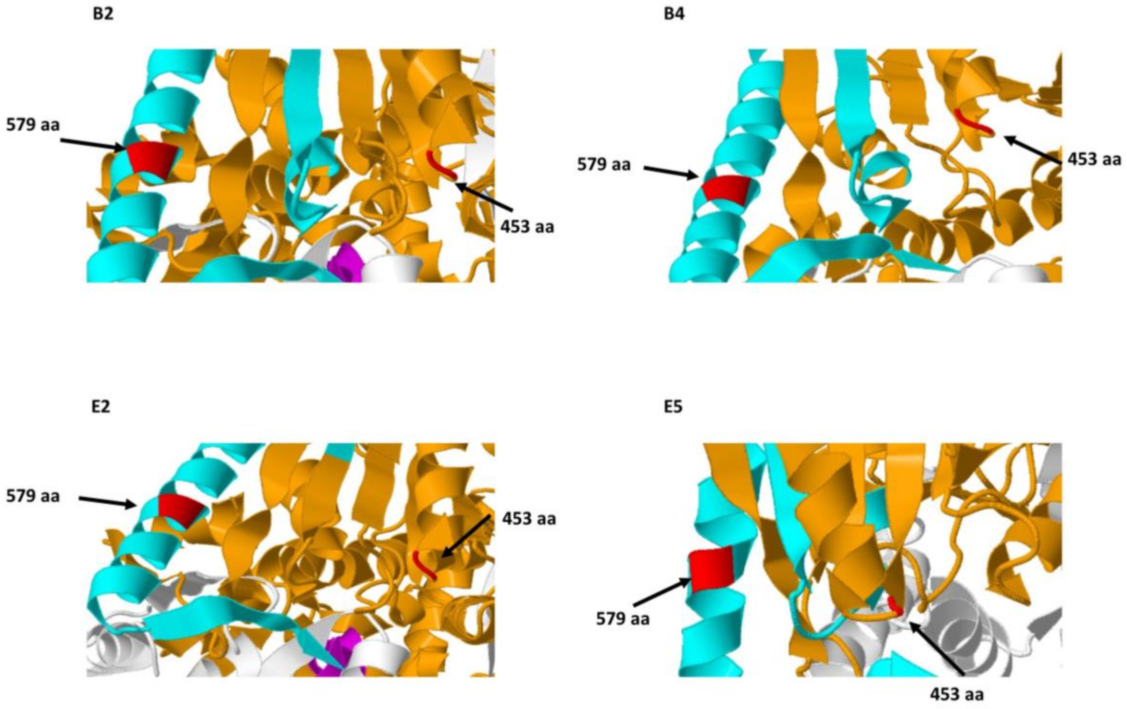


Figure 6.27: Close-up of the DNA binding region of the predicted UvrB protein tertiary structure of the Namib isolates. Colours: UvrA binding site (magenta), DNA binding and bending site (orange), UvrA/UvrC binding site (cyan), identified motif at 453 and 579 aa (red). Tertiary structures predicted using ‘normal modelling’ in Phyre2 V2.0 (Kelley et al. 2015) and coloured using Geneious version 2019.0.4 by Biomatters.

As seen in Figure 6.27 above, the motifs identified in Figure 6.25 at 453 and 579 aa do not appear to change the predicted tertiary structure of the UvrB protein in E2.

As mentioned, E5 had two UvrB genes. When aligning the second UvrB gene (locus tag GHIKLHBA_01404) to a UvrB model using Phyre2 (Kelley et al. 2015), only 593 aa were modelled out of the 1,051 aa, indicating that the UvrB protein model is not a good fit for analysing this protein. A BLASTP was carried out on the second E5 UvrB sequence. The protein shares a 90.82% similarity (99% query cover) with *P. chlorophenolicus* DUF3427 domain-containing protein (WP_015936305.1). The DUF3427 protein is functionally uncharacterised and is usually between 243 to 275 aa (EMBL-EBI 2019). Further analysis of the DUF3427 protein shows that several *Arthrobacter*, *Pseudarthrobacter* and *Paenarthrobacter* appear to have these proteins attached to the end of a PLDc_2 – ResIII – Helicase C complex like UvrB (data not shown) (EMBL-EBI 2019). As such, it cannot be assumed that the second E5 ‘UvrB’ will act in the same way as UvrB’s of typical length (~695 aa) and further classification of this protein is required to establish the role of the DUF3427 domain.

6.3.3.3 UvrC

The final protein of the UvrABC complex is UvrC. The UvrC protein is approximately 680 aa long and catalyses 3' and 5', incisions using two catalytic sites (Goosen and Moolenaar 2008). The binding regions of UvrC are: 3' incision end (1 – 100 aa), UvrB binding (205 – 239 aa), 5' incision end (341 – 494 aa), and a helix-hairpin-helix (HhH) domain (497 – 557 aa) (Karakas et al. 2007). Goosen and Moolenaar (2008) note that the UvrC protein is the least conserved of the UvrABC proteins, often only sharing 30% homology. The protein alignment and secondary structure of UvrC protein of the Namib isolates can be seen in Figure 6.28 below.

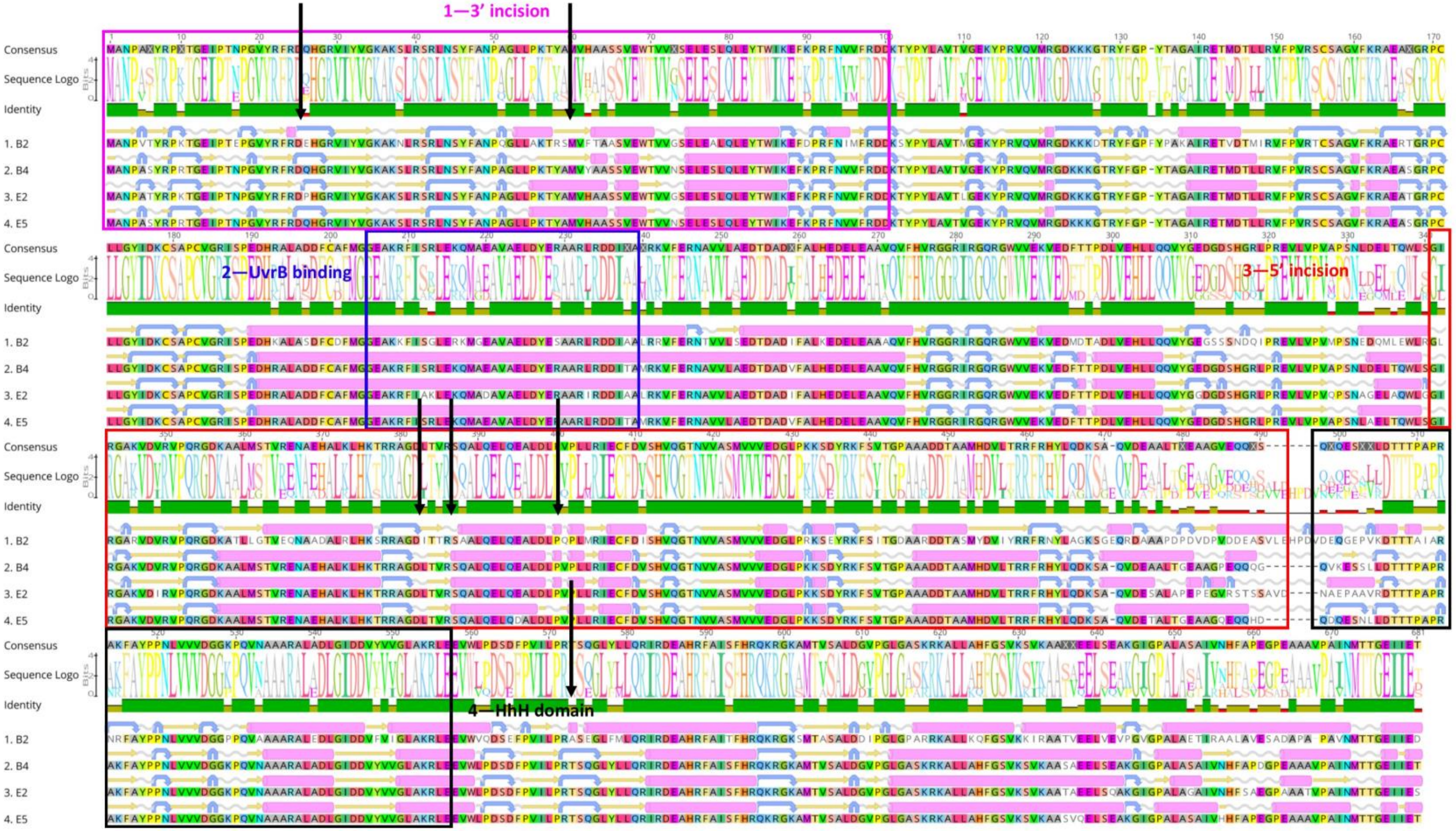


Figure 6.28: Primary and secondary structure alignment of the UvrC protein of the Namib isolates. Signature 1: 3' incision region (1 – 100 aa) is shown in pink, signature 2: UvrB binding (205 – 239 aa) is shown in blue, signature 3: 5' incision region (341 – 494 aa) is shown in red, signature 4: helix-hairpin-helix domain (497 – 557 aa) is shown in black. The coded amino acids are shown above the alignment and the consensus sequence is shown above the amino acid sequence logo. Black arrows indicate a motif that is different between the UVC resistant and UVC sensitive isolates. Coils are shown as grey lines, turns as blue U-turn arrows, α-helices are pink cylinders and β-sheets are as yellow block arrows. Secondary structures created using created using Geneious version 2019.0.4 by Biomatters.

As seen in Figure 6.28, the UvrC protein appears to be less conserved than the UvrA and UvrB proteins, as is expected. Several motifs of interest were detected, as indicated by the black arrows. Two motifs in the 3' incision signature at 25 aa and 38 – 42 aa were identified. Three motifs in the 5' incision signature at 383 aa, 386 aa and 400 aa were also identified. Finally, one motif at 573 aa was identified, however, this motif is outside of the known active regions of this protein.

The UvrC protein of the Namib isolates is between 672 and 680 aa in length. When uploading the UvrC proteins to the Phyre2 server under normal settings, the resulting modelled protein PBV file was between 220 and 221 aa long. This supports the findings of Goosen and Moolenaar (2008) whereby approximately 30% homology is observed within the UvrC protein. The proteins were then uploaded again using the 'intensive' modelling mode to ensure greater modelling coverage. As discussed previously (Section 2.7.3) Phyre2 'intensive mode' was used to model the UvrC proteins. The 'intensive mode' of Phyre2 predicts the protein structure *ab initio*, making the resulting model unreliable (Kelley et al. 2015). Due to this, less than 40% of the aa of these proteins were modelled to a reference when checking for the modelling quality. It was therefore not possible to accurately calculate the average quality score for the UvrC proteins. Instead, the residues modelled at >90% confidence by Phyre2 are reported in Appendix 8. The resulting UvrC proteins of the Namib isolates should therefore be analysed with caution.

The predicted tertiary structures for the UvrC protein for the Namib isolates can be seen in Figure 6.29 below. The 3' incision end (magenta), UvrB binding (yellow), 5' incision end (cyan), and HhH domain (red) are indicated.

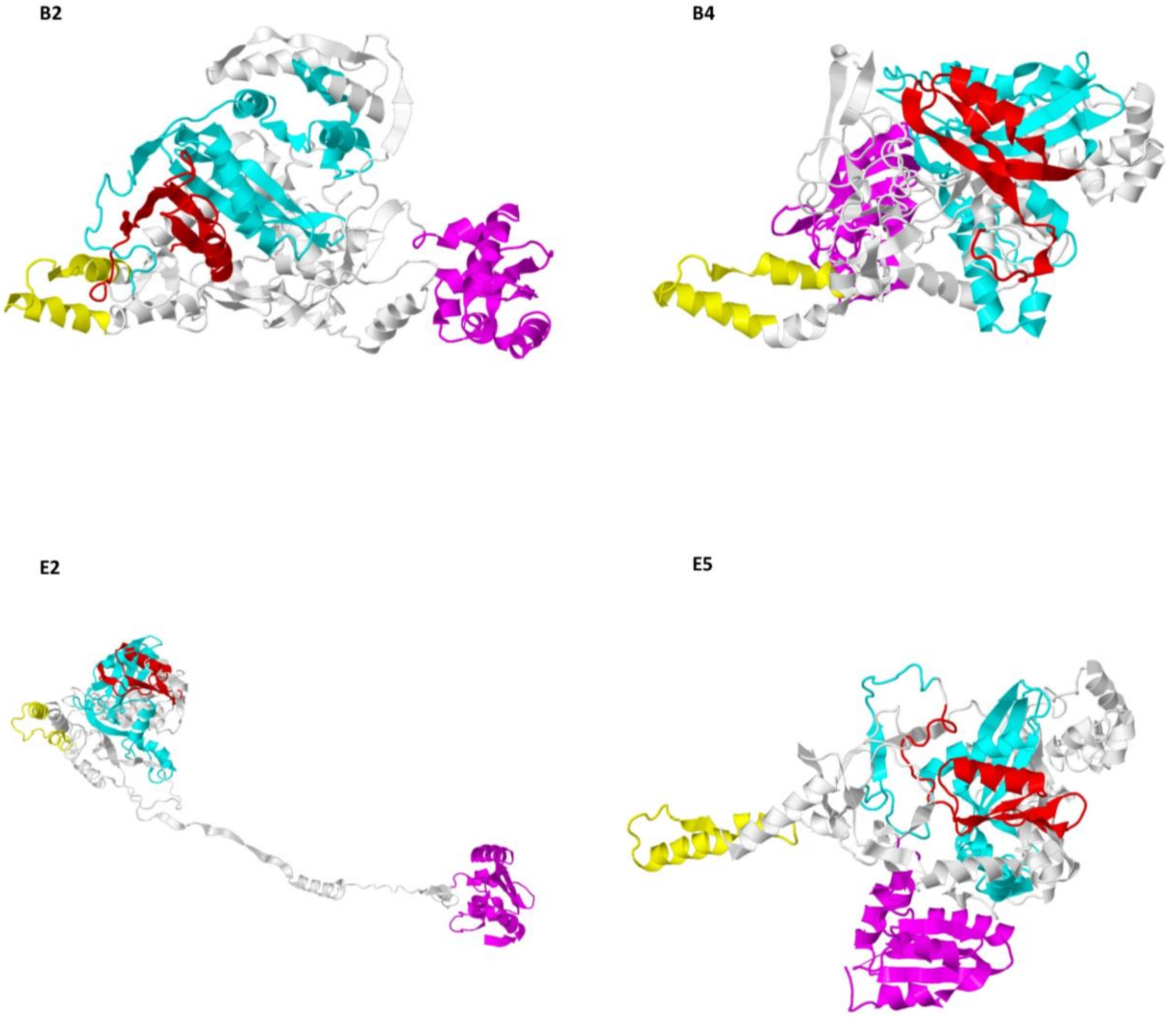


Figure 6.29: Predicted tertiary structure of the UvrC protein of the Namib isolates. Colours: 3' incision region (magenta), UvrB binding (yellow), 5' incision region (cyan), helix-hairpin-helix domain (red). A is B2, B is B4, C is E2 and D is E5. Tertiary structures predicted using ‘intensive modelling’ in Phyre2 V2.0 (Kelley et al. 2015) and coloured using Geneious version 2019.0.4 by Biomatters.

As seen in Figure 6.29 above, the predicted UvrC protein structure of the Namib isolates is highly variable compared to the UvrA and UvrB proteins. The 3' incision region (magenta) appears to be in different regions on each of the Namib isolates. Notably, the UvrC protein of isolate E2 appears as a very different model from the other UvrC proteins. Following the 3' incision region, to protein appears to have an ‘unfolded’ chain of aa, leading to the unusual shape. Observation of the region between the 3' incision and the UvrB binding region of UvrC in E2 (Figure 6.28) does not appear to show any differences in the secondary structure between B4, E5 and E2, making the prediction of this protein interesting. If this folding is accurate, this may be a contributing factor to the sensitivity of the E2 isolate. The unusual folding of the E2 protein is most likely caused by a point mutation that results in a lack of a salt bridge. Salt bridges are created between the anionic carboxylate ($RCOO^-$) of glutamic acid (protein symbol E) or aspartic acid (protein symbol D) and the cationic ammonium (RNH_3^+) of lysine (protein symbol K), arginine (protein symbol R), histine (protein symbol H), tyrosine (protein symbol Y) or serine (protein symbol S) (Pagni 2006). Analysis of the protein alignment between B2, B4, E2 and E5 revealed seven potential salt bridge-forming aa differences between E2 and the other Namib isolates. These are shown as yellow blocks below in Figure 6.30.

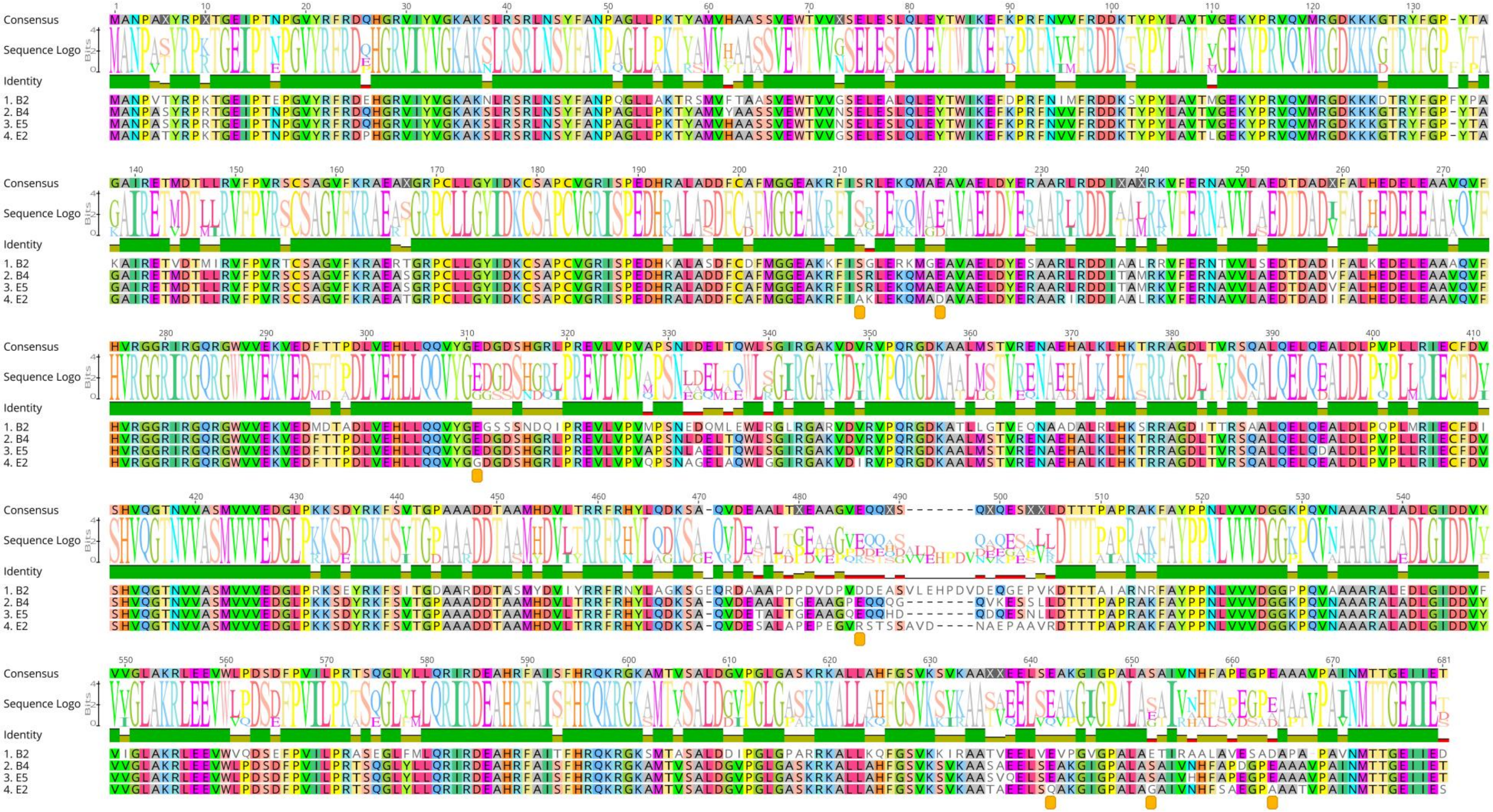


Figure 6.30: Primary structure alignment of the UvrC protein of the Namib isolates. Changes in salt-bridge forming amino acids between E2 and the other Namib isolates is indicated as yellow blocks. Yellow blocks are at 212, 220, 311, 486, 642, 652 and 664 aa of the protein alignment.

As shown in Figure 6.30 above, seven motifs are highlighted to show changes to salt bridge forming amino acids that may explain the overall predicted shape of the E2 UvrC protein. Two motifs (212 and 220 aa) are within the UvrB binding region and one motif (486 aa) is within the 5' incision region of the UvrC protein. The seven motifs at 212, 220, 311, 486, 642, 652 and 664 aa are highlighted in Figure 6.31 below.

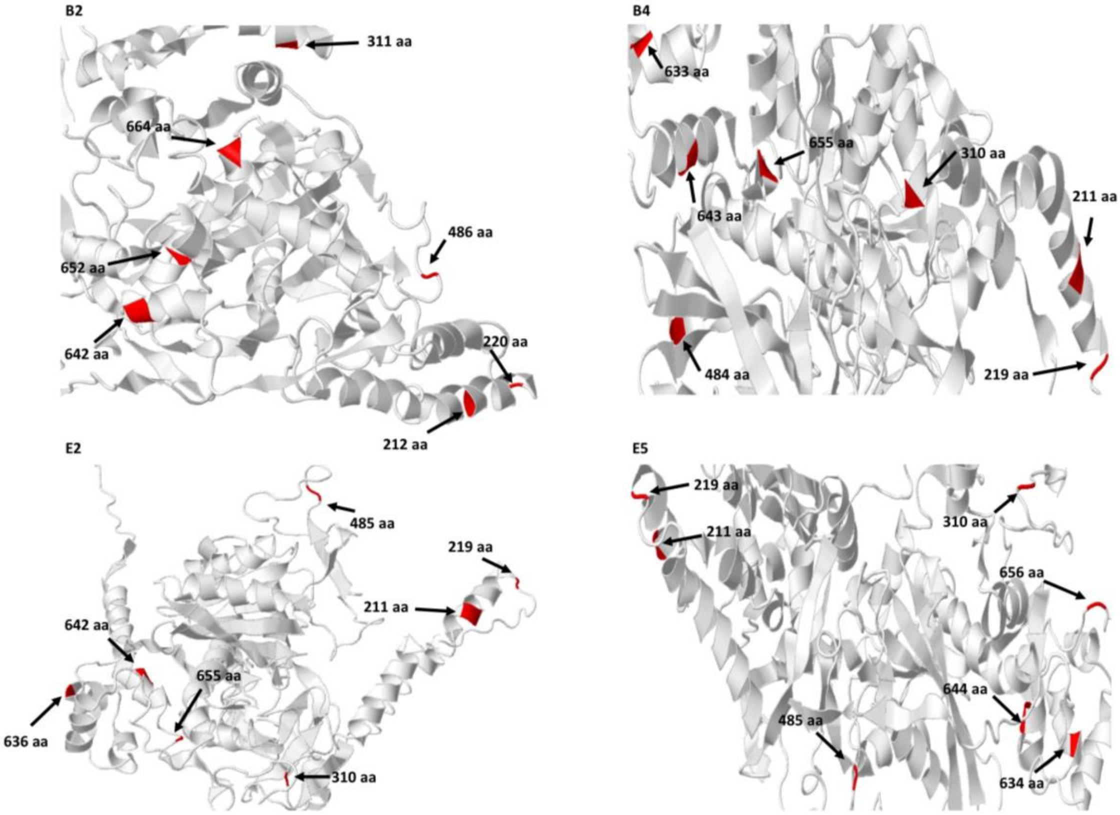


Figure 6.31: Close-up of identified motifs that are different between E2 and the other Namib isolates. Tertiary structures predicted using ‘intensive modelling’ in Phyre2 V2.0 (Kelley et al. 2015) and coloured using Geneious version 2019.0.4 by Biomatters.

As seen in Figure 6.31 above, the seven identified motifs of Figure 6.30 do not appear to be responsible for the unusual folding of the E2 UvrC protein themselves. However, a further motif at 167 aa (Figure 6.30) reveals that E2 has a threonine (T) substitution instead of an S. B2 also has this T substitution at 167 aa, however, B2 also has a salt bridge forming R at 166 aa. Further analysis of the UvrC protein revealed that the aa at 166 (167 in B2) does not form a suspected salt bridge with the same aa in each of the Namib isolates. The 166 aa appears to form a suspected salt bridge with 652 aa for E5, 264 aa for B4 and 448 aa for B2. As previously discussed, the UvrC protein has low homology. As identified, there is a motif at 664 where B2, B4 and E5 all have an anionic carboxylate (RCOO$^-$) amino acid, while E2 does not. E2 does have two anionic carboxylates at 638 and 639, which may be able to form a salt bridge in place of the missing anionic carboxylate at 664 aa.

It is hypothesised that the point mutation at 167 aa in E2 causes a lack of salt bridge with the aa at 664, resulting in an unusual protein structure. To further investigate this hypothesis, the aa at 166 in the E2 protein was *in silico* ‘reverse mutated’ to a salt bridge-forming S (Figure 6.32b) and uploaded to the Phyre2 server. The two predicted tertiary structures of E2 with a T (Figure 6.32c) and an S (Figure 6.32d) can be seen below.

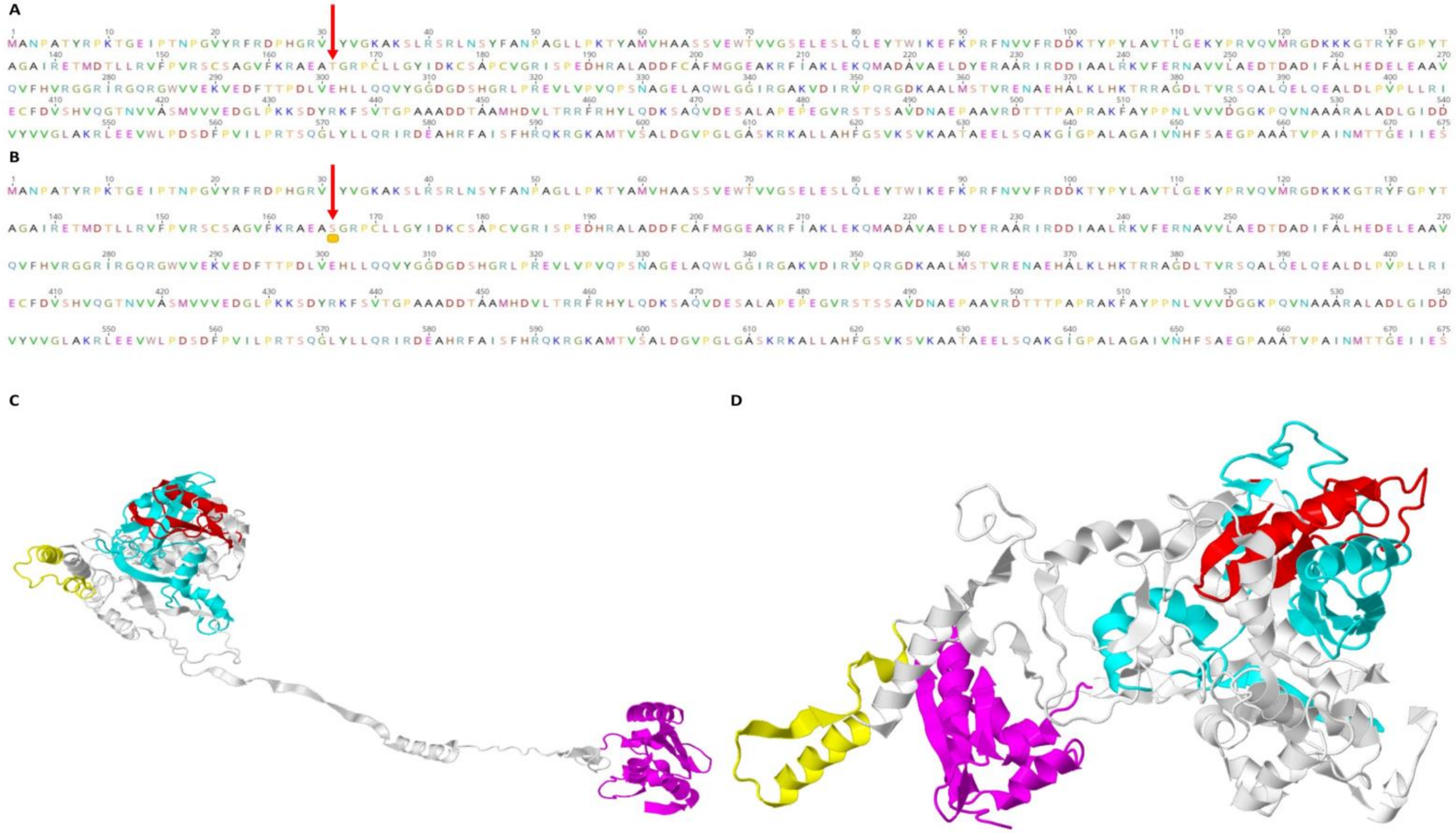

A
MANPATYRPKTGEIPTNPGVYRFRDPHGRV YVGKAKSLRSRLNSYFANPAGLLPKTYAMVHAASSVEWTVVGSELESLQLEYTWIKEFKPRFNVVFRDDKTYPYLAVTLGEKYPRVQVMRGDKKKGTRYFGPYT
AGAIRETMDTLLRVFPVRSCSAGVFKRAEATGRPCLLGYIDKCSAPCVGRISPEDHRALADDFCAFMGGEAKRFIAKLEKQMADAVAELDYERAARIRDDIAALRKVFERNAVVLAEDTDADIFALHEDELEAAV
QVFHVRGGRIRGQRGWVVEKVEDFTTPDLVEHLLQQVYGGDGDSHGRLPREVLVPVQPSNAGELAQWLGGIRGAKVDIRVPQRGDKAALMSTVRENAEHALKLHKTRRAGDLTVRSQALQELQEALDLPVPLLRI
ECFDVSHVQGTNVVASMVVVEDGLPKKSDYRKFSVTGPAAADDTAAMHDVLTRRFRHYLQDKSAQVDESALAPEPEGVRSTSSAVDNAEPAAVRDTTTPAPRAKFAYPPNLVVVDGGKPQVNAAARALADLGIDD
VYVVGLAKRLEEVWLPDSDFPVILPRTSQGLYLLQRIRDEAHRFAISFHRQKRGKAMTVSALDGVPGLGASKRKALLAHFGSVKSVKAATAEELSQAKGIGPALAGAIVNHFSAEGPAAATVPAINMTTGEIIES
B
MANPATYRPKTGEIPTNPGVYRFRDPHGRV YVGKAKSLRSRLNSYFANPAGLLPKTYAMVHAASSVEWTVVGSELESLQLEYTWIKEFKPRFNVVFRDDKTYPYLAVTLGEKYPRVQVMRGDKKKGTRYFGPYT
AGAIRETMDTLLRVFPVRSCSAGVFKRAEASGRPCLLGYIDKCSAPCVGRISPEDHRALADDFCAFMGGEAKRFIAKLEKQMADAVAELDYERAARIRDDIAALRKVFERNAVVLAEDTDADIFALHEDELEAAV
QVFHVRGGRIRGQRGWVVEKVEDFTTPDLVEHLLQQVYGGDGDSHGRLPREVLVPVQPSNAGELAQWLGGIRGAKVDIRVPQRGDKAALMSTVRENAEHALKLHKTRRAGDLTVRSQALQELQEALDLPVPLLRI
ECFDVSHVQGTNVVASMVVVEDGLPKKSDYRKFSVTGPAAADDTAAMHDVLTRRFRHYLQDKSAQVDESALAPEPEGVRSTSSAVDNAEPAAVRDTTTPAPRAKFAYPPNLVVVDGGKPQVNAAARALADLGIDD
VYVVGLAKRLEEVWLPDSDFPVILPRTSQGLYLLQRIRDEAHRFAISFHRQKRGKAMTVSALDGVPGLGASKRKALLAHFGSVKSVKAATAEELSQAKGIGPALAGAIVNHFSAEGPAAATVPAINMTTGEIIES
C
D

Figure 6.32: Primary (A and B) and predicted tertiary structure (C and D) of the E2 UvrC protein. A) Original primary structure of the E2 UvrC protein with T at aa 166; B) mutated primary structure of the E2 UvrC protein with S at aa 166; C) predicted tertiary structure of the E2 UvrC protein with T at aa 166; D) predicted tertiary structure of the E2 UvrC protein with S at aa 166. Colours: 3' incision region (magenta), UvrB binding (yellow), 5' incision region (cyan), helix-hairpin-helix domain (red). Tertiary structures predicted using 'intensive modelling' in Phyre2 V2.0 (Kelley et al. 2015) and coloured using Geneious version 2019.0.4 by Biomatters.

As seen in Figure 6.32d the ‘reverse mutated’ E2 UvrC protein with an S folds in a similar way to the UvrC protein of the other Namib isolates, which is assumed to be the correct folding. As such, the point mutation at 166 aa from a salt bridge-forming amino acid to a non-salt bridge-forming amino acid, appears to result in an incorrectly folded protein that is assumed dysfunctional. To further confirm which aa in the E2 UvrC protein forms a salt bridge with aa 166, the E2 protein were observed again with the 166 aa highlighted in green. This can be seen below in Figure 6.33.

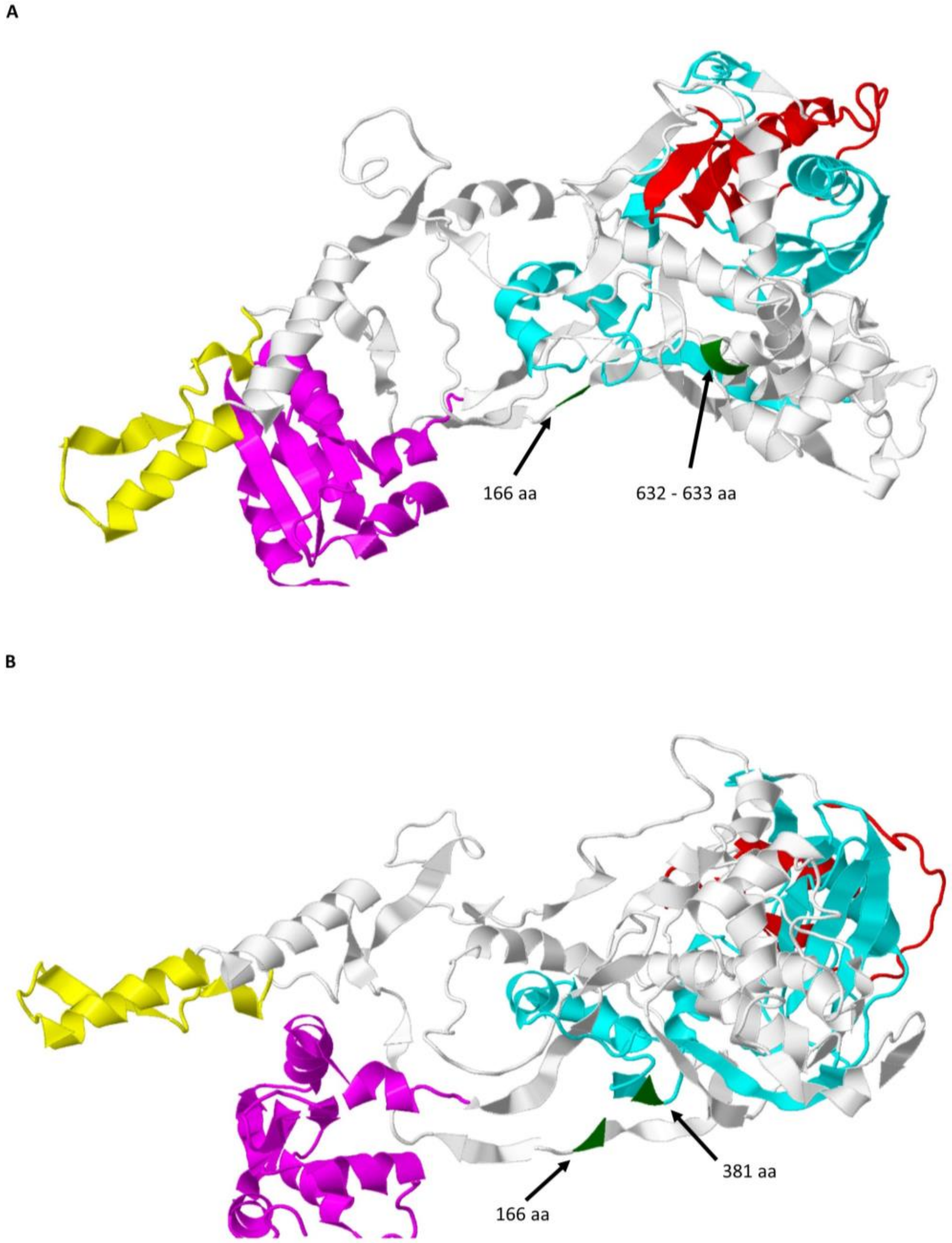


Figure 6.33: Predicted tertiary structure of the of the E2 UvrC protein. A) Predicted tertiary structure of the E2 UvrC protein with S at aa 166 and suspected salt bridge-forming 632 – 633 aa motifs indicated in green. B) Predicted tertiary structure of the E2 UvrC protein with S at aa 166 and suspected salt bridge-forming 381 aa motifs indicated in green. Colours: 3' incision region (magenta), UvrB binding (yellow), 5' incision region (cyan), helix-hairpin-helix domain (red). Tertiary structures predicted using ‘intensive modelling’ in Phyre2 V2.0 (Kelley et al. 2015) and coloured using Geneious version 2019.0.4 by Biomatters.

It was initially hypothesised that aa 166 may form a salt bridge with the 632 – 633 aa. However, as seen in Figure 6.33a, the 166 aa and 632 – 633 aa are too far apart to form a salt bridge. Further analysis of the mutated E2 UvrC protein showed that aa 381 appeared close to aa 166 (Figure 6.33b) and may be responsible for the new salt bridge formation in the predicted tertiary structure.

However, as mentioned, since there is low homology with the UvrC protein, accurate modelling is made more difficult. This may lead to an inaccurate representation of the UvrC protein within the Namib isolates. Further analysis using X-ray crystallography may resolve the observed structures of the UvrC protein in the Namib isolates.

Further analysis of the two motifs in the 3' incision region at 25 aa and 38 – 42 aa were identified in Figure 6.28 was also carried out. Figure 6.34 below shows a close-up of these motifs.

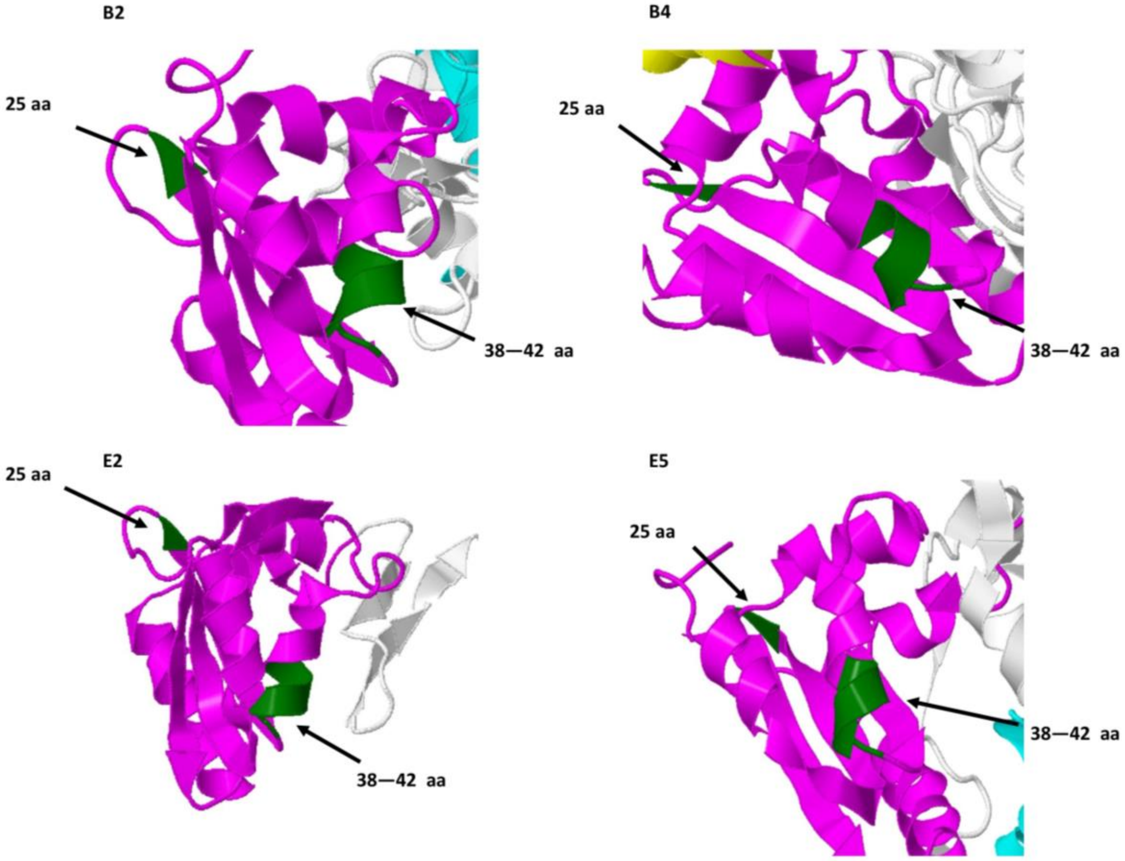


Figure 6.34: Close-up of the 3' incision region of the predicted tertiary structure of the UvrC protein of the Namib isolates. Colours: 3' incision region (magenta), UvrB binding (yellow), 5' incision region (cyan), identified motifs (green). Tertiary structures predicted using 'intensive modelling' in Phyre2 V2.0 (Kelley et al. 2015) and coloured using Geneious version 2019.0.4 by Biomatters.

As seen in Figure 6.34, the identified motifs in the 3' incision region at 25 aa and 38 – 42 aa do not appear to affect the predicted tertiary structure of the UvrC protein in the Namib isolates. Figure 6.35 below shows the three motifs in the 5' incision region at 383 aa, 386 aa and 400 aa, and the motif outside of the active sites at 573 aa.

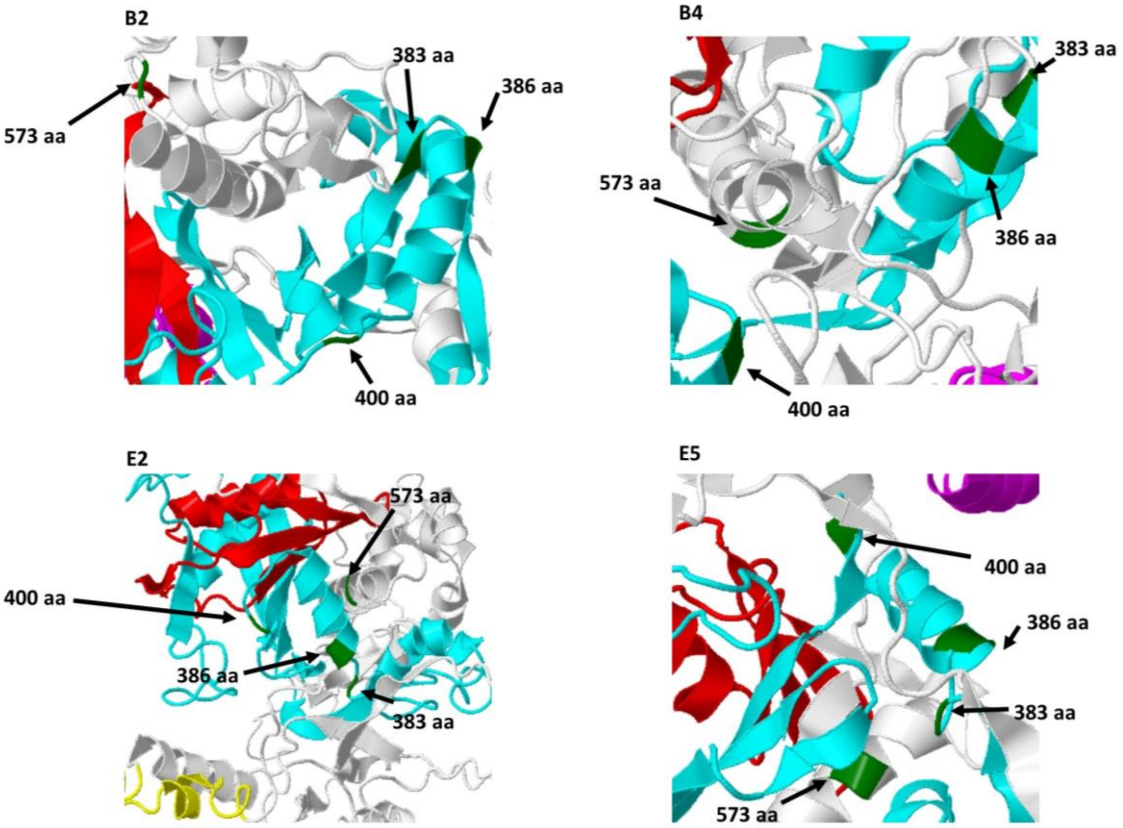


Figure 6.35: Close-up of the 5' incision region of the predicted tertiary structure of the UvrC protein of the Namib isolates. Colours: 3' incision region (magenta), UvrB binding (yellow), 5' incision region (cyan), helix-hairpin-helix domain (red), identified motifs (green). A is B2, B is B4, C is E2 and D is E5. Tertiary structures predicted using 'intensive modelling' in Phyre2 V2.0 (Kelley et al. 2015) and coloured using Geneious version 2019.0.4 by Biomatters.

As seen in Figure 6.35, the 386 motif is an α-helix in each of the Namib isolates. Both the 400 aa and 573 aa motifs are a coil in B2 and E2, but they are an α-helix in B4 and E5. This does not indicate that these motifs may be responsible for UV resistance or sensitivity, as they are found in one UV resistant and one UVC sensitive pair. However, the 386 aa motif is an α-helix in the UVC resistant isolates B2 and B4, but it is a coil in the UVC sensitive isolates E2 and E5. The 5' incision region plays an important part in excising the nucleotides surrounding the damaged DNA (Goosen and Moolenaar 2008). If this motif plays a key role in either binding to the 5' region of the damaged DNA, or in the excision of damaged DNA, this may cause the protein to become less efficient, and may result in a UVC sensitive phenotype. As noted by Jaciuk et al. (2011) mutations in only a few residues can cause non-binding within a protein.

Previous studies have identified strictly conserved residues that are important for UvrA, UvrB and UvrC function in *T. maritima*, *Bacillus* sp. and *D. radiodurans* (Goosen and Moolenaar 2008; Jaciuk et al. 2011; Van Houten et al. 2005). However, to date, no studies have focused on the UvrABC proteins of Actinobacteria. The important residues for protein-DNA binding and protein-protein binding can therefore not be determined or investigated in this study. As noted, the UvrC protein in particular has a homology of approximately 30%, highlighting a need for a larger database of the UvrABC proteins with the Actinobacteria phylum.

6.3.4 Summary of results

The UVC resistant isolates B2 and B4 do not appear to share orthologous UV repair genes that are not also shared by the UVC sensitive isolates E2 and E5. However, it was found that the UV resistant isolates B2 and B4 share CYP with each other and *A. agilis* 4041: a protein that was not detected in isolate E2, nor was a homologous protein found in E5. While another CYP protein was found in E5, it was determined that the E5 CYP protein was not orthologous with the B2 and B4 CYP protein as verified by OrthoVenn and a BLAST search of the E5 genome. The author theorises that the CYP protein of B2 and B4 may help in repairing DNA damage in these two UV resistant isolates. While the protein similarity of the B2 and B4 CYP was low with other UV resistant isolates (Figure 6.10), the protein similarity was conserved between B2, B4 and other *Arthrobacter* that had the CYP protein (Figure 6.11).

Genome analysis revealed several genes associated with BER, NER and recombinational repair in the four Namib isolates (Figure 6.13). However, these genes were also identified in the reference *Arthrobacter* and *Pseudarthrobacter* genomes of which the UV resistance or sensitivity is unknown. Many these genes were also found in *M. luteus* NCTC 2665and *K. rosea* ATCC 186, suggesting that these genes are not specific to desert microorganisms. Isolate E2 does not appear to encode for *splB*; a protein that repairs UV induced photodamage. The absence of this protein may account for the phenotype observed in Chapter 4.

The primary, secondary and predicted tertiary structure of the UvrABC proteins of the NER pathway were investigated within the Namib isolates. It was found that isolates B2, B4 and E5 had three UvrA proteins [called UvrA1, UvrA2a 1 and UvrA2a 2], while isolate E2 had two UvrA proteins [called UvrA1 and UvrA2a 2]. Isolate B2, B4 and E2 had one UvrB protein while E5 had two UvrB proteins. Further investigation found that one of the UvrB proteins in E5 (locus tag GHIKLHBA_01404) is more likely to be a DUF3427 protein instead of UvrB. Currently, the DUF3427 protein has an unknown function. The UvrA and UvrB proteins appeared to be well conserved in the Namib isolates. UvrC was less conserved, as described in 6.3.3.3.

6.3.5 Reliability of results

This study used three different approaches for identifying genes that may be involved in bacterial UV resistance within *Arthrobacter* and *Pseudarthrobacter* spp. OrthoVenn and Mauve were employed to distinguish differences between the UV resistant B2 and B4 genomes and the UVC sensitive E2 and E5 genomes. Prokka files were used to identify the presence or absence of known UV repair genes, and Phyre2 was then used to observe the predicted tertiary structure of the UvrABC proteins of the Namib isolates. Each technique has advantages and disadvantages.

OrthoVenn uses an algorithm to identify orthologous genes between organisms. The e-value used during OrthoVenn analysis was set at a default of 1e-2. This e-value was set at a low limit to capture as many shared proteins between the isolates as possible. However, this may have excluded proteins only present in B2 and B4, but not E2 or E5, from detection. As the e-value was set at a low accuracy value, it is possible that some genes may have been mis-assigned as orthologous. Following OrthoVenn carried out during this research, OrthoVenn2 was released (Xu et al. 2019). OrthoVenn2 offers a comparative analysis of orthologous clusters from up to 12 species, while the original OrthoVenn (Wang et al. 2015) was only able to compare six species at one time. As such, only two reference genomes: *A. agilis* 4041 and *P. phenanthrenivorans* Sphe3, could be used for comparison against the Namib isolates using OrthoVenn (now OrthoVenn1). However, while OrthoVenn2 allows more species to be compared at one time, the Venn diagram feature (Figure 6.1, Section 6.3) still only allows for visual comparison of six species. Thus, the use of OrthoVenn1 in this thesis is still contemporary.

The use of OrthoVenn revealed that isolates B2 and B4 shared a homologue with *A. agilis* 4041 called cytochrome P450 (CYP). As discussed, CYPs are highly variable and have a range of different biological functions. To examine if the CYPs found in B2 and B4 were truly absent in E2 and E5, Mauve was utilised. Mauve is able to align homologous sites within a genome and identify the genes within the region as local collinear blocks. Through Mauve analysis, it was found that both isolate E2 and E5 were missing the CYP genes found in the B2 and B4 CYP regions. This supports the findings of the OrthoVenn1 analysis in which CYP genes in E2 and E5 were not found. A limitation of this analysis is that both OrthoVenn1 and Mauve utilise homology with a reference genome to identify orthologous genes. OrthoVenn1 will compare the homology of uploaded species while Mauve is used for a comparative analysis of a query sequence against a reference genome. To counteract this, the eggNOG annotation files obtained in Chapter 5 were used to search for cytochrome P450 in isolates E2 and E5. While E2 did not return any genes called cytochrome P450, E5 had two genes belonging to the cytochrome P450 superfamily. However, these genes (locus tags GHIKLHBA_00506 and GHIKLHBA_00507) shared low homology with the B2 and B4 CYPs. While eggNOG also used the homology algorithm HMMER, it is based on all available Actinobacteria protein sequences, making the results more specific to *Arthrobacter* and *Pseudarthrobacter* (Section 5.2.6, Chapter 5).

Phyre2 was used to examine the predicted tertiary structure of the UvrABC proteins. This software also relies on protein similarity. If similarity is not detected between the submitted protein sequence and a sequence of known structure, the Phyre2 modelling becomes unreliable (Kelley et al. 2015). This issue was encountered during analysis of the UvrC protein of the Namib isolates. The UvrC protein shares low sequence similarity with members of different species (Goosen and Moolenaar 2008). As such, the use of intensive modelling was able to predict the tertiary structure of UvrC; however, the reliability of the predicted structure when using ‘intensive’ modelling is based solely on the primary sequence (Kelley et al. 2015). In addition, while Phyre2 is able to predict the phenotypic result of a point mutation, the software is unable to predict the structural consequences of a point mutation (Kelley et al. 2015). Despite this, it was shown in Figure 6.30 (Section 6.3, Chapter 6) that a single *in silico* ‘reverse’ point mutation at aa 166 in the UvrC protein of E2 resulted in a normal predicted tertiary structure using Phyre2. Further examination using X-ray crystallography would help to reconcile the true structure of the E2 UvrC protein.

The present study can provide genomic insights into how isolates B2 and B4 may be able to resist UVC radiation over isolates E2 and E5, but it presents no information on whether the proteins are actually expressed and functional. This would require further analysis at a proteomics level and checking for enzyme activity.

6.4 Conclusions

This study was conducted to determine if soil bacteria from arid deserts were resistant to UV radiation and if a genomics approach could be used to provide insights into possible molecular DNA repair mechanisms. A range of organisms were found to be resistant to UV radiation (Table 4.4, Section 4.3); however, the UVC resistant isolates B2 and B4 and the UVC sensitive isolates E2 and E5 were ultimately chosen for further analysis. The Namib isolates belong to either the *Arthrobacter* or *Pseudarthrobacter* genera; members of both genera were used as reference genomes in this study (Table 5.2). All reference genomes and the Namib isolates had at least one complete NER and recombinational repair pathway. However, variability within the *ligC*, *ssb* and *uvrA* gene copy number indicates that gene repair using these pathways may differ between organisms. Within the Namib isolates the copy number of repair proteins was similar; however, the protein sequence similarity was variable. The absence of the *splB* gene in isolate E2 may contribute to this organisms UV sensitivity. Analysis of the genomes of the four Namib isolates revealed three possible candidate genes for the differences in UV resistance and sensitivity observed in Chapter 4. Firstly, it is theorised that the presence of the homologous CYP in B2 and B4 may help these two isolates to repair DNA damage at a more efficient rate than the UVC sensitive isolates E2 and E5. This is supported by the presence of CYP in other known UV resistant genomes. Secondly, the absence of the *splB* gene in the E2 genome may cause E2 to be more sensitive to UV radiation that the other isolates. Finally, the predicted tertiary structure of the UvrC protein in isolate E2 appears to be misfolded. If this predicted tertiary structure of the E2 UvrC protein is accurate, this may contribute towards the observed UVC sensitive phenotype.

While there is evidence to suggest some genomic differences between the UVC resistant and UVC sensitive genotypes of the Namib isolates, at this stage these theories are unable to be proven. Further application of transcriptomics to determine the up and down regulation of UV resistance genes and the activity of these proteins needs to be investigated.

Chapter 7: General discussion and future directions

This thesis has made an original contribution to science by examining both the phenotypic diversity of bacteria from the Dry Valleys and the Namib Desert, as well as providing a genomics insight into the UV repair genes of novel *Arthrobacter* and *Pseudarthrobacter* isolated from the Namib Desert.

7.1 Isolation and diversity of UV resistant isolates from desert systems

In Chapter 3, the bacterial 16S rRNA gene-defined community of the Dry Valleys and the Namib Desert were investigated. Actinobacteria and Proteobacteria were the dominant phyla in most of the desert locations. The soil chemistry of sites T4 and T6 prevented DNA extraction. The ASVs found within the Namib Desert appeared to be quite distinct from one another, compared to the communities of the Dry Valleys, which appeared to be more consistent across the six locations. Site V2 had a distinct population of ASVs that did not appear to be in any of the other Dry Valley sample sites. The α-diversity of the sites showed that the Chao1 (species abundance) and Shannon's index (species diversity) were variable, but not distinct between the two deserts. PCoA revealed that the difference in communities could be explained by PCoA 1, with the Dry Valley sample sites clustering separately from the Namib Desert sample sites. PCoA 2 could not explain the difference between the Dry Valleys and the Namib Desert. Finally, RDA was conducted using the ASV communities and the soil chemistry obtained from the desert samples. There was no significant variability in microbial community structure observed due to the tested environmental variables. This suggests that stochasticity plays a role in the community assembly of the Namib Desert and the Dry Valleys.

Culturable bacteria were isolated onto agar in Chapter 4. The majority of bacteria isolated in this study were isolated and grew at 15°C and 20°C. The isolated bacteria were then grown in broth and exposed to UV radiation using a developed rapid screening method. From this trial, 16 isolates were deemed UVC resistant. Using the 16S rRNA gene, these isolates were preliminary identified, and their phylogenetic relationships were investigated and compared with UVC sensitive isolates. It was found that several UVC resistant and UVC sensitive isolates shared close phylogeny based on their 16S rRNA gene. As Actinobacteria were the dominant phyla in the Namib Desert locations, the UV resistance and sensitivity of this phyla was of interest for this research. To determine if any phenotypic differences could be linked to genotypic differences, the UVC resistant isolates B2, B4 and the UVC sensitive isolates E2 and E5 were selected from the *Arthrobacter* and *Pseudarthrobacter* genera for further analysis. The UV survivability of two phylogenetically similar species has not previously been reported. As the 16S rRNA gene phylogeny did not display strong support for the phylogenetic relationship between *Arthrobacter* and *Pseudarthrobacter*, it was crucial to assign accurate taxonomy to isolates B2, B4, E2 and E5 using whole genome comparative analysis.

7.2 Taxonomic classification of the Namib isolates

Following the selection of B2, B4, E2 and E5 for further analysis, whole genome phylogenomic comparisons were conducted to confirm the relationship between the four isolates and other *Arthrobacter* and *Pseudarthrobacter* species. Comparative genomic analysis using a range of whole genome techniques revealed that all four Namib isolates are genetically distinct from each other, and from the named reference genomes of *Arthrobacter* and *Pseudarthrobacter*. Using PhyloPhlAn, B2 was found to group in the same clade as *A. agilis,* while B4, E2 and E5 were found to group with *A. nitrophenolicus*. However, according to the OrthoANI and *in silico* DDH values, both B4 and E5 shared more similarity with *P. phenanthrenivorans* Sphe3 than *A. nitrophenolicus* SJCon. This analysis highlights an interesting problem with the classification of *Arthrobacter* and *Pseudarthrobacter*. Using whole genome sequencing, it is difficult to phylogenetically classify *Arthrobacter* and *Pseudarthrobacter* due to the non-monophyletic clades formed by the two genera. This has previously been reported (Nouioui et al. 2018) however, to the authors knowledge, this is the first study to report on phylogenetic analysis of the *Arthrobacter* and *Pseudarthrobacter* genera using the whole genome information. Due to the issue of B4, E2 and E5 clustering with *A. nitrophenolicus* and having high similarity with *P. phenanthrenivorans*, it is not possible to assume the genus of B4, E2 or E5 currently. As the only difference in classification between *Arthrobacter* and *Pseudarthrobacter* is the peptidoglycan structure, peptidoglycan analysis must be conducted before assigning these isolates to a genus.

Mauve analysis of the Namib isolates showed that B2 appears to have some gene loss when compared to *A. agilis* 4041, resulting in a smaller genome (3.87 Mb for *A. agilis* 4041 and 3.26 Mb for B2). Similarly, Mauve analysis showed that B4 (3.81 Mb), E2 (4.03 Mb) and E5 (3.83 Mb [genome only]) were smaller than *P. phenanthrenivorans* Sphe3 (4.25 Mb), *P. chlorophenolicus* A6 (4.4 Mb), *A. enclensis* NIO-1008 (4.23 Mb) and *A. nitrophenolicus* SJCon (4.39 Mb). It appears that B4, E2 and E5 have all experienced some gene loss. This may be due to the restrictive nature of the Namib Desert, resulting in the loss of genes not required for immediate survival in the arid soil. Alternatively, genomic coverage may have been missed due to the Namib isolates being draft genomes.

7.3 Genomic insights into UV survivability and the UvrABC protein structure

The genome sequences of the four Namib isolates were further analysed to determine if there were survival genes present on the UVC resistant isolates that were absent on the UVC sensitive isolates in Chapter 6. OrthoVenn analysis of the Namib isolates and *A. agilis* 4041 and *P. phenanthrenivorans* Sphe3 revealed that B2, B4 and *A. agilis* 4041 share the protein cytochrome P450 (CYP). Additionally, the absence of this protein was further confirmed by analysis of the B2 and B4 CYP regions in E2 and E5. While the function of this protein cannot be determined currently, CYPs were also observed in other UV resistant organisms, providing an avenue for further study using transcriptome and functional analysis.

The presence and absence of known DNA repair proteins were observed across the Namib isolates and the reference genomes (Figure 6.13) and the UvrABC proteins of the NER pathway were selected for further analysis. Isolates B2, B4 and E5 had one UvrA1 protein and two UvrA2a proteins, while E2 had one UvrA1 protein and one UvrA2a protein. Through Phyre2 protein folding analysis, it was found that the UvrA and UvrB proteins are highly conserved between the Namib isolates. While there were some motifs that varied between the UVC resistant and UVC sensitive isolates, visual analysis of these motifs did not appear to hinder the overall function of the protein. Further proteomic analysis of all four isolates is required to confirm this.

The structure of the UvrC protein however, appeared quite different between the Namib isolates. In particular, isolate E2 had an unusual UvrC protein structure, and it is assumed by the structure that this protein may have difficulty functioning. This may contribute to the phenotype of UV sensitivity observed in E2. The predicted tertiary structure of the E2 UvrC protein changed to ‘normal’ folding when 166 aa was exchanged from non-salt bridge forming threonine (T) to a salt-bridge forming serine (S). This indicates that there is likely a salt-bridge forming amino acid around 165 to 166 aa contributing to the protein fold. However, this would have to be confirmed using X-ray crystallography.

At this stage, phenotypic differences cannot be linked to genotypic differences in the Namib isolates. This thesis has only focused on one DNA repair pathway. Further tertiary structure analysis of other repair pathways in the Namib isolates would provide a more comprehensive picture of the genotypic relationship to the observed phenotype.

7.4 Future directions

This study is the first to investigate UV resistance in *Arthrobacter* and *Pseudarthrobacter* from the Namib Desert. Investigations into UV resistance and DNA repair in microorganisms from arid locations is important in continuing our understanding the diversity and adaptation of UV repair systems. However, several aspects require further investigation or work. These include:

- **Generation of a complete genome of the Namib isolates using MinION or PacBio sequencing.** The generation of a complete genome will help to close the genomes and may provide more information regarding the genomes of the Namib isolates.

- **Peptidoglycan, respiratory quinone and polar lipid analysis of the Namib isolates to establish and formally describe them as new species.** As noted in Chapter 5, the genomes of the Namib isolates appear to be novel, as determined by OrthoANI and *in silico* DDH. Differentiating between *Arthrobacter* and *Pseudarthrobacter* relies on the peptidoglycan structure (A3α (Lys-Ala$_{2-3}$) for *Arthrobacter* and A3α (Lys-Ser-Thr-Ala) for *Pseudarthrobacter*). As noted in Chapter 5, B4, E5 and E2 share homology with *A. nitrophenolicus* SJCon and *P. phenanthrenivorans* Sphe3. Both *A. nitrophenolicus* SJCon and *P. phenanthrenivorans* Sphe3 appear in the same clade, but have different peptidoglycan structures, as per their genus criteria. It is therefore essential to conduct peptidoglycan, respiratory quinone and polar lipid analysis to establish if B4, E5 and E2 belong to the *Arthrobacter* or *Pseudarthrobacter* genera.

- **Environment vs agar.** During this thesis, UV exposure was conducted on plate medium. While the agars used were low nutrient agars, the presence of additional nutrients may have influenced some bacteria to act differently than in their natural environment. In addition, it is possible that other bacteria from the Namib Desert and the Dry Valleys could have greater UVC resistance than the four Namib isolates, but these were not identified since they could not be cultured. Further studies into the transcriptome of the Namib isolates within desert soil during UV exposure may provide a more accurate representation of the UV resistance of these isolates.

- **Identification of transcription factors that may bind to and regulate UV repair genes.** The binding of transcription factors influences gene regulation. The differences between the UVC resistant B2 and B4 and the UVC sensitive E2 and E5 may be due to a mutation in a transcription factor binding site. Further investigation using transcriptomics using RNA sequencing, promoter tags and promoter exchange experiments may reveal this. Investigations into the up and down regulation of genes through RNA sequencing may help determine key differences between the UV resistant and UVC sensitive isolates.

- **Functional specificity of repair proteins.** Establishing the functional specificity of repair proteins could be determined through proteomic studies. This would involve using affinity purification coupled with mass spectrometry to characterise the functions of UV repair proteins. Furthermore, investigation of the promoter regions using fluorescent tags, or promoter mutations, may help to determine the efficiency of a particular repair gene.

Damage to the promoter region may cause the protein to go unexpressed and could be a deterministic factor in the UVC sensitivity observed in E2 and E5. This could be confirmed by genetically modifying a model organism, such as *E. coli*, to determine if a different promoter before specific UV repair genes increases the ability of *E. coli* to survive UV radiation.

- **Identification of the role of CYP within the B2 and B4 genomes**. RNA sequencing would assist this investigation. If CYP is upregulated during UV exposure, further investigations into the role of CYP could be achieved firstly through gene knock out or mutation of the CYP gene. If B2 and B4 become more sensitive to UV radiation following this, this may indicate that CYP plays a role in UV resistance.

- **X-ray crystallography of the UvrABC proteins in *Arthrobacter* and *Pseudarthrobacter***. As noted in Chapter 6, the binding sites of UvrABC are known in *T. maritima*, *Bacillus* sp. and *D. radiodurans*, but currently there is no X-ray crystallography structure for these proteins in Actinobacteria. As also noted in Chapter 6, the predicted tertiary structure of the UvrC protein in the Namib isolates was highly variable, with the E2 UvrC protein appearing to have an 'unfolded' chain following the 3' incision region. Using X-ray crystallography may help determine if the E2 UvrC protein was predicted incorrectly by Phyre2, or if the predicted tertiary structure is accurate and the E2 UvrC is dysfunctional.

Summary

In summary, this thesis demonstrated that closely related organisms from desert environments can vary phenotypically and genotypically. A comparative genomics approach was able to identify differences between the Namib isolates and between the available reference genomes. Gaining insight into UV resistance using a genomics approach was challenging and as such, genomic differences cannot be linked to phenotype at this stage. DNA repair genes were found in the Namib isolates, as well as in the reference *Arthrobacter* and *Pseudarthrobacter* genomes, suggesting that these genes are beneficial for growth in a range of environments. Further work is required, as detailed in Section 7.4, to understand how B2 and B4 came to be UVC resistant isolates while E2 and E5 are UVC sensitive.

References

Abrevaya XC, Paulino-Lima IG, Galante D, Rodrigues F, Mauas PJD, Cortón E, Lage CAS. 2011. Comparative Survival Analysis of Deinococcus radiodurans and the Haloarchaea natrialba magadii and Haloferax volcanii, Exposed to Vacuum Ultraviolet Irradiation. Astrobiology.11(10):1034-1040.

Aislabie JM, Chhour K-L, Saul DJ, Miyauchi S, Ayton J, Paetzold RF, Balks MR. 2006. Dominant bacteria in soils of Marble Point and Wright Valley, Victoria Land, Antarctica. Soil Biology and Biochemistry. 38(10):3041-3056.

Aislabie JM, Jordan S, Barker GM. 2008. Relation between soil classification and bacterial diversity in soils of the Ross Sea region, Antarctica. Geoderma. 144(1):9-20.

Albarracín VH, Gärtner W, Farias ME. 2013. UV Resistance and Photoreactivation of Extremophiles from High-Altitude Andean Lakes.

Albarracín VH, Simon J, Pathak GP, Valle L, Douki T, Cadet J, Borsarelli CD, Farías ME, Gärtner W. 2014. First characterisation of a CPD-class I photolyase from a UV-resistant extremophile isolated from High-Altitude Andean Lakes. Photochemical and Photobiological Sciences. 13(5):739-750.

Anderson MJ, Willis TJ. 2003. Canonical Analysis of Principal Coordinates: A Useful Method of Constrained Ordination for Ecology. Ecology. 84(2):511-525.

Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M et al. 2004. UniProt: the Universal Protein knowledgebase. Nucleic Acids Res. 32(Database issue):D115-119.

Arahal DR, Pujalte MJ, Rodrigo-Torres L. 2016. Draft genomic sequence of Nereida ignava CECT 5292T, a marine bacterium of the family Rhodobacteraceae. Standards in Genomic Sciences. 11(1):21.

Aranda J, Bardina C, Beceiro A, Rumbo S, Cabral MP, Barbé J, Bou G. 2011. Acinetobacter baumannii RecA Protein in Repair of DNA Damage, Antimicrobial Resistance, General Stress Response, and Virulence Journal of Bacteriology. 193(15):3740-3747.

Archer SDJ. Forthcoming 2017. UV recordings in Antarctica, 2017.

Archer SDJ, de los Ríos A, Lee KC, Niederberger TS, Craig CS, Coyne KJ, Douglas S, Lacap-Bugler DC, Pointing SB. 2017. Endolithic microbial diversity in sandstone and granite from the McMurdo Dry Valleys, Antarctica. Polar Biology. 40(5):997-1006.

Archer SDJ, Lee KC, Caruso T, Maki T, Lee CK, Cary SC, Cowan DA, Maestre FT, Pointing SB. 2019. Airborne microbial transport limitation to isolated Antarctic soil habitats. Nature Microbiology. 4(3).

Archer SDJ, McDonald IR, Herbold CW, Lee CK, Cary CS. 2015. Benthic microbial communities of coastal terrestrial and ice shelf Antarctic meltwater ponds. Frontiers in Microbiology. 6(485).

Armstrong A. 2014. Seasonal Dynamics of Edaphic Bacterial Communities in the Hyper-Arid Namib Desert. [South Africa]: University of the Western Cape.

Armstrong A, Valverde A, Ramond J-B, Makhalanyane TP, Jansson JK, Hopkins DW, Aspray TJ, Seely M, Trindade MI, Cowan DA. 2016. Temporal dynamics of hot desert microbial communities reveal structural and functional responses to water input. Scientific Reports. 6.

Arora PK, Jain RK. 2013. Arthrobacter nitrophenolicus sp. nov. a new 2-chloro-4-nitrophenol degrading bacterium isolated from contaminated soil. 3 Biotech. 3(1):29-32.

Aslam SN, Dumbrell AJ, Sabir JS, Mutwakil MHZ, Baeshen MMN, Abo-Aba SEM, Clark DR, Yates SA, Baeshen NA, Underwood GJC et al. 2016. Soil compartment is a major determinant of the impact of simulated rainfall on desert microbiota. Environmental Microbiology. 18(12):5048-5062.

Auch AF, von Jan M, Klenk H-P, Göker M. 2010. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Standards in Genomic Sciences. 2(1):117-134.

Ayora S, Carrasco B, Cárdenas PP, César CE, Cañas C, Yadav T, Marchisone C, Alonso JC. 2011. Double-strand break repair in bacteria: a view from Bacillus subtilis FEMS Microbiology Reviews. 35(6):1055-1081.

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M et al. 2008. The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics. 9(1):75.

Balasubramanian S. 2015. Solexa Sequencing: Decoding Genomes on a Population Scale. Clinical Chemistry. 61(1):21-24.

Baliga NS, Bjork SJ, Bonneau R, Pan M, Iloanusi C, Kottemann MCH, Hood L, DiRuggiero J. 2004. Systems Level Insights Into the Stress Response to UV Radiation in the Halophilic Archaeon Halobacterium NRC-1. Genome Research. 14(6):1025-1035.

Ball MM, Gómez W, Magallanes X, Rosales R, Melfo A, Yarzábal LA. 2014. Bacteria recovered from a high-altitude, tropical glacier in Venezuelan Andes. World Journal of Microbiology and Biotechnology. 30(3):931-941.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19(5):455-477.

Bao H, Marchant D. 2006. Quantifying sulfate components and their variations in soils of the McMurdo Dry Valleys, Antarctica. J Geophys Res. 111.

Barnard RL, Osborne CA, Firestone MK. 2014. Changing precipitation pattern alters soil microbial community response to wet-up under a Mediterranean-type climate. The ISME journal. 9(4):946-957.

Bates ST, Berg-Lyons D, Caporaso JG, Walters WA, Knight R, Fierer N. 2010. Examining the global distribution of dominant archaeal populations in soil. The Isme Journal. 5:908.

Beckmann M, Václavík T, Manceur AM, Šprtová L, von Wehrden H, Welk E, Cord AF. 2014. glUV: a global UV-B radiation data set for macroecological studies. Methods in Ecology and Evolution. 5(4):372-383.

Benedict RC, Kowalczykowski SC. 1988. Increase of the DNA Strand Assimilation Activity of recA Protein by Removal of the C Terminus and Structure-Function Studies of the Resulting Protein Fragmen. The Journal of biological chemistry. 263:15513-15520.

Bentchikou E, Servant P, Coste G, Sommer S. 2010. A Major Role of the RecFOR Pathway in DNA Double-Strand-Break Repair through ESDSA in Deinococcus radiodurans. PLOS Genetics. 6(1):e1000774.

Bhasin M, Raghava GPS. 2006. Computational Methods in Genome Research. In: Arora DK, Berka RM, Singh GB, editors. Applied Mycology and Biotechnology. Elsevier.

Blakemore LC, Searle PL, Daly BK. 1981. Methods for chemical analysis of soils.

Bockheim JG. 2008. Functional diversity of soils along environmental gradients in the Ross Sea region, Antarctica. Geoderma. 144(1):32-42.

Boor KJ. 2006. Bacterial Stress Responses: What Doesn't Kill Them Can Make Them Stronger. PLoS Biology. 4(1).

Brown SD, Nagaraju S, Utturkar S, De Tissera S, Segovia S, Mitchell W, Land ML, Dassanayake A, Köpke M. 2014. Comparison of single-molecule sequencing and hybrid approaches for finishing the genome of Clostridium autoethanogenum and analysis of CRISPR systems in industrial relevant Clostridia. Biotechnology for Biofuels. 7(1):40.

Büdel B, Darienko T, Deutschewitz K, Dojani S, Friedl T, Mohr KI. 2009. Southern African Biological Soil Crusts are Ubiquitous and Highly Diverse in Drylands, Being Restricted by Rainfall Frequency. Microbial Ecology. 57(2):229-247.

Buscardo E, Geml J, Schmidt SK, Silva ALC, Ramos RTJ, Barbosa SMR, Andrade SS, Dalla Costa R, Souza AP, Freitas H et al. 2018. Of mammals and bacteria in a rainforest: Temporal dynamics of soil bacteria in response to simulated N pulse from mammalian urine. Functional Ecology. 32(3):773-784.

Busse H-J. 2016. Review of the taxonomy of the genus Arthrobacter, emendation of the genus Arthrobacter sensu lato, proposal to reclassify selected species of the genus Arthrobacter in the novel genera Glutamicibacter gen. nov., Paeniglutamicibacter gen. nov., Pseudoglutamicibacter gen. nov., Paenarthrobacter gen. nov. and

Pseudarthrobacter gen. nov., and emended description of Arthrobacter roseus. International Journal of Systematic and Evolutionary Microbiology. 66:9-37.

Busse H-J, Schumann P. 2019. Reclassification of Arthrobacter enclensis as Pseudarthrobacter enclensis comb. nov., and emended descriptions of the genus Pseudarthrobacter, and the species Pseudarthrobacter phenanthrenivorans and Pseudarthrobacter scleromae. International Journal of Systematic and Evolutionary Microbiology.

Busse H-J, Wieser M, Buczolits S. 2012. Genus III. Arthrobacter. In: William B. Whitman, Michael Goodfellow, Peter Kämpfer, Hans-Jürgen Busse, Martha E. Trujillo, Wolfgang Ludwig, Suzuki K-i, editors. Bergey's Manual of Systematic Bacteriology. 2nd ed. New York: Springer. p. 578-624.

Busse H-J, Wieser M, Buczolits S. 2015. Arthrobacter. In: Whitman WB, Rainey F, Kämpfer P, Trujillo M, Chun J, DeVos P, Hedlund B, Dedysh S, editors. Bergey's Manual of Systematics of Archaea and Bacteria.

Cabrol NA, Feister U, Häder DP, Piazena H, Grin EA, Klein A. 2014. Record solar UV irradiance in the tropical Andes. Frontiers in Environmental Science. 2.

Caliz J, Vila X, Martí E, Sierra J, Cruañas R, Garau MA, Montserrat G. 2011. Role of bacterial isolates in enhancing the bud induction in the industrially important red alga Gracilaria dura. FEMS Microbiology Ecology. 78(1):150-164.

Callahan BJ, McMurdie PJ, Holmes SP. 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. The Isme Journal. 11:2639.

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods. 13:581.

Camargo FA, Bento FM, Okeke BC, Frankenberger WT. 2004. Hexavalent chromium reduction by an actinomycete, Arthrobacter crystallopoietes ES 32. Biol Trace Elem Res. 97(2):183-194.

Cameron RE, King J, David CN. 1970. Soil microbial ecology in Wheeler Valey, Antarctica. Soil Science. (2):110-120.

Campbell EA, Westblade LF, Darst SA. 2008. Regulation of bacterial RNA polymerase σ factor activity: a structural perspective. Current Opinion in Microbiology. 11(2):121-127.

Cannone N, Wagner D, Hubberten HW, Guglielmin M. 2008. Biotic and abiotic factors influencing soil properties across a latitudinal gradient in Victoria Land, Antarctica. Geoderma. 144(1):50-65.

Caporaso J, G, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M et al. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME Journal. 6:1621-1624.

Caruso T, Chan Y, Lacap DC, Lau MCY, McKay CP, Pointing SB. 2011. Stochastic and deterministic processes interact in the assembly of desert microbial communities on a global scale. The ISME Journal. 5(9):1406-1413.

Cary SC, McDonald IR, Barrett JE, Cowan DA. 2010. On the rocks: the microbiology of Antarctic Dry Valley soils. Nature Reviews Microbiology. 8(2):129-138.

Chan Y, Lacap DC, Lau MCY, Ha KY, Warren-Rhodes KA, Cockell CS, Cowan DA, McKay CP, Pointing SB. 2012. Hypolithic microbial communities: between a rock and a hard place. Environmental Microbiology. 14(9):2272-2282.

Chan Y, Van Nostrand JD, Zhou J, Pointing SB, Farrell RL. 2013. Functional ecology of an Antarctic Dry Valley. Proceedings of the National Academy of Sciences, USA. 110(22):8990-8995.

Chauhan A, Pathak A, Jaswal R, Edwards B, III, Chappell D, Ball C, Garcia-Sillas R, Stothard P, Seaman J. 2018. Physiological and Comparative Genomic Analysis of Arthrobacter sp. SRS-W-1-2016 Provides Insights on Niche Adaptation for Survival in Uraniferous Soils. Genes (Basel). 9(1):31.

Chun J, Lee J-H, Jung Y, Kim M, Kim S, Kim BK, Lim Y-W. 2007. EzTaxon: a web-based tool for the identification of prokaryotes based on 16S ribosomal RNA gene sequences. International Journal of Systematic and Evolutionary Microbiology. 57:2259-2261.

Chuvikovsky DV, Esipov RS, Skoblov YS, Chupova LA, Muravyova TI, Miroshnikov AI, Lapinjoki A, Mikhailopulo IA. 2006. Ribokinase from E. coli: Expression, purification, and substrate specificity. Bioorganic and Medicinal Chemistry. 14(18):6327-6332.

Clauß M. 2006. Higher effectiveness of photoinactivation of bacterial spores,UV resistant vegetative bacteria and mold spores with 222 nmcompared to 254 nm wavelength. Clean - Soil Air Water. 34(6):525-532.

Cockell CS, McKay CP, Warren-Rhodes KA, Horneck G. 2008. Ultraviolet radiation-induced limitation to epilithic microbial growth in arid deserts – Dosimetric experiments in the hyperarid core of the Atacama Desert. Journal of Photochemistry and Photobiology B: Biology. 90(2):79-87.

Coelho FJRC, Santos AL, Coimbra J, Almeida A, Cunha Â, Cleary DFR, Calado R, Gomes NCM. 2013. Interactive effects of global climate change and pollution on marine microbes: the way ahead. Ecology and Evolution. 3(6):1808-1818.

Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR, Tiedje JM. 2013. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Research. 42(D1):D633-D642.

Conn HJ, Dimmick I. 1947. Soil bacteria similar in morphology to Mycobacterium and Corynebacterium. J Bacteriol. 54:291-303.

Cordero RR, Seckmeyer G, Damiani A, Riechelmann S, Rayas J, Labbe F, Laroze D. 2014. The world’s highest levels of surface UV. Photochemical & Photobiological Sciences. 13(70).

Courcelle J, Crowley DJ, Hanawalt PC. 1999. Recovery of DNA Replication in UV-Irradiated Escherichia coli Requires both Excision Repair and RecF Protein Function. Journal of Bacteriology. 181(3):916-922.

Cowan DA, Makhalanyane TP, Dennis PG, Hopkins DW. 2014. Microbial ecology and biogeochemistry of continental Antarctic soils. Frontiers in Microbiology. 5(154).

Cowan DA, Pointing SB, Stevens MI, Cary CS, Stomeo F, Tuffin MI. 2010. Distribution and abiotic influences on hypolithic microbial communities in an Antarctic Dry Valley. Polar Biology. 34(2):307-311.

Cowan DA, Ramond J-B, Makhalanyane TP, De Maayer P. 2015. Metagenomics of extreme environments. Current Opinion in Microbiology. 25:97-102.

Cowan DA, Russell NJ, Mamais A, Sheppard DM. 2002. Antarctic Dry Valley mineral soils contain unexpectedly high levels of microbial biomass. Extremophiles. 6(5):431-436.

Cowan DA, Sohm JA, Makhalanyane TP, Capone DG, Green TGA, Cary SC, Tuffin IM. 2011. Hypolithic communities: important nitrogen sources in Antarctic desert soils. Environmental Microbiology Reports. 3(5):581-586.

Cowan DA, Tow LA. 2004. Endangered antarctic environments. Annual Reviews of Microbiology. 58:649-690.

Cregger MA, Schadt CW, McDowell NG, Pockman WT, Classen AT. 2012. Response of the Soil Microbial Community to Changes in Precipitation in a Semiarid Ecosystem. Applied and Environmental Microbiology. 78(24):8587-8594.

Crovadore J, Grizard D, Chablais R, Cochard B, Blanc P, Lefort F. 2018. Whole-Genome Sequences of Two Arthrobacter sp. Strains, 4041 and 4042, Potentially Usable in Agriculture and Environmental Depollution. Microbiology resource announcements. 7(10):e01054-01018.

Crowley DJ, Boubriak I, Berquist BR, Clark M, Richard E, Sullivan L, DasSarma S, McCready S. 2006. The uvrA, uvrB and uvrC genes are required for repair of ultraviolet light induced DNA photoproducts in Halobacterium sp. NRC-1. Saline Systems. 2(11).

Daly MJ. 2009. A new perspective on radiation resistance based on Deinococcus radiodurans. Nature Reviews Microbiology. 7:237-245.

Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. Genome Research. 14(7):1394-1403.

Darling AE, Mau B, Perna NT. 2010. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. PLoS One. 5(6):e111147.

Darling AE, Miklós I, Ragan MA. 2008. Dynamics of Genome Rearrangement in Bacterial Populations. PLoS Genetics. 4(7).

Dartnell LR, Hunter SJ, Lovell KV, Coates AJ, Ward JM. 2010. Low-Temperature Ionizing Radiation Resistance of Deinococcus radiodurans and Antarctic Dry Valley Bacteria. Astrobiology. 10(7):717-732.

Dastager SG, Liu Q, Tang SK, Krishnamurthi S, Lee JC, Li WJ. 2014. Arthrobacter enclensis sp. nov., isolated from sediment sample. Archives of Microbiology. 196(11):775-782.

de Groot A, Dulermo R, Ortet P, Blanchard L, Guérin P, Fernandez B, Vacherie B, Dossat C, Jolivet E, Siguier P et al. 2009. Alliance of Proteomics and Genomics to Unravel the Specificities of Sahara Bacterium Deinococcus deserti. PLoS Genetics. 5(3).

de la Fuente G, Belanche A, Girwood SE, Pinloche E, Wilkinson T, Newbold CJ. 2014. Pros and Cons of Ion-Torrent Next Generation Sequencing versus Terminal Restriction Fragment Length Polymorphism T-RFLP for Studying the Rumen Bacterial Community. PLOS ONE. 9(7):e101435.

Deaconescu AM, Sevostyanova A, Artsimovitch I, Grigorieff N. 2012. Nucleotide excision repair (NER) machinery recruitment by the transcription-repair coupling factor involves unmasking of a conserved intramolecular interface. Proceedings of the National Academy of Sciences of the United States of America. 109(9):3353-3358.

Delgado-Baquerizo M, Oliverio AM, Brewer TE, Benavent-González A, Eldridge DJ, Bardgett RD, Maestre FT, Singh BK, Fierer N. 2018. A global atlas of the dominant bacteria found in soil. Science. 359(6373):320-325.

Demple B, Halbrook J, Linn S. 1983. Escherichia coli xth mutants are hypersensitive to hydrogen peroxide. Journal of Bacteriology. 153(2):1079-1082.

Demple B, Harrison L. 1994. Repair of Oxidative Damage to DNA: Enzymology and Biology. Annual Review of Biochemistry. 63:915-948.

Deng Y, Cui X, Hernández M, Dumont MG. 2014. Microbial Diversity in Hummock and Hollow Soils of Three Wetlands on the Qinghai-Tibetan Plateau Revealed by 16S rRNA Pyrosequencing. PLoS One. 9(7).

Dhar G, Sanders ER, Johnson RC. 2004. Architecture of the Hin Synaptic Complex during Recombination. Cell. 119(1):33-45.

Di Capua C, Bortolotti A, Farías ME, Cortez N. 2011. UV-resistant Acinetobacter sp. isolates from Andean wetlands display high catalase activity FEMS Microbiology Letters. 317(2):181-189.

Dib J, Motok J, Fernández Zenoff V, Ordoñez O, Farías ME. 2008. Occurrence of Resistance to Antibiotics, UV-B, and Arsenic in Bacteria Isolated from Extreme Environments in High-Altitude (Above 4400 m) Andean Wetlands. Current Microbiology. 56(5):510-517.

Dieser M, Greenwood M, Foreman CM. 2010. Carotenoid Pigmentation in Antarctic Heterotrophic Bacteria as a Strategy to Withstand Environmental Stresses. Arctic, Antarctic, and Alpine Research. 42(4):396-405.

Diffey BL. 2002. Sources and measurement of ultraviolet radiation. Methods. 28(1):4-13.

Doran PT, McKay CP, Clow GD, Dana GL, Fountain AG, Nylen T, Lyons WB. 2002. Valley floor climate observations from the McMurdo dry valleys, Antarctica, 1986–2000. Journal of Geophysical Research: Atmospheres. 107(D24):ACL13-11-12.

Frequently Asked Questions. 2019. Leibniz Institute DSMZ; [accessed 2019 26 March]. http://ggdc.dsmz.de/faq.php#qggdc6.

Dsouza M, Taylor MW, Turner SJ, Aislabie J. 2015. Genomic and phenotypic insights into the ecology of Arthrobacter from Antarctic soils. BMC Genomics. 16(36).

Dummer AM, Bonsall JC, Cihla JB, Lawry S, Johnson GC, Peck R. 2011. Bacterioopsin-Mediated Regulation of Bacterioruberin Biosynthesis in Halobacterium salinarum.

Dunlap CA, Schisler DA, Perry EB, Connor N, Cohan FM, Rooney AP. 2017. Bacillus swezeyi sp. nov. and Bacillus haynesii sp. nov., isolated from desert soil. International Journal of Systematic and Evolutionary Microbiology. 67(8):2720-2725.

Eckardt F, Drake N. 2010. Introducing the Namib Desert Playas. In: Öztürk M, Böer B, Barth H-J, Breckle S-W, Clüsener-Godt M, Khan MA, editors. Sabkha Ecosystems: Volume III: Africa and Southern Europe. New York, USA: Springer. p. 19-25.

Eckardt FD, Spiro B. 1999. The origin of sulphur in gypsum and dissolved sulphate in the Central Namib Desert, Namibia. Sedimentary Geology. 123(3):255-273.

Ee R, Yong D, Lim YL, Yin W-F, Chan K-G. 2015. Complete genome sequence of oxalate-degrading bacterium Pandoraea vervacti DSM 23571T. Journal of Biotechnology. 204:5-6.

Eisen JA, Hanawalt PC. 1999. A phylogenomic study of DNA repair genes, proteins, and processes. Mut Res. 435:171-213.

El Houmami N, Bakour S, Bzdrenga J, Rathored J, Seligmann H, Robert C, Armstrong N, Schrenzel J, Raoult D, Yagupsky P et al. 2017. Isolation and characterization of Kingella negevensis sp. nov., a novel Kingella species detected in a healthy paediatric population. International Journal of Systematic and Evolutionary Microbiology. 67(7):2370-2376.

Family: DUF3427 (PF11907). 2019. [accessed]. http://pfam.xfam.org/family/PF11907#familySummaryBlock.

Ensign JC, Rittenberg SC. 1963. A crystalline pigment produced from 2-hydroxypyridine by Arthrobacter crystallopoietes n.sp. Archiv für Mikrobiologie. 47(2):137-153.

Eschbach M, Möbitz H, Rompf A, Jahn D. 2003. Members of the genus Arthrobacter grow anaerobically using nitrate ammonification and fermentative processes: anaerobic adaptation of aerobic bacteria abundant in soil FEMS Microbiology Letters. 223(2):227-230.

Fadrosh DW, Ma B, Gajer P, Sengamalay N, Ott S, Brotman RM, Ravel J. 2014. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. Microbiome. 2(6).

Fernández Zenoff V, Heredia J, Ferrero M, Siñeriz F, Farías ME. 2006a. Diverse UV-B Resistance of Culturable Bacterial Community from High-Altitude Wetland Water. Current Microbiology. 52(5):359-362.

Fernández Zenoff V, Siñeriz F, Farías ME. 2006b. Diverse Responses to UV-B Radiation and Repair Mechanisms of Bacteria Isolated from High-Altitude Aquatic Environments. Applied Environmental Microbiology. 72(12):7857-7863.

Ferrari BC, Binnerup SJ, Gillings M. 2005. Microcolony Cultivation on a Soil Substrate Membrane System Selects for Previously Uncultured Soil Bacteria. Applied and Environmental Microbiology. 71(12):8714-8720.

Fierer N, Leff JW, Adams BJ, Nielsen UN, Bates ST, Lauber CL, Owens S, Gilbert JA, Wall, D. H, Caporaso JG. 2012. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. PNAS. 109(52):21390-21395.

Filippidou S, Wunderlin T, Junier T, Jeanneret N, Dorador C, Molina V, Johnson DR, Junier P. 2016. A Combination of Extreme Environmental Conditions Favor the Prevalence of Endospore-Forming Firmicutes. Frontiers in Microbiology. 7(1707).

Fleming A. 1922. On a remarkable bacteriolytic element found in tissues and secretions. Proceedings of the Royal Society B.306-317.

Flores MR, Ordoñez OF, Maldonado MJ, Farías ME. 2009. Isolation of UV-B resistant bacteria from two high altitude Andean lakes (4,400 m) with saline and non saline conditions. The Journal of General and Applied Microbiology. 55(6):447-458.

Fong N, Burgess M, Barrow K, Glenn D. 2001. Carotenoid accumulation in the psychrotrophic bacterium Arthrobacter agilis in response to thermal and salt stress. Applied Microbiology and Biotechnology. 56(5):750-756.

Forsberg KJ, Patel S, Gibson MK, Lauber CL, Knight R, Fierer N, Dantas G. 2014. Bacterial phylogeny structures soil resistomes across habitats. Nature. 509:612.

Fox GE, Wisotzkey JD, Jurtshuk P. 1992. How Close Is Close: 16S rRNA Sequence Identity May Not Be Sufficient To Guarantee Species Identity. International Journal of Systematic and Evolutionary Microbiology. 42:166-170.

Fredrickson JK, Zachara JM, Balkwill DL, Kennedy D, Li S-mW, Kostandarithes HM, Daly MJ, Romine MF, Brockman FJ. 2004. Geomicrobiology of high-level nuclear waste-contaminated vadose sediments at the hanford site, washington state. Applied and environmental microbiology. 70(7):4230-4241.

Frossard A, Ramond J-B, Seely M, Cowan DA. 2015. Water regime history drives responses of soil Namib Desert microbial communities to wetting events. Scientific Reports. 5:12263.

Fu H, Wei Y, Zou Y, Li M, Wang F, Chen J, Zhang L, Liu Z, Ding L. 2014. Research Progress on the Actinomyces arthrobacter. Advances in Microbiology. 4:747-753.

Fukui K. 2010. DNA Mismatch Repair in Eukaryotes and Bacteria. Journal of Nucleic Acids. 2010.

Funke G, Hutson RA, Bernard KA, Pfyffer GE, Wauters G, Collins MD. 1996. Isolation of Arthrobacter spp. from Clinical Specimens and Description of Arthrobacter cumminsii sp. nov. and Arthrobacter woluwensis sp. nov. Journal of Clinical Microbiology. 34(10):2356-2363.

Ganzert L, Bajerski F, Mangelsdorf K, Lipski A, Wagner D. 2011. Arthrobacter livingstonensis sp. nov. and Arthrobacter cryotolerans sp. nov., salt-tolerant and psychrotolerant species from Antarctic soil. International Journal of Systematic and Evolutionary Microbiology. 61(4):979-984.

García-Gómez C, Parages ML, Jiménez C, Palma A, Mata MT, Segovia M. 2012. Cell survival after UV radiation stress in the unicellular chlorophyte Dunaliella tertiolecta is mediated by DNA repair and MAPK phosphorylation. Journal of Experimental Botany. 63(14):5259-5274.

García-Ortiz M-V, Ariza RR, Roldán-Arjona T. 2001. A Chemiluminescent Method for the Detection of DNA Glycosylase/Lyase Activity. Analytical Biochemistry. 298(1):127-129.

Garcia-Pichel F, Loza V, Marusenko Y, Mateo P, Potrafka RM. 2013. Temperature Drives the Continental-Scale Distribution of Key Microbes in Topsoil Communities. Science. 340(6140):1574-1577.

Ghosh T, Bose D, Zhang X. 2010. Mechanisms for activating bacterial RNA polymerase FEMS Microbiology Reviews. 34(5):611-627.

Glaeser SP, Kämpfer P. 2014. The Family Sphingomonadaceae. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F, editors. The Prokaryotes: Alphaproteobacteria and Betaproteobacteria. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 641-707.

Google Earth. 2018a. Antarctica. In: U.S. Geological Survey CA, © 2018 DigitalGlobe, editor.

Google Earth. 2018b. Namibia. Data SIO, NOAA, U.S. Navy, NGA, GEBCO.

Goordial J, Davila A, Lacelle D, Pollard W, Marinova MM, Greer CW, DiRuggiero J, McKay CP, Whyte LG. 2016. Nearing the cold-arid limits of microbial life in permafrost of an upper dry valley, Antarctica. The ISME Journal. 10(7):1613-1624.

Goosen N, Moolenaar GF. 2008. Repair of UV damage in bacteria. DNA Repair. 7(3):353-379.

Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. 2007. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. International Journal of Systematic and Evolutionary Microbiology. 57:81-91.

Goudie A, Viles H. 2014. Landscapes and Landforms of Namibia. Springer Netherlands.

Graf H-F, Rohekar SS, Oppenheimer C, Jarvis M, R P, Jacob D. 2010. Continental Scale Antarctic deposition of sulphur and black carbon from anthropogenic and volcanic sources. Atmospheric Chemistry and Physics. 10.

Graumann PL. 2007. Cytoskeletal Elements in Bacteria. Annual Review of Microbiology. 61(1):589-618.

Groves JT. 2005. Models and Mechanisms of Cytochrome P450 Action. In: Ortiz de Montellano PR, editor. Cytochrome P450: Structure, Mechanism, and Biochemistry, Page 450. 3rd ed. New York, USA: Kluwer Academic/Plenum Publishers.

Gruber N, Galloway JN. 2008. An Earth-system perspective of the global nitrogen cycle. Nature. 451:293.

Guengerich FP. 2005. Models and Mechanisms of Cytochrome P450 Action. In: Ortiz de Montellano PR, editor. Cytochrome P450: Structure, Mechanism, and Biochemistry, Page 450. 3rd ed. New York, USA: Kluwer Academic/Plenum Publishers.

Gunina A, Kuzyakov Y. 2015. Sugars in soil and sweets for microorganisms: Review of origin, content, composition and fate. Soil Biology and Biochemistry. 90:87-100.

Gutierrez T. 2017. Aerobic Hydrocarbon-Degrading Gammaproteobacteria: Xanthomonadales. In: McGenity TJ, editor. Taxonomy, Genomics and Ecophysiology of Hydrocarbon-Degrading Microbes. p. 1-15.

Hanada K, Iwasaki M, Ihashi S, Ikeda H. 2000. UvrA and UvrB suppress illegitimate recombination: Synergistic action with RecQ helicase. Proceedings of the National Academy of Sciences. 97(11):5989-5994.

Handelsman J. 2004. Metagenomics: Application of Genomics to Uncultured Microorganisms. Microbiology and Molecular Biology Reviews. 68(4):669-685.

Haubold B, Wiehe T. 2004. Comparative genomics: methods and applications. Naturwissenschaften. 91:405-421.

Heather JM, Chain B. 2016. The sequence of sequencers: The history of sequencing DNA. Genomics. 107(1):1-8.

Heyrman J, Verbeeren J, Schumann P, Swings J, De Vos P. 2005. Six novel Arthrobacter species isolated from deteriorated mural paintings. International Journal of Systematic and Evolutionary Microbiology. 55(4):1457-1464.

Hill R, Saetnan ER, Scullion J, Gwynn-Jones D, Ostle N, Edwards A. 2016. Temporal and spatial influences incur reconfiguration of Arctic heathland soil bacterial community structure. Environmental Microbiology. 18(6):1942-1953.

Hirsch P. 1986. Microbial life at extremely low nutrient levels. Advances in Space Research. 6(12):287-298.

Hirsch P, Gallikowski CA, Siebert J, Peissl K, Kroppenstedt R, Schumann P, Stackebrant E, Anderson R. 2004. Deinococcus frigens sp. nov., Deinococcus saxicola sp. nov., and Deinococcus marmoris sp. nov., Low Temperature and Draught-tolerating, UV-resistant Bacteria from Continental Antarctica. System Appl Microbiol. 27(6):636-645.

Hodkinson BP, Grice EA. 2015. Next-Generation Sequencing: A Review of Technologies and Tools for Wound Microbiome Research. Advances in wound care. 4(1):50-58.

Hong C, Si Y, Xing Y, Li Y. 2015. Illumina MiSeq sequencing investigation on the contrasting soil bacterial community structures in different iron mining areas. Environmental Science and Pollution Research. 22(14):10788-10799.

Hooper DU, Adair EC, Cardinale BJ, Byrnes JEK, Hungate BA, Matulich KL, Gonzalez A, Duffy JE, Gamfeldt L, O'Connor MI. 2012. A global synthesis reveals biodiversity loss as a major driver of ecosystem change. Nature. 486:105-108.

Horn H, Keller A, Hildebrandt U, Kämpfer P, Riederer M, Hentschel U. 2016. Draft genome of the Arabidopsis thaliana phyllosphere bacterium, Williamsia sp. ARP1. Standards in Genomic Sciences. 11(1):8.

Horneck G, Klaus DM, Mancinelli RL. 2010. Space Microbiology. Microbiology and Molecular Biology Reviews. 74(1):121-156.

Horner-Devine MC, Bohannan BJM. 2006. Phylogenetic Clustering and Overdispersion in Bacterial Communities. Ecology. 87(sp7):S100-S108.

Hsu DS, Kim ST, Sun Q, Sancar A. 1995. Structure and function of the UvrB protein. The Journal of biological chemistry. 270(14):8319-8327.

Hu Q-W, Chu X, Xiao M, Li C-T, Yan Z-F, Hozzein WN, Kim C-J, Zhi X-Y, Li W-J. 2016. Arthrobacter deserti sp. nov., isolated from a desert soil sample. International Journal of Systematic and Evolutionary Microbiology. 66:2035-2040.

Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M et al. 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Research. 44(D1):D286-D293.

Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 11(1):119.

Ii KM, Kono N, Paulino-Lima IG, Tomita M, Rothschild LJ, Arakawa K. 2019. Complete Genome Sequence of Arthrobacter sp. Strain MN05-02, a UV-Resistant Bacterium from a Manganese Deposit in the Sonoran Desert. Journal of genomics. 7:18-25.

Ikehata H, Ono T. 2011. The Mechanisms of UV Mutagenesis. Journal of Radiation Research. 52:115-125.

İnce İA, Demirbağ Z, Katı H. 2014. Arthrobacter pityocampae sp. nov., isolated from Thaumetopoea pityocampa (Lep., Thaumetopoeidae). International Journal of Systematic and Evolutionary Microbiology. 64:3384-3389.

Jaciuk M, Nowak E, Skowronek K, Tańska A, Nowotny M. 2011. Structure of UvrA nucleotide excision repair protein in complex with modified DNA. Nat Struct Mol Biol. 18(2):191-197.

Jain M, Olsen HE, Paten B, Akeson M. 2016. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biology. 17(1):239.

Janda JM, Abbott SL. 2007. 16S rRNA Gene Sequencing for Bacterial Identification in the Diagnostic Laboratory: Pluses, Perils, and Pitfalls. Journal of Clinical Microbiology. 45(9):2761-2764.

Janion C. 2001. Some aspects of the SOS response system — A critical survey. Acta Biochimica Polonica. 48(3):599-610.

Johnson RC. 2015. Site-specific DNA Inversion by Serine Recombinases. Microbiology spectrum. 3(3):1-36.

Johnson RM, Ramond J-B, Gunnigle E, Seely M, Cowan DA. 2017. Namib Desert edaphic bacterial, fungal and archaeal communities assemble through deterministic processes but are influenced by different abiotic parameters. Extremophiles. 21(2):381-392.

Jørgensen BB, Boetius A. 2007. Feast and famine — microbial life in the deep-sea bed. Nature Reviews Microbiology. 5(10):770-781.

Jünemann S, Prior K, Szczepanowski R, Harks I, Ehmke B, Goesmann A, Stoye J, Harmsen D. 2012. Bacterial Community Shift in Treated Periodontitis Patients Revealed by Ion Torrent 16S rRNA Gene Amplicon Sequencing. PLOS ONE. 7(8):e41606.

Jungfer C, Schwartz T, Obst U. 2007. UV-induced dark repair mechanisms in bacteria associated with drinking water. Water Research. 41(1):188-196.

Kageyama A, Morisaki K, Omura S, Takahashi Y. 2008. Arthrobacter oryzae sp. nov. and Arthrobacter humicola sp. nov. International Journal of Systematic and Evolutionary Microbiology. 58:53-56.

Kallimanis A, Kavakiotis K, Perisynakis A, Spröer C, Pukall R, Drainas C, Koukkou AI. 2009. Arthrobacter phenanthrenivorans sp. nov., to accommodate the phenanthrene-degrading bacterium Arthrobacter sp. strain Sphe3. International Journal of Systematic and Evolutionary Microbiology. 59:275-279.

Kallimanis A, LaButti KM, Lapidus A, Clum A, Lykidis A, Mavromatis K, Pagani I, Liolios K, Ivanova N, Goodwin L et al. 2011. Complete genome sequence of Arthrobacter phenanthrenivorans type strain (Sphe3). Standards in Genomic Sciences. 4(2):123-130.

Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M. 2006. From genomics to chemical genomics: new developments in KEGG. Nucleic acids research. 34(Database issue):D354-D357.

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 44(D1):D457-D462.

Kaplan H, Ratering S, Felix-Henningsen P, Schnell S. 2019. Stability of in situ immobilization of trace metals with different amendments revealed by microbial 13C-labelled wheat root decomposition and efflux-mediated metal resistance of soil bacteria. Science of The Total Environment. 659:1082-1089.

Karakas E, Truglio JJ, Croteau D, Rhau B, Wang L, Van Houten B, Kisker C. 2007. Structure of the C-terminal half of UvrC reveals an RNase H endonuclease domain with an Argonaute-like catalytic triad. The EMBO Journal. 26(2):613-622.

Karlsson E, Lärkeryd A, Sjödin A, Forsman M, Stenberg P. 2015. Scaffolding of a bacterial genome using MinION nanopore sequencing. Scientific Reports. 5:11996.

Kavakli IH, Baris I, Tardu M, Gul S, Oner H, Cal S, Bulut S, Yarparvar D, Berkel C, Ustaoglu P et al. 2017. The Photolyase/Cryptochrome Family of Proteins as DNA Repair Enzymes and Transcriptional Repressors. Photochem Photobiol. 93(1):93-103.

EC 6.4.1.1. 1961. [accessed]. https://www.genome.jp/dbget-bin/www_bget?ec:6.4.1.1.

KEGG Pathway Database. 2017. Carbon Fixation in Prokaryotes. Kanehisa Laboratories.

KEGG Pathway Database. 2018a. Citrate cycle (TCA cycle). Kanehisa Laboratories.

KEGG Pathway Database. 2018b. Glycolysis/Gluconeogenesis. Kanehisa Laboratories.

KEGG Pathway Database. 2018c. Nitrogen Metabolism. Kanehisa Laboratories.

KEGG Pathway Database. 2018d. Pentose Phosphate Pathway. Kanehisa Laboratories.

KEGG Pathway Database. 2019. Sulfur Metabolism. Kanehisa Laboratories.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols. 10:845.

Kim J-S, Makama M, Petito J, Park N-H, Cohan FM, Dungan RS. 2012. Diversity of Bacteria and Archaea in hypersaline sediment from Death Valley National Park, California. Microbiology Open. 1(2):135-148.

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

Kiskeer C, Kuper J, Van Houten B. 2013. Prokaryotic Nucleotide Excision Repair. Cold Spring Harb Perspect Biol. 5(3).

Klelner A. 2005. Experiments on Photoreactivation with Bacteria and Other Microorganisms. Journal of Cellular and Comparative Physiology. 39(1):115-117.

Klenk HP, Göker M. 2010. En route to a genome-based classification of Archaea and Bacteria? Systematic and Applied Microbiology. 33(4):175-182.

Koren S, Phillippy AM. 2015. One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. Current Opinion in Microbiology. 23:110-120.

Kougias PG, Campanaro S, Treu L, Zhu Z, Angelidaki I. 2017. A novel archaeal species belonging to Methanoculleus genus identified via de-novo assembly and metagenomic binning process in biogas reactors. Anaerobe. 46:23-32.

Kouskoumvekaki I, Shublaq N, Brunak S. 2014. Facilitating the use of large-scale biological data and tools in the era of translational bioinformatics. Briefings in Bioinformatics. 15(6):942-952.

Kowalczykowski SC, Dixon DA, Eggleston AK, Lauder SD, Rehrauer WM. 1994. Biochemistry of Homologous Recombination in Escherichia coli. Microbiological Reviews. 58(3):401-465.

Krishnapillai V. 1975. Resistance to ultraviolet light and enhanced mutagenesis conferred by Pseudomonas aeruginosa plasmids. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis. 29(3):363-371.

Krokan HE, Standal R, Slupphaug G. 1997. DNA glycosylases in the base excision repair of DNA. Biochemical Journal. 325(1):1-16.

Krwawicz J, Archzewska KD, Speina E, Maciejewska A, Grzesiuk E. 2007. Bacterial DNA repair genes and their eukaryotic homologues: 1. Mutations in genes involved in base excision repair (BER) and DNA-end processors and their implication in mutagenesis and human disease. Acta Biochimica Polonica. 54(3):413-434.

Kuhlman KR, Allenbach LB, Ball CL, Fusco WG, La Duc MT, Kuhlman GM, Anderson RC, Stuecker T, Erickson IK, Benardini J et al. 2005. Enumeration, isolation, and characterization of ultraviolet (UV-C) resistant bacteria from rock varnish in the Whipple Mountains, California. Icarus. 174(2):585-595.

Kumar R, Singh D, Swarnkar MK, Singh AK, Kumar S. 2016. Complete genome sequence of Arthrobacter alpinus ERGS4:06, a yellow pigmented bacterium tolerant to cold and radiations isolated from Sikkim Himalaya. Journal of Biotechnology. 220:86-87.

Kurth D, Belfiore C, Gorriti MF, Cortez N, Farias ME, Albarracín VH. 2015. Genomic and proteomic evidences unravel the UV-resistome of the poly-extremophile Acinetobacter sp. Ver3. Frontiers in Microbiology. 6(328).

Kyriakis JM, Avruch J. 2001. Mammalian Mitogen-Activated Protein Kinase Signal Transduction Pathways Activated by Stress and Inflammation. Physiological Reviews. 81(2):807-869.

Lacap-Bugler DC, Lee KK, Archer S, Gillman LN, Lau MCY, Leuzinger S, Lee CK, Maki T, McKay CP, Perrott JK et al. 2017. Global Diversity of Desert Hypolithic Cyanobacteria. Frontiers in Microbiology. 8(867).

Lacap DC, Warren-Rhodes KA, McKay CP, Pointing SB. 2011. Cyanobacteria and chloroflexi-dominated hypolithic colonization of quartz at the hyper-arid core of the Atacama Desert, Chile. Extremophiles. 15(1):31-38.

Lambrechts S, Willems A, Tahon G. 2019. Uncovering the Uncultivated Majority in Antarctic Soils: Toward a Synergistic Approach. Frontiers in Microbiology. 10(242).

Lane DJ. 1991. 16S/23S rRNA Sequencing. In: Stackebrandt E, Goodfellow M, editors. Nucleic Acid Techniques in Bacterial Systematic. New York: John Wiley and Sons. p. 115-175.

Lange V, Böhme I, Hofmann J, Lang K, Sauter J, Schöne B, Paul P, Albrecht V, Andreas JM, Baier DM et al. 2014. Cost-efficient high-throughput HLA typing by MiSeq amplicon sequencing. BMC Genomics. 15(63).

Langenheder S, Berga M, Östman Ö, Székely AJ. 2011. Temporal variation of β-diversity and assembly mechanisms in a bacterial metacommunity. The Isme Journal. 6:1107.

Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Research. 32(1):11-16.

Lee CK, Barbier BA, Bottos EM, McDonald IR, Cary SC. 2012. The Inter-Valley Soil Comparative Survey: the ecology of Dry Valley edaphic microbial communities. The ISME journal. 6(5):1046-1057.

Lee CK, Laughlin DC, Bottos EM, Caruso T, Joy K, Barrett JE, Brabyn L, Nielsen UN, Adams BJ, Wall DH et al. 2019. Biotic interactions are an unexpected yet critical control on the complexity of an abiotically driven polar ecosystem. Communications Biology. 2(1):62.

Lee I, Kim YO, Park S-C, Chun J. 2016. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. International Journal of Systematic and Evolutionary Microbiology. 66:1100-1103.

Lee J-H, Bae S-H, Choi B-S. 2000. The Dewar photoproduct of thymidylyl(3′→5′)- thymidine (Dewar product) exhibits mutagenic behavior in accordance with the “A rule”. Proceedings of the National Academy of Sciences, USA. 97(9):4591-4596.

Lee KC, Caruso T, Archer SDJ, Gillman LN, Lau MCY, Cary CS, Lee CK, Pointing SB. 2018. Stochastic and Deterministic Effects of a Moisture Gradient on Soil Microbial Communities in the McMurdo Dry Valleys of Antarctica. Frountiers of Microbiology. 9(2619).

Legendre P, Gallagher ED. 2001. Ecologically meaningful transformations for ordination of species data. Oecologia. 129:271-280.

Li W, Lv X, Ruan J, Yu M, Song Y-B, Yu J, Dong M. 2019. Variations in Soil Bacterial Composition and Diversity in Newly Formed Coastal Wetlands. Frontiers in microbiology. 9:3256-3256.

Liao Y-C, Lin S-H, Lin H-H. 2015. Completing bacterial genome assemblies: strategy and performance comparisons. Scientific Reports. 5:8747.

Lin Z, Kong H, Nei M, Ma H. 2006. Origins and evolution of the recA/RAD51 gene family: Evidence for ancient gene duplication and endosymbiotic gene transfer. Proceedings of the National Academy of Sciences, USA. 103(2):10328-10333.

Lindström ES, Langenheder S. 2012. Local and regional factors influencing bacterial community assembly. Environmental Microbiology Reports. 4(1):1-9.

Liu Y, Zhou J, Omelchenko MV, Beliaev AS, Venkateswaran A, Stair J, Wu L, Thompson DK, Xu D, Rogozin IB et al. 2003. Transcriptome dynamics of Deinococcus radiodurans recovering from ionizing radiation. Proceedings of the National Academy of Sciences. 100(7):4191-4196.

Lorence MC, Maika SD, Rupert CS. 1990. Physical Analysis of phr Gene Transcription in Escherichia coli K-12. Journal of Bacteriology. 172(11):6551-6556.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 15(12):550-550.

Loveland-Curtze J, Sheridan PP, Gutshall KR, Brenchley JE. 1999. Biochemical and phylogenetic analyses of psychrophilic isolates belonging to the Arthrobacter subgroup and description of Arthrobacter psychrolactophilus, sp. nov. Archives of Microbiology. 171:355-363.

Lowe CD, Martin LE, Roberts EC, Watts PC, Wootton EC, Montagnes DJS. 2010. Collection, isolation and culturing strategies for Oxyrrhis marina. Journal of Plankton Research. 33(4):569-578.

Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: Search and Contextual Analysis of Transfer RNA Genes. Nucleic Acids Research. 8(44 (W1)):W54-57.

Lu H, Giordano F, Ning Z. 2016a. Oxford Nanopore MinION Sequencing and Genome Assembly. Genomics, Proteomics & Bioinformatics. 14(5):265-279.

Lu Y, Shen Y, Warren W, Walter R. 2016b. Next Generation Sequencing in Aquatic Models. In: Kulski J, editor. Next Generation Sequencing.

Lucas RM, Norval M, Wright CY. 2016. Solar ultraviolet radiation in Africa: a systematic review and critical evaluation of the health risks and use of photoprotection. Photochemical & Photobiological Sciences. 15(1):10-23.

Lud D, Moerdijk TCW, Van de Poll WH, Buma AGJ, Huiskes AHL. 2002. DNA damage and photosynthesis in Antarctic and Arctic Sanionia uncinata (Hedw.) Loeske under ambient and enhanced levels of UV-B radiation. Plant, Cell & Environment. 25(12):1579-1589.

Luecke S, Wincent E, Backlund M, Rannug U, Rannug A. 2010. Cytochrome P450 1A1 gene regulation by UVB involves crosstalk between the aryl hydrocarbon receptor and nuclear factor κB. Chemico-Biological Interactions. 184(3):466-473.

Luo C, Tsementzi D, Kyrpides N, Read T, Konstantinidis KT. 2012. Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. PloS one. 7(2):e30087-e30087.

Ma B, Lv X, Cai Y, Chang SX, Dyck MF. 2018. Liming does not counteract the influence of long-term fertilization on soil bacterial community structure and its co-occurrence pattern. Soil Biology and Biochemistry. 123:45-53.

Mace GM, Norris K, Fitter AH. 2012. Biodiversity and ecosystem services: a multilayered relationship. Trends in Ecology & Evolution. 27(1):19-26.

Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD et al. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic acids research. 47(W1):W636-W641.

Madigan MT, Bender KS, Buckley DH, Stattley WM, Stahl DA. 2006. Brock Biology of Microorganisms. Harlow, United Kingdom: Pearson Education Ltd.

Maestre FT, Escolar C, Ladrón de Guevara M, Quero JL, Lázaro R, Delgado-Baquerizo M, Ochoa V, Berdugo M, Gozalo B, Gallardo A. 2013. Changes in biocrust cover drive carbon cycle responses to climate change in drylands. Global Change Biology. 19(12):3835-3847.

Makhalanyane TP. 2012. Microbial ecology of hot and cold desert edaphic communities. [South Africa]: Univesity of the Western Cape.

Makhalanyane TP, Valverde A, Gunnigle E, Frossard A, Ramond J-B, Cowan DA. 2015. Microbial ecology of hot desert edaphic systems. FEMS Microbiology Reviews. 39(2):203-221.

Makhalanyane TP, Valverde A, Lacap DC, Pointing SB, Tuffin MI, Cowan DA. 2013. Evidence of species recruitment and development of hot desert hypolithic communities. Environmental Microbiology Reports. 5(2).

Malard LA, Šabacká M, Magiopoulos I, Mowlem M, Hodson A, Tranter M, Siegert MJ, Pearce DA. 2019. Spatial Variability of Antarctic Surface Snow Bacterial Communities. Frontiers in Microbiology. 10(461).

Mandlik J, Shah N, Sharma A, Desai M. 2016. Microbial Identification in Endodontic Infections with an emphasis on Molecular Diagnostic Methods: A Review. The IIOAB Journal. 7(6):60-70.

Manzanera M, Narváez-Reinaldo JJ, García-Fontana C, Vílchez JI, González-López J. 2015. Genome Sequence of Arthrobacter koreensis 5J12A, a Plant Growth-Promoting and Desiccation-Tolerant Strain. Genome Announcements. 3(3):e00648-00615.

Marszałkowska M, Bil M, Kreft Ł, Olszewski M. 2014. A new division of bacterial UvrA homologues. BioTechnologia. 94(1):54-56.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011. 17(1):3.

Martins-Pinheiro M, Marques RCP, Menck CFM. 2007. Genome analysis of DNA repair genes in the alpha proteobacterium Caulobacter crescentus. BMC microbiology. 7:17-17.

Martiny JBH, Eisen JA, Penn K, Allison SD, Horner-Devine MC. 2011. Drivers of bacterial β-diversity depend on spatial scale. Proceedings of the National Academy of Sciences. 108(19):7850-7854.

Matallana-Surget S, Meador JA, Joux F, Douki T. 2008. Effect of the GC content of DNA on the distribution of UVB-induced bipyrimidine photoproducts. Photochemical & Photobiological Sciences. 7(7):794-801.

Matallana-Surget S, Wattiez R. 2013. Impact of Solar Radiation on Gene Expression in Bacteria. Proteomes. 1(2):70-86.

McKay CP. 2008. An Approach to Searching for Life on Mars, Europa, and Enceladus. Space Science Review. 135:49-54.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M et al. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research. 20:1297-1303.

McLeod M, Bockheim JG, Balks MR. 2008. Glacial geomorphology, soil development and permafrost features in central-upper Wright Valley, Antarctica. Geoderma. 144(1):93-103.

McMurdie PJ, Holmes S. 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLOS ONE. 8(4):e61217.

Meehl GAS, T. F, Collins WD, Friedlingstein P, Gaye AT, Gregory JM, Kitoh A, Knutti R, Murphy JM, Noda A, Raper SCB et al. 2007. 2007: Global Climate Projections. Cambridge, United Kingdom.

Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics. 14(60).

Méndez-García C, Bargiela R, Martínez-Martínez M, Ferrer M. 2018. Chapter 2: Metagenomic Protcols and Strategies. In: Nagarajan M, editor. Metagenomics: Perspective, Methods and Applications. London, UK: Elsevier Science and Technology. p. 15-54.

Michel B, Grompone G, Florès M-J, Bidnenko V. 2004. Multiple pathways process stalled replication forks. Proceedings of the National Academy of Sciences of the United States of America. 101(35):12783-12788.

Milani C, Hevia A, Foroni E, Duranti S, Turroni F, Lugli GA, Sanchez B, Martín R, Gueimonde M, van Sinderen D et al. 2013. Assessing the Fecal Microbiota: An Optimized Ion Torrent 16S rRNA Gene-Based Analysis Protocol. PLOS ONE. 8(7):e68739.

Miller RA, Beno SM, Kent DJ, Carroll LM, Martin NH, Boor KJ, Kovac J. 2016. Bacillus wiedmannii sp. nov., a psychrotolerant and cytotoxic Bacillus cereus group species isolated from dairy foods and dairy environments. International journal of systematic and evolutionary microbiology. 66(11):4744-4753.

Minas K, McEwan NR, Newbold CJ, Scott KP. 2011. Optimization of a high-throughput CTAB-based protocol for the extraction of qPCR-grade DNA from rumen fluid, plant and bacterial pure cultures. FEMS Microbiology Letters. 325(2):162-169.

Miranda J, Mattar S, Puerta-González A, Muskus C, Oteo JA. 2019. Genome Sequence of Candidatus Rickettsia colombianensi, a Novel Tick-Associated Bacterium Distributed in Colombia. Microbiology Resource Announcements. 8(14):e01433-01418.

Mirshad JK, Kowalczykowski SC. 2003. Biochemical Characterization of a Mutant RecA Protein Altered in DNA-Binding Loop 1. Biochemistry. 42:5945-5954.

Mitra S, Forster-Fromme K, Damms-Machado A, Scheurenbrand T, Biskup S, Huson DH, Bischoff SC. 2013. Analysis of the intestinal microbiota using SOLiD 16S rRNA gene sequencing and SOLiD shotgun sequencing. BMC Genomics. 14 Suppl 5:S16.

Mohammadi M, Burbank L, Roper MC. 2012. Biological Role of Pigment Production for the Bacterial Phytopathogen Pantoea stewarti. Applied and Environmental Microbiology. 78(19):6859-6865.

Mongodin EF, Shapir N, Daugherty SC, T DeBoy RT, Emerson JB, Shvartzbeyn A, Radune D, Vamathevan J, Riggs F, Grinberg V et al. 2006. Secrets of Soil Survival Revealed by the Genome Sequence of Arthrobacter aurescens TC1. PLoS Genetics. 2(12):e214.

Morales-Ruiz T, Ortega-Galisteo AP, Ponferrada-Marín MI, Martínez-Macías MI, Ariza RR, Roldán-Arjona T. 2006. Demeter and Repressor of Silencing 1 encode 5-methylcytosine DNA glycosylases. Proceedings of the National Academy of Sciences, USA. 103(18):6853-6858.

Morimatsu K, Kowalczykowski SC. 2003. RecFOR Proteins Load RecA Protein onto Gapped DNA to Accelerate DNA Strand Exchange. Molecular Cell. 11:1337-1347.

Murra M, Lützen L, Barut A, Zbinden R, Lund M, Villesen P, Nørskov-Lauritsena N. 2018. Whole-Genome Sequencing of Aggregatibacter Species Isolated from Human Clinical Specimens and Description of Aggregatibacter kilianii sp. nov. Journal of Clinical Microbiology. 56(7):e00053-00018.

Murray HC, Maltby VE, Smith DW, Bowden NA. 2015. Nucleotide excision repair deficiency in melanoma in response to UVA. Experimental Hematology & Oncology. 5(6).

Musilova M, Wright G, Ward JM, Dartnell LR. 2015. Isolateion of Radiation-Resistant Bacteria from Mars Analog Antarctic Dry Valleys by Preselection, and the Correlation between Radiation and Desiccation Resistance. Astrobiology. 15(12):1076-1090.

All-Time Weather Records, Namib Desert Lodge, Solitaire, Namib Naukluft. 2019. [accessed].

Navarro-González R, Rainey FA, Molina P, Bagaley DR, Hollen BJ, de la Rosa J, Small AM, Quinn RC, Grunthaner FJ, Cáceres L et al. 2003. Mars-like soils in the Atacama Desert, Chile, and the dry limit of microbial life. Science. 302(5647):1010-1021.

Nichio BTL, Marchaukoski JN, Raittz RT. 2017. New Tools in Orthology Analysis: A Brief Review of Promising Perspectives. Frontiers in Genetics. 8(165).

NIWA. 2016. UV Atlas 2.2.

Nouioui I, Carro L, García-López M, Meier-Kolthoff JP, Woyke T, Kyrpides NC, Pukall R, Klenk H-P, Goodfellow M, Göker M. 2018. Genome-Based Taxonomic Classification of the Phylum Actinobacteria. Frountiers of Microbiology. 9(2007).

Nouspikel T. 2019. Techniques used in the Nouspikel lab.

Obryk MK, Fountain AG, Doran PT, Lyons WB, Eastman R. 2018. Drivers of solar radiation variability in the McMurdo Dry Valleys, Antarctica. Scientific Reports. 8(1):5002.

Oka M, Uchida Y. 2018. Heavy metals in slag affect inorganic N dynamics and soil bacterial community structure and function. Environmental Pollution. 243:713-722.

Olsen SR, Cole CV, Watanabe FS, Dean LA. 1954. Estimation of available phosphorus in soils by extraction with sodium bicarbonate. Circular. 939:19.

Órdenes-Aenishanslins N, Anziani-Ostuni G, Vargas-Reyes M, Alarcón J, Tello A, Pérez-Donoso JM. 2016. Pigments from UV-resistant Antarctic bacteria as photosensitizers in Dye Sensitized Solar Cells. Journal of Photochemistry and Photobiology B: Biology. 162:707-714.

Ordoñez OF, Flores MR, Dib JR, Paz A, Farías ME. 2009. Extremophile Culture Collection from Andean Lakes: Extreme Pristine Environments that Host a Wide Diversity of Microorganisms with Tolerance to UV Radiation. Environmental Microbiology. 58(3):461-473.

Osman S, Peeters Z, La Duc MT, Mancinelli R, Ehrenfreund P, Venkateswaran K. 2008. Effect of Shadowing on Survival of Bacteria under Conditions Simulating the Martian Atmosphere and UV Radiation. Applied Environmental Microbiology. 74(4):959-970.

Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M et al. 2013. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Research. 42(D1):D206-D214.

Pagni R. 2006. Modern Physical Organic Chemistry. In: Anslyn EV, Dougherty DA, editors. Journal of Chemical Education. New York: American Chemical Society. p. 387.

Parker J. 2018. Modified mirror drop plate method.

Parker J. 2019. Schematic diagram of bacterial growth before UV exposure.

Patel M, Jiang Q, Woodgate R, Cox MM, Goodman MF. 2010. A New Model for SOS-induced Mutagenesis: How RecA Protein Activates DNA Polymerase V. Critical Reviews in Biochemistry and Molecular Biology. 45(3):171-184.

Paulino-Lima IG, Azua-Bustos A, Vicuña R, González-Silva C, Salas L, Teixeira L, Rosado A, da Costa Leitao AA, Lage C. 2013. Isolation of UVC-Tolerant Bacteria from the Hyperarid Atacama Desert, Chile. Environmental Microbiology. 35(2):325-335.

Paulino-Lima IG, Fujishima K, Navarrete JU, Galante D, Rodrigues F, Azua-Bustos A, Rothschild LJ. 2016. Extremely high UV-C radiation resistant microorganisms from desert environments with different manganese concentrations. Journal of Photochemistry and Photobiology B: Biology. 163:327-336.

Pavlopoulou A, Savva GD, Louka M, Bagos PG, Vorgias CE, Michalopoulos I, Georgakilas AG. 2016. Unraveling the mechanisms of extreme radioresistance in prokaryotes: Lessons from nature. Mutation Research/Reviews in Mutation Research. 767:92-107.

Pehnec G, Klasinc L, Šorgo G. 2009. Estimation of biologically effective UV radiation in Croatia. Periodicum Biologorum.111(1):65-71.

Pei A, Nossa CW, Chokshi P, Blaser MJ, Yang L, Rosmarin DM, Pei Z. 2009. Diversity of 23S rRNA genes within individual prokaryotic genomes. PloS one. 4(5):e5437-e5437.

Pham VHT, Kim J. 2012. Cultivation of unculturable soil bacteria. Trends in Biotechnology. 30(9):475-484.

Philippot L, Spor A, Hénault C, Bru D, Bizouard F, Jones CM, Sarr A, Maron P-A. 2013. Loss in microbial diversity affects nitrogen cycling in soil. The ISME Journal. 7(8):1609-1619.

Piersen CE, Prince MA, Augustine ML, Dodson ML, Lloyd RS. 1995. Purification and Cloning of Micrococcus luteus Ultraviolet Endonuclease, an N-Glycosylase/Abasic Lyase That Proceeds via an Imino Enzyme-DNA Intermediate. Journal of Biological Chemistry. 270(40):23475-23484.

Pointing SB. 2016. Hypolithic communities. In: Weber B, Budel B, Belnap J, editors. Biological Soil Crusts: An Organising Principle in Drylands, Ecological Studies Series 226. Berlin: Springer.

Pointing SB, Belnap J. 2012. Microbial colonization and controls in dryland systems. Nature Reviews Microbiology.10(8):551-562. http://www.nature.com.ezproxy.aut.ac.nz/nrmicro/journal/v10/n8/pdf/nrmicro2831.pdf.

Pointing SB, Chan Y, Lacap DC, Lau MCY, Jurgens JA, Farrell RL. 2009. Highly specialized microbial diversity in hyper-arid polar desert. Proceedings of the National Academy of Sciences, USA. 106(47):19964-199969.

Ponferrada-Marín MI, Martínez-Macías MI, Morales-Ruiz T, Roldán-Arjona T, Ariza RR. 2010. Methylation-independent DNA Binding Modulates Specificity of Repressor of Silencing 1 (ROS1) and Facilitates Demethylation in Long Substrates. The Journal of biological chemistry. 285(30):23032-23039.

Pootakham W, Mhuantong W, Yoocha T, Putchim L, Sonthirod C, Naktang C, Thongtham N, Tangphatsornruang S. 2017. High resolution profiling of coral-associated bacterial communities using full-length 16S rRNA sequence data from PacBio SMRT sequencing system. Scientific Reports. 7(1):2774.

Porter TM, Hajibabaei M. 2018. Scaling up: A guide to high-throughput genomic approaches for biodiversity analysis. Molecular Ecology. 27(2):313-338.

Prabha R, Singh DP, Gupta SK, Rai A. 2014. Whole genome phylogeny of Prochlorococcus marinus group of cyanobacteria: genome alignment and overlapping gene approach. Interdisciplinary Sciences: Computational Life Sciences. 6(2):149-157.

Prestel E, Salamitou S, DuBow MS. 2008. An Examination of the Bacteriophages and Bacteria of the Namib Desert. Journal of Microbiology. 46(4):364-372.

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PlOs One. 5(3):e9490.

Pulschen AA, Bendia AG, Fricker AD, Pellizari VH, Galante D, Rodrigues F. 2017. Isolation of Uncultured Bacteria from Antarctica Using Long Incubation Periods and Low Nutritional Media. Frontiers in Microbiology. 8(1346).

Pulschen AA, Rodrigues F, Duarte RTD, Araujo GG, Santiago IF, Paulino-Lima IG, Rosa CA, Kato MJ, Pellizari VH, Galante D. 2015. UV-resistant yeasts isolated from a high-altitude volcanic area on the Atacama Desert as eukaryotic models for astrobiology. MicrobiologyOpen. 4(4):574-588.

Puspita ID, Kamagata Y, Tanaka M, Asano K, Nakatsu CH. 2012. Are Uncultivated Bacteria Really Uncultivable? Microbes and Environments. 27(4):356-366.

Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics. 13(1):1-13.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research. 41(D1):D590-D596.

R Core Team. 2017. R: A language and environment for statistical computing. In: Computing RFfS, editor. Vienna, Austria.

Rajasekar A, Sekar R, Medina-Roldán E, Bridge J, Moy CKS, Wilkinson S. 2018. Next-generation sequencing showing potential leachate influence on bacterial communities around a landfill in China. Canadian Journal of Microbiology. 64(8):537-549.

Rampelotto PH. 2013. Extremophiles and Extreme Environments. Life (Basel). 3(3):482-485.

Rappé MS, Giovannoni SJ. 2003. The Uncultured Microbial Majority. Annual Reviews of Microbiology. 57:369-394.

Rasmussen LD, Zawadsky C, Binnerup SJ, Oregaard G, Sørensen SJ, Kroer N. 2008. Cultivation of hard-to-culture subsurface mercury-resistant bacteria and discovery of new merA gene sequences. Applied and environmental microbiology. 74(12):3795-3803.

Rastogi RP, Richa, Kumar A, Tyagi MB, Sinha RP. 2010. Molecular Mechanisms of Ultraviolet Radiation-Induced DNA Damage and Repair. Journal of Nucleic Acids. 2010(592980).

Rasuk MC, Ferrer GM, Kurth D, Portero LR, Farías ME, Albarracín VH. 2017. UV-Resistant Actinobacteria from High-Altitude Andean Lakes: Isolation, Characterization and Antagonistic Activities. Photochemistry and Photobiology. 93(3):865-880.

Ray A, Lindahl E, Wallner B. 2012. Improved model quality assessment using ProQ2. BMC Bioinformatics. 13(1):224.

Rayment GE, Higginson FR. 1992. Australian laboratory handbook of soil and water chemical methods. Port Melbourne.

Reed SC, Coe KK, Sparks JP, Housman DC, Zelikova TJ, Belnap J. 2012. Changes to dryland rainfall result in rapid moss mortality and altered soil fertility. Nature Climate Change. 2:752.

Rice KP, Cox MM. 2001. Recombinational DNA Repair in Bacteria: Postreplication. Chichester: John Wiley & Sons, Ltd.

Richardson EJ, Watson M. 2013. The automatic annotation of bacterial genomes. Briefings in bioinformatics. 14(1):1-12.

Ricker N, Qian H, Fulthorpe RR. 2012. The limitations of draft assemblies for understanding prokaryotic adaptation and evolution. Genomics. 100(3):167-175.

Riesenfeld CS, Schloss PD, Handelsman J. 2004. Metagenomics: Genomic Analysis of Microbial Communities. Annual Review of Genetics. 38:525-552.

Rissman AI, Mau B, Biehl BS, Darling AC, Glasner JD, Perna NT. 2009. Reordering contigs of draft genomes using the Mauve Aligner. Bioinformatics. 25(16):2071-2073.

Roberts FA, Darveau RP. 2015. Microbial protection and virulence in periodontal tissue as a function of polymicrobial communities: symbiosis and dysbiosis. Periodontology 2000. 69:18-27.

Rocha EPC, Cornet E, Michel B. 2005. Comparative and evolutionary analysis of the bacterial homologous recombination systems. PLoS genetics. 1(2):e15-e15.

Rodriguez-R LM, Castro JC, Kyrpides NC, Cole JR, Tiedje JM, Konstantinidis KT. 2018. How much do rRNA gene surveys underestimate extant bacterial diversity? Applied Environmental Microbiology. 84(6):e00014-00018.

Roldán-Arjona T, Ariza RR. 2009. Repair and tolerance of oxidative DNA damage in plants. Mutation Research/Reviews in Mutation Research. 681(2-3):169-179.

Ronaghi M. 2001. Pyrosequencing Sheds Light on DNA Sequencing. Genome Research. 11:3-11.

Ronca S, Ramond J-B, Jones BE, Seely M, Cowan DA. 2015. Namib Desert dune/interdune transects exhibit habitat-specific edaphic bacterial communities. Frountiers of Microbiology. 6(845).

Russell DA, Hatfull GF. 2016. Complete Genome Sequence of Arthrobacter sp. ATCC 21022, a Host for Bacteriophage Discovery. Genome Announcements. 4(2):e00168-00116.

Russell JA, León-Zayas R, Wrigton K, Biddle JF. 2016. Deep Subsurface Life from North Pond: Enrichment, Isolation, Characterization and Genomes of Heterotrophic Bacteria. Frontiers in Microbiology. 7(678).

Sader A, Oliveira S, Berchielli T. 2004. Application of Kjeldahl and Dumas combustion methods for nitrogen analysis. Archives of Veterinary Science. 9(2).

Sancar GB, Smith FW, Sancar A. 1983. Identification and amplification of the E. coli phr gene product. Nucleic Acids Res. 11(19).

Sang W, Ma W-H, Qiu L, Zhu Z-H, Lei C-L. 2012. The involvement of heat shock protein and cytochrome P450 genes in response to UV-A exposure in the beetle Tribolium castaneum. Journal of Insect Physiology. 58(6):830-836.

Sangal V, Jones AL, Goodfellow M, Hoskisson PA, Kämpfer P, Sutcliffe IC. 2015. Genomic analyses confirm close relatedness between Rhodococcus defluvii and Rhodococcus equi (Rhodococcus hoagii). Archives of Microbiology. 197(1):113-116.

Sanschagrin S, Yergeau E. 2014. Next-generation Sequencing of 16S Ribosomal RNA Gene Amplicons. Journal of Visualized Experiments. 90.

Satou K, Shiroma A, Teruya K, Shimoji M, Nakano K, Juan A, Tamotsu H, Terabayashi Y, Aoyama M, Teruya M et al. 2014. Complete Genome Sequences of Eight <span class="named-content genus-species" id="named-content-1">Helicobacter pylori</span> Strains with Different Virulence Factor Genotypes and Methylation Profiles, Isolated from Patients with Diverse Gastrointestinal Diseases on Okinawa Island, Japan, Determined Using PacBio Single-Molecule Real-Time Technology. Genome Announcements. 2(2):e00286-00214.

Sayed M, Sayed WF, Hatti-Kaul R, Pyo S-H. 2017. Complete Genome Sequence of Mycobacterium sp. MS1601, a Bacterium Performing Selective Oxidation of Polyols. Genome Announcements. 5(15):e00156-00117.

Schimel J, Balser TC, Wallenstein M. 2007. Microbial Stress-Response Physiology and its Implications for Ecosystem Function. Ecology. 88(6):1386-1394.

Schippers-Lammertse AF, Muijsers AO, Klatser-Oedekerk KB. 1963. Arthrobacter polychromogenes nov.spec., its pigments, and a bacteriophage of this species. Antonie van Leeuwenhoek. 29(1):1-15.

Schleifer KH. 2009. Classification of Bacteria and Archaea: Past, present and future. Systematic and Applied Microbiology. 32(8):533-542.

Exposure to Sunlight and Development of Different Types of DNA Lesions. 2014. [accessed 2016 12 February]. http://www.esciencecentral.org/ebooks/oxidatively/exposure-sunlight-development-different-dna-lesions.php.

Schuerger AC, Mancinelli RL, Kern RG, Rothschild LJ, McKay CP. 2003. Survival of endospores of Bacillus subtilis on spacecraft surfaces under simulated martian environments: implications for the forward contamination of Mars. Icarus. 165(2):253-276.

Scola V, Ramond J-B, Frossard A, Zablocki O, Adriaenssens EM, Johnson RM, Seely M, Cowan DA. 2017. Namib Desert Soil Microbial Community Diversity, Assembly, and Function Along a Natural Xeric Gradient. Microbial Ecology.

Sedlar K, Kolek J, Provaznik I, Patakova P. 2017. Reclassification of non-type strain Clostridium pasteurianum NRRL B-598 as Clostridium beijerinckii NRRL B-598. Journal of Biotechnology. 244:1-3.

See-Too W-S, Ee R, Lim Y-L, Convey P, Pearce DA, Mohidin TBM, Yin W-F, Gan K. 2017. Complete genome of Arthrobacter alpinus strain R3.8, bioremediation potential unraveled with genomic analysis. Standards in Genomic Sciences. 12(52).

Seeberga E, Eidec L, Bjørås M. 1995. The base excision repair pathway. Trends in Biochemical Sciences. 20(10):391-397.

Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 15;30(14):2068-2068.

Segata N, Börnigen D, Morgan XC, Huttenhower C. 2013. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. Nature Communications. 4:2304.

Serre L, Pereira de Jésus K, Boiteux S, Zelwer C, Castaing B. 2002. Crystal structure of the Lactococcus lactis formamidopyrimidine-DNA glycosylase bound to an abasic site analogue-containing DNA. The EMBO journal. 21(12):2854-2865.

Shaik S, Kumar N, Lankapalli AK, Tiwari SK, Baddam R, Ahmed N. 2016. Contig-Layout-Authenticator (CLA): A Combinatorial Approach to Ordering and Scaffolding of Bacterial Contigs for Comparative Genomics and Molecular Epidemiology. PLOS ONE. 11(6):e0155459.

Shen L, Yao T, Liu Y, Jiao N, Kang S, Xu B, Zhang S, Liu X. 2014. Downward-Shifting Temperature Range for the Growth of Snow-Bacteria on Glaciers of the Tibetan Plateau. Geomicrobiology Journal. 31(9):779-787.

Shintani M, Sanchez ZK, Kimbara K. 2015. Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. Frontiers in microbiology. 6:242-242.

Shokralla S, Spall JL, Gibson JF, Hajibabaei M. 2012. Next-generation sequencing technologies for environmental DNA research. Molecular Ecology. 21(8):1794-1805.

Singh RN, Gaba S, Yadav AN, Gaur P, Gulati S, Kaushik R, Saxena AK. 2016. First high quality draft genome sequence of a plant growth promoting and cold active enzyme producing psychrotrophic Arthrobacter agilis strain L77. Standards in Genomic Sciences. 11(1).

Singleton MR, Dillingham MS, Gaudier M, Kowalczykowski SC, Wigley DB. 2004. Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. Nature. 432(7014):187-193.

Sinha RP, Häder DP. 2002. UV-induced DNA damage and repair: a review. Photochemical and Photobiological Sciences. 1(4):225-236.

Skennerton CT. 2006. Minced—mining CRISPRs in environmental datasets.

Skerman VBD, McGowan V, Sneath PHA. 1980. Approved Lists of Bacterial Names. International Journal of Systematic and Evolutionary Microbiology. 30:225-420.

Smith DP, Peay KG. 2014. Sequence Depth, Not PCR Replication, Improves Ecological Inference from Next Generation DNA Sequencing. PLOS ONE. 9(2):e90234.

Smith GR. 2012. How RecBCD Enzyme and Chi Promote DNA Break Repair and Recombination: a Molecular Biologist's View. Microbiology and Molecular Biology Reviews. 76(2):217-228.

Sonowal R, Nandimath K, Kulkarni SS, Koushika SP, Nanjundiah V, Mahadevan S. 2013. Hydrolysis of aromatic β-glucosides by non-pathogenic bacteria confers a chemical weapon against predators. Proc Biol Sci. 280(1762):20130721-20130721.

Soon WW, Hariharan M, Snyder MP. 2013. High-throughput sequencing for biology and medicine. Molecular Systems Biology. 9:640.

Spies M, Kowalczykowski SC. 2005. Homologous Recombination by the RecBCD and RecF Pathways. In: Higgins P, editor. Bacterial Chromosomes. Washington D.C: ASM Press.

Srinivasan S, Lee S-Y, Kim MK, Jung H-Y. 2017. Complete genome sequence of Hymenobacter sp. DG25A, a gamma radiation-resistant bacterium isolated from soil. Mol Cell Toxicol. 13(1):65-72.

Stackebrandt E, Koch C, Gvozdiak O, Schumann P. 1995. Taxonomic dissection of the genus Micrococcus: Kocuria gen. nov., Nesterenkonia gen. nov., Kytococcus gen. nov., Dermacoccus gen. nov., and Micrococcus Cohn 1872 gen. emend. International Journal of Systematic Microbiology. 45(682-692).

Stewart EJ. 2012. Growing Unculturable Bacteria. Journal of Bacteriology. 194(16):4151-4160.

Stomeo F, Makhalanyane TP, Valverde A, Pointing SB, Stevens MI, Cary CS, Tuffin MI, Cowan DA. 2012. Abiotic factors influence microbial diversity in permanently cold soil horizons of a maritime-associated Antarctic Dry Valley. FEMS Microbiology Ecology. 82(2):326-340.

Stomeo F, Valverde A, Pointing SB, McKay CP, Warren-Rhodes KA, Tuffin MI, Seely M, Cowan DA. 2013. Hypolithic and soil microbial community assembly along an aridity gradient in the Namib Desert. Extremophiles. 17(2):329-337.

Stracy M, Jaciuk M, Uphoff S, Kapanidis AN, Nowotny M, Sherratt DJ, Zawadzki P. 2016. Single-molecule imaging of UvrA and UvrB recruitment to DNA lesions in living Escherichia coli. Nature Communications. 7:12568.

Streit WR, Schmitz RA. 2004. Metagenomics – the key to the uncultured microbes. Current Opinion in Microbiology. 7(5):492-498.

Subramanian S, Di Pierro V, Shah H, Jayaprakas AD, Weisberger I, Shim J, George A, Gelb BD, Sachidanandam R. 2013. MiST: A new approach to variant detection in deep sequencing datasets. Nucleic Acids Research. 41(16).

Sutherland BM. 1981. Photoreactivation in Bacteria and in Skin. The Journal of Inversigative Dermatology. 77(1):91-95.

Sutthiwong N, Fouillaud M, Valla A, Caro Y, Dufossé L. 2014. Bacteria belonging to the extremely versatile genus Arthrobacter as novel source of natural pigments with extended hue range. Food Research International. 65:156-162.

Suzuki H, Nishizawa T. 2014. Oxidative Stress and Stomach Cancer. In: Preedy VR, editor. Cancer: Oxidative Stress and Dietary Antioxidants. London: Elsevier.

Swanson AL, Wang J, Wang Y. 2012. Accurate and Efficient Bypass of 8,5'-Cyclopurine-2'-Deoxynucleosides by Human and Yeast DNA Polymerase η. Chemical Research in

Toxicology.25(8):1682-1691. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3423583/. doi:10.1021/tx3001576.

Tanaka M, Narumi I, Funayama T, Kikuchi M, Watanabe H, Matsunaga T, Nikaido O, Yamamoto K. 2005. Characterization of Pathways Dependent on the uvsE, uvrA1, or uvrA2 Gene Product for UV Resistance in Deinococcus radiodurans. Journal of Bacteriology. 187(11):3693-3697.

Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 28(1):33-36.

Taylor J-S. 2005. Structure and Properties of DNA Photoproducts. In: Kutter JP, editor. Separation Methods In Microanalytical Systems. Florida, USA: CRC Press.

Tiao G, Lee CK, McDonald IR, Cowan DA, Cary SC. 2012. Rapid microbial response to the presence of an ancient relic in the Antarctic Dry Valleys. Nature Communications. 3:660.

Timmins J, Gordon E, Caria S, Leonard G, Acajjaoui S, Kuo M-S, Monchois V, McSweeney S. 2009. Structural and Mutational Analyses of Deinococcus radiodurans UvrA2 Provide Insight into DNA Binding and Damage Recognition by UvrAs. Structure. 17(4):547-558.

Tong SYC, Schaumburg F, Ellington MJ, Corander J, Pichon B, Leendertz F, Bentley SD, Parkhill J, Holt DC, Peters G et al. 2015. Novel staphylococcal species that form part of a Staphylococcus aureus-related complex: the non-pigmented Staphylococcus argenteus sp. nov. and the non-human primate-associated Staphylococcus schweitzeri sp. nov. International journal of systematic and evolutionary microbiology. 65(Pt 1):15-22.

Tortora GJ, Funke BR, Case CL. 2009. Microbiology: An introduction. United States of America: Benjamin Cummings.

Tringe SG, von Mering c, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Detter JC et al. 2005. Comparative Metagenomics of Microbial Communities. Science. 308(5721):554-557.

Tyson JR, O’Neil NJ, Jain M, Olsen HE, Hieter P, Snutch TP. 2017. Whole genome sequencing and assembly of a Caenorhabditis elegans genome with complex genomic rearrangements using the MinION sequencing device. bioRxiv.099143.

Basic Genetic Mechanisms - DNA Damage & Repair. 2016. [accessed 2016 10 February]. http://mb207.blogspot.co.nz/2012/01/part-ii-basic-genetic-mechanisms-dna_25.html.

Ugolini FC, Bockheim JG. 2008. Antarctic soils and soil formation in a changing environment: A review. Geoderma. 144(1):1-8.

UNEP. 1992. World Atlas of Desertification. London: Edward Arnold.

Valdespino-Castillo PM, Cerqueda-García D, Espinosa AC, Batista S, Merino-Ibarra M, Taş N, Alcántara-Hernández RJ, Falcón LI. 2018. Microbial distribution and turnover in Antarctic microbial mats highlight the relevance of heterotrophic bacteria in low-nutrient environments. FEMS Microbiology Ecology. 94(9).

Valverde A, Makhalanyane TP, Seely M, Cowan DA. 2015. Cyanobacteria drive community composition and functionality in rock–soil interface communities. Molecular Ecology. 24(4):812-821.

van der Veen S, Tang CM. 2015. The BER necessities: the repair of DNA damage in human-adapted bacterial pathogens. Nature Reviews Microbiology. 13:83-94.

Van Goethem MW, Makhalanyane TP, Valverde A, Cary CS, Cowan DA. 2016. Characterization of bacterial communities in lithobionts and soil niches from Victoria Valley, Antarctica. FEMS Microbiology Ecology. 92(4).

Van Horn DJ, Van Horn ML, Barrett, J. E, Gooseff MN, Altrichter AE, Geyer KM, Zeglin LH, Takacs-Vesbach CD. 2013. Factors Controlling Soil Microbial Biomass and Bacterial Diversity and Community Composition in a Cold Desert Ecosystem: Role of Geographic Scale. PLoS One. 8(6).

Van Houten B, Croteau DL, DellaVecchia MJ, Wang H, Kisker C. 2005. ‘Close-fitting sleeves’: DNA damage recognition by the UvrABC nuclease system. Mut Res. 577(1-2):92-117.

Vasileiadis S, Puglisi E, Arena M, Cappa F, Cocconcelli PS, Trevisan M. 2012. Soil Bacterial Diversity Screening Using Single 16S rRNA Gene V Regions Coupled with Multi-Million Read Generating Sequencing Technologies. PLOS ONE. 7(8):e42671.

Vazquez S, Ruberto L, Mac Cormack W. 2005. Properties of extracellular proteases from three psychrotolerant Stenotrophomonas maltophilia isolated from Antarctic soil. Polar Biology. 28(4):319-325.

Venkatachalam KV, Akita H, Strott CA. 1998. Molecular cloning, expression, and characterization of human bifunctional 3'-phosphoadenosine 5'-phosphosulfate synthase and its functional domains. The Journal of biological chemistry. 273(30):19311-19320.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA et al. 2001. The Sequence of the Human Genome. Science. 291(5507):1304-1351.

Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W et al. 2004. Environmental Genome Shotgun Sequencing of the Sargasso Sea. Science. 304(5667):66-74.

Větrovský T, Baldrian P. 2013. The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. PLoS One.

Viles HA, Goudie AS. 2013. Weathering in the central Namib Desert, Namibia: Controls, processes and implications. Journal of Arid Environments. 93:20-29.

Walker VK, Palmer GR, Voordouw G. 2006. Freeze-Thaw Tolerance and Clues to the Winter Survival of a Soil Community. Applied and Environmental Microbiology. 72(3):1784-1792.

Wallenstein MD, Hall EK. 2012. A trait-based framework for predicting when and where microbial adaptation to climate change will affect ecosystem functioning. Biogeochemistry. 109(1):35-47.

Walterson AM, Stavrinides J. 2015. Pantoea: insights into a highly versatile and diverse genus within the Enterobacteriaceae. FEMS Microbiology Reviews. 39(6):968-984.

Wang Y, Coleman-Derr D, Chen G, Gu YQ. 2015. OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res. 1(43):W78–W84.

Warnes G, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, Lumley T, Mächler M, Magnusson A, Möller S. 2015. gplots: Various R programming tools for plotting data.

Warren-Rhodes KA, Lee KC, Archer SDJ, Cabrol N, Ng-Boyle L, Wettergreen D, Zacny K, Pointing SB, TNLitAPT. 2019. Subsurface Microbial Habitats in an Extreme Desert Mars-Analog Environment. Frontiers in Microbiology. 10(69).

Wei STS, Lacap-Bugler DC, Lau MCY, Caruso T, Rao S, de los Rios A, Archer SK, Chiu JMY, Higgins C, Van Nostrand JD et al. 2016. Taxonomic and Functional Diversity of Soil and Hypolithic Microbial Communities in Miers Valley, McMurdo Dry Valleys, Antarctica. Frountiers of Microbiology. 7:1642.

Westerberg K, Elväng AM, Stackebrandt E, Jansson JK. 2000. Arthrobacter chlorophenolicus sp. nov., a new species capable of degrading high concentrations of 4-chlorophenol. Int J Syst Evol Microbiol. 50(6):2083-2092.

Wick RR, Judd LM, Gorrie CL, Holt KE. 2017a. Completing bacterial genome assemblies with multiplex MinION sequencing. Microbial genomics. 3(10):e000132-e000132.

Wick RR, Judd LM, Gorrie CL, Holt KE. 2017b. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Computer Biology. 13(6):e1005595.

Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies Bioinformatics. 31(20):3350-3352.

Wickham H. 2009. ggplot2: Elegant Graphics for Data Analysis. NY: Springer-Verlag.

Wikimedia Commons. 2019. Spore photoproduct lyase.

Wilkins D. 2019. gggenes: Draw Gene Arrow Maps in 'ggplot2'. R package version 0.4.0. ed.

Winter C, Moeseneder MM, Herndl GJ. 2001. Impact of UV Radiation on Bacterioplankton Community Composition. Applied and Environmental Microbiology. 67(2):665-672.

Woo PCY, Lau SKP, Teng JLL, Tse H, Yuen K-Y. 2008. Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. Clinical Microbiology and Infection. 14(10):908-934.

Xu L, Dong Z, Fang L, Luo Y, Wei Z, Guo H, Zhang G, Gu YQ, Coleman-Derr D, Xia Q et al. 2019. OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Research. 47(W1):W52-W58.

Xu L, Shi W, Zheng X-C, Yang Y, Zhou L, Mu Y, Liu Y. 2017. Draft genome sequence of Arthrobacter sp. strain B6 isolated from the high-arsenic sediments in Datong Basin, China. Standards in Genomic Sciences. 12(11).

Yang L, Li L. 2015. Spore Photoproduct Lyase: The Known, the Controversial, and the Unknown. Journal of Biological Chemistry. 13290(7):4003-4009.

Yang X-W, He Y, Xu J, Xiao X, Wang F-P. 2013. The Regulatory Role of Ferric Uptake Regulator (Fur) during Anaerobic Respiration of Shewanella piezotolerans WP3. PLOS ONE. 8(10):e75588.

Yang Y, Yatsunami R, Ando A, Miyoko N, Fukui T, Takaichi S, Nakamura S. 2015. Complete Biosynthetic Pathway of the C50 Carotenoid Bacterioruberin from Lycopene in the Extremely Halophilic Archaeon Haloarcula japonica. Journal of Bacteriology. 197(9):1614-1623.

Yao Y, Tang H, Su F, Xu P. 2015a. Comparative genome analysis reveals the molecular basis of nicotine degradation and survival capacities of Arthrobacter. Scientific Reports. 5.

Yao Y, Tang H, Su F, Xu P. 2015b. Comparative genome analysis reveals the molecular basis of nicotine degradation and survival capacities of Arthrobacter. Scientific Reports. 5:8642.

Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer K-H, Whitman WB, Euzéby J, Amann R, Rosselló-Móra R. 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nature Reviews Microbiology. 12:635.

Yergeau E, Bokhorst S, Kang S, Zhou J, Greer CW, Aerts R, Kowalchuk GA. 2012. Shifts in soil microorganisms in response to warming are consistent across a range of Antarctic environments. The ISME journal. 6(3):692-702.

Yoon SH, Ha SM, Kwon S, Lim J, Kim Y, Seo H, Chun J. 2017. Introducing EzBioCloud: A taxonomically united database of 16S rRNA and whole genome assemblies. Int J Syst Evol Microbiol. 67:1613-1617.

Yoshinaka T, Yano K, Yamaguchi H. 1973. Isolation of Highly Radioresistant Bacterium, Arthrobacter radiotolerans nov. sp. Agricultural and Biological Chemistry. 37(10):2269-2275.

Yu H, Si P, Shao W, Qiao X, Yang X, Gao D, Wang Z. 2016. Response of enzyme activities and microbial communities to soil amendment with sugar alcohols. MicrobiologyOpen. 5(4):604-615.

Yu X-Y, Zhang L, Ren B, Yang N, Liu M, Liu X-T, Zhang L-X, Ding L-X. 2015. Arthrobacter liuii sp. nov., resuscitated from Xinjiang desert soil. International Journal of Systematic and Evolutionary Microbiology. 65:896-901.

Yuan M, Chen M, Zhang W, Lu W, Wang J, Yang M, Zhao P, Tang R, Li X, Hao Y et al. 2012. Genome Sequence and Transcriptome Analysis of the Radioresistant Bacterium Deinococcus gobiensis: Insights into the Extreme Environmental Adaptations. PLoS One. 7(3).

Yung CCM, Chan Y, Lacap DC, Pérez-Ortega S, de los Rios-Murillo A, Lee CK, Cary SC, Pointing SB. 2014. Characterization of Chasmoendolithic Community in Miers Valley, McMurdo Dry Valleys, Antarctica. Microbial Ecology. 68(2):351-359.

Zaikova E, Goerlitz DS, Tighe SW, Wagner NY, Bai Y, Hall BL, Bevilacqua JG, Weng MM, Samuels-Fair MD, Johnson SS. 2019. Antarctic Relic Microbial Mat Community Revealed by Metagenomics and Metatranscriptomics. Frontiers in Ecology and Evolution. 7(1).

Zengler K. 2009. Central Role of the Cell in Microbial Ecology. Microbiology and Molecular Biology Reviews. 73(4):712-729.

Zhang J, Ma Y, Yu H. 2012a. Arthrobacter cupressi sp. nov., an actinomycete isolated from the rhizosphere soil of Cupressus sempervirens. International Journal of Systematic and Evolutionary Microbiology. 62(11):2731-2736.

Zhang Z, Theurkauf WE, Weng Z, Zamore PD. 2012b. Strand-specific libraries for high throughput RNA sequencing (RNA-Seq) prepared without poly(A) selection. Silence. 3(9).

Appendix 1: Photographs of isolated bacteria exposed to UV radiation

Please see Appendix 1 as a Supplementary File.

Appendix 2: In-house script for Prokka file generation

The commands used to generate the Prokka files for each of the Namib isolates is described below using B2 as an example. A description for each step is included where the sentence starts with ##.

The raw reads were subsampled using the seqtk function and assembled using SPAdes

```
~/Bioinformatics/seqtk/seqtk sample -s100 B2_L5_1.fq.gz 0.25 |
gzip > B2_L5_1_0.25.fq.gz
```

Kmer normalisation to reduce sequencing depth using BBnorm

```
bbnorm.sh in1=R1.fastq in2=R2.fastq out1=R1_norm.fastq
out2=R2_norm.fastq target=100 min=5
```

```
bbnorm.sh in1=B2_L5_1.fq in2=B2_L5_2.fq out1=B2_L5_1_norm.fq
out2=B2_L5_2_norm.fq target=100 min=5
```

Unicycler used for *de novo* assembly

```
unicycler -1 B2_L5_1_norm.fq.gz -2 B2_L5_2_norm.fq.gz --out
B2_unicycler -t 12
```

The generated Prokka files have been provided with this thesis. Please see the Prokka files for each of the Namib genomes in the attached Dropbox file:

https://www.dropbox.com/sh/u0dsnzr7op8mfqb/AAAl6I1uv4NbDVeakZfATpGva?dl=0

Appendix 3: Quality of raw reads

The quality of raw reads for the Namib isolates genome sequences can be seen below. The base composition and distribution analysis, the Phred quality score and a sequence data quality pie chart was generated for each of the Namib isolates by Genewiz (Suzhou, China).

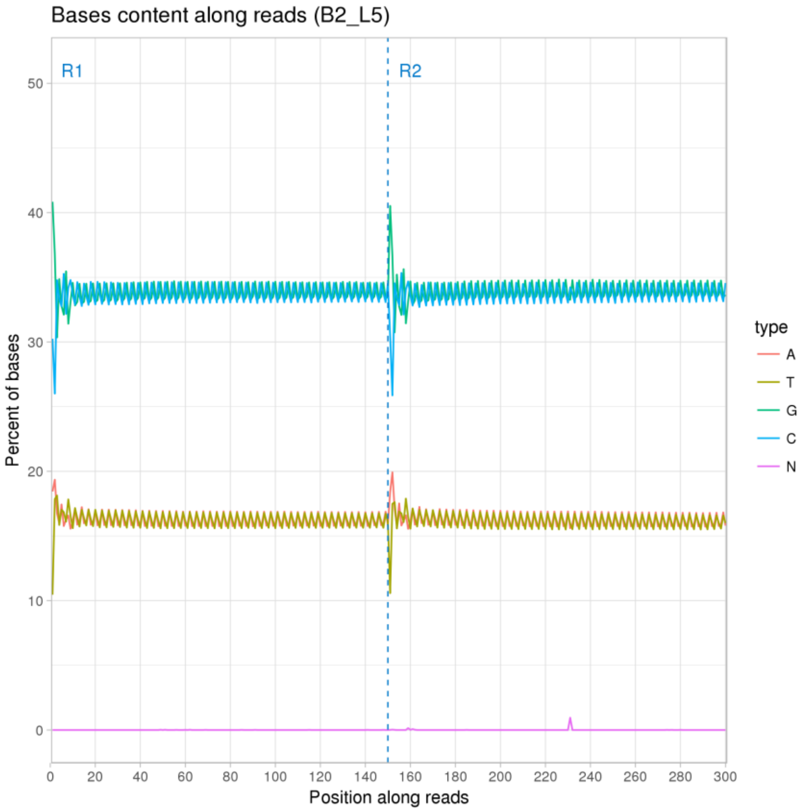


Figure S1: Base composition and distribution analysis for isolate B2

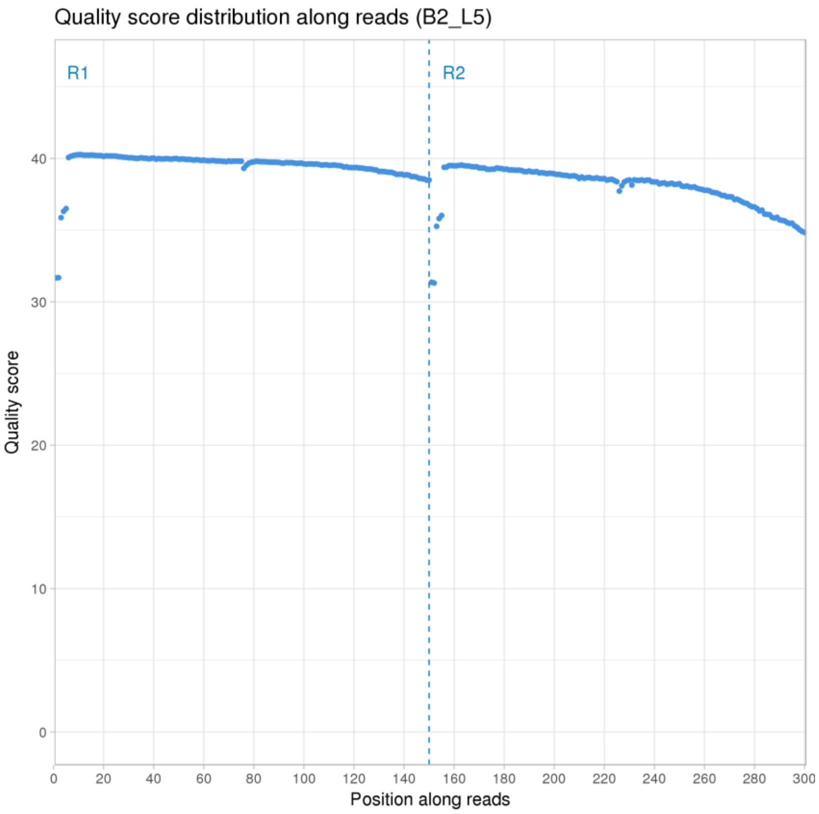


Figure S2: Phred quality score for isolate B2

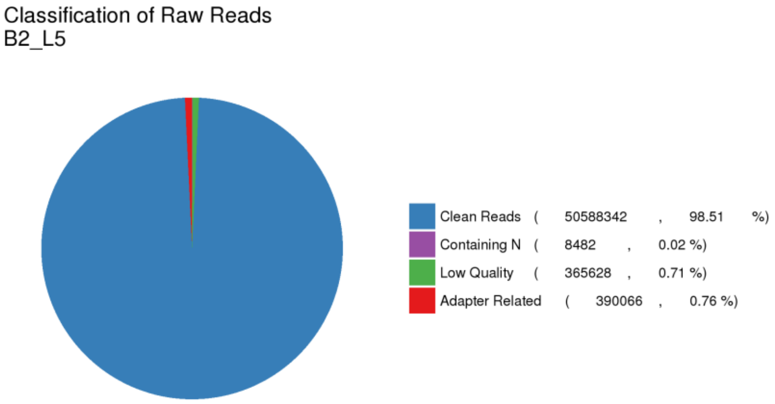


Figure S3: Classification of raw reads for isolate B2

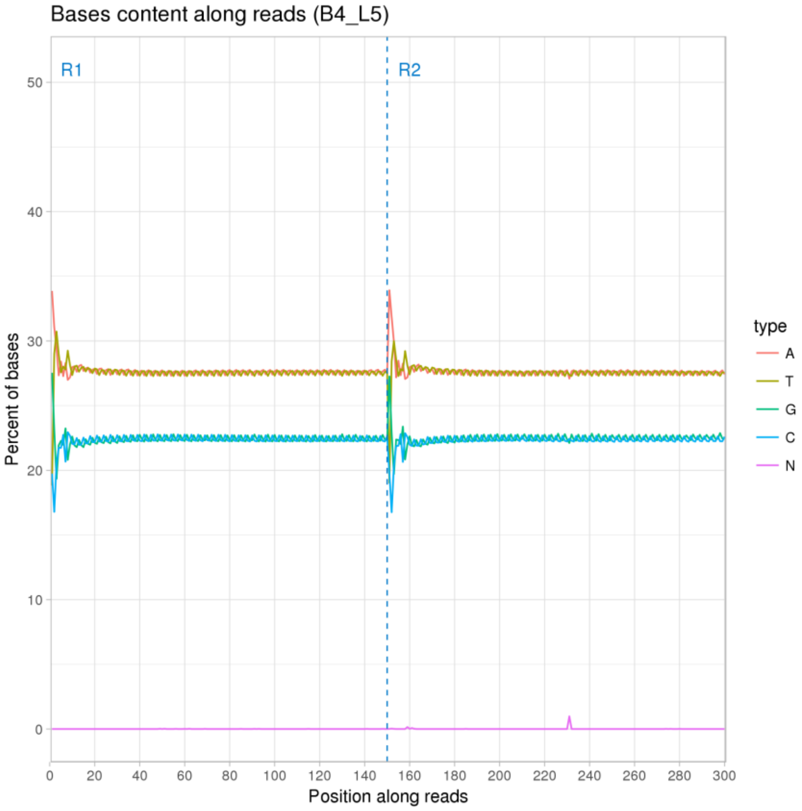


Figure S4: Base composition and distribution analysis for isolate B4

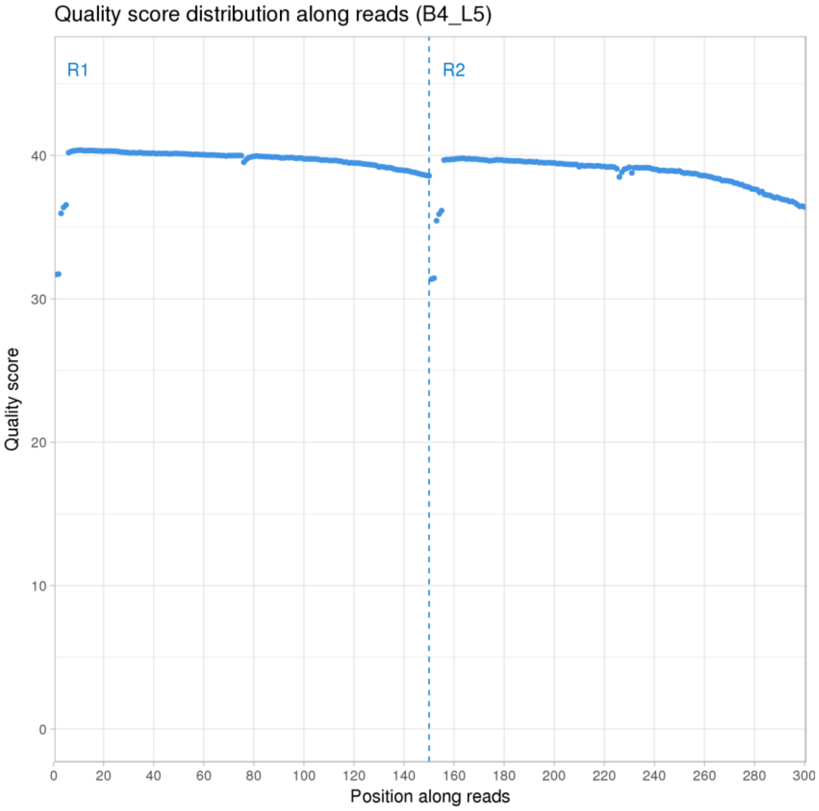


Figure S5: Phred quality score for isolate B4

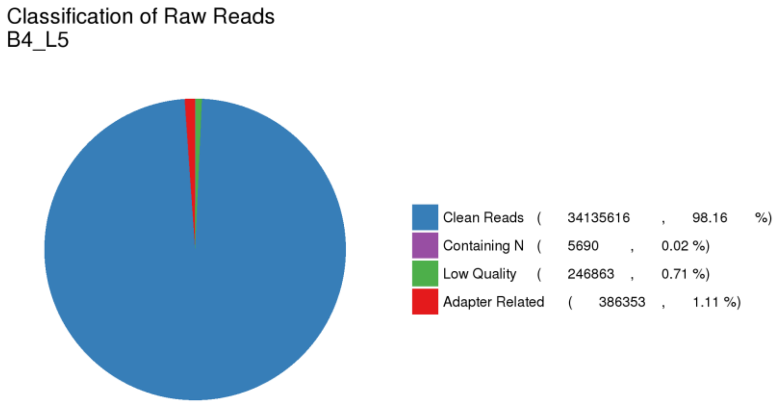


Figure S6: Classification of raw reads for isolate B4

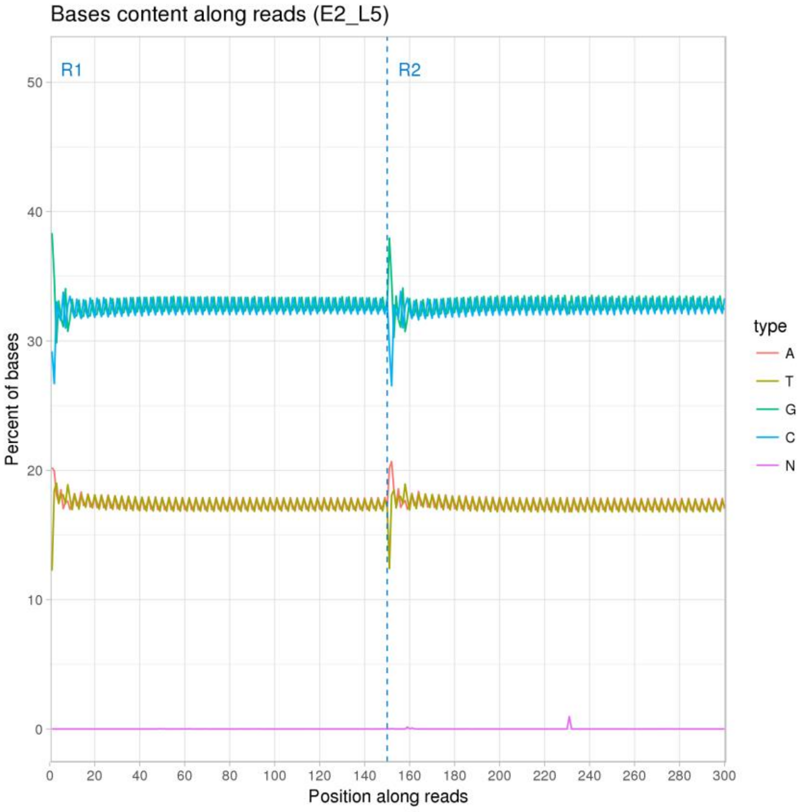


Figure S7: Base composition and distribution analysis for isolate E2

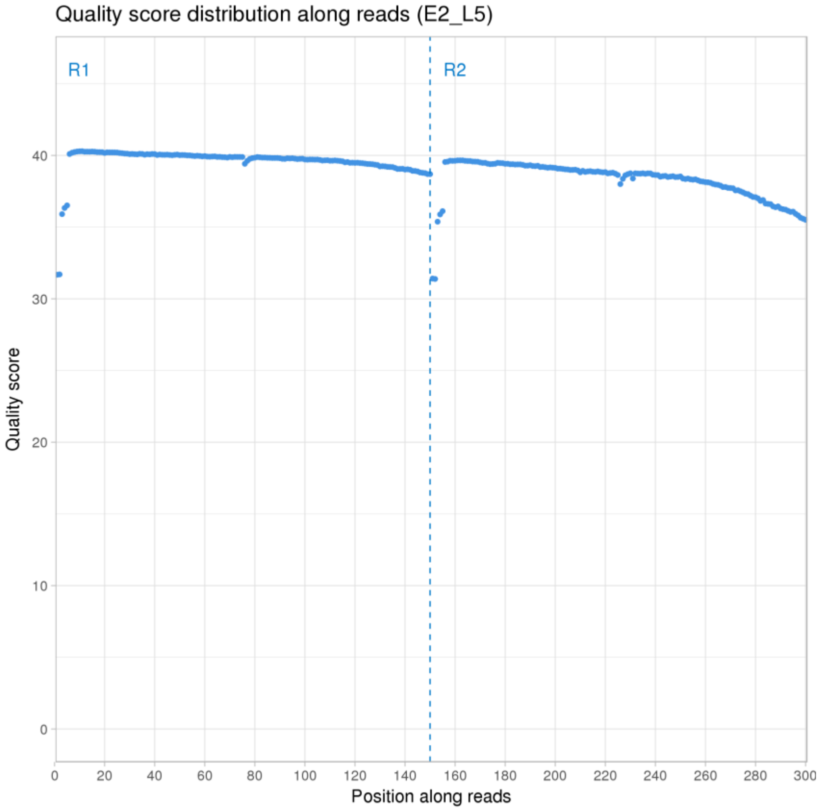


Figure S8: Phred quality score for isolate E2

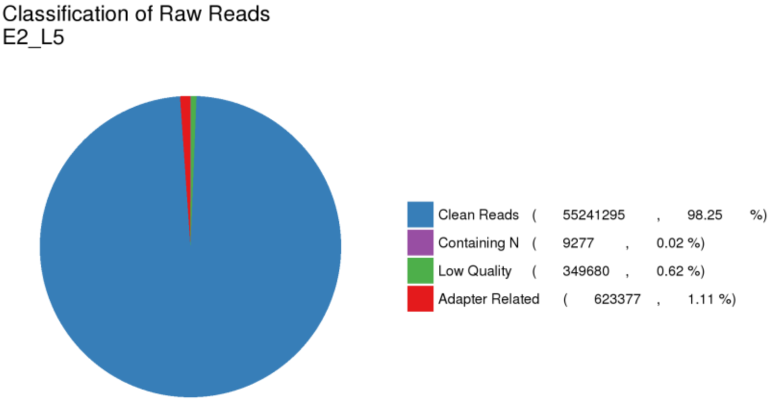


Figure S9: Classification of raw reads for isolate E2

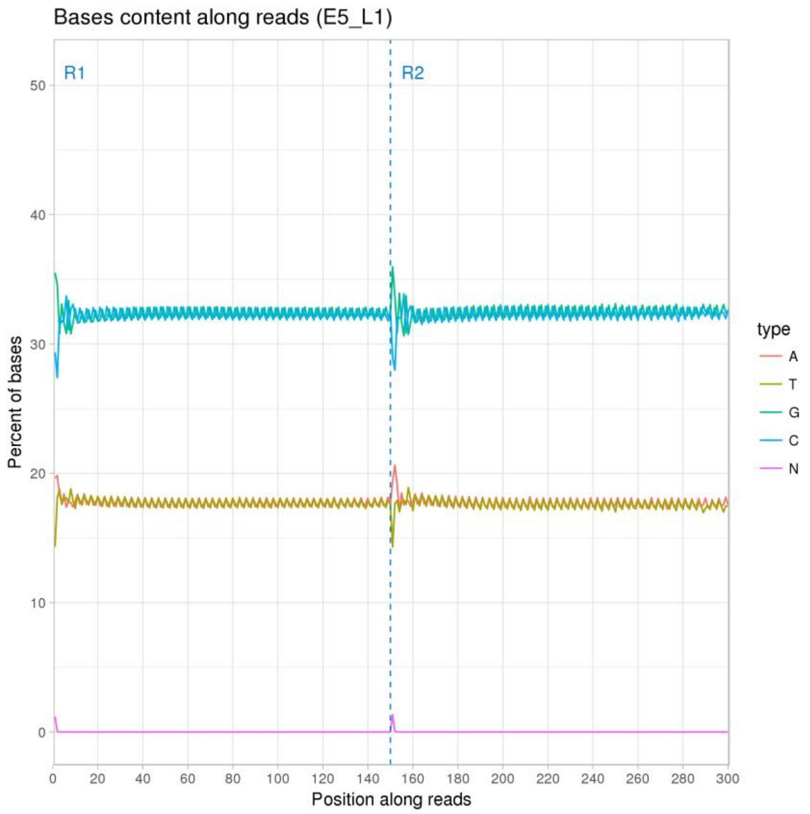


Figure S10: Base composition and distribution analysis for isolate E5

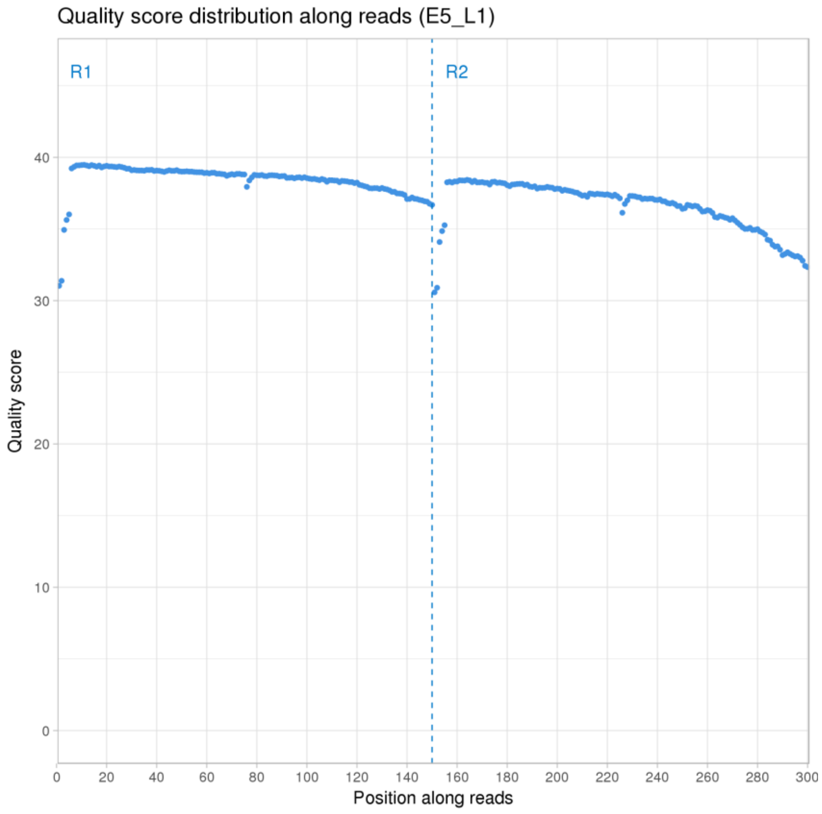


Figure S11: Phred quality score for isolate E5

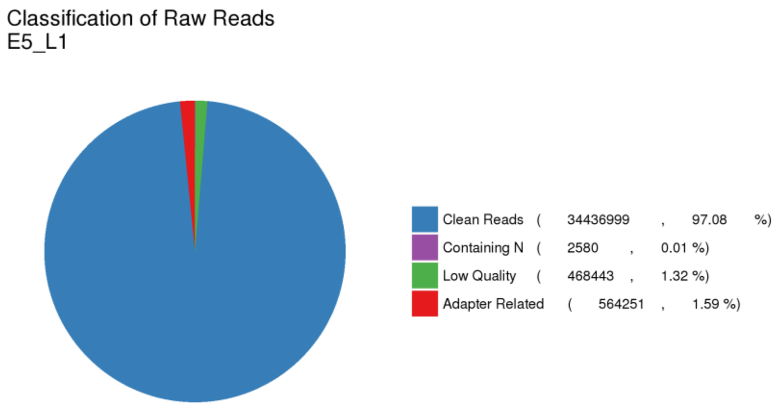


Figure S12: Classification of raw reads for isolate E5

Appendix 4: Mauve alignments

Mauve alignments were created for each of the Namib isolates against each reference genome (Table 5.2). The alignments for each of the isolates with each reference genome can be seen in Supplementary File.

Please see Appendix 4 as a Supplementary File.

Appendix 5: Locus tags for DNA repair genes

The locus tags for DNA repair genes found on the Namib isolates B2, B4, E2 and E5 and the annotated reference genomes used in this study (Table 5.2) can be found in the Supplementary File. Genes were identified from searching the RefSeq files for each genome.

Please see Appendix 5 as a Supplementary File.

Appendix 6: Identifying orthologous clusters using eggNOG

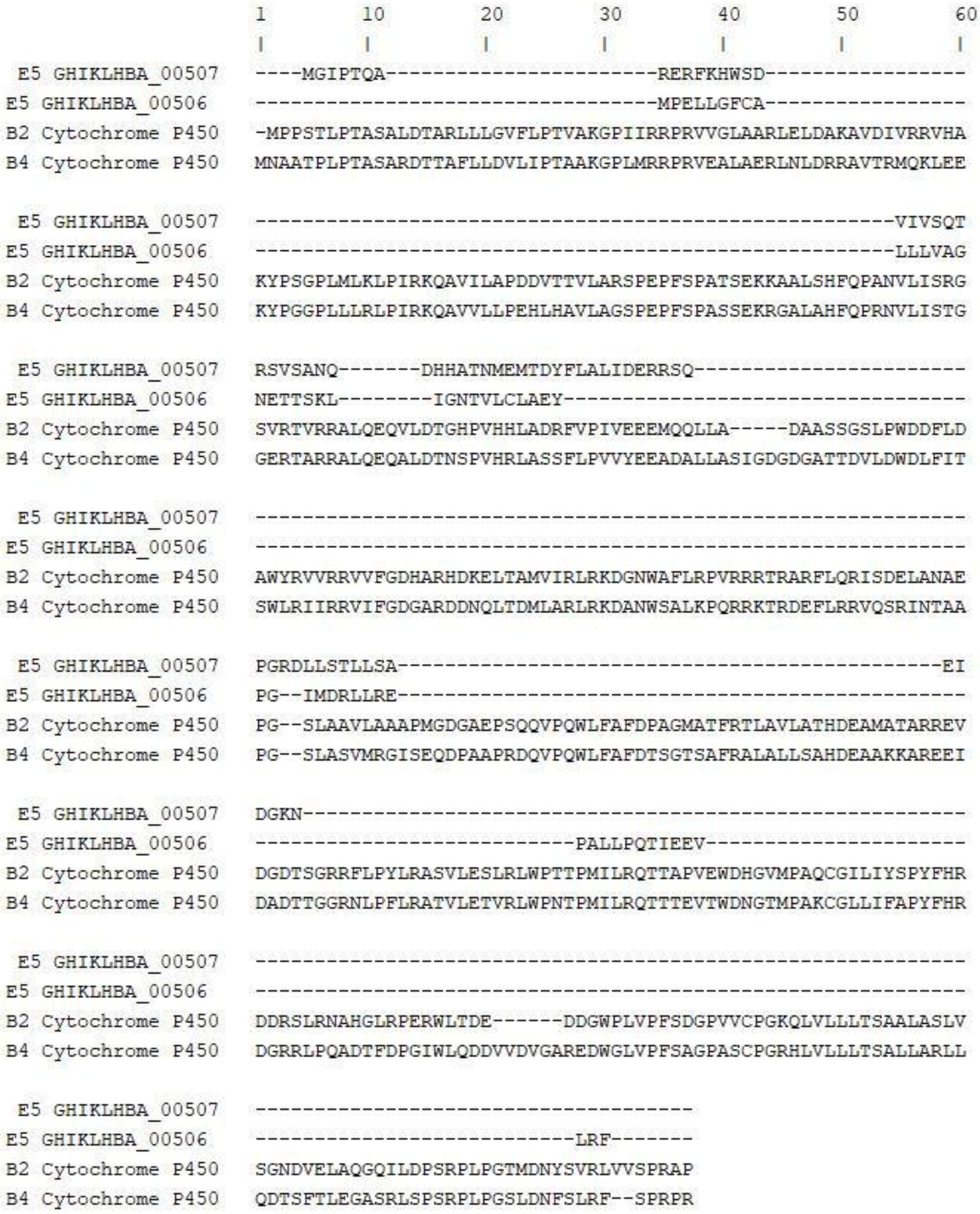
eggNOG annotation files were generated from the eggNOG website using the Prokka .faa files for each isolate. The eggNOG annotations for isolates B2, B4, E2 and E5, as well as the annotated reference genomes (Table 5.2), can be found in the Supplementary File.

Please see Appendix 6 as a Supplementary File.

These files can be viewed using Sublime Text 3 or an Excel spreadsheet.

Appendix 7: Protein alignment of E5 CYPs with B2 and B4 CYP

Two CYP sequences were identified in E5 (locus tags GHIKLHBA_00506 and GHIKLHBA_00507) after searching the eggNOG annotation file. Below is a protein alignment of the two E5 CYPs with the B2 and B4 CYP protein.



```
                     1        10        20        30        40        50        60
                     |        |         |         |         |         |         |
 E5 GHIKLHBA_00507   ----MGIPTQA-----------------------RERFKHWSD-----------------
E5 GHIKLHBA_00506    ----------------------------------MPELLGFCA-----------------
B2 Cytochrome P450   -MPPSTLPTASALDTARLLLGVFLPTVAKGPIIRRPRVVGLAARLELDAKAVDIVRRVHA
B4 Cytochrome P450   MNAATPLPTASARDTTAFLLDVLIPTAAKGPLMRRPRVEALAERLNLDRRAVTRMQKLEE

 E5 GHIKLHBA_00507   ------------------------------------------------------VIVSQT
E5 GHIKLHBA_00506    ------------------------------------------------------LLLVAG
B2 Cytochrome P450   KYPSGPLMLKLPIRKQAVILAPDDVTTVLARSPEPFSPATSEKKAALSHFQPANVLISRG
B4 Cytochrome P450   KYPGGPLLLRLPIRKQAVVLLPEHLHAVLAGSPEPFSPASSEKRGALAHFQPRNVLISTG

 E5 GHIKLHBA_00507   RSVSANQ-------DHHATNMEMTDYFLALIDERRSQ-----------------------
E5 GHIKLHBA_00506    NETTSKL--------IGNTVLCLAEY----------------------------------
B2 Cytochrome P450   SVRTVRRALQEQVLDTGHPVHHLADRFVPIVEEEMQQLLA-----DAASSGSLPWDDFLD
B4 Cytochrome P450   GERTARRALQEQALDTNSPVHRLASSFLPVVYEEADALLASIGDGDGATTDVLDWDLFIT

 E5 GHIKLHBA_00507   ------------------------------------------------------------
E5 GHIKLHBA_00506    ------------------------------------------------------------
B2 Cytochrome P450   AWYRVVRRVVFGDHARHDKELTAMVIRLRKDGNWAFLRPVRRRTRARFLQRISDELANAE
B4 Cytochrome P450   SWLRIIRRVIFGDGARDDNQLTDMLARLRKDANWSALKPQRRKTRDEFLRRVQSRINTAA

 E5 GHIKLHBA_00507   PGRDLLSTLLSA----------------------------------------------EI
E5 GHIKLHBA_00506    PG--IMDRLLRE------------------------------------------------
B2 Cytochrome P450   PG--SLAAVLAAAPMGDGAEPSQQVPQWLFAFDPAGMATFRTLAVLATHDEAMATARREV
B4 Cytochrome P450   PG--SLASVMRGISEQDPAAPRDQVPQWLFAFDTSGTSAFRALALLSAHDEAAKKAREEI

 E5 GHIKLHBA_00507   DGKN--------------------------------------------------------
E5 GHIKLHBA_00506    ---------------------------PALLPQTIEEV----------------------
B2 Cytochrome P450   DGDTSGRRFLPYLRASVLESLRLWPTTPMILRQTTAPVEWDHGVMPAQCGILIYSPYFHR
B4 Cytochrome P450   DADTTGGRNLPFLRATVLETVRLWPNTPMILRQTTTEVTWDNGTMPAKCGLLIFAPYFHR

 E5 GHIKLHBA_00507   ------------------------------------------------------------
E5 GHIKLHBA_00506    ------------------------------------------------------------
B2 Cytochrome P450   DDRSLRNAHGLRPERWLTDE------DDGWPLVPFSDGPVVCPGKQLVLLLTSAALASLV
B4 Cytochrome P450   DGRRLPQADTFDPGIWLQDDVVDVGAREDWGLVPFSAGPASCPGRHLVLLLTSALLARLL

 E5 GHIKLHBA_00507   -------------------------------------
E5 GHIKLHBA_00506    ---------------------------LRF-------
B2 Cytochrome P450   SGNDVELAQGQILDPSRPLPGTMDNYSVRLVVSPRAP
B4 Cytochrome P450   QDTSFTLEGASRLSPSRPLPGSLDNFSLRF--SPRPR
```

Figure S13: Protein alignment of the two E5 CYPs with the B2 and B4 CYP protein

Appendix 8: Quality metrics for Phyre2 analysis

The quality metrics of the UvrA and UvrB proteins were calculated using ProQ2 quality assessment tool (Ray et al. 2012) through PhyreInvestigator (Kelley et al. 2015). ProQ2 is a single-model method for predicting the quality of a predicted tertiary protein model. Each of the UvrA proteins had >90% of the query amino acids modelled (Table S1). The average quality score for the UvrA proteins was between 0.194293 (UvrA2a 1) and 0.300904 (E5 UvrA2a 2), where 1=bad modelling and 0=good modelling. The UvrB proteins had >87% of the query amino acids modelled (Table S1). The average quality score for the UvrA proteins was between 0.194293 (UvrA2a 1) and 0.300904 (E5 UvrA2a 2), where 1=bad modelling and 0=good modelling.

Table S1: Quality metrics for predicted tertiary modelling using Phyre2.

Isolate	Protein name	Number of aa modelled (%)	Average quality score*
B2	UvrA1	904 (93%)	0.278509
B4	UvrA1	886 (91%)	0.253961
E2	UvrA1	913 (94%)	0.277416
E5	UvrA1	886 (91%)	0.282224
B2	UvrA2a 1	801 (95%)	0.194293
B4	UvrA2a 1	802 (94%)	0.273909
E5	UvrA2a 1	803 (95%)	0.213694
B2	UvrA2a 2	747 (94%)	0.264164
B4	UvrA2a 2	744 (93%)	0.28373
E2	UvrA2a 2	747 (94%)	0.257966
E5	UvrA2a 2	744 (93%)	0.300904
B2	UvrB	618 (89%)	0.2165
B4	UvrB	618 (89%)	0.234706
E2	UvrB	618 (88%)	0.220984
E5	UvrB	618 (89%)	0.243144

*Note: 1=bad model, 0=good model

Phyre2 ‘intensive mode’ was used to model the UvrC proteins. The ‘intensive mode’ of Phyre2 predicts the protein structure *ab initio*, making the resulting model unreliable (Kelley et al. 2015). This modelling method is necessary for UvrC because these proteins have low conservation among genera (Goosen and Moolenaar 2008). Due to this, less than 40% of the aa of these proteins were modelled to a reference when checking for the modelling quality. It was therefore not possible to accurately calculate the average quality score for the UvrC proteins. Instead, the residues modelled at >90% confidence by Phyre2 are reported in Table S2 below.

Table S2: Residues modelled at >90% confidence for the predicted tertiary structure of the UvrC protein using the 'intensive modelling' function of Phyre2.

Isolate	Protein name	Number of aa modelled (%)	Residues modelled at >90% confidence
B2	UvrC	680 (100%)	56%
B4	UvrC	672 (100%)	71%
E2	UvrC	675 (100%)	62%
E5	UvrC	672 (100%)	63%