

Exploring horizontal gene transfer and phage infections in a South African deep subsurface bacterial population

By

Cumisa Manzikazi Mlandu

June 2017



UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA

Exploring horizontal gene transfer and phage infections in a South African deep subsurface bacterial population

By

Cumisa Manzikazi Mlandu

BSc. Hons. (Rhodes University)

Submitted in fulfillment of the requirements for the degree

MAGISTER SCIENTIAE

In the

Department of Microbial, Biochemical and Food Biotechnology

Faculty of Natural and Agricultural Sciences

University of the Free State

Bloemfontein

South Africa

June 2017

Supervisor: Prof. E. van Heerden

Co-Supervisor: Dr. E. D. Cason

I hereby dedicate this dissertation to my parents Mandisi Limo Mlandu and Tozama Mlandu, my siblings Noncedo Mlandu Luvuno, Sisanda Dalasile Nkukwana and Wandumzi Mlandu for their love and support during my studies. A special dedication to my late sister Tumeka Queen Mlandu, from the day you passed I promised you I would work hard to reach the goals you helped me set out.

“Science never solves a problem without creating ten more”

- George Bernard Shaw

ACKNOWLEDGMENTS

I would like to express my gratitude to the following contributors:

- **God**, The Almighty, for giving me the strength to prevail at times when I thought of giving up.
- **Prof. E. van Heerden** for being the supervisor I needed. Thank you for believing in me and giving me a chance without knowing me, given I was a student from another university. You clearly saw potential in me even when times were tough academically and personally. Thank you for being an extraordinary supervisor as your level of knowledge and coaching skills have left a mark in my growth as a scientist.
- **Dr. E. Cason** for being a very supportive co-supervisor and sharing your knowledge and advice whenever I needed it. Thank you for all the time you spent coaching and teaching me all the Metagenomics work and for all your patience. I have learned so much from you.
- The team from **Star Diamonds Mine, Frontier Mining** especially **Ben Visser** for accommodating me, providing the study site and helping me carry the sampling material in and out of the mine during both sampling trips.
- **Dr. W.J. van Rensburg** for allowing me to use the flow cytometry facilities at the Medical Faculty Haematology unit at the University of the Free State. Most importantly for your willingness to assist me with the planning, running and analyses of the flow cytometry samples.
- **Prof. P. van Wyk, Hanlie Grobler** and **Dr. C. Swart-Pistor** from the Centre for Microscopy for helping me plan and troubleshoot the SEM and TEM studies in order to obtain the final excellent images.
- **Dr. M. Erasmus** and **Prof. D. Litthauer** for helping me setup the tangential flow filtration setup. **Dr. M. Erasmus** and **Christo van Vuuren** for accompanying me on the sampling

trips and helping me setup on site. **Dr O'Neill** thank you for all your assistance with the move from Rhodes University to UFS and help with proofing the final draft.

- **My parents Mandisi and Tozama Mlandu** for believing in my dreams without any hesitations, for the sacrifices to get me to where I am and for answering the phone allowing me to vent when times got tough.
- **Members of the TIA/Biosaense research group** for all the support you gave me, especially **Elizabeth Ojo** and **Marcele Vermeulen** for all your assistance with some experiments and data analysis. **Maleke, Reitumetse Molaoa, Karabelo Moloantoa** and **Nqobile Radebe** for your friendship, all round assistance and much needed advice through the good and bad times.
- The **people** in the **Department of Biochemistry, Microbiology and Food Biotechnology** who helped where they could and offered suggestions and ideas.
- **Technology Innovation Agency (TIA)** for all the financial support.

DECLARATION

I hereby declare that this dissertation is submitted by me for the Magister Scientiae degree at the University of the Free State. This work is solely my own and has not been previously submitted by me at any other University or Faculty, and the other sources of information used have been acknowledged. I further grant copyright of this dissertation in favour of the University of the Free State.

C.M Mlandu (2010076079)

Date: June 2017

Table of Contents

LIST OF FIGURES	xii
LIST OF TABLES.....	xviii
LIST OF ABBREVIATIONS.....	xix

CHAPTER 1: LITERATURE REVIEW

1.1. Introduction	2
1.2. Categories of extremophiles in extreme environments	6
1.3. Deep terrestrial subsurface environment	7
1.4. Deep sea environment	12
1.5. Prokaryotes and eukaryotes from extreme environments	16
1.6. Microbial diversity in the South African mines.....	16
1.7. Viruses in subsurface environments	17
1.8. Techniques used to enumerate and characterize viruses in extreme environments	20
1.9. Host-phage interactions in the subsurface.....	21
1.9.1. Marine host-phage interactions.....	22
1.9.2. Hydrothermal vent phages.....	23
1.9.3. Terrestrial subsurface host-phage interactions	24
1.9.4. Horizontal gene transfer.....	25
1.10. Host-phage interactions in other extreme environments	26
1.10.1. Hot springs	26
1.10.2. Solfataric fields	29
1.11. Conclusions.....	29
1.12. References	31

CHAPTER 2: INTRODUCTION TO STUDY

2.1. Introduction	46
2.2. Main objectives.....	47
2.3. References.....	48

CHAPTER 3: SAMPLING, MICROBIAL ENUMERATION AND CHARACTERIZATION OF PHAGES USING MICROSCOPY AND SANGER SEQUENCING

3.1.	Introduction	50
3.1.2.	Aims of chapter.....	53
3.2.	Methods and materials	54
3.2.1.	Study site.....	54
3.2.2.	Physiochemical analyses.....	54
3.2.3.	Concentration of water using tangential flow filtration	54
3.2.4.	Microbial enumeration using DAPI (4',6-diamidino-2-phenylindole, dihydrochloride) staining and acridine orange staining.....	56
3.2.5.	Microbial enumeration using flow cytometry.....	57
3.2.6.	Scanning electron microscopy (SEM) of the morphology of the bacterial cells.....	57
3.2.7.	Microbial diversity assessment using Sanger sequencing.....	58
3.2.7.1.	Genomic DNA isolation.....	58
3.2.7.2.	Polymerase chain reaction (PCR).....	58
3.2.7.3.	Denaturing gradient gel electrophoresis.....	59
3.2.7.4.	Gel extraction, re-amplification and cloning into pGEM [®] -T Easy vector.....	60
3.2.7.5.	Sanger sequencing and analysis	61
3.2.8.	Purifying viral-like particles	62
3.2.9.	Epifluorescent microscopy of free viral-like particles.....	63
3.2.10.	Transmission Electron Microscopy (TEM).....	63
3.2.11.	Viral genomic DNA isolation	64
3.2.12.	Amplification of phage genes.....	65
3.2.13.	Cloning into the pSMART [®] HCKan vector and Sanger sequencing.....	66
3.3.	Results and discussions	67
3.3.1.	Physiochemical analysis.....	67
3.3.2.	Microbial enumeration using DAPI (4',6-diamidino-2-phenylindole, dihydrochloride) staining and acridine orange staining.....	68
3.3.3.	Microbial enumeration using flow cytometry.....	70
3.3.4.	Scanning electron microscopy (SEM) of the morphology of the bacterial cells.....	74
3.3.5.	Microbial diversity assessments using Sanger sequencing.....	75
3.3.6.	Identification of viral-like particles using EFM and TEM	79
3.3.7.	Identification of phage genes using Sanger sequencing	82
3.4.	Conclusions.....	86

3.5. References.....	87
----------------------	----

CHAPTER 4: MICROBIAL DIVERSITY IDENTIFICATION OF PHAGE GENES AND HORIZONTAL GENE TRANSFERS

4.1. Introduction	96
4.1.2. Aims of chapter.....	98
4.2. Methods and materials	99
4.2.1. Bacterial diversity assessment using next generation 16S rRNA sequencing	99
4.2.1.1. Genomic DNA isolation.....	99
4.2.1.2. Sequencing of the targeted 16S rRNA of Bacteria using the Illumina MiSeq sequencer and analyses	99
4.2.2. Microbial diversity assessment using shotgun whole metagenome sequencing.....	100
4.2.2.1. Genomic DNA isolation.....	100
4.2.2.2. Illumina NextSeq 500 shotgun sequencing of the whole metagenome and analyses	100
4.2.3. Preliminary identification and annotation of phage genes and prophages in the bacterial population.....	101
4.2.3.1. Identifying phage genes using VirSorter.....	101
4.2.3.2. Annotating prophages using PHASTER.....	102
4.2.4. Binning of metagenome data and annotation of phage genes, prophages and genes related to horizontal gene transfers using RAST and Island Viewer 3.....	102
4.2.4.1. Binning of contigs using MetaBAT 2 and bin completeness and identification using CheckM.....	102
4.2.4.2. Annotation of the purified bin genomes using RAST	103
4.2.4.3. Prediction of genomic islands/horizontal gene transfer related genes using Island Viewer 3.....	103
4.3. Results and discussions	103
4.3.1. 16S rRNA next generation sequencing of bacterial diversity assessment using the Illumina MiSeq next generation sequencer	103
4.3.2. Microbial diversity studies using MG-RAST	106
4.3.3. Identification and annotation of phage genes and prophages in the bacterial population.....	113

4.3.4.	Annotation of phage genes and horizontal gene transfer related genes within the binned genomes using RAST and Island Viewer 3	116
4.4.	Conclusions.....	133
4.5.	Suppliment A.....	135
4.6.	References.....	135

CHAPTER 5: SUMMARY

5.1.	Summary.....	144
------	--------------	-----

LIST OF FIGURES

- Figure 1.1: New updated tree of life highlighting the recovered lineages from the bacterial domain (Taken from Hug *et al.*, 2016).
- Figure 1.2: Water radiolysis within the deep terrestrial subsurface. Highlighted in the red box is the process of water radiolysis and its ability to generate the oxidation of hydrogen which when coupled to the reduction of sulfate and carbon fixation produces energy and nutrients for the *Candidatus Desulforudis audaxviator* (Adapted from Chivian *et al.*, 2008).
- Figure 1.3: Map highlighting the Witwatersrand Basin area (Adapted from McCarthy, 2011).
- Figure 1.4: Diagram showing the microbial energy, temperature and chemical circulations in the hydrothermal vent, plumes and surrounding environments (Taken from Dick *et al.*, 2013).
- Figure 1.5: The reproductive life cycle of viruses. (1) The bacteriophage enters the host cell. (2-3a) and integrates with the hosts genome and hijacks the bacterial cell's replication and translation machinery by forcing the host metabolism to produce new phages. (4a.) Phages enzymes such as lysins and proteins such as holin's destabilize the bacterial membrane allowing for the cell to release the bacteriophages. (3b) If the phage chooses the lysogenic cycle it enters a dormant stage where it becomes a prophage that replicates with the hosts genome. (4b.) The infected bacterial cell divides at the same rate as normal bacterial cells (Adapted from Anderson *et al.*, 2013).
- Figure 1.6: Typical Siphoviruses with long non contractile tails (Taken from Suttle, 2005).
- Figure 1.7: Transmission electron micrographs of phages found in the deep sea hydrothermal vent environments. (A) *Bacillus* virus W1 (BVW1), (B) *Geobacillus* virus E1 (GVE1) (Taken from Liu *et al.*, 2006), and (C) *Geobacillus* bacteriophage, D6E (Taken from Zhang & Wang, 2010).
- Figure 1.8: Map showing the studied alkaline hot springs (circled in purple) in Limpopo Province, South Africa (Adapted from Tekere *et al.*, 2015).

- Figure 1.9: Transmission electron micrograph of the *Acidianus* bottle-shaped virus isolated in the crater of the Solfatara volcano at Pozzuoli, Italy (Taken from Haring *et al.*, 2005). Bar = 100 nm.
- Figure 3.1: On-site tangential flow filtration setup. (A) Shows the taps and tubing on the 220 L drums that were used to fill the 50 L Nalgene Carboy drums. (B) Shows the TFF setup of the two 100 kDa filters.
- Figure 3.2: Diagram of the combined overall tangential flow filtration concentration method.
- Figure 3.3: pGEM[®]-T Easy vector system circle map (Promega).
- Figure 3.4: pSMART[®] HCKan vector system circle map (Lucigen).
- Figure 3.5: DAPI stain images of the fissure water from the Star Diamonds mine. (A) Sample before concentration showing a low count of stained cells in macro and biofilm-like structures. (B) Sample after concentration using the 100 kDa TFF filter showing a higher concentration of stained cells in macro and biofilm-like structures. (C) TFF (100 kDa filter) concentrated samples after vortexing for 10 minutes resulting in the observed dislodged cells from the biofilm-like structures. Scale for A is equal to 10 μm and 2 μm for B and C.
- Figure 3.6: Acridine orange stain images of the fissure water from the Star Diamonds mine. (A) Sample before concentration with the TFF 100 kDa filter showing the macro-like structures. (B) Sample after concentration and vortexing for 10 minutes. The dislodged bacterial cells and macro-like structures are stained in green and the debris is stained in orange. Scale for A is equal to 10 μm and scale for B is equal to 2 μm .
- Figure 3.7: Cell counts using flow cytometry of the 2015 fissure water samples from the Star Diamonds mine. The dot plots include gated areas for the microspheres and bacterial cells. The background noise is represented as the area below the bacterial gate. (A) Sample pre-concentration. (B) Sample after concentration. (C) Sample after concentration and vortexing for 10 minutes.
- Figure 3.8: Cell counts using flow cytometry of the 2016 fissure water samples from the Star Diamonds mine. (A) Unstained control sample. (B) Sample pre-concentration. (C)

Water from the filtrate of the TFF filter. (D) Sample after concentration and vortexing for 10 minutes. (E) Sample after concentration. (F) Sample from the viral fraction of the water.

Figure 3.9: Scanning electron microscopy images of the fissure water from Star Diamonds mine. (A) Different bacterial morphological structures where the filamentous bacteria is indicated by the red and green while the coccobacillus is indicated by the blue purple and orange bacteria. (B) A biofilm-like structure indicated in red. Scale bars are equal to 1 μm .

Figure 3.10: Isolated genomic from the bacterial fraction of the Star Diamonds mine fissure water visualized on an ethidium bromide-stained agarose gel 0.8% (w/v): lane M; MassRuler™ DNA ladder (Thermo Scientific), lanes G:genomic DNA .

Figure 3.11: PCR amplicons of the rRNA gene fragments. The amplicons were visualized on an ethidium bromide-stained agarose gel 1% (w/v): lane M: GeneRuler™ DNA ladder (Thermo Scientific), lane C: negative control, lane A: archaea, lane B: bacteria and lane C: eukarya.

Figure 3.12: DGGE diversity profiles for (a) Bacteria, (b) Eukarya and (c) Archaea. Lane B: bacteria, lane N: negative control, lane E: eukarya and lane A: archaea.

Figure 3.13: EFM images of the bacterial and viral fraction of the fissure water from Star Diamonds mine. (A) Bacterial fraction sample, (B) viral fraction sample prior to CsCl gradient ultracentrifugation and (C) viral fraction sample after CsCl gradient ultracentrifugation. Scale bar is equal to 2 μm for A and 1 μm for B and C.

Figure 3.14: TEM images of the phage morphologies found in the fissure water from Star Diamonds mine. A and B are *Podoviridae*-like phages with head diameters of 59 nm and C is a *Myoviridae*-like phage with head diameter of 113.3 nm and a short non-contractile tale. Scale bars are equal to 100 nm.

Figure 3.15: Extracted viral genomic DNA from the viral fraction of the Star Diamonds mine fissure water visualized on an ethidium bromide-stained agarose gel 0.8% (w/v): lane M; MassRuler™ DNA ladder (Thermo Scientific), lanes V: viral genomic DNA/virome.

Figure 3.16: Gradient PCR amplification of the 1F/1R primers which amplify the conserved loci integrase gene of lysogenic phages. Amplification was performed on the viral genomic DNA extracted from the viral fraction of the Star Diamonds mine fissure water. The amplicons were visualized on an ethidium bromide-stained agarose gel 1% (w/v): lane M; GeneRuler™ DNA ladder (Thermo Scientific), lanes 1: amplification at an annealing temperature of 55°C and lane 2: at an annealing temperature of 56°C.

Figure 3.17: Gradient PCR amplification performed on the genomic DNA from the bacterial fraction of the Star Diamonds mine fissure water. The amplicons were visualized on an ethidium bromide-stained agarose gel 1% (w/v). (A): lane M; GeneRuler™ DNA ladder (Thermo Scientific), lanes 1-5 amplification using the 1F/1R primers which amplify the conserved loci integrase gene of lysogenic phage at gradient annealing temperatures 55-60°C respectively, lanes 6-9 8F/8R primers amplify the conserved loci integrase gene of lysogenic phage at gradient annealing temperatures 55-59°C respectively and lanes 10-14 T4g23F/T4g23R primers amplify the g23 (major capsid protein) from T4 type phages at gradient annealing temperatures 55-60°C respectively. (B): lane M; GeneRuler™ DNA ladder, lanes 1-5 MZIAbis/MZIA6 primers which amplify the g23 (major capsid protein) from T4 type phages at gradient annealing temperatures 55-60°C respectively, lanes 6-10 HECTORPol29F/HECTORPol711R primers amplifying the DNA polymerase from uncultured podophages, lanes 11-15 HECTORPol29F/HECTORPol500R primers amplifying the DNA polymerase from uncultured podophages and lanes 16-19 CPS3/CPS8 primers amplifying the gene encoding major capsid proteins from Cyanophages.

Figure 4.1: Isolated genomic for 16S rRNA next generation sequencing visualized on an ethidium bromide-stained agarose gel 0.8% (w/v): lane M; MassRuler™ DNA ladder (Thermo Scientific), lanes G: genomic DNA.

Figure 4.2: Phylum-level taxonomic distribution of the 2015 and 2016 samples respectively. The legend shows the phylum identification and also the percentage of Illumina tag composition represented by each phylum.

- Figure 4.3: Isolated genomic for shotgun whole metagenome sequencing visualized on an ethidium bromide-stained agarose gel 0.8% (w/v): lane M; MassRuler™ DNA ladder (Thermo Scientific), lanes G: genomic DNA.
- Figure 4.4: Krona diagram of the Metagenome data microorganisms present from MG-RAST.
- Figure 4.5: Krona diagram of the Archaea at order level from MG-RAST.
- Figure 4.6: Krona diagram of the Bacteria at order level from MG-RAST.
- Figure 4.7: Krona diagram of the Eukarya at order level from MG-RAST.
- Figure 4.8: Krona diagram of the Caudovirales found using MG-RAST.
- Figure 4.9: Incomplete prophage annotations of 6 prophages from the metagenome data contigs using PHASTER.
- Figure 4.10: Circular genome view of bin 11 aligned with a reference genome. The colour blocks represent the prediction methods used where blue represents IslandPath-DIMOB and orange SIGI-HMM. The table represents the genomic island product genes along with their prediction methods.
- Figure 4.11: Circular genome view of bin 12 aligned with a reference genome. Table represents the genomic island product genes and prediction methods.
- Figure 4.12: Circular genome view of bin 14 aligned with a reference genome. Table represents the genomic island product genes and prediction methods.
- Figure 4.13: Circular genome view of bin 18 aligned with a reference genome. Table represents the genomic island product genes and prediction methods.
- Figure 4.14: Circular genome view of bin 21 aligned with a reference genome. Table represents the genomic island product genes and prediction methods.
- Figure 4.15: Circular genome view of bin 22 aligned with a reference genome. Table represents the genomic island product genes and prediction methods.
- Figure 4.16: Circular genome view of bin 23 aligned with a reference genome. Table represents the genomic island product genes and prediction methods.

Figure 4.17: Circular genome view of bin 26 aligned with a reference genome. Table represents the genomic island product genes and prediction methods.

Figure 4.18: Circular genome view of bin 27 aligned with a reference genome. Table represents the genomic island product genes and prediction methods.

Figure 4.19: Circular genome view of bin 36 aligned with a reference genome. Table represents the genomic island product genes and prediction methods.

Figure 4.20: Circular genome view of bin 42 aligned with a reference genome. Table represents the genomic island product genes and prediction methods.

LIST OF TABLES

- Table 1.1: Microorganisms isolated in deep-South African mines.
- Table 1.2: Microorganisms isolated in deep-sea hydrothermal vents.
- Table 3.2.7.2: Oligonucleotide primers used for amplification of the V3/V4 hypervariable regions for Archaea, Bacteria and Eukarya.
- Table 3.2.12: Oligonucleotide primers used for the amplification of T4 and T7 phage genes.
- Table 3.3.1: Analyses of the borehole water sampled in 2015 and 2016.
- Table 3.3.2: Cell counts using flow cytometry for both 2015 and 2016 samples.
- Table 3.3.3: Estimated cell counts using DAPI, acridine orange stains and flow cytometry
- Table 3.3.5: Taxonomy of the pGEM[®]-T Easy cloned and sequenced bacteria DGGE bands.
- Table 3.3.7: DNA and RNA concentrations of the three extraction methods measured using the Qubit[™].
- Table 4.3.3: Phage genes from category 1 and 2 detected from the metadata contigs using VirSorter.
- Table 4.3.4.1: Binned genomes from the metagenome data contigs showing contamination and completeness percentages and taxonomy.
- Table 4.3.4.2: Bins annotation of phage associated subsystem features and the functions and specific phage genes identified.

ABBREVIATIONS

°C	Degrees Celsius
%	Percentage
~	More or less
>	Greater than
≥	Greater than/Equal to
µg/µL	Microgram per microlitre
µg/mL	Microgram per millilitre
µL	Microlitre
µm	Micrometer
µM	Micromolar
µS/cm	Microsiemens per centimeter
ABV	<i>Acidianus</i> bottle-shaped virus
AFV1	<i>Acidanus</i> filamentous virus 1
APS	Ammonium persulfate
ATP	Adenosine triphosphate
ATV	<i>Acidanus</i> two tailed virus
BLAST	Basic Local Alignment Search Tool
BLASTN	Nucleotide Basic Local Alignment Search Tool
bp	Base pair
BSA	Bovine serum albumin
BVW1	<i>Bacillus</i> virus W1

cells/mL	cells per millilitre
cm	Centimeter
CPGR	Centre for Proteomic and Genomic Research
CPR	Candidate Phyla Radiation
CRISPRs	Clustered Regularly Interspaced Short Palindromic Repeats
D6E	<i>Geobacillus</i> phage D6E
DAPI	4', 6-Diamidino-2-Phenylindole, Dihydrochloride
DGGE	Denaturing gradient gel electrophoresis
DIC	Dissolved inorganic carbon
dH ₂ O	Distilled water
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleotide triphosphates
ds	double-stranded
EC	Electrical conductivity
EDTA	Ethylenediaminetetraacetic acid
EFM	Epifluorescence microscopy
<i>et al.</i>	et alii/and others
EtBr	Ethidium bromide
EtOH	Ethanol
g	Gram
g/mL	Gram per millilitre
g/L	Gram per litre

gDNA	Genomic DNA
GI	Genomic island
GVE1/2	<i>Geobacillus</i> virus 1 or 2
HGT	Horizontal gene transfer
hr	Hour
IGS	Institute of Ground water Studies
<i>in situ</i>	On-site
IPTG	Isopropyl β -D-1-thiogalactopyranoside
Kb	Kilobase
km	Kilometer
kmbs	Kilometer below surface
kPa	Kilopascal
KDa	Kilodaltons
L	Liter
LB	Luria-bertani
LTP	Life tree project
M	Molar
mg	Milligram
mg/L	Milligram per litre
mg/mL	Milligram per millilitre
MG-RAST	Metagenomics Rapid Annotation using Subsystems Technology
min	Minute

mL	Millilitre
mm	Millimeter
mM	Millimolar
MPa	Megapascal
mS/m	Millisiemens per meter
mV	Millivolts
MWCO	Molecular weight cut-off
ng	Nanogram
nm	Nanometer
ng/μL	Nanogram per microlitre
NGS	Next Generation Sequencing
nt	Nucleotide
ORP	Oxidation reduction potential
OUT	Operational Taxonomic Units
PBS	Phosphate buffered saline
PES	Polyethersulone
PCR	Polymerase Chain Reaction
pH	Measure of acidity or basicity of a solution
PHASTER	Phage Search Tool – Enhanced Release
pmol/μL	Picomole per microliter
RAST	Rapid Annotation using Subsystems Technology
RNA	Ribonucleic acid

rRNA	Ribosomal ribonucleic acid
rpm	Revolutions per minute
SAG	Single amplification genomes
SDS	Sodium dodecyl sulfate
SEM	Scanning electron microscope
SIRV1/2	<i>S. islandicus</i> rod-shaped virus 1 and 2
SIFV	<i>S. islandicus</i> filamentous virus
sp.	Specie
ss	singe-stranded
TAE	Tris-acetate-ethylenediaminetetraacetic acid
TDS	Total dissolved solids
TE	Tris-ethylenediaminetetraacetic acid
TEM	Transmission electron microscope
TEMED	N,N,N',N'-tetramethyl-ethylene-diamine
TFF	Tangential flow filtration
TNF	Tetranucleotide frequency
TOC	Total organic carbon
tRNA	Transfer ribonucleic acid
UFS	University of the Free State
UV/Vis	Ultraviolet-visible spectrophotometry
V	Volts
v/v	Volume per volume

w/v	Weight per volume
$x g$	Acceleration due to gravity
X-Gal	5-Bromo-4-chloro-3-indoly-beta-D-galactopyranosidehosphate

CHAPTER 1

LITERATURE REVIEW

1.1. Introduction

Until about thirty years ago life in the deep subsurface was not considered a possibility due to the harsh conditions such as the high pressure and temperature, limited energy and nutrient supply, extreme acidity and alkalinity, metal toxicity and even radioactivity (Reith, 2011). This phenomenon was disproved in the 1980s when life was discovered in deep ocean hydrothermal vents (Gold, 1992). The deep terrestrial subsurface is also believed to have an extensively widespread microbial community with groundwater cell densities falling between 1×10^3 and 1×10^6 cells/mL (Akob & Küsel, 2011; Onstott *et al.*, 1998).

The tree of life consists of the domains Bacteria, Archaea and Eukarya of which the bacterial domain outweighs the other two domains put together (Woese *et al.*, 1990). The archaea and bacteria share similar cellular structures, genome organization and structure, and the function of their enzymes that are involved in basic metabolism (Erdmann, 2013). As similar as these two prokaryotes may be, they do have differences that distinguish one from the other. The archaea for instance have a different lipid composition which lacks fatty acids and their cell wall lacks a peptidoglycan (Bullock, 2000). The metabolic pathways in archaea are also different and are more complex than those of bacteria (Bullock, 2000). Differences in key genetic sequences such as the 16S rRNA gene sequence (universal gene marker found in all prokaryotes) of the two prokaryotes are also present (Casamayor *et al.*, 2002; Muyzer *et al.*, 1993; Woese *et al.*, 1990). The main difference between the two prokaryotes and the eukarya is the universal gene marker found in eukarya which is the 18S rRNA gene sequences (Diez *et al.*, 2001).

Recently however, an updated tree of life has been published by Hug and co-workers (2016) (Figure 1.1). The noticeable change to the tree of life is the addition of new phyla such as the Candidate Phyla Radiation (CPR) as a lineage of the bacterial domain (Hug *et al.*, 2016). These phyla previously undetected in diversity studies became noticeable due to the use of next-generation sequencing (NGS) techniques such as single cell genomics (SCG) and whole metagenome sequencing (Hug *et al.*, 2016). Both SCG and whole metagenome sequencing are considered to be far more advanced NGS techniques as they are able to sequence the whole genome of a cell by performing shotgun sequencing (Blainey, 2014; Hug *et al.*, 2016). SCG performs shotgun whole genome sequencing on isolated cells of interest in the whole community of a given complex sample while whole metagenome sequencing performs shotgun whole genome sequencing on the whole community present in a given complex sample (Blainey, 2014). These sequence techniques allow for comprehensive sequencing in order to

identify and assemble genomes of species that are present at low frequencies in environmental samples thereby giving rise to previously uncategorized phyla (Gawad *et al.*, 2016; Hug *et al.*, 2016). SCG and whole metagenome sequencing are genome based techniques that are able to provide information about the metabolic potential and a variety of phylogenetically informative sequences that can be used to classify organisms (Hug *et al.*, 2016). The CPR lineage has become of interest due to the characteristics of the phyla as all members have relatively small genomes and most have restricted metabolic capabilities with many being inferred as symbionts (Hug *et al.*, 2016). Examples of such symbionts include the following CPR lineages TM6, TM7 and Parcubacteria (Brown *et al.*, 2015; He *et al.*, 2015; Nelson & Stegen, 2015).

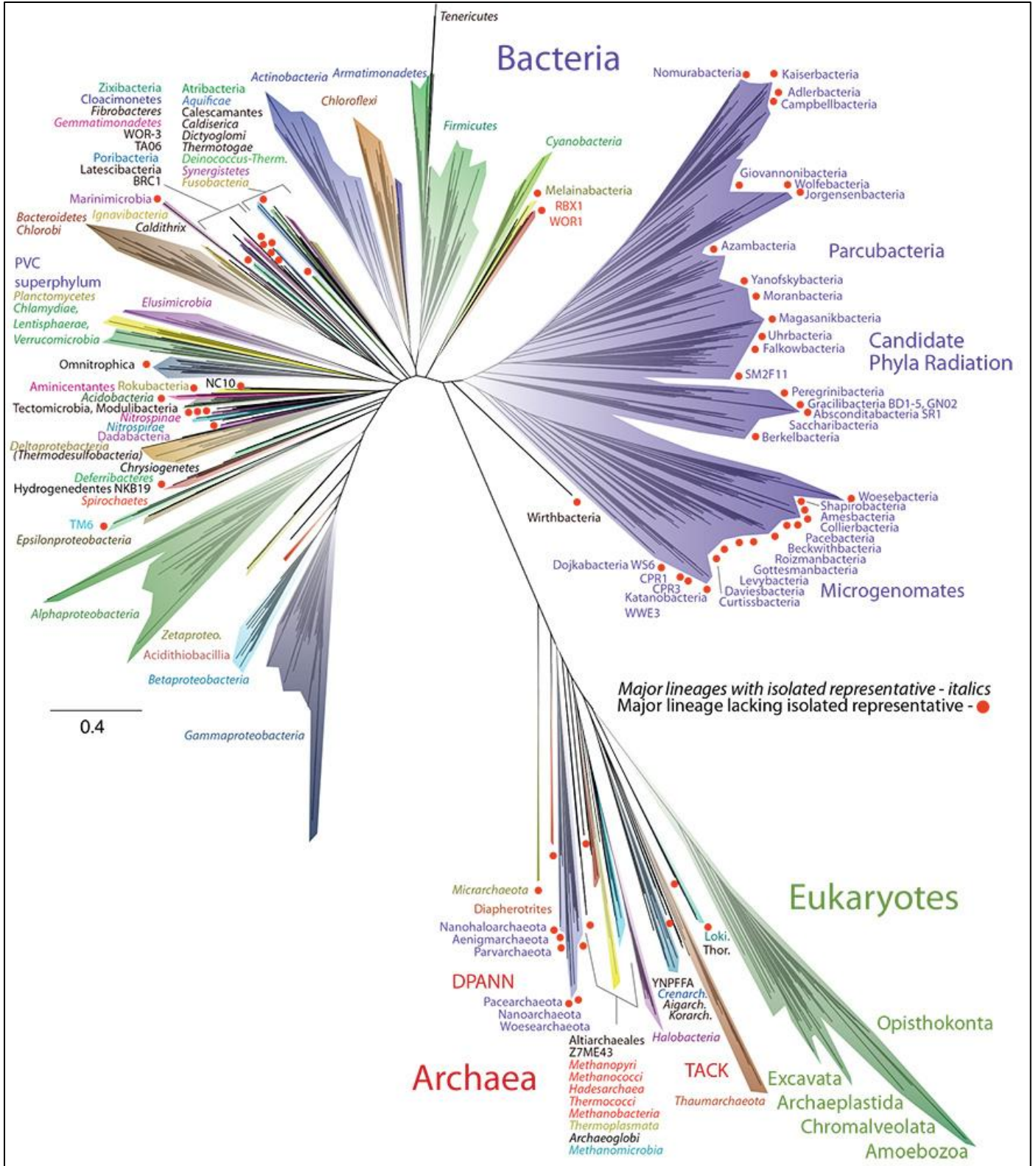


Figure 1.1: New updated tree of life highlighting the recovered lineages from the bacterial domain (Taken from (Hug et al., 2016)).

Even with the recent update to the tree of life, viruses are still not recognized as entities of life even though they are even considered as the most abundant organisms on the Earth's surface (Breitbart & Rohwer, 2005). For many years there has been a big debate to whether viruses are living or non-living entities and if they should be included into the tree of life. Numerous evidence from research has disproved the theory that viruses are living entities and that they should be a domain on the tree of life (Fitch, 2000; Koonin *et al.*, 2006; Moreira & López-García, 2009). Viruses are considered non-living because unlike the domains on the tree of life, they lack some of the enzymes needed for their expression and replication and therefore depend on the host they infect for reproduction (Moreira & López-García, 2009). In other words they neither replicate nor evolve as they are evolved by the host cell (Moreira & López-García, 2009). Viruses also lack genes for energy and carbon metabolism and they do not grow or produce waste (Moreira & López-García, 2009).

Phylogenetic trees are inferred based on characteristics that have been inherited from previous common ancestors of the taxa, but viruses do not have genes that are shared by all viruses (Fitch, 2000; Moreira & López-García, 2009). As such no phylogenetic tree can be drawn for viral lineages, a point that also makes them unsuitable for inclusion on the tree of life. There are a few genes that are shared between a specific viral lineage and their host cells, but these are acquired genes from the host to the virus during horizontal gene transfer (Fitch, 2000; Moreira & López-García, 2009). Therefore, viruses cannot be compared amongst themselves as there is an absence of common characteristics like gene contents among viral families nor can they be compared with the domains on the tree of life suggesting that viruses have various evolutionary origins thereby making them polyphyletic (Fitch, 2000; Koonin *et al.*, 2006; Moreira & López-García, 2009). Even though viruses are neither seen as living entities nor included in the tree of life, it does not imply that they do not have a significant role in the evolution of life (Moreira & López-García, 2009). From a general view environmental viruses play a major role in the biogeochemical processes especially in the deep subsurface and other extreme environments where there is little to no human modification (Hambly & Suttle, 2005).

Even though viruses have not been considered onto the tree of life, attempts have been made to phylogenetically describe their diversity. These attempts have been limited, as viruses, unlike the prokaryotes and eukaryotes, do not have a universally conserved locus making it possible to perform phylogenetic classification (Maniloff, 1995). The conserved loci for the prokaryotes and eukaryotes are the 16S and 18S rRNA gene sequences respectively and are PCR amplified and sequenced for the identification and phylogenetic classification of bacteria, archaea and eukarya

(Diez *et al.*, 2001; Lane *et al.*, 1985). In order to overcome the limitations for the identification of environmental viruses Rowher and Edwards in 2002 developed a phage proteomic tree based on 105 available complete viral/phage genomes which highlights genes or sequence fragments that are conserved in specific regions for each phage group or family. This led to the development of PCR degenerate primers specific for the gene of interest in specific phage groups (Rohwer & Edwards, 2002). Some of the primers that have been used in environmental samples include primer sets that detect the major capsid proteins from Cyanophages (Zhong *et al.*, 2002), the g23 (major capsid protein) from T4-type phages (Filée *et al.*, 2005), lysogenic phages using the integrase gene as the conserved loci (Balding *et al.*, 2005) and primer sets that detect DNA polymerases from uncultured podophages (Breibart *et al.*, 2004). The increase in the identification of viruses that do not fall into known phage families and the fact that the primers are universal/degenerate thereby limiting their specificity for a specific environment has caused limitations to the above solution.

1.2. Categories of extremophiles in extreme environments

Extremophiles are microorganisms that are able to survive under extreme physio-chemical environments (Le Romancer *et al.*, 2007). Some bacterial families are extremophiles while archaea are primarily known to be extremophiles even though they have been found to exist in moderate environments (Le Romancer *et al.*, 2007). Extremophiles are microorganisms that thrive at extreme temperature and pH conditions, high salinity, desiccation, hydrostatic pressure, radiation, anaerobiosis and very low water activity environments (Le Romancer *et al.*, 2007). Some of these extremophiles have been classified into different classes depending on the different environmental conditions they are able to thrive in. Microorganisms that thrive at relatively high temperatures (between 45°C and 80°C) are known as thermophiles while hyperthermophiles are those that thrive at temperatures above 80°C (Rampelotto, 2010). Psychrophiles are microorganisms that grow in cold environments below or at 0°C and have an optimum growth temperature of 15°C and an upper limit of 20°C (Rampelotto, 2010). Extremophiles that thrive at certain pH conditions are known as acidophiles and/or alkaliphiles. Acidophiles grow optimally at pH values of 2.0, while alkaliphiles grow optimally at pH values above 9.0, often with pH optima around 10.0 (Rampelotto, 2010). Halophiles thrive in environments that have elevated salt concentrations starting from approximately 10% sodium chloride to saturation (Rampelotto, 2010). Another extremophile class is the piezophiles which

are able to grow in high-pressure environments up to 110 Mpa and grow easily under hydrostatic pressure conditions than at atmospheric pressure (Rampelotto, 2010).

1.3. Deep terrestrial subsurface environment

Depending on the geology in a given location, the temperature within the Witwatersrand Basin of South Africa increases by approximately 8-10°C per km of depth (Lin *et al.*, 2006a). The upper limit for life is approximately 121°C which would then allow for the deep subsurface life to extend up to 12km (Kashefi & Lovley, 2003; Lin *et al.*, 2006a). The hydrostatic pressure in the terrestrial deep subsurface is dependent on the depth of the groundwater and microorganisms, which generally tolerate hydrostatic pressure of 10-1000 MPa (with the barophilic communities being able to tolerate up to 1000 MPa) (Fredrickson & Onstott, 1996).

The deep subsurface is a reducing environment that relies on chemical energy in order to fuel primary production of microorganisms in the biosphere (Gold, 1992). Chemical energy is the main source of survival in the deep subsurface as opposed to solar energy and photosynthesis due to the temperatures being too hot for photosynthetic machinery to operate and the environment being too deep for light or photosynthetic derived carbon substrates to penetrate (Chapelle *et al.*, 2002; Kieft *et al.*, 2005; Lau *et al.*, 2014; Lin, *et al.*, 2006b; Yim *et al.*, 2006). Due to the lack in photosynthetic energy the microbial communities derive their energy from the small amounts of chemicals (chemoorganotrophy) or the inorganic chemicals (chemolithotrophy). The methanogens along with the acetogens, sulfate-reducers and iron reducers are chemolithotrophs as they are all able to utilize dissolved inorganic carbon (DIC) autotrophically (Lollar *et al.*, 2006). The possible primary producers of the deep subsurface could be the lithoautotrophic microbial ecosystems as they are able to produce energy from the inorganic chemicals. In the shallow subsurface some microorganisms, through metabolism, are able to get their energy via photosynthesis derived from the surface (Lin, *et al.*, 2006b). The metabolic reactions involve electron acceptors and donors such as organic carbon and molecular oxygen (Lin, *et al.*, 2006b). The most abundant energy source to the deep subsurface communities is the organic matter and the H₂ which are most abundant in the sediments (Reith, 2011). The chemical energy sources are generated by radiolysis, thermogenesis, water-rock interactions or microbial activity and include hydrogen, methane, sulfate and hydrocarbons (Kieft *et al.*, 2005; Lin *et al.*, 2006b; Onstott *et al.*, 2006).

An increase in subsurface depth results in an increase in radioactivity caused by the increase in radioactive elements such as Uranium (^{238}U), Thorium (^{232}T) and Potassium (^{40}K) within the rocks (Blair *et al.*, 2007). The interaction of water with the rocks, in water-rock systems such as water-saturated sediments or water filled fractures, result in a transfer of energy by alpha, beta or gamma radiation which excites and ionizes the water molecules (Blair *et al.*, 2007). The major products of water radiolysis are hydrogen protons, hydrogen radicals, hydroxyl radicals, hydrogen peroxide and molecular hydrogen (H_2) (Blair *et al.*, 2007). The production of H_2 is most abundant in water-saturated sediments with high radioactive element concentrations, 50% porosity and has small grain sizes (Blair *et al.*, 2007). Water radiolysis (Figure 1.2) therefore results in an increase in abiogenic H_2 production with the increase in subsurface depth and allows for the continuous flux of H_2 as an energy source to subsurface microorganism (Blair *et al.*, 2007; Chivian *et al.*, 2008). Lithoautotrops such as methanogens, acetogens, sulphate, nitrogen and iron reducing bacteria and hydrogen oxidizing bacteria are examples of subsurface microorganisms that can use H_2 as an energy source (Chapelle *et al.*, 2002; Lin *et al.*, 2006b). These microorganisms gain their energy by coupling the oxidation of hydrogen to the reduction of compounds such as oxygen, nitrate iron, sulfate and carbon dioxide (Chapelle *et al.*, 2002). The oxidation of H_2 is activated by the enzyme hydrogenase (Park *et al.*, 2011).

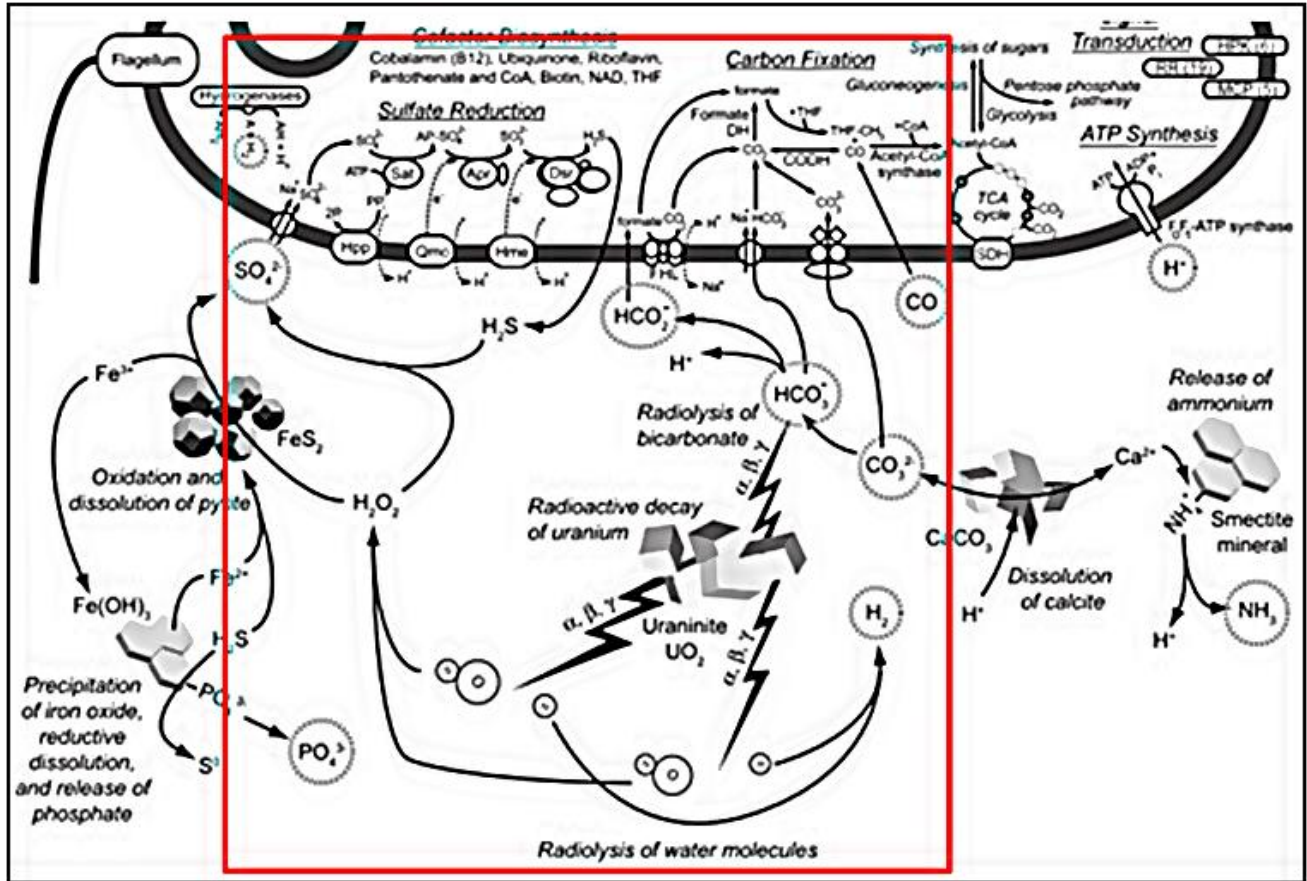


Figure 1.2: Water radiolysis within the deep terrestrial subsurface. Highlighted in the red box is the process of water radiolysis and its ability to generate the oxidation of hydrogen, which when coupled to the reduction of sulfate and carbon fixation produces energy and nutrients for the *Candidatus Desulfurudis audaxviator* (Adapted from Chivian *et al.*, 2008).

Our knowledge of the deep subsurface microorganisms is still very limited due to the difficulties in accessing the environment without introducing contaminants and the fact that the microorganisms isolated are unculturable (Colwell & D'Hondt, 2013). Research suggests that the deep subsurface microbial communities differ from those isolated on the surface with regards to their taxonomic composition, energy limitations, energy production and extremely low metabolic rates that span hundreds to thousands of years all due to nutrient deprivation in the deep subsurface environment (Chivian *et al.*, 2008; Lin *et al.*, 2006b; Phelps *et al.*, 1994; Stevens & McKinley, 1995). The microorganisms that live in such extreme environments do not only adapt but are also able to change the environmental conditions to suit their needs (Reith, 2011).

It is difficult to access deep subsurface communities as it involves extensive drilling which requires specialized equipment and this is extremely expensive; however a solution to this has been the use of deep mines and their network of tunnels to provide access. Since deep mining sites serve as access to the terrestrial subsurface, they can therefore be exploited for their boreholes/water-filled fractures as a source for sampling microbial communities. The boreholes/water-filled fractures provide an abundance of water, which serve as microbial habitats (Onstott *et al.*, 2006; Onstott *et al.*, 1998). Biofilms are another source of microbial inhabitants and these can be found hand-in-hand with the water filled fracture systems as they form close to the water filled fractures. Microbial communities form biofilms as a response to unfavourable environmental conditions such as extreme environments and low nutrient availability (Toole *et al.*, 2000). These biofilms are formed when the communities attach to hard surfaces such as hard rock, vent chimney deposits or sediments.

In South African mine rock samples it has been discovered that the dominant community is the thermophilic sulfate-reducing bacteria (Onstott *et al.*, 2003). South African deep saline fracture waters have been found to have high concentrations of hydrogen, methane and other higher hydrocarbon gases (both produced autotrophically by carbon dioxide reduction i.e. methanogenesis) (Lollar *et al.*, 2008). The methane (which is a key potential carbon and energy source) and hydrogen can be produced abiotically through mantle out-gassing and water rock interactions (Sherwood Lollar *et al.*, 2002). Mantle outgassing occurs when mantle-based rocks outgas/degas by releasing volatiles such as CO₂, H₂ and H₂S in neutral or slightly acidic fluids (Nealson *et al.*, 2005).

The Witwatersrand Basin of South Africa (Figure 1.3), is 2.9 Ga and contain some of the world's deepest mines (0.6- >4 kmbls) of which most are gold mines. The principal formation of the basin is as follows, the 2.9 Ga quartzites of the Witwatersrand Supergroup, the 2.7 Ga metamorphosed basalt, basaltic andesite of the Ventersdorp Supergroup and the sediments and volcanic strata of the 2.45 Ga Transvaal Supergroup (Bau *et al.*, 1999). The gold mines extend greater than 3 km below the surface and research on the geochemistry and microbiology of the fracture water and rocks has recently become of increased interest. Shown in Table 1.1 are some examples of microorganisms previously isolated from the deep subsurface South African mines.

Table 1.1: Microorganisms isolated in deep-South African mines.

Isolated microorganisms	South African mine	Reference
<i>Candidatus Desulforudis audaxviator</i>	Mponeng	(Chivian <i>et al.</i> , 2008)
<i>Thermus scotoductus</i> sp. strain SA-01	3.2 kmbls South African gold mine	(Kieft <i>et al.</i> , 1999)
<i>Geobacillus thermoleovorans</i> sp. strain GE-7	Driefontein	(DeFlaun <i>et al.</i> , 2007)
<i>Pyrococcus abyssi</i> sp	Kloof	(Takai <i>et al.</i> , 2001)
<i>Desulfotomaculum</i> and <i>Methanobacterium</i> sp.	Driefontein	(Moser <i>et al.</i> , 2005)
<i>Tepidibacillus infernus</i> sp.	Tau Tona	(Podosokorskaya <i>et al.</i> , 2016)

The 2.7-billion-year-old Ventersdorp Supergroup fracture water ages tens of millions of years making the water adequately old (Lippmann *et al.*, 2003; Lin *et al.*, 2006b). The water has an abundance of abiogenetic hydrocarbons and radiolytically produced H₂ which serve as potential energy sources for microorganisms within the ecosystem (Lin *et al.*, 2006b). The microbial diversity studies done in the mines from the Supergroup have shown Firmicutes, Proteobacteria, Nitrospira, Chlorobi and Thermus, Actinobacteria and Bacteroidetes phyla to be part of the bacterial diversity (Erasmus, 2015; Lin, *et al.*, 2006a; Lin *et al.*, 2006b). The archaeal diversity includes the Crenarchaeota and Euryarchaeota phyla (Erasmus, 2015; Lin, *et al.*, 2006a) and the eukaryal diversity includes the Arthropoda, Streptophyta, Nematoda, Apicomplexa and Chlorophyta phyla (Erasmus, 2015). Mines in the Ventersdorp Supergroup have become of interest to research studies due to the age of the water and microbial diversity.

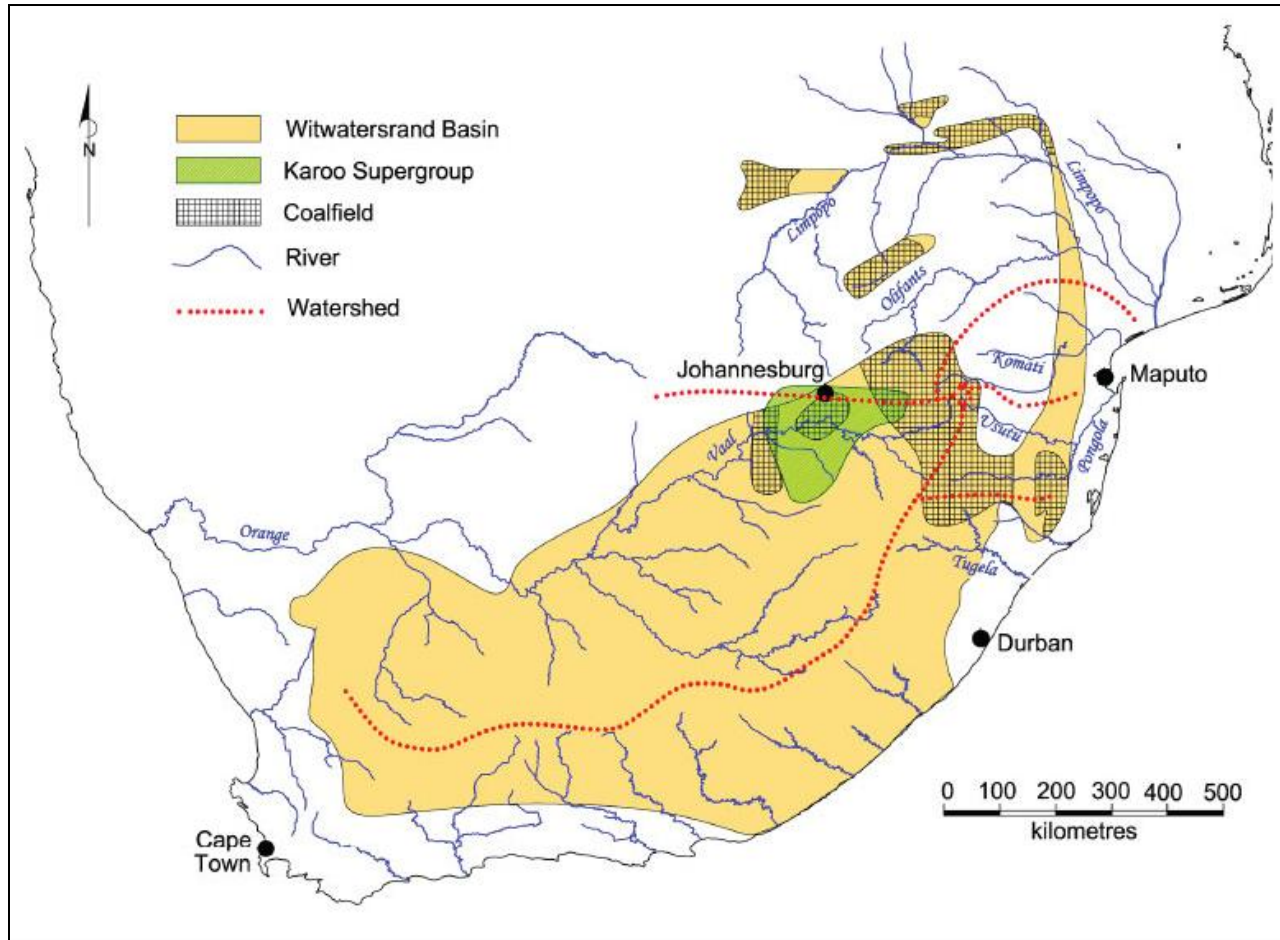


Figure 1.3: Map highlighting the Witwatersrand Basin area (Adapted from McCarthy, 2011).

1.4. Deep sea environment

The ecosystem in the deep sea covers about 65% of the Earth's surface and plays an important role in biomass production and biogeochemical cycles on a global scale which are largely mediated by benthic prokaryotes that use organic detritus for biomass production and respiration (Danovaro *et al.*, 2008). The average bacterial densities in deep marine benthic levels are between 1×10^8 and 1×10^9 cells/mL (Danovaro *et al.*, 2002.). These densities when compared to those of terrestrial subsurface counts (1×10^3 to 1×10^6 cells/mL) mentioned in Section 1.1 shows that the deep marine has a higher abundance of prokaryotic microorganisms. The benthic level/zone is the lowest level of water in the ocean or lake. The environment in the deep sea is the same as the deep terrestrial subsurface as it is characterized as dark, extreme and lacks photosynthetic primary production (Danovaro *et al.*, 2008). The deep sea ecosystem is dependent on the photosynthesis produced in surface waters which provides it with organic

matter that is needed for its sustainability (Dell'Anno *et al.*, 2015). The organic concentrations in the deep sea are 10-20 times lower than those in coastal systems resulting in low organic resources (Danovaro *et al.*, 2008). The ocean floor uses the organic matter to sustain the metabolism of benthic food webs and is a highly oligotrophic ecosystem (Dell'Anno *et al.*, 2015). The top 10 cm of the deep sea sediments has a nitrogen and phosphate rich biomass, which is equivalent to 30-45% of the total microbial carbon on Earth (Danovaro *et al.*, 2008). Chemolithoautotrophs are abundant and ubiquitous in the deep dark ocean, where they fix CO₂ (exported from the photic zone) through primary production (Anantharaman *et al.*, 2014). These microorganisms use the same metabolic reaction as lithoautotrophs, which has been mentioned previously, as they are a subset of the group. Chemolithoautotrophs, such as methanogens produce methane by oxidizing hydrogen while reducing carbon dioxide. (Akob & Küsel, 2011).

The most extreme environments in the deep sea, with regards to temperature, where microorganisms have been discovered are hydrothermal vents and plumes. Hydrothermal vents were first discovered in 1977 as areas associated with tectonically active mid-ocean ridges and basins near volcanic island arcs (Corliss *et al.*, 1979; Williamson *et al.*, 2008; Liu & Zhang, 2008; Rogers *et al.*, 2012). The vents are a result of high temperature water-rock reactions that occur when the water comes into contact with the magma (Anderson *et al.*, 2013). Hydrothermal vents are found in the deep sea subsurface, are strictly anoxic ecosystems and are the most extreme habitats on Earth as they have high hydrostatic pressures and temperatures (Le Romancer *et al.*, 2007). The ecosystem is highly acidic, reduced and enriched with chemicals including heavy metals, methane and hydrogen sulfide. The majority of energy in the environment is derived from the oxidation of hydrogen sulfide (Le Romancer *et al.*, 2007; Rogers *et al.*, 2012). The hydrothermal vents (Figure 1.4) are significant sources of Fe(II) and Mn(II) as they are 10⁶ times higher in concentration compared to the surrounding deep sea environment and are also significant sources of CH₄, H₂S and H₂ (Dick *et al.*, 2013). The water from the hydrothermal vents is chemically altered and reaches temperatures as high as 400°C (Liu & Zhang, 2008).

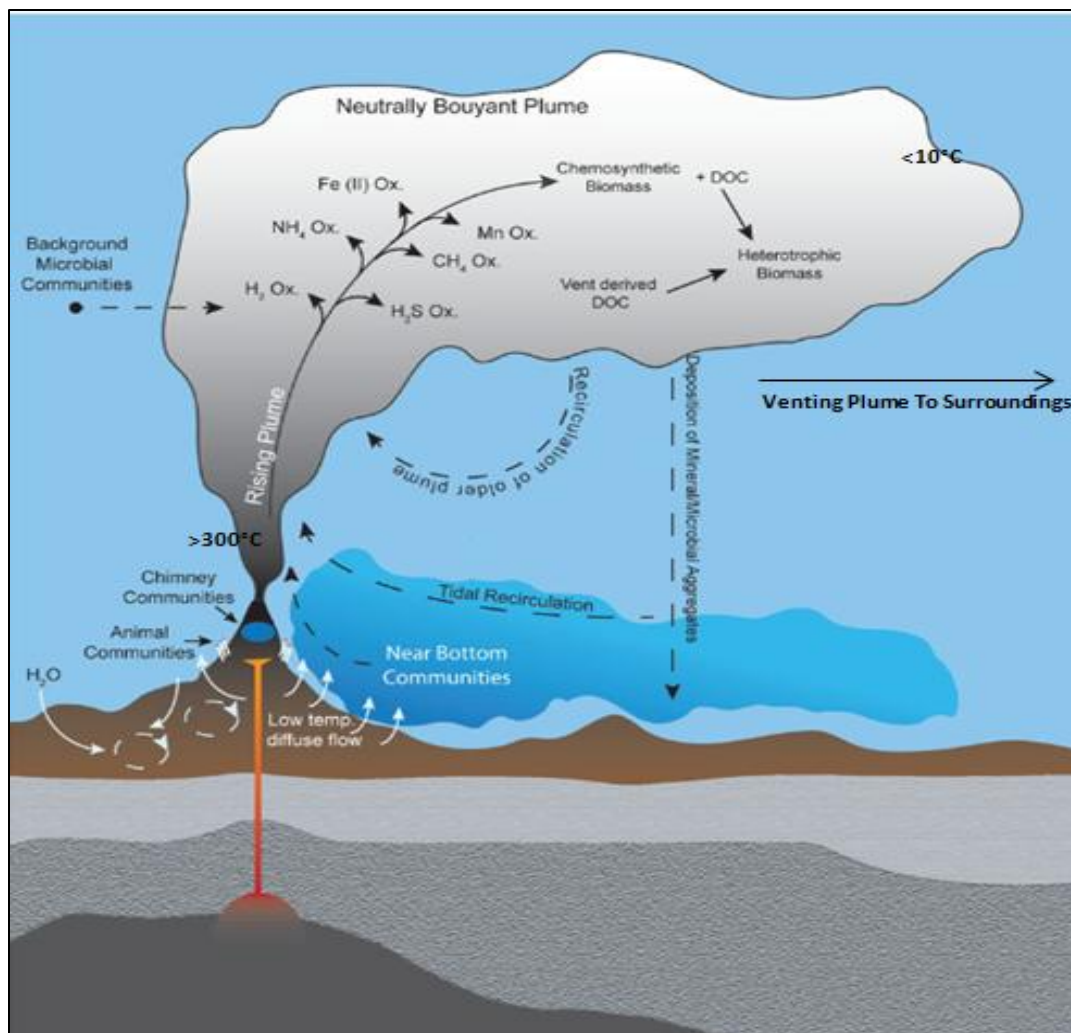


Figure 1.4: Diagram showing the microbial energy, temperature and chemical circulations in the hydrothermal vent, plumes and surrounding environments (Taken from Dick *et al.*, 2013).

The water is vented laterally and is diluted by the surrounding deep seawater forming plumes that can be carried by the currents (Ortmann & Suttle, 2005). The plumes can span up to hundreds of kilometers laterally and rise up to hundreds of meters therefore dispersing seafloor microbes from site to site (Dick *et al.*, 2013). The microbial concentrations within the area of the vents are approximately 1×10^9 /mL and in the plumes it is between 1×10^3 - 1×10^6 /mL (Ortmann & Suttle, 2005). The source of the microbial communities within the hydrothermal plumes can be derived from the sea floor communities, background deep seawater or from within the plume itself (Dick *et al.*, 2013). The plumes are characterized by milky precipitates of sulfide minerals which are a result of the hot water and cooler water mixing (Anderson *et al.*, 2013; Corliss *et al.*, 1979). The organisms in deep sea hydrothermal vents, unlike most deep ocean organisms, do

not depend on the low quality photosynthesis from the surface water, but instead depends on microbial chemoautotrophic production (Liu & Zhang, 2008; Ortmann & Suttle, 2005). The thermophilic chemosynthetic microorganisms extract energy from reduced inorganic compounds and are therefore primary producers which form the basis of the food chain in the vents (Liu & Zhang, 2008). Most of the microbial life in the deep-sea hydrothermal vents is believed to be composed of heterotrophic microorganisms, even though the primary production is chemosynthetic (Liu & Zhang, 2008). Some examples of the microorganisms that have been isolated in hydrothermal vents are shown in Table 1.2 below.

Table 1.2: Microorganisms isolated in deep-sea hydrothermal vents.

Isolated microorganisms	Hydrothermal vent Location	Reference
<i>Vibrio</i> sp.	East Pacific Rise, Mexico	(Hasan <i>et al.</i> , 2015)
Archaeal orders: <i>Thermococcales</i> and <i>Archaeoglobales</i> . Bacterial: <i>Aquificales</i> (order), Proteobacteria (phylum), and <i>Desulfobacterium</i> (genus)	Snake Pit, Mid-Atlantic Ridge	(Reysenbach <i>et al.</i> , 2000)
Isolates from genus <i>Cytophaga-Flavobacterium</i> and <i>Acidobacterium</i>	Aegean Sea, Greece	(Sievert <i>et al.</i> , 2000)
<i>Deferribacter desulfuricans</i> strain SSM1	Seamount Izu-Bonin Arc, Japan	(Takaki <i>et al.</i> , 2010)
<i>Geobacillus</i> and <i>Bacillus</i> sp.	East Pacific	(Liu <i>et al.</i> , 2006)
<i>Thiomicrospira crunogena</i> sp.	East Pacific Rise, Mexico	(Scott <i>et al.</i> , 2006)
<i>Nautilia profundicola</i> strain Am-H	East Pacific Rise Axial Caldera	(Campbell <i>et al.</i> , 2009)

1.5. Prokaryotes and eukaryotes from extreme environments

Bacteria and archaea are prokaryotic organisms meaning that they lack membrane-bound organelles (Anderson *et al.*, 2013; Woese & Fox, 1977). Most bacteria are 0.2 μm in diameter, 2-8 μm in length and have three basic cell shapes, coccus, bacillus and spiral. Archaea are generally between 0.1 μm and 15 μm in diameter, 200 μm in length and their shape is similar to that of bacteria with the exception of some thermophilic archaea. Archaea are able to survive in extreme environmental conditions and are classified as ancient life forms due to their ability to mimic conditions thought to have existed 3-4 billion years ago on Earth (Woese *et al.*, 1990). Archaea were first believed to be a separate group of bacteria (Woese & Fox, 1977) until 1990 where they were reclassified as a separate domain with methanogens as the first members. Later on, the halophiles, thermophiles and acidophiles were also added (Erdmann, 2013; Woese *et al.*, 1990). Woese and co-workers, (1990) defined the archaea into two kingdoms, the Euryarchaeota and the Crenarchaeota. There are three new phyla that have been proposed to the archaea namely the Korarchaeota, Nanoarchaeota and Thaumarchaeota (Erdmann, 2013). Bacteria are ubiquitous in all environments and unlike archaea they are generally thought to thrive in moderate environmental conditions.

Eukaryotes, which only consist of one domain which is the Eukarya, are organisms with membrane bound organelles (Doolittle, 1998; Vellai & Vida, 1999). These organisms are multicellular and have a cell size ranging from 10-100 μm . The abundance and diversity of eukarya in extreme environments such as the deep subsurface is not as prominent as that of bacteria or even archaea and this has been supported by the lack of data published prior to the discovery of the first eukarya (from the phylum Nematoda) in South African gold mines at a depth of 1.3 km below surface (Borgonie *et al.*, 2011). The nematodes were discovered to be thriving in the palaeometric fissure/fracture mine water of up to 12 300 year old (Borgonie *et al.*, 2011). The absence of lower order eukarya led to the further exploration by Borgonie and co-workers resulting in the discovery of Protozoa, Fungi, Platyhelminthes, Rotifera, Annelida and Arthropoda species in the Driefontein and Kopanang South African gold mines at depths of 1-1.4 km below surface (Borgonie *et al.*, 2015).

1.6. Microbial diversity in the South African deep mines

All three domains of life (Archaea, Bacteria and Eukarya) have been identified in diversity studies done in most of the South African deep mines. The majority of these microbial studies

have focused mainly on the bacterial diversity where the Firmicutes, Proteobacteria, Actinobacteria, Bacteroidetes, Nitrospirae, Chloroflexi, Aquificae, Cyanobacteria, Planctomycetes, Spirochaetes, Deinococcus-Thermus, Fusobacteria, Acidobacteria, OD1, OP3 and OP9 phyla have been identified (Chivian *et al.*, 2008; Erasmus, 2015; Labonté *et al.*, 2015; Lau *et al.*, 2014; Magnabosco *et al.*, 2014). Although the archaeal diversity hasn't been as extensively studied in the deep mines, the phyla Euryarchaeota and Crenarchaeota have been previously identified (Erasmus, 2015; Takai *et al.*, 2001). The Eukarya in the deep mines became of interest with the first isolation of the Nematoda in deep mines (Borgonie *et al.*, 2011; 2015). The Nematoda phylum is not the only phylum that has been identified in the deep mines as other phyla include the Streptophyta, Arthropoda, Annelida, Rotifera, Platyhelminthes, Apicomplexa, and Chlorophyta (Borgonie *et al.*, 2011; 2015; Erasmus, 2015).

1.7. Viruses in subsurface environments

There are about 10^{31} viruses on Earth and most of these viruses are phages that infect bacteria (Breitbart & Rohwer, 2005). Viruses are abundant in all of the Earth's biosphere ranging from the oceans, soils and the air (Anderson *et al.*, 2013). They are amongst the smallest biological entities and are responsible for up to 10^{23} infections per second in the ocean (Suttle, 2007). Viruses range between 20 nm to well over 800 nm in size and have ubiquitous shapes and genomic sizes ranging from a few bases to 100 kb (Anderson *et al.*, 2013). Viruses can have their genetic material made up of single-stranded DNA (ssDNA), double-stranded DNA (dsDNA), single-stranded RNA (ssRNA) or double-stranded (dsRNA) (Anderson *et al.*, 2013). The smallest viruses normally include those made up of RNA genetic material of which can be as small as 2 kb (Anderson *et al.*, 2013). The ssDNA viruses have recently been acknowledged as important members of the marine viral community, an example of this being the *Microviridae* which has been found to be one of the most common viral family in marine environments (Angly *et al.*, 2006).

Numerous studies have shown that viruses play major roles in processes ranging from microbial mortality to global geochemical cycles (Anderson *et al.*, 2013; Danovaro, 2000; Dell'Anno *et al.*, 2015; Hambly & Suttle, 2005; Li *et al.*, 2014; Liu *et al.*, 2006; Roux *et al.*, 2014). Viral metagenomic studies have shown that the environmental viruses harbor the largest genetic diversity on the planet (Breitbart & Rohwer, 2005; Hambly & Suttle, 2005). It has been estimated that in 200 liters of seawater there are about 5000 viral genotypes and that one kilogram of

marine sediment has approximately 1 million viral genotypes (Breitbart & Rohwer, 2005). The abundance of viruses in a given environment depends on the abundance of the host organism they infect (Hambly & Suttle, 2005). The Bank model is a theory based on there being two different types of viruses in a given environment, those which are active and those which are inactive/banked (Breitbart & Rohwer, 2005). The active viruses are those with hosts that are abundant in the environment and are therefore susceptible to infection (Breitbart & Rohwer, 2005). Active viruses remain active as long as the host grows (Breitbart & Rohwer, 2005). Once the environment changes, different hosts grow and the banked viruses become active while the previously active viruses start to decay and enter the banked fraction (Breitbart & Rohwer, 2005). With the Bank model the active virus-host pair behave in a Kill-the-Winner manner where the most abundant host is reduced by its viral partner therefore giving rise to another virus-host pair (Breitbart & Rohwer, 2005).

Viruses have two major reproductive lifestyle strategies (Figure 1.5), they either go through the lysogenic or lytic cycles (Weinbauer *et al.*, 2003). In the lytic cycle the bacteriophage enters the host cell and hijacks the bacterial cells replication and translation machinery by forcing the host metabolism to produce new phages (Weinbauer *et al.*, 2003). The phages then diffuse into the bacterial cell wall resulting in the lysis of the cell to release the bacteriophages (Weinbauer *et al.*, 2003). The lysogenic cycle involves the integration of the phage genome into the host genome entering a dormant stage where it becomes a prophage that replicates with the hosts genome (Anderson *et al.*, 2013). This lifecycle does not cause lysis of the cell unless induction occurs where the synthesis of the repressor, which prevents the lytic cycle from occurring, is stopped (Anderson *et al.*, 2013). This results in the prophage encoding enzymes which excise viral DNA from the bacterial chromosome resulting in lytic behavior (Anderson *et al.*, 2013).

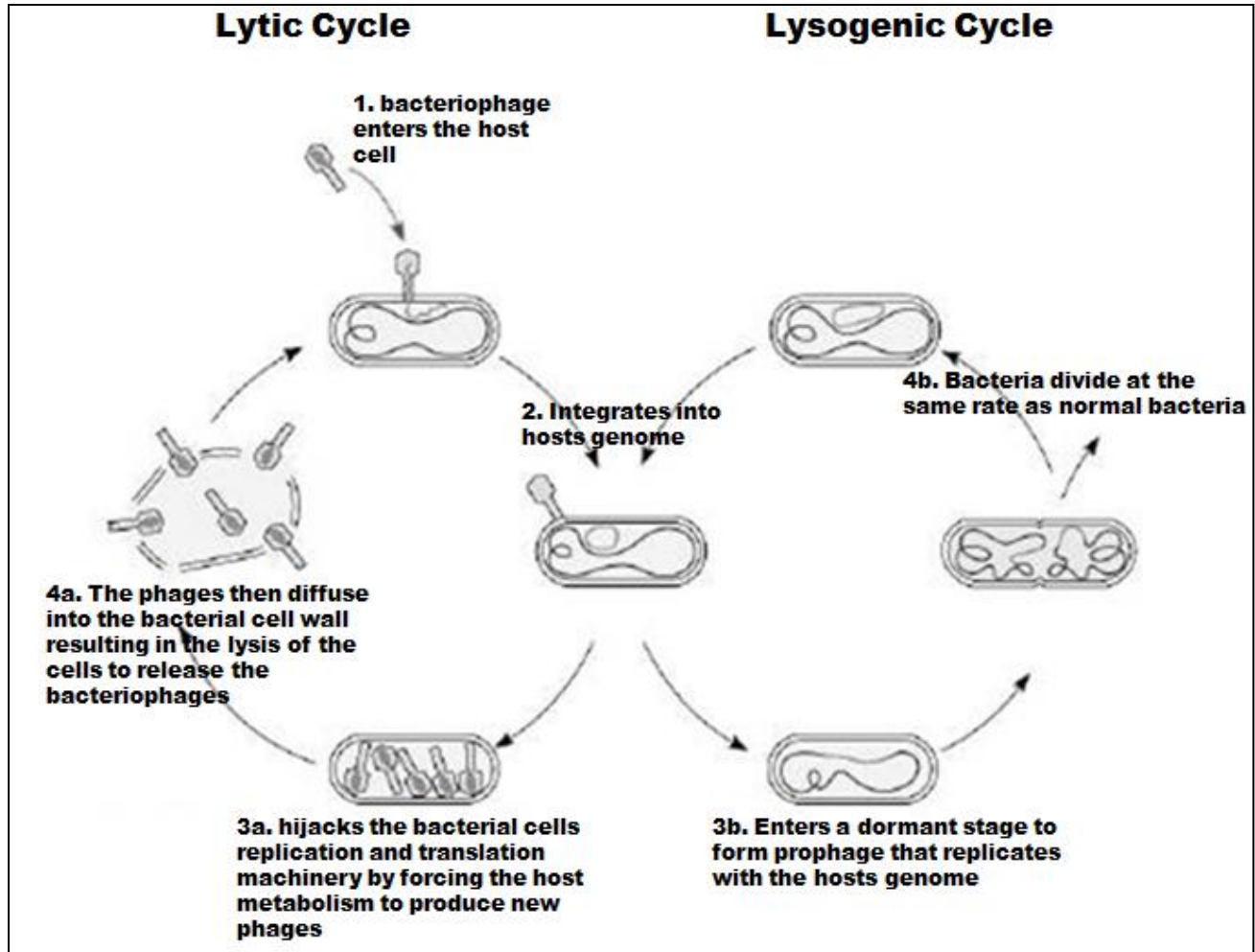


Figure 1.5: The reproductive life cycle of viruses. (1) The bacteriophage enters the host cell. (2-3a) and integrates with the hosts genome and hijacks the bacterial cell's replication and translation machinery by forcing the host metabolism to produce new phages. (4a.) Phages enzymes such as lysins and proteins such as holin's destabilize the bacterial membrane allowing for the cell to release the bacteriophages. (3b) If the phage chooses the lysogenic cycle it enters a dormant stage where it becomes a prophage that replicates with the hosts genome. (4b.) The infected bacterial cell divides at the same rate as normal bacterial cells (Adapted from Anderson *et al.*, 2013).

Extremophiles, just like all other organisms, are susceptible to viral infection and are therefore hosts for viral replication (Le Romancer *et al.*, 2007). Archaeal and bacterial viruses possess a wide range of morphologies which include filamentous, icosahedral and head-tail shapes (Anderson *et al.*, 2013). Most archaeal viruses found in extreme environments have ubiquitously unusual shapes of which some of these shapes allow the viruses to possess the ability to change shape when outside of the host in order to adapt to the harsh extracellular environment (Haring *et al.*, 2005). This change in morphology also allows some of the viruses to have

unusual mechanisms of releasing their virions from the host's cell (Bize *et al.*, 2009; Brumfield *et al.*, 2009).

The thermophilic archaeal viruses consist of 9 different families including the *Fuselloviridae*, *Bicaudaviridae*, *Ampullaviridae*, *Clavaviridae*, *Lipothrixiviridae*, *Rudiviridae*, *Glubulaviridae*, *Myoviridae* and *Siphoviridae* of which have been isolated from several genera including the *Sulfolobus*, *Acidianus*, *Pyrobaculum*, *Thermoproteus*, *Aeropyrum*, *Stygiolobus*, *Methanobacterium*, *Pyrococcus* and *Thermococcus* (Satyanarayana *et al.*, 2013). The thermophilic bacteriophages are classified into the following families which consist of the *Myoviridae*, *Siphoviridae*, *Tectiviridae* and *Inoviridae* of which have been isolated from 6 different bacterial genera including *Bacillus*, *Geobacillus*, *Thermus*, *Meiothermus*, *Rhodothermus* and *Thermonospora* (Satyanarayana *et al.*, 2013).

There have been recent discoveries of genomes greater than 100 kb which belong to giant viruses like the Mimivirus which has a genome size of 1,185 kb (La Scola *et al.*, 2003; Raoult *et al.*, 2004). The Mimivirus (infecting amoeba) was isolated from water and was the first ever discovered giant virus (Verneau *et al.*, 2016). Giant viruses have particles and genome sizes that are similar to those of small bacteria (Katzourakis & Aswad, 2014). Their DNA is double stranded and has several genes (which are uncharacteristic of viruses) that are similar to cellular genes that are involved in DNA repair, translation, protein folding and polysaccharide synthesis (Katzourakis & Aswad, 2014). These viruses are too big to infect bacteria and those that have been discovered to date, infect amoeba instead (Katzourakis & Aswad, 2014). Since the discovery of the Mimivirus there have been new discoveries of other giant viruses of which most have been isolated from soil and water (Verneau *et al.*, 2016). The discovery of these giant viruses has resulted in the introduction of two families namely the *Mimiviridae* and the *Marseilleviridae* and two putative families including pandovirus isolates and *Pithovirus sibericum* (Verneau *et al.*, 2016).

1.8. Techniques used to enumerate and characterize viruses in extreme environments

The concentration of viruses in environmental samples is diluted; hence the use of ultrafiltration and ultracentrifugation techniques to concentrate the viruses is necessary for further analysis (Sambrook & Russel, 2001). The high abundance of viruses in deep marine environments has allowed for direct cell counts and morphological characterization, but the use of ultra-

centrifugation techniques such as gradient centrifugation and ultrafiltration techniques such as tangential flow filtration is key in order to reduce background nucleic acids from prokaryotic and eukaryotic cells (Li *et al.*, 2014; Sambrook & Russel, 2001). Epifluorescence microscopy EFM is the preferred method for viral counts and has been used extensively for the enumeration of viral particles in the marine environments (Dell'Anno *et al.*, 2015; Danovaro, 2000; Danovaro *et al.*, 2008; Ortmann & Suttle, 2005; Suttle & Fuhrman, 2010). The visual morphological characterization of viruses in all deep marine environments, hot spring and solfataric environments have been conducted using transmission electron microscopy (TEM). Viruses in hot springs and solfataric environments are generally concentrated through culture enrichments (Arnold *et al.*, 2000; Haring *et al.*, 2005; Prangishvili *et al.*, 1999), but direct enumeration of viral particles in hot springs has been previously conducted using EFM where the concentrations ranged from 0.07×10^6 to 7.0×10^6 particles/mL (Breitbart *et al.*, 2004).

In the deep subsurface mine fissure water the identification of free phages using TEM has not been previously published. Instead the presence of phage related genes have been detected in bacterial metagenomic related studies (Chivian *et al.*, 2008; Labonté *et al.*, 2015).

1.9. Host-phage interactions in the subsurface

Viruses have been shown to play a role in altering the biogeochemical cycles of the ecosystem, the structure and the genetic content within the deep subsurface (Anderson *et al.*, 2011; Prangishvili & Garrett, 2004). There is an increase in the amount of research with regards to the role of viruses in the surface marine water and how they play a role in important marine processes; however, the role of viruses in the deep subsurface marine and terrestrial biospheres has been rarely considered (Anderson *et al.*, 2013; Breitbart & Rohwer, 2005; Labonté *et al.*, 2015; Prangishvili & Garrett, 2004). The abundance of phages in an environment is dependent on the abundance of the host they infect (Breitbart & Rohwer, 2005). For instance, as mentioned in Section 1.4, the prokaryotic abundances in the marine subsurface is greater than the abundances in the terrestrial subsurface; therefore, the viral abundance should be expected to be higher in the marine environments. Viruses are able to control the population in the deep subsurface from top-down through inducing cell mortality during their lytic lifecycle (Anderson *et al.*, 2013). By phages killing their hosts they in turn play an important role in nitrogen and phosphorus cycling (Dell'Anno *et al.*, 2015). According to literature, as the depth

increases, viruses become the main source of prokaryote mortality due to the lack of virus grazers (Anderson *et al.*, 2013); therefore the viruses act as environmental regulators.

1.9.1. Marine host-phage interactions

The most abundant life form in the marine ecosystems are viruses (approximately 4×10^{30} viruses with the assumption that the volume of the oceans is 1.3×10^{21} /L and the average abundance of viruses is 3×10^9 /L) and exceeding the abundance of the prokaryotic organisms by one order of magnitude (Danovaro *et al.*, 2008; Suttle, 2005). Prokaryotes dominate the deep ocean floors where they use the small amounts of organic matter from the surface photosynthesis in their metabolic reaction as a source of food and energy (Dell'Anno *et al.*, 2015). Studies using TEM for morphological characterization have shown that the most abundant phage groups in the all marine communities are the T7-like Podophage, λ -like Siphophages and the T4-like Myophage (Breitbart & Rohwer, 2005). An example of some of the phages isolated in marine environments is shown in Figure 1.6.



Figure 1.6: Typical Siphoviruses with long non contractile tails (Taken from Suttle, 2005).

Viruses influence the marine environment by altering its biogeochemical cycles by shunting their hosts through lysis, modifying the diversity and abundance of their host and by altering the host genetic content (Danovaro *et al.*, 2008). Viral lysis of the infected prokaryotes results in development of viral shunts where the cell content and biomass is transformed into organic detritus which in turn is used by the non-infected prokaryotes for respiration and biomass production (Danovaro *et al.*, 2008). This results in a proposed relationship whereby increases in

viral shunts results in higher prokaryotic growth rates (Dell'Anno *et al.*, 2015). In marine environments the viral lysis results in the removal of approximately 20-40% of the prokaryotic biomass by converting it into dissolved organic matter (DOM) through viral shunts (Anderson *et al.*, 2013) After viral lysis the viruses can either infect new hosts or go through viral decomposition of which the latter is important in both viral assemblage and the flow of energy and nutrients in the marine ecosystem (Dell'Anno *et al.*, 2015). Viruses exceed the bacterial densities in aquatic environments and are therefore considered to be important components of the aquatic microbial communities (Danovaro, 2000). Viral production in deep-ocean benthic ecosystems worldwide is extremely high and this in turn increases the viral-induced prokaryotic mortality with increase in water depth (Danovaro *et al.*, 2008).

1.9.2. Hydrothermal vent phages

Viruses that infect thermophiles are important components involved in ecological and geochemical processes in the deep-sea vent ecosystems as they play a vital part in the mortality of the thermophiles (Liu & Zhang, 2008). The first occurrence of viruses isolated through culture enrichments from thermophilic bacteria in the deep sea was shown with the isolation of two lytic bacteriophages, *Bacillus* virus W1 and *Geobacillus* virus E1 (Liu *et al.*, 2006). The *Geobacillus* virus 1 (GVE1), which infects *Geobacillus* sp E26323, was isolated in the east-Pacific hydrothermal fields, while the *Bacillus* virus W1 (BVW1), which infects *Bacillus* sp W13, was isolated from the west-Pacific hydrothermal field (Liu *et al.*, 2006). The BVW1 and GVE1 viruses both have double stranded genomic DNA of size 18 kb and 41 kb respectively. Both viruses have hexagonal heads with tails, the tail of BVW1 is longer than that of the GVE1 (Liu *et al.*, 2006). Viruses use the tails to attach to the surface of the bacterial cells in order to transport their DNA into the cell (Liu & Zhang, 2008). The *Geobacillus* virus 2 (GVE2) is a *Siphoviridae* bacteriophage like GVE1 with a hexagonal head and a tail (Liu & Zhang, 2008). The virus is virulent, infects a deep sea thermophilic *Geobacillus* sp E263 and has a double-stranded linear genomic DNA of 40,863 bp (Liu & Zhang, 2008). The deep-sea thermophilic phage D6E is a *Geobacillus* bacteriophage which was isolated from the east-Pacific deep sea hydrothermal field (Zhang & Wang, 2010). The virus has an icosahedral capsid and a contractile tail with a tail fiber (Zhang & Wang, 2010). Phage D6E belongs to the *Myoviridae* family and has circular double-stranded genomic DNA of 49,335 bp (Zhang & Wang, 2010). The BVW1, GVE1 and D6E phages are shown in Figure 1.7 where they were morphologically characterized using TEM.

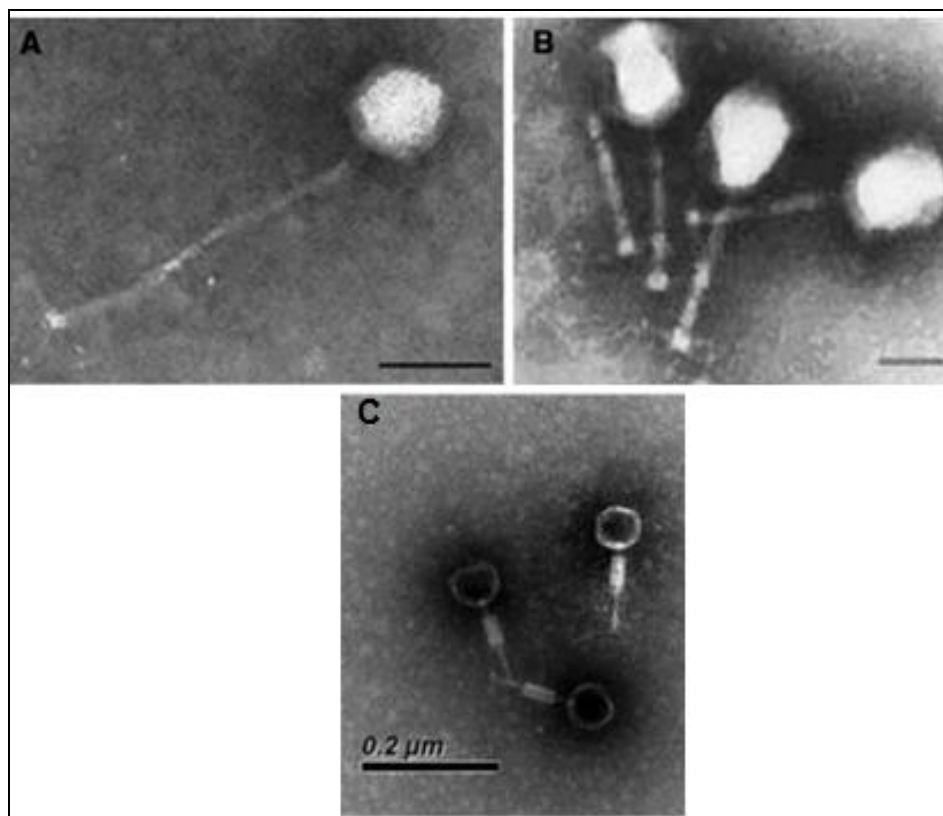


Figure 1.7: Transmission electron micrographs of phages found in the deep sea hydrothermal vent environments. (A) *Bacillus* virus W1 (BVW1), (B) *Geobacillus* virus E1 (GVE1) (Taken from Liu *et al.*, 2006), and (C) *Geobacillus* bacteriophage, D6E (Taken from Zhang & Wang, 2010).

1.9.3. Terrestrial subsurface host-phage interactions

Lysogenic lifestyles of phages are more common in the deep terrestrial subsurface (specifically those in fracture waters), this is due to phages selecting for lifestyles that will limit the necessity for finding hosts in the harsh environmental conditions (Anderson *et al.*, 2013). The lysogenic lifestyle therefore accounts for the low number of viruses in thermophilic/hyperthermophilic environments as these viruses need to protect their viral particles by reducing their exposure to extreme conditions (Prangishvili & Garrett, 2004). This lysogenic lifestyle is beneficial for the archaea and bacteria that are under energy-limited environments such as those found in the deep subsurface as it minimizes the amount of energy they use for certain metabolic pathways (Anderson *et al.*, 2013). The phages allow for the repression of wasteful bacterial and archaeal metabolic pathways and this results in a symbiotic relationship between the host and the phage rather than a parasitic relationship (Anderson *et al.*, 2013). The microbes that have formed biofilms have high cell density and therefore viruses tend to accumulate in them. The purpose of

the biofilm structure is to protect the microbes from unfavourable conditions such as viral infection, but the phages are able to penetrate the biofilm structure. The viral lifestyle in the biofilms is generally lytic and therefore results in cell death (Webb et al., 2003), but little is known about the phage-biofilm interaction to further validate why the viral lifestyle is lytic and how the phages are able to infect the biofilms (Anderson *et al.*, 2013).

1.9.4. Horizontal gene transfer

Phages are able to alter the genetic content and expression of genes in the host by the process of horizontal gene transfer through transduction (Thomas & Nielsen, 2005). Horizontal gene transfer occurs when the host's genetic material is incorporated into the viral genome (during virion synthesis) and this genetic material can then be transferred to a new host when the newly synthesized phage infects a new host (Thomas & Nielsen, 2005). By moving genetic material from one host to another the phages effectively alter the microbial evolution of the microbial species (Angly *et al.*, 2006). Phages are able to horizontally transfer genes between distantly related lineages and between domains which are relatively common with about 3% of genes in bacteria and 4-8% in archaeal genomes (Anderson *et al.*, 2011). Phages have genes that improve the hosts fitness and that can be expressed during integration of viral genome to the hosts genome (Lindell *et al.*, 2005). An example of this phenomena is the photosynthetic genes expressed by cyanophages infecting *Prochlorococcus* and *Synechococcus* which improves the photosynthesis of the host (Lindell *et al.*, 2005).

Metagenomic data, using single cell genomics, so far shows partial genomes of Mu-like transposable phages, retrons, Cluster Regulatory Interspaced Short Palindromic Repeats (CRISPRs) and lambda-like prophages in the bacterial genomes in the deep terrestrial subsurface (Labonté *et al.*, 2015). Mu-like transposable phages are temperate phages and reproduce by transposition and have been discovered within the genomes of several Firmicutes and Proteobacteria (Labonté *et al.*, 2015). Retrongs encode for reverse transcriptase and are found in most bacterial genomes where they could play a role in bacterial fitness (Labonté *et al.*, 2015). CRISPRs are sequences in the host genome that act as defense system degrading phage genomic material (Anderson *et al.*, 2013). They can be identified by their short repeat sequences separated by spacers, which are used as recognition elements to find matching viral genomes (Anderson *et al.*, 2013). In a previous study by Chivian and co-workers, (2008) the *Candidatus Desulfurudis audaxviator*, a sulfate-reducing bacterium dependent on carbon fixation and sporulation to serve as energy and food source, was isolated from the Mponeng gold mine (AngloGold Ashanti). The bacterium shared genes of archaeal origin in the nitrogen

and carbon fixation pathways, cobalamin synthesis and sulfide reduction pathways and with motility and sporulation genes. These genes were believed to be products of horizontal gene transfer (HGT) events from other bacteria and archaea and this was validated by the presence of transposase and CRISPRs in its genome (Chivian *et al.*, 2008).

1.10. Host-phage interactions in other extreme environments

1.10.1 Hot Springs

One of the extreme environments microorganisms can be found in are hot springs and South Africa has over 90 of them (Olivier *et al.*, 2011) In general hot springs are of extremely high pH (10-12) and can only support alkaliphilic microorganisms (Magnabosco *et al.*, 2014). Hot springs are a result of either volcanic activity or deep circulation of groundwater (Olivier *et al.*, 2008). Since, South Africa does not have active volcanoes, the hot springs are therefore attributed to deep circulation of groundwater (Olivier *et al.*, 2008). The following studied hot springs; Mphephu, Tshipise, Sagole, Siloam, Souting and Eiland (Figure 1.8) have temperatures ranging from 41-63°C and pH ranging from 7.2-9.5 which consider them to be alkaline (Tekere *et al.*, 2015). The dominant genera detected in these hot springs were the *Stenotrophomonas*, *Hydrogenophaga*, *Flectobacillus*, *Rheinheimera*, *Pseudomonas*, *Zavarzinella*, *Aquaspirillum* and *Limnobacter* (Tekere *et al.*, 2015). Acidic hot springs also exist and the most abundant phyla found in these hot springs is the Crenarchaeota phyla of the archaea domain, with hypothermophilic viruses isolated in the genera *Sulfolobus*, *Acidianus*, *Thermoproteus* and *Pyrobaculum* of this phyla (Takai & Sako, 1999; Prangishvili & Garrett, 2004).

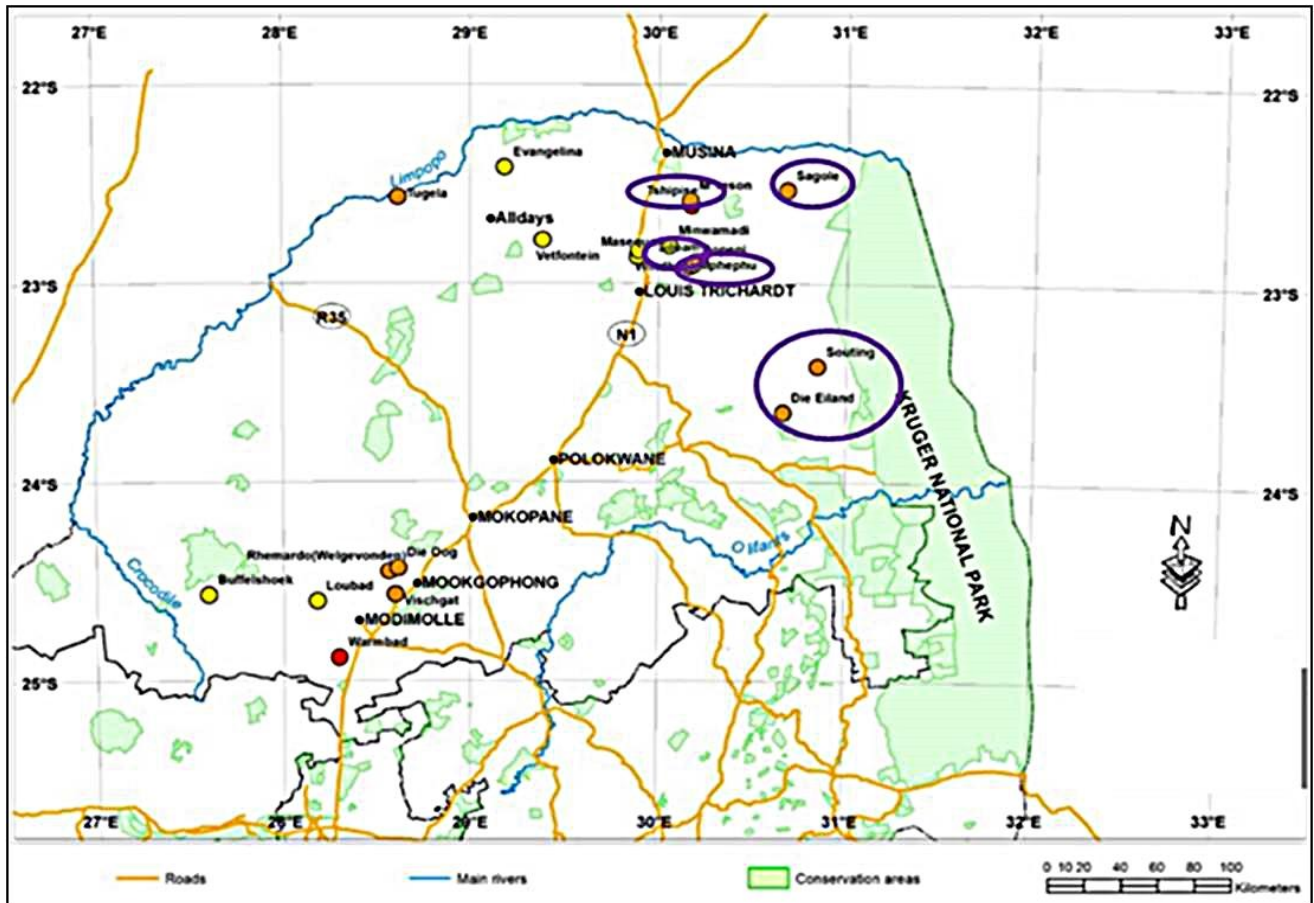


Figure 1.8: Map showing the studied alkaline hot springs (circled in purple) in Limpopo Province, South Africa (Adapted from Tekere *et al.*, 2015).

Unfortunately, there is currently no published literature on the prevalence of viruses within South African hot springs, but they have been identified in other hot springs across the globe making their presence in South African hot springs a fair assumption. The Iceland and Yellowstone National Park hot springs have temperatures above 80°C with numerous different viruses of which many have not been previously observed in nature (Haring *et al.*, 2005). Some of the novel virus families isolated have been the spindle-shaped *Fuselloviridae*, filamentous *Lipothrixviridae*, rod-shaped *Rudiviridae*, droplet-shaped *Guttaviridae* and the spherical *Globuloviridae* (Haring *et al.*, 2005). A portion of the isolated viruses have double-stranded DNA and also infect hyperthermophilic archaea from the Crenarchaeota phylum (Haring *et al.*, 2005). A novel virus assigned to a new family, was discovered in a hot acidic spring (with temperatures ranging from 87-93°C and a pH ranging from 1.5-2.) in the crater of the Solfatara volcano at Pozzuoli, Italy (Haring *et al.*, 2005). The virus was named the *Acidianus* bottle-shaped virus, ABV, as it infects strains of the hyperthermophilic archaeal genus *Acidianus* (Haring *et al.*, 2005).

The ABV virus is unique due to its previously unreported morphotype, as it has a bottle-shaped morphology (Figure 1.9) of the virions and this has led to it being assigned to a new viral family, *Ampullaviridae* (“ampulla” meaning bottle in Latin) (Haring *et al.*, 2005).

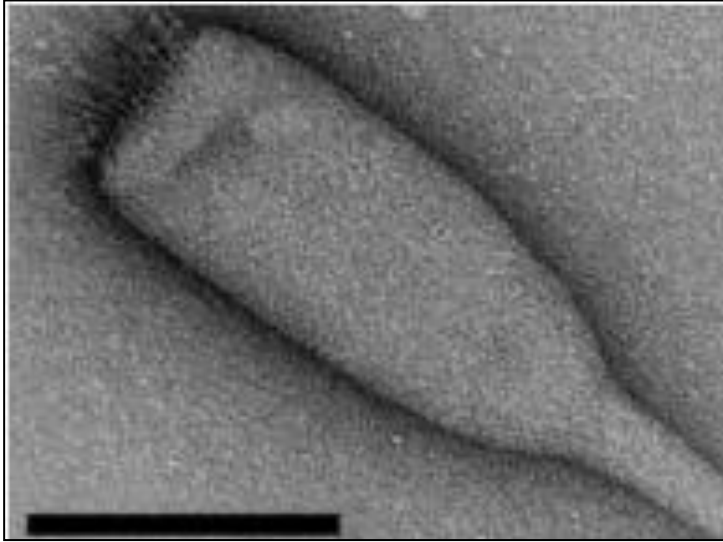


Figure 1.9: Transmission electron micrograph of the *Acidianus* bottle-shaped virus isolated in the crater of the Solfatara volcano at Pozzuoli, Italy (Taken from Haring *et al.*, 2005). Bar = 100 nm.

The ATV (*Acidianus* two tailed virus) is another virus that infects the genus *Acidianus* and has also been isolated from the hot acidic springs (87-93°C and pH 1.5-2) in Pozzuoli, Naples, Italy (Prangishvili *et al.*, 2006). The virus has a double-stranded circular DNA of 62,730 bp with 4 putative transposable elements (Prangishvili *et al.*, 2006). The virus can undergo both lytic and lysogenic reproduction and inside the host cell the virus is characterized by its lemon-shaped tailless structure but outside the host, the ATV goes under morphological development (Prangishvili *et al.*, 2006). Outside of the host the ATV develops a protrusion of tail-like appendices from the virion body (on both ends) and this morphological development is fastest at temperatures close to 85°C (Prangishvili *et al.*, 2006). Both the development of the 2 appendages and the fact that the virus goes through extracellular morphological development independent of the host, are unique and previously unknown in the viral community (Prangishvili *et al.*, 2006). The AFV1 (*Acidianus* filamentous virus 1) has been isolated from the hot springs in the Crater Hill region of Yellowstone Nation Park where the temperature is 85°C and the pH is 2 (Bettstetter *et al.*, 2003). The virus has been assigned to the *Lipothrixviridae* as it has filamentous virions with both ends carrying claw-like structures (Bettstetter *et al.*, 2003) The virions are covered with a lipid envelope and the genome consists of a linear double-stranded DNA of 20.8 kb in size (Bettstetter *et al.*, 2003). The virus chooses to enter the lysogenic cycle

and not to kill its host and therefore persists in the host in a stable carrier state (Bettstetter *et al.*, 2003).

1.10.2 Solfataric fields

The genera *Sulfolobus* is found in solfataric fields that are terrestrial habitats that mainly contain sulfur (Erdmann, 2013). These fields come in the form of soils, mud and surface waters that have been heated by volcanic activities (Erdmann, 2013). The temperature growth conditions for the *Sulfolobus* genera range from 76-90°C and the pH ranges from 1.5-2.4 (Erdmann, 2013). The archaea are able to grow autotrophically by oxidizing H₂S or sulfur to H₂SO₄ and can use CO₂ as their sole carbon source by fixing it (Erdmann, 2013). *Sulfolobus Islandicaus* serves as a host for many viruses (Erdmann, 2013) including and not limited to the SIFV, SIRV1 and SIRV2 viruses. The SIFV (*S. islandicus* filamentous virus) has been isolated from solfataric fields in Iceland (Arnold *et al.*, 2000). The virus has a linear virion and a linear double-stranded DNA genome and has been classified as a *Lipotrixvirus* (Arnold *et al.*, 2000). The SIFV has a core which is formed by a zipper-like array of DNA-associated protein subunits (Arnold *et al.*, 2000). The core is covered by a lipid containing coat which comprises of the hosts lipids (Arnold *et al.*, 2000). The SIRV 1 and 2 (*S. islandicus* rod-shaped virus 1 and 2) are part of the *Rudiviridae* family and have also been isolated from the solfataric fields in Iceland (Prangishvili *et al.*, 1999). The genomes of both viruses is linear and consists of double-stranded DNA (Prangishvili *et al.*, 1999). The size and length of the two viruses show a correlation, where SIRV1 has a genome size of 32.3 kb and SIRV2 is 35.8 kb and this correlates with the fact that SIRV1 is 70 nm shorter than SIRV2 (Prangishvili *et al.*, 1999). Both viruses are stiff rods but unlike the *Lipotrixviridae*, they lack the lipid envelope covering their core (Prangishvili *et al.*, 1999). SIRV1 and SIRV2 are non lytic viruses as they are present in their hosts in a stable carrier state (Prangishvili *et al.*, 1999).

1.11. Conclusions

Life in the deep subsurface environments does exist with all three domains being represented. Although viruses are not considered to be entities of life, they are still the most abundant microorganisms especially in biospheres such as oceans and soil.

Viruses infect their host using the lysogenic or the lytic cycle of which the lysogenic cycle is more dominant in the deep subsurface terrestrial environment. Deep subsurface viruses play a major role in altering the biogeochemical cycles, ecosystems, structure and genetic material of

the microbial diversity. Bacteria and Archaea are ubiquitous and present in all ecosystems and serve as hosts for viruses. The prokaryotic density in deep marine environments is greater than that in the terrestrial subsurface and the virus densities are the highest in marine environments. Therefore it is expected that the viral counts would be greater in marine environments due to the abundance in hosts for the viruses. The high viral densities in the deep marine environments has made it easier to perform studies on the viruses in these ecosystems. Even with the current research in the deep subsurface environments, little is understood about the viral host interactions and how they influence the deep terrestrial biosphere. Horizontal gene transfers are a result of viruses transferring genetic material from host to host. Recent research has focused on horizontal gene transfers in the deep terrestrial subsurface and how they could influence survival and evolvability of hosts such as bacteria. This influence has been seen with isolates such as the *Candidatus Desulforudis audaxviator*, which has obtained genes via horizontal gene transfers from archaeal origin and these genes have allowed the bacterium to produce its own energy and food sources in order to survive in such an extreme environment.

Accessing the deep terrestrial subsurface has been previously a challenge, but the use of mines has been a breakthrough to access the microbial matter through boreholes and fracture water. Microorganisms isolated in the deep subsurface environments are predominantly unculturable due to their extreme growth conditions, therefore culture-independent techniques like metagenomic platforms and microscopy have been used to study these microorganisms. The bacterial diversity studies done on some of the South African mines include Firmicutes, Proteobacteria, Nitrospira, Chlorobi and Thermus, Actinobacteria and Bacteroidetes phyla. These phyla could serve as hosts for the phages within the environment.

Research done by Chivian and co-workers, (2008) has suggested that the deep terrestrial subsurface is more selective than the deep marine subsurface, as 99% of the microorganisms in Mponeng mine (AngloGold Ashanti) fracture water was composed of the viral infections in only the *Candidatus Desulforudis audaxviator*, resulting in the phage related genes being evident in the isolate genome alone. In contrary other deep mines like the Tau Tona mine (AngloGold Ashanti) studied by Labonté and co-workers, (2015) have shown a broader bacterial diversity within the fracture water which in turn has shown evidence of phage infections in isolates within the phylum Firmicutes and other phyla. The effect of the host-phage interactions on the survival and evolution of the bacterial diversity in the deep mine environment is expanding with the use of highly sensitive culture-independent NGS techniques which allow for the sequencing of single genomes.

Viruses in environmental samples are concentrated using ultra-centrifugation techniques such as gradient centrifugation and ultrafiltration techniques such as tangential flow filtration which can also reduce background nucleic acids from prokaryotic and eukaryotic cells. In order to perform viral counts EFM has been used as the preferred method. Visualization and characterization of the free phages using TEM has been published for deep marine environments, hot springs and solfataric fields, but has not been published for deep mine fracture water. Instead the presence of phages and their interactions within their host has been analyzed through the use of the metagenome studies to identify prophages and phage related genes within the host genome. The use of TEM in order to visualize and characterize the phages in the deep terrestrial environment would give insight into the relative abundance of free phages and the type of phages (previously characterized in other environments or novel) found in the deep terrestrial environment.

1.12. References

Akob, D. M. and Küsel, K. (2011). Where microorganisms meet rocks in the Earth's Critical Zone. *Biogeosciences*. **8**: 3531–3543.

Anantharaman, K., Duhaime, M. B., Breier, J. A., Wendt, K. A., Toner, B. M. and Dick, G. J. (2014). Sulfur oxidation genes in diverse deep-sea viruses. *Science*. **344**: 757–760.

Anderson, R. E., Brazelton, W. J. and Baross, J. A. (2011). Is the genetic landscape of the deep subsurface biosphere affected by viruses? *Frontiers in Microbiology*. **2**: 1–16.

Anderson, R. E., Brazelton, W. J. and Baross, J. A. (2013). The deep virosphere : assessing the viral impact on microbial community dynamics in the deep subsurface. *Reviews in Mineral Geochemistry*. **75**: 649–675.

Angly, F. E., Felts, B., Breitbart, M., Salamon, P., Edwards, R. A., Carlson, C., Chan, A. M., Haynes, M., Kelley, S. Liu, H., Mahaffy, J. M., Mueller, J. E., Nulton, J., Olson, R., Parson, R., Rayhawk, S., Suttle, C. A. and Rohwer, F. (2006). The marine viromes of four oceanic regions. *PLoS Biology*. **4(11)**: 368.

Arnold, H. P., Zillig, W., Ziese, U., Holz, I., Crosby, M., Utterback, T., Weidmann, J. F., Kristjanson, J. K., Klenk, H. P. Nelson, K. E. and Fraser, C.M. (2000). A novel lipothrixvirus, SIFV, of the extremely thermophilic crenarchaeon *Sulfolobus*. *Virolog.* **267**: 252–266.

Balding C, Bromley S A, Pickup R W and Saunders J R, (2005). Diversity of phage integrases in *Enterobacteriaceae*: development of markers for environmental analysis of temperate phages. *Environmental Microbiology.* **7**: 1558–1567.

Bau, M., Romer, R. L., Luders, V. and Beukes, N. J. (1999). Pb , O , and C isotopes in silici ed Moidraai dolomite (Transvaal Supergroup , South Africa): implications for the composition of Paleoproterozoic seawater and dating the increase of oxygen in the Precambrian atmosphere. *Earth and Planetary Science Letters.* **174**: 43–57.

Bettstetter, M., Peng, X., Garrett, R. A. and Prangishvili, D. (2003). AFV1, a novel virus infecting hyperthermophilic archaea of the genus *Acidianus*. *Virology* **315**: 68–79.

Bize, A., Karlsson, E. A., Ekefjard, K., Quax, T. E. F., Pina, M., Prevost, M-C., Forterre, P., Tenailon, O., Bernander, R. and Prangishvili, D. (2009). A unique virus release mechanism in Archaea. *PNAS.* **106**: 11306–11311.

Blainey, P. C. (2014). The future is now: single-cell genomics of bacteria and archaea. *FEMS Microbial Review.* **37**: 1574–6976.

Blair, C. C., Hondt, S. D., Spivack, A. J., Kingsley, R. H. and Al, B. E. T. (2007). Radiolytic Hydrogen and Microbial Respiration in Subsurface Sediments. *Astrobiology.* **7**: 951–970.

Borgonie, G., García-Moyano, A., Litthauer, D., Bert, W., Bester, A., van Heerden, E., Möller, C., Erasmus, M. and Onstott, T. C. (2011). Nematoda from the terrestrial deep subsurface of South Africa. *Nature.* **474**: 79–82.

Borgonie, G., Linage-Alvarez, B., Ojo, A., Mundle, S., Freese, L., Van Rooyen, C., Kuloyo, O., Albertyn, J., Pohl, C., Cason, E. D., Vermeulen, J., Pienaar, C., Litthauer, D., van Niekerk, H., van Eeden, J., Sherwood Lollar, B., Onstott, T. C. and van Heerden, E. (2015). Eukaryotic opportunists dominate the deep-subsurface biosphere in South Africa. *Nature Communications.* **6**: 1–12.

Breitbart, M. and Rohwer, F. (2005). Here a virus , there a virus , everywhere the same virus ? *TRENDS in Microbiology*. **13(6)**: 276–284.

Breitbart, M., Wegley, L., Leeds, S., Rohwer, F. and Schoenfeld, T. (2004a). Phage community dynamics in hot springs. *Applied and Environmental Microbiology*. **70(3)**: 1633–1640.

Breitbart M, Miyake J H and Rohwer F, (2004b). Global distribution of nearly identical phage-encoded DNA sequences. *FEMS Microbiology Letters*. **236**: 249-256.

Brown, C. T., Hug, L. A., Thomas, B. C., Sharon, I., Castelle, C. J., Singh, A., Wilkins, M. J., Wrighton, K. C., Williams, K. H. and Banfield, J. F. (2015). Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature*. **000**: 1–18.

Brumfield, S. K., Ortmann, A. C., Ruiggrok, V., Suci, P., Douglas, T. and Young, M. J. (2009). Particle assembly and ultrastructural features associated with replication of the lytic archaeal virus Sulfolobus turreted icosahedral virus. *Journal of Virology*. **83**: 5964–5970.

Bullock, C. (2000). The Archaea- a biochemical perspective. *Biochemistry Molecular Biology Education*. **28**: 186–191.

Campbell, B. J., Smith, J. L., Hanson, T. E., Klotz, M. G., Stein, L. Y., Lee, C. K., Wu, D., Robinson, J. M., Khouri, H. M., Eisen, J. A. and Cary, S. C. (2009). Adaptations to submarine hydrothermal environments exemplified by the genome of *Nautilia profundicola*. *PLoS Genetics*. **5**: 22–25.

Casamayor, E.O., Massana, R., Benlloch, S., ØvreåS, L., Díez, B., Goddard, V.J., Gasol, J.M., Joint, I., Rodríguez-Valera, F. and Pedrós-Alió, C. (2002). Changes in Archaeal, Bacterial and Eukaryal assemblages along a salinity gradient by comparison of genetic fingerprinting methods in a multipond solar system. *Environmental Microbiology*. **4(6)**: 338-348

Chapelle, F. H., O'Neill, K., Bradley, P. M., Methé, B. A., Ciuffo, S. A., Knobel, L. L. & Lovley, D. R. (2002). A hydrogen-based subsurface microbial community dominated by methanogens. *Letters to Nature*. **415**: 312–315.

Chivian, D., Brodie, E. L., Alm, E. J., Culley, D. E., Dehal, P. S., Desantis, T. Z., Gihring, T. M., Lapidus, A., Lin, L., Lowry, S. R., Moser, D. P., Richardson, P. M., Southam, G., Wanger, G., Pratt, L. M., Andersen, G. L., Hazen, T. C., Brockman, F. J., Arkin, A. P. and Onstott, T.C. (2008). Environmental genomics reveals a single-species ecosystem deep within earth. *Science Reports*. **322**: 275–278.

Colwell, F. S. and D'Hondt, S. (2013). Nature and extent of the deep biosphere. *Reviews in Mineralogy and Geochemistry*. **75**: 547–574.

Corliss, J. B., Dymond, J., Gordon, L. I., Edmond, J. M., von Herzen, R. P., Ballard, R. D., Green, K., Williams, D., Bainbridge, A., Crane, K. and van Andel, T. H. (1979). Submarine thermal springs on the Galapagos Rift. *Science*. **203**: 1073–1083.

Danovaro, R. (2000). Viral density and virus-to-bacterium ratio in deep-sea sediments of the Eastern Mediterranean. *Applied and Environmental Microbiology*. **66(5)**: 1857–1861.

Danovaro, R., Manini, E. and Anno, A. D. (2002). Higher abundance of bacteria than of viruses in deep Mediterranean sediments. *Applied and Environmental Microbiology*. **68(3)**: 1468–1472.

Danovaro, R., Dell'Anno, A., Corinaldesi, C., Magagnini, M., Noble, R., Tamburini, C. and Weinbauer, M. (2008). Major viral impact on the functioning of benthic deep-sea ecosystems. *Nature*. **454**: 1084–1088.

Deflaun, M. F., Fredrickson, J. K., Dong, H., Pfiffner, S. M., Onstott, T. C., Balkwill, D. L., Streger, S. H., Stackebrandt, E., Knoessen, S. and van Heerden, E. (2007). Isolation and characterization of a *Geobacillus thermoleovorans* strain from an ultra-deep South African gold mine. *Systems in Applied Microbiology*. **30**: 152–164

Dell'Anno, A., Corinaldesi, C. and Danovaro, R. (2015). Virus decomposition provides an important contribution to benthic deep-sea ecosystem functioning. *PNAS*. e2014–e2019.

Dick, G. J., Anantharaman, K., Baker, B. J., Li, M., Reed, D. C. and Sheik, C. S. (2013). The microbiology of deep-sea hydrothermal vent plumes: Ecological and biogeographic linkages to seafloor and water column habitats. *Frontiers in Microbiology*. **4**: 1–16.

Díez, B., Pedrós-Alió, C., Marsh, T. L. and Massana, R. (2001). Application of denaturing gradient gel electrophoresis (DGGE) to study the diversity of marine picoeukaryotic assemblages and comparison of DGGE with other molecular techniques. *Applied and Environmental Microbiology*. **67**: 2942–2951.

Doolittle, W. F. (1998). A paradigm gets shifty. *Nature News and Views*. **392**: 15–16

Erasmus, M. (2015). *Geological Carbon Sequestration and its Influence on Subsurface Microbial Diversity and Metabolic Carbon Cycling*. (PhD thesis). University of the Free State.

Erdmann, S. (2013). *Studies of archaeal virus-host systems in thermal environments*. (PhD thesis). University of Copenhagen.

Filée, J., Tétart, F., Suttle, C. A. and Krisch H M, (2005). Marine T4-Type bacteriophages, a ubiquitous component of the dark matter of the biosphere. *Proceedings of the National Academy of Sciences*. **102**: 12471-12476.

Fredrickson, J.K. and Onstott, T.C. (1996). Microbes deep inside the Earth. *Scientific American*. **275**: 68-73.

Fitch, W. M. (2000). Homology a personal view on some of the problems. *Trends in Genetics*. **16**: 227–231.

Gawad, C., Koh, W. and Quake, S. R. (2016). Single-cell genome sequencing : current state of the science. *Nature Reviews*. **17**: 175–188.

Gold, T. (1992). The deep , hot biosphere. *Proceedings of the National Academy of Science of the U S A*. **89**: 6045–6049.

Hambly, E. and Suttle, C. A. (2005). The virosphere , diversity , and genetic exchange within phage communities. *Current Opinion Microbiology*. **8**: 444–450.

Haring, M., Rachel, R., Peng, X., Garrett, R. A. and Prangishvili, D. (2005). Viral diversity in hot springs of Pozzuoli , Italy , and characterization of a unique archaeal virus , Acidianus Bottle-Shaped Virus , from a New Family , the Ampullaviridae. *Journal of Virology*. **79**: 9904–9911.

Hasan, N. a, Grim, C. J., Grim, C. J., Lipp, E. K., Lipp, E. K., Rivera, I. N. G., Rivera, I. N. G., Chun, J., Haley, B. J. and Taviani, E., Choi, S. Y., Hoq, M., Munk, A. C., Brettin, T. S., Bruce, D., Challacombe, J. F. Detter, J.C., Han, C. S., Eisen, J. A., Huq, A. and Colwell, R. R. (2015). Deep-sea hydrothermal vent bacteria related to human pathogenic *Vibrio* species. *Proceedings of the National Academy of Science of the U S A.* E2813–E2819.

He, X., Mclean, J. S., Edlund, A., Yooseph, S., Hall, A. P., Liu, S., Dorrestein, P. C., Esquenazi, E., Hunter, R. C., Genhong, C., Nelson, K. E., Renate, L. and Wenyuan, S. (2015). Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. *PNAS.* **112(1)**: 244–249.

Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., Butterfield, C. N., Herndorf, A. W., Amano, Y., Ise, K., Suzuki, Y., Dudek, N., Relman, D. A., Finstad, K. M., Amundson, R., Thomas, B. C. and Banfield, J. F. (2016). A new view of the tree of life. *Nature Microbiology* **1**: 1–6.

Kashefi, K. and Lovley, D. R. (2003). Extending the upper temperature limit for life. *Science.* **301**: 934.

Katzourakis, A. and Aswad, A. (2014). The origins of giant viruses, virophages and their relatives in host genomes. *BMC Biology* **12(51)**: e1–e4.

Kieft, T. L., Fredrickson, J. K., Onstott, T. C., Gorby, Y. a., Kostandarithes, H. M., Bailey, T. J., Kennedy, D. W., Li, S. W., Plymale, A. E., Spadoni, C. M. and Gray, M. S. (1999). Dissimilatory reduction of Fe(III) and other electron acceptors by a *Thermus* isolate. *Applied Environmental Microbiology* **65**: 1214–1221.

Kieft, T.L., McCuddy, S.M., Onstott, T.C., Davidson, M., Lin, L., Mislouack, B., Pratt, L., Boice, E., Sherwood Lollar, B., Lippmann-Pipke, J., Pfiffner, S.M., Phelps, T.J., Gihring, T., Moser, D. and van Heerden, E. (2005). Geochemically generated, energy-rich substrates and indigenous microorganisms in deep, ancient groundwater. *Geomicrobiology. Journal.* **22**: 325–335.

Koonin, E. V., Senkevich, T. G. and Dolja, V. V. (2006). The ancient virus world and evolution of cells. *Biology. Direct.* **1**: 29.

La Scola, B., Audic, S., Robert, C., Jungang, L., de Lamballerie, X., Drancourt, M., Birtles, R., Claverie, J-M. and Raoult, D. (2003). A giant virus in amoeba. *Science*. **299**: 2033.

Labonté, J. M., Field, E. K., Lau, M., Chivian, D., van Heerden, E., Wommack, K. E., Kieft, T. L., Onstott, T. C. and Stepanauskas, R. (2015). Single cell genomics indicates horizontal gene transfer and viral infections in a deep subsurface Firmicutes population. *Frontiers in Microbiology* **6**: 1–11.

Lane D J, Pace B, Olson G J, Stahl D A, Sogin M L and Pace N R, (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proceedings of the National Academy of Sciences*. **82**: 6955–6959.

Lau, M. C. Y., Cameron, C., Magnabosco, C., Brown, C. T., Schilkey, F., Grim, S., Hendrickson, S., Pullin, M., Lollar, B. S., van Heerden, E., Kieft, T. L. and Onstott, T. C.. (2014). Phylogeny and phylogeography of functional genes shared among seven terrestrial subsurface metagenomes reveal N-cycling and microbial evolutionary relationships. *Frontiers in Microbiology*. **5**: 1–17.

Li, Y., Luo, T., Sun, J., Cai, L., Liang, Y., Jiao, N. and Zhang, R. (2014). Lytic viral infection of bacterioplankton in deep waters of the Western Pacific Ocean. *Biogeosciences*. **11**: 2531–2542.

Lin, L., Hall, J., Onstott, T. C., Gihring, T., Lollar, B. S., Boice, E., Pratt, L., Lippmann-pipke, J. and Bellamy, R. E. S. (2006a). Planktonic microbial communities associated with fracture-derived groundwater in a deep gold mine of South Africa. *Geomicrobiology Journal* **23**: 475–497.

Lin, L., Wang, P., Rumble, D., Lippmann-pipke, J., Boice, E., Pratt, L. M., Lollar, B. S., Brodie, E. L., Hazen, T. C., Andersen, G., DeSantis, T. Z., Moser, D. P., Kershaw, D. and Onstott, T. C. (2006b). Long-term sustainability of a high-energy, low-diversity crustal biome. *Science*. **314**: 479–482.

Lindell, D., Jaffe, J. D., Johnson, Z. I., Church, G. M. and Chisholm, S. W. (2005). Photosynthesis genes in marine viruses yield proteins during host infection. *Nature*. **438**: 86–89.

- Lippmann, J., Stute, M., Torgersen, T., Moser, D. P., Hall, J. A., Lin, L., Borcsik, M., Bellamy, R. E. S. and Onstott, T. C.** (2003). Dating ultra-deep mine waters with noble gases and ^{36}Cl , Witwatersrand Basin, South Africa. *Geochimica et Cosmochimica Acta*. **67**: 4597–4619.
- Liu, B., Wu, S., Song, Q., Zhang, X. and Xie, L.** (2006). Two novel bacteriophages of thermophilic bacteria isolated from deep-sea hydrothermal fields. *Current Microbiology*. **53**: 163–166.
- Liu, B. and Zhang, X.** (2008). Deep-sea thermophilic *Geobacillus* bacteriophage GVE2 transcriptional profile and proteomic characterization of virions. *Applied Microbiological Biotechnology*. **80**: 697–707.
- Lollar, B. S., Lacrampe-Couloume, G., Slater, G. F., Ward, J., Moser, D. P., Gihring, T. M., Lin, L. H. and Onstott, T. C.** (2006). Unravelling abiogenic and biogenic sources of methane in the Earth's deep subsurface. *Chemical Geology*. **226**: 328–339.
- Lollar, B. S., Lacrampe-Couloume, G., Voglesonger, K., Onstott, T. C., Pratt, L. M. and Slater, G. F.** (2008). Isotopic signatures of CH_4 and higher hydrocarbon gases from Precambrian Shield sites: A model for abiogenic polymerization of hydrocarbons. *Geochimica et Cosmochimica Acta*. **72**: 4778–4795.
- Magnabosco, C., Tekere, M., Lau, M. C. Y., Linage, B., Kuloyo, O., Erasmus, M., Cason, E., van Heerden, E., Borgonie, G., Kieft, T. L., Olivier, J. and Onstott, T. C.** (2014). Comparisons of the composition and biogeographic distribution of the bacterial communities occupying South African thermal springs with those inhabiting deep subsurface fracture water. *Frontiers in Microbiology*. **5**: 1–17.
- Maniloff J.** (1995). Identification and classification of viruses that have not been propagated. *Archives of Virology*. **140**: 1515-1520.
- McCarthy, T. S.** (2011). The impact of acid mine drainage in South Africa. *South African Journal of Science*. **107**: 1–7.
- Moreira, D. and López-García, P.** (2009). Ten reasons to exclude viruses from the tree of life. *Nature Reviews*. **7**: 306–311.

Moser, D. P., Gihring, T., Fredrickson, J. K., Brockman, F. J., Balkwill, D., Dollhopf, M. E., Sherwood-Lollar, B., Pratt, L. M., Boice, E., Southam, G., Wagner, G., Baker, B. J., Pfiffner, S. M., Lin, L. and Onstott, T. C. (2005). Desulfotomaculum spp. and Methanobacterium spp. dominate 4-5 km deep fault. *Applied Environmental Microbiology* **71(12)**: 8773–8783.

Muyzer, G., De Waal, E.C. and Uitterlinden, A.G. (1993). Profiling of complex microbial populations by denaturing gradient Gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Applied and Environmental Microbiology*. **59(3)**: 695-700.

Nealson, K. H., Inagaki, F. and Takai, K. (2005). Hydrogen-driven subsurface lithoautotrophic microbial ecosystems (SLiMEs): do they exist and why should we care? *TRENDS in Microbiology*. **13**: 405–410

Nelson, W. C. & Stegen, J. C. (2015). The reduced genomes of Parcubacteria (OD1) contain signatures of a symbiotic lifestyle. *Frontiers in Microbiology* **6**: 1–14.

Olivier, J., van Niekerk, H. and van der Walt, I. (2008). Physical and chemical characteristics of thermal springs in the Waterberg area in Limpopo Province , South Africa. *Water SA*. **34(2)**: 163–174.

Olivier, J., Venter, J. S. & Jonker, C. (2011). Thermal and chemical characteristics of hot water springs in the northern part of the Limpopo Province , South Africa. *Water SA*. **37(4)**: 427–436.

Onstott, T. C., Phelps, T. J., Colwell, F. S., Ringelberg, D., White, D. C., Boone, D. R., Mckinley, J. P., Stevens, T. O., Long, P. E., Balkwill, D. L., Friffin, W. T. and Kieft, T. (1998). Observations pertaining to the origin and ecology of microorganisms recovered from the deep subsurface of Taylorsville Basin, Virginia. *Geomicrobiology Journal*. **15**: 353–385.

Onstott, T. C., Moser, D. P., Pfiffner, S. M., Fredrickson, J. K., Brockman, F. J., Phelps, T. J., White, D. C., Peacock, a., Balkwill, D., Hoover, R., Krumholz, L. R., Borscik, M. and Kieft, T.L. (2003). Indigenous and contaminant microbes in ultradeep mines. *Environmental Microbiology*. **5(11)**: 1168–1191.

Onstott, T. C., Ward, J., Lollar, B. S., Boice, E., Pratt, L. M., Pfiffner, S., Moser, D., Gihring, T., Kieft, T. L., Phelps, T.J., van Heerden, E., Litthaur, D., DeFlaun, M., Rothmel, R.,

Wanger, G. and Southam, G. (2006). The Origin and Age of biogeochemical trends in deep fracture water of the Witwatersrand Basin , South Africa. *Geomicrobiology Journal*. **23**: 369–414.

Ortmann, A. C. and Suttle, C. A. (2005). High abundances of viruses in a deep-sea hydrothermal vent system indicates viral mediated microbial mortality. *Deep-Sea Research Part I*. **52**: 1515–1527.

Park, J. M., Kim, T. Y. and Lee, S. Y. (2011). Genome-scale reconstruction and in silico analysis of the *Ralstonia eutropha* H16 for polyhydroxyalkanoate synthesis, lithoautotrophic growth, and 2-methyl citric acid production. *BMC Systems Biology*. **5(101)**: 1–11.

Phelps, T., Murphy, E., Pfiffner, S. and White, D. (1994). Comparison between geochemical and biological estimates of subsurface microbial activities. *Microbial Ecology*. 335–349.

Podosokorskaya, O .A., Merkel, A. Y., Gavrilov, S. N., Fedoseev, I., van Heerden, E., Cason, E. D., Novikov, A. A., Kolganova, T. V., Aleksei A Korzhenkov, A. A., Bonch-Osmolovskaya, E. A. and Kublanov, I. V. (2016). *Tepidibacillus infernus* Sp. Nov., a moderately Thermophilic, selenate- and arsenate-respiring hydrolytic bacterium isolated from a gold mine, and emended description of the genus *Tepidibacillus*. *International Journal Systematic and Evolutionary Microbiology*. **66**: 3189–3194

Prangishvili, D., Arnold, H. P., Götz, D., Ziese, U., Holz, I., Kristjansson, J. K. and Zillig, W. (1999). A novel virus family, the Rudiviridae: Structure, virus-host interactions and genome variability of the *Sulfolobus* viruses SIRV1 and SIRV2. *Genetics Society of America*. **152**: 1387–1396.

Prangishvili, D. and Garrett, R. A. (2004). Exceptionally diverse morphotypes and genomes of crenarchaeal hyperthermophilic viruses. *Biochemical Society Transactions*. **32**: 204–208.

Prangishvili, D., Vestergaard, G., Häring, M., Aramayo, R., Basta, T., Rachel, R. and Garrett, R. A. (2006). Structural and genomic properties of the hyperthermophilic archaeal Virus ATV with an extracellular stage of the reproductive cycle. *Journal of Molecular Biology*. **359**: 1203–1216.

Rampelotto, P. H. (2010). Resistance of microorganisms to extreme environmental conditions and its contribution to astrobiology. *Sustainability* **2**: 602–1623.

Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M. and Claverie J-M. (2004). The 1.2-megabase genome sequence of Mimivirus. *Science*. **306**: 1344–1350.

Reith, F. (2011). Life in the deep subsurface. *Geology Society America*. **39(3)**: 287–288.

Reysenbach, A. L., Longnecker, K. and Kirshtein, J. (2000). Novel bacterial and archaeal lineages from an in situ growth chamber deployed at a mid-atlantic ridge hydrothermal vent. *Applied Environmental Microbiology* **66(9)**: 3798–3806.

Rogers, A. D., Tyler, P. a., Connelly, D. P., Copley, J. T., James, R., Larter, R. D., Linse, K., Mills, R. a., Garabato, A. N., Pancost, R. D., Pearce, D. A., Poulunin, N. V. C., German, C. R., Shank, T., Boersch-Supan, P. H., Alker, B. J., Aquilina, A., Bennett, S. A., Clarke, A., Dinley, R. J. J., Graham, A. G. C., Green, D. R. H., Hawkes, J. A., Hepburn, L., Hilario, A., Huvenne, V. A. I., Marsh, L., Ramirez-Llodra, E., Reid, W. D. K., Roterman, C. N., Sweeting, C. J., Thatje, S. and Zwirgmaier, K. (2012). The discovery of new deep-sea hydrothermal vent communities in the Southern Ocean and implications for biogeography. *PLoS Biology*. **10(1)**: e1001234.

Rohwer F and Edwards, (2002). The Phage Proteomic Tree: a genome-based taxonomy for phage. *Journal of Bacteriology*. **184**: 4529-35.

Le Romancer, M., Gaillard, M., Geslin, C. and Prieur, D. (2007). Viruses in extreme environments. *Reviews in Environmental Science and Biotechnology*. **6**: 17–31.

Roux, S., Hawley, A. K., Beltran, M. T., Scofield, M., Schwientek, P., Stepanauskas, R., Woyke, T., Hallam, S. J. and Sullivan, M. B. (2014). Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and meta-genomics. *Ecology: Microbiology Infectious Disease*. **3**: 1–20.

Sambrook, J. and Russell, D. W. (2001). Molecular cloning: a laboratory manual, (2rd ed.). *Cold Spring Harbor Laboratory Press*. Cold Spring Harbor, NY.

Satyanarayana, T., Littlechild, J. and Kawarabayasi, Y. (2013). Thermophilic Microbes in environmental and industrial biotechnology: Biotechnology of thermophiles, (2nd ed.). *Springer Dordrecht Heidelberg NY*.

Scott, K. M., Sievert, S. M., Abril, F. N., Ball, L. A., Barrett, C. J., Blake, R. A., Boller, A. J., Chain, P. S. G., Clark, J. A., Davis, C. R., Detter, C., Do, K. F., Dobrinski, K. P., Faza, B. I., Fitzpatrick, K. A., Freyermuth, S. K., Harmer, T. L., Hauser, L. J., Hugler, M., Kerfeld, C. A., Koltz, M. G., Kong, W. W., Land, M., Lapidus, A., Larimer, F. W., Longo, D. L., Lucas, S., Malfatti, S. A., Massey, S. E., Martin, D. D., McCuddin, Z., Meyer, F., Moore, J. L., Ocampo, L. H., Paul, J. H., Paulsen, I. T., Reep, D. K., Ren, Q., Ross, R. L., Sato, P. Y., Thomas, P., Tinkham, L. E. and Zeruth, G. T. (2006). The genome of deep-sea vent chemolithoautotroph *Thiomicrospira crunogena* XCL-2. *PLoS Biology*. **4(12)**: 2196–2212.

Sherwood Lollar, B., Westgate, T. D., Ward, J. a, Slater, G. F. and Lacrampe-Couloume, G. (2002). Abiogenic formation of alkanes in the Earth's crust as a minor source for global hydrocarbon reservoirs. *Nature*, **416**: 522–524.

Sievert, S. M., Kuever, J. A N. and Muyzer, G. (2000). Identification of 16S Ribosomal DNA-defined bacterial populations at a shallow submarine hydrothermal vent near Milos Island (Greece). *Applied and Environmental Microbiology*. **66**: 3102–3109.

Stevens, T. O. nad McKinley, J. P. (1995). Lithoautotrophic microbial ecosystems in deep basalt aquifers. *Science*. **270**: 450–455.

Suttle, C. A. (2005). Viruses in the sea. *Nature*. **437**: 356–361.

Suttle , C. A. (2007). Marine virus- major players in the global ecosystem. *Nature Reviews Microbiology*. **5**: 801–812.

Suttle, C. A. and Fuhrman, J. A. (2010). Enumeration of virus particles in aquatic or sediment samples by epifluorescence microscopy. *Limnology and Oceanography*. **145–153**.

Takai, K. & Sako, Y. (1999). A molecular view of archaeal diversity in marine and terrestrial hot water environments. *FEMS Microbiology Ecology*. **28**: 177–188.

Takai, K., Moser, D. P., Flaun, M. D. E., Onstott, T. C. and Fredrickson, J. K. (2001). Archaeal diversity in waters from deep South African gold mines. *Applied and Environmental Microbiology*. **67(12)**: 5750–5760.

Takaki, Y., Shimamura, S., Nakagawa, S., Fukuhara, Y., Horikawa, H., Ankai, A., Harada, T., Hosoyama, A., Oguchi, A., Fukui, S., Fujita, N., Takami, H. and Takai, K. (2010). Bacterial lifestyle in a deep-sea hydrothermal vent chimney revealed by the genome sequence of the thermophilic bacterium *Deferribacter desulfuricans* SSM1. *DNA Research*. **17**: 123–137.

Tekere, M., Lötter, A., Olivier, J. and Venter, S. (2015). Bacterial Diversity in Some South African Thermal Springs : A Metagenomic Analysis. *Proceeding World Geothermal Congress*. 19–25.

Thomas, C. M. and Nielsen, K. A. (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature Reviews in Microbiology*. **3**: 711–721.

Toole, G. O., Kaplan, H. B. and Kolter, R. (2000). Biofilm formation as microbial development. *Annual Reviews Microbiology*. **54**: 49–79.

Vellai, T. and Vida, G. (1999). The origin of eukaryotes: the difference between prokaryotic and eukaryotic cells. *Proceedings of the Royal Society B*. **266**: 1571–1577.

Verneau, J., Levasseur, A., Raoult, D., Scola, B. La & Colson, P. (2016). MG-Digger : An Automated Pipeline to Search for Giant Virus-Related Sequences in Metagenomes. *Frontiers in Microbiology*. **7**: 1–11.

Webb, J. S., Givskov, M. and Kjelleberg, S. (2003). Bacterial biofilms: prokaryotic adventures in multicellularity. *Current Opinion in Microbiology*. **6**: 578–585.

Weinbauer, M. G., Brettar, I. and Ho, M. G. (2003). Lysogeny and virus-induced mortality of bacterioplankton in surface , deep , and anoxic marine waters. *Limnology and Oceanography*. **48(4)**: 457–1465.

Williamson, S. J., Cary, S. C., Williamson, K. E., Helton, R. R., Bench, S. R., Winget, D. and Wommack, K. E. (2008). Lysogenic virus – host interactions predominate at deep-sea diffuse-flow hydrothermal vents. *Hydrothermal Vent lysogeny*. **2**: 1112–1121.

Woese, C. R. and Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceeding of the National Academy of Science U S A.* **74**: 5088–5090.

Woese, C. R., Kandler, O. and Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceeding of the National Academy of Science U S A.* **87**: 4576–4579.

Yim, L. C., Hongmei, J., Aitchison, J. C. and Pointing, S. B. (2006). Highly diverse community structure in a remote central Tibetan geothermal spring does not display monotonic variation to thermal stress. *FEMS Microbiology Ecology.* **57**: 80–91.

Zhang, X. and Wang, Y. (2010). Genome analysis of deep-sea thermophilic phage D6E. *Applied and Environmental Microbiology.* **76(23)**: 7861–7866.

Zhong Y, Chen F, Wilhelm S W, Poorvin L and Hodson R E, (2002). Phylogenetic diversity of marine cyanophage isolates and natural virus communities as revealed by sequences of viral capsid assembly protein gene *g20*. *Applied and Environmental Microbiology.* **8**: 1576–1584.

CHAPTER 2

INTRODUCTION TO STUDY

2.1. Introduction

As discussed in Chapter 1, predominantly viruses in the deep terrestrial subsurface environments prefer lysogenic lifecycles (Anderson *et al.*, 2013). Anderson and co-workers, (2013) stated that the lysogenic lifecycle is due to harsh environmental conditions being a limiting factor especially for phages that necessarily need to find a host.

Several studies in South African deep subsurface mining areas, using metagenomics platforms, have revealed genes believed to be products of viral infections in the genomes of indigenous bacteria from the genus *Desulforudis* and *Thermincola* (Chivian *et al.*, 2008; Labonté *et al.*, 2015; Gounder *et al.*, 2011; Lau *et al.*, 2014).

One of the leading *Science* publications from 2008 by Chivian and co-workers, (2008) already showed the presence of a Firmicute *Candidatus Desulforudis audaxviator* MP104C, from the 2.8 km deep South African Mponeng mine (AngloGold Ashanti), that had features believed to be products of horizontal gene transfer (HGT). A complete genome assembly was done using a combination of shotgun Sanger and 454 Sequencing which revealed CRISPR regions, mobile elements, transposons, intergrases and phage associated genes (Chivian *et al.*, 2008). Interestingly, the genome of the *Ca. D. audaxviator* MP104C shared genes of archaeal origin in the nitrogen and carbon fixation pathways, cobalamin synthesis and sulfide reduction pathways and with motility and sporulation genes (Chivian *et al.*, 2008). Furthermore, the motility, sporulation and carbon fixation genes were complemented by horizontally acquired systems frequently found in archaea (Chivian *et al.*, 2008).

In a first attempt by Dr L. Mabizela (2009), during her PhD study in the TIA/Biosaense research group, showed phage morphologies of viruses using transmission electron Microscopy (TEM) from a characterized biofilm mat in a 1.176 kmbs deep gold mine from the Free State area, South Africa (Beatrix, Sibanye Gold). The study also used PCR amplification to successfully detected T4-like phages and T7-like podoviruses using the major capsid protein (g23) and DNA polymerase fragment primers, respectively from fissure water from the Beatrix mine. The T4-like phage morphologies observed in the TEM images correlated with sequence-based results. Furthermore, Mabizela, (2009) sequenced the viral metagenomes and this revealed 7 prophages and insertion sequences believed to be transposase, in bacterial tRNA cluster regions becoming the first study to demonstrate these occurrences.

Thereafter, in 2015 Labonté and co-workers published a study where they used Single Cell Genomics and specifically 5 SAGs (single amplification genomes) were sequenced from *Desulforudis* isolates from different deep gold mine sample (3.14 kmbls deep South African Tau Tona (Anglo Gold Ashant) (Labonté *et al.*, 2015). These species were 99.5% identical to the *Ca. D. audaxviator* MP104C previously isolated by Chivian and co-workers, (2008) (Labonté *et al.*, 2015). The SAGs revealed phage-like sequences such as Mu-like transposable prophages, transposase, CRISPRs and retrons (Labonté *et al.*, 2015). This study was the first to demonstrate the presence of retrons in the deep subsurface environments and it confirmed the work from Mabizela in demonstrating the presence of transposable phages and their interactions with the indigenous microorganisms (Labonté *et al.*, 2015). The unusually high frequency of transposase suggested that HGTs and viral infections may play a role in the evolution of indigenous microbes in the subsurface (Labonté *et al.*, 2015).

These studies have not fully characterized subsurface viromes and this study identified the necessity to study the *in situ* bacteria-phage interactions and to evaluate their impact/ contribution to microbial evolution.

2.2. Main objectives

The objectives of this study were to:

Identify viral communities in deep subsurface water and attempt to sequence the bacterial metagenomes.

Identify phage genes and the presence of horizontal gene transfer in fissure water metagenomes.

In this study fracture water from Star diamonds mine, Frontiers mining (0.6 kmbs in South Africa) was sampled and the viral and bacterial fractions were separated. Sanger sequencing, 16S rDNA next generation and whole metagenome sequencing have been used to characterize the microbial diversity and catalogue gene transfer.

2.3. References

Anderson, R. E., Brazelton, W. J. and Baross, J. A. (2013). The deep virosphere : assessing the viral impact on microbial community dynamics in the deep subsurface. *Reviews in Mineral Geochemistry*. **75**: 649–675.

Chivian, D., Brodie, E. L., Alm, E. J., Culley, D. E., Dehal, P. S., Desantis, T. Z., Gihring, T. M., Lapidus, A., Lin, L., Lowry, S. R., Moser, D. P., Richardson, P. M., Southam, G., Wanger, G., Pratt, L. M., Andersen, G. L., Hazen, T. C., Brockman, F. J., Arkin, A. P. and Onstott, T.C. (2008). Environmental genomics reveals a single-species ecosystem deep within earth. *Science Reports*. **322**: 275–278.

Gounder, K., Brzuszkiewicz, E., Liesegang, H., Wollherr, A., Daniel, R., Gottschalk, G., Reva, O., Kumwenda, B., Srivastava, M., Bricio, C., Berenguer, J., van Heerden, E. & Litthauer, D. (2011). Sequence of the hyperplastic genome of the naturally competent *Thermus scotoductus* SA-01. *BioMed Central Genomics*. **12(577)**: 1–14.

Labonté, J. M., Field, E. K., Lau, M., Chivian, D., van Heerden, E., Wommack, K. E., Kieft, T. L., Onstott, T. C. and Stepanauskas, R. (2015). Single cell genomics indicates horizontal gene transfer and viral infections in a deep subsurface Firmicutes population. *Frontiers in Microbiology* **6**: 1–11.

Lau, M. C. Y., Cameron, C., Magnabosco, C., Brown, C. T., Schilkey, F., Grim, S., Hendrickson, S., Pullin, M., Lollar, B. S., van Heerden, E., Kieft, T. L. and Onstott, T. C. (2014). Phylogeny and phylogeography of functional genes shared among seven terrestrial subsurface metagenomes reveal N-cycling and microbial evolutionary relationships. *Frontiers in Microbiology*. **5**: 1–17.

Mabizela, L. (2009). *A metagenomic investigation of phage communities from South African mines* (PhD thesis). University of the Free State.

CHAPTER 3

SAMPLING, MICROBIAL ENUMERATION AND CHARACTERIZATION OF PHAGES USING MICROSCOPY AND SANGER SEQUENCING

3.1. Introduction

Mines have become the breakthrough solution to accessing the terrestrial deep subsurface by providing access to fracture and borehole water to study the microorganisms in the environment, that would have otherwise been inaccessible (Onstott *et al.*, 2006). Most South African gold mines are situated in the Witwatersrand Basin, which consists of the 2.9 Ga quartzites of the Witwatersrand Supergroup, the 2.7 Ga metamorphosed basalt, basaltic andesite of the Ventersdorp Supergroup and the sediments and volcanic strata of the 2.45 Ga Transvaal Supergroup (Bau *et al.*, 1999).

The Star Diamonds mine (Frontiers mining) is a 0.6 kmbs deep mine which is part of the Ventersdorp Supergroup. The mine is situated east of a dolerite hill on the Adelaide Subgroup of the Beaufort Group (Vermaak, 2012). The fissure water in this mine represents a subsurface residence time consistent with the isotopic signatures that indicate water of paleo-meteoric origin (Erasmus, 2015). The calculated radiocarbon age of the water is 24 320 (~131) years BP making it adequately older water (Erasmus, 2015). This mine is a unique study site due to the open boreholes being possibly near or falling on the fracture where the quartzites of the Witwatersrand Supergroup intersects with the basaltic andesite of the Ventersdorp Supergroup. It is also due to its adequately older water.

As mentioned previously, the microorganisms isolated in the deep subsurface environments are predominantly unculturable due to their unique growth conditions, therefore culture-independent techniques such as shotgun metagenomics sequencing and microscopy have been used to study these microorganisms (Chivian *et al.*, 2008; Danovaro, 2000; Hambly & Suttle, 2005; Labonté *et al.*, 2015).

With the knowledge that the cell densities in groundwater range from 1×10^3 - 1×10^6 cell/mL of ground water and decreases with depth and that some culture independent techniques such as transmission electron microscopy need cell densities of at least 10^6 cell/mL, it is vital to use concentration techniques during or after sampling in order to achieve viable results (Akob & Küsel, 2011; Onstott *et al.*, 1998; Vale *et al.*, 2010). Methods used to concentrate microorganisms in water include centrifugation based techniques, membrane filtration, ultrafiltration, adsorption elution (which is specific for viruses) and flocculation (Koster *et al.*, 2003). Most, if not all, of the filtration techniques share the same drawback which is that they are liable to destroying fragile cells and tangential flow filtration (TFF) (an ultrafiltration

technique) might be the most threatening due to the shear forces that are generated by the repeat passage of cells through the filter unit (Giovannoni *et al.*, 1990). However in a study by Petrusevski and co-workers, (1995) they showed that the use of TFF did not damage the ciliate structure, flagella and the general structure of bacterial cells hence TFF has since the 90's been used widely in the field of medicine for the concentration of human cells, viruses and bacterial cells.

TFF allows the cells to be kept in suspension by allowing the water to flow tangentially across the membrane (Petrusevski *et al.*, 1995; van Reis, 1993). Particles that fall in the range of the membrane pores in size are able retained within the pores and delivered back to the feed stream/retentate (Petrusevski *et al.*, 1995; van Reis, 1993). The particles that are smaller than the membrane pores pass through the membrane (Petrusevski *et al.*, 1995; van Reis, 1993). The retentate is concentrated by the continuous high velocity recirculation along the filter and the elimination of the particles smaller than the target size (Petrusevski *et al.*, 1995; van Reis, 1993). The pros of using TFF as a filter based concentration technique is that the direction of the fluid is parallel to the filter membrane; therefore preventing clogging of the membrane (van Reis, 1993). In a study done Giovannoni and co-workers, (1990) the efficiency of TFF was tested by checking cell concentrations of marine picoplankton before and after using TFF which were 5×10^5 and 4.7×10^{12} cells/mL respectively. Therefore, the cells were concentrated 7 orders of magnitude higher.

Once the retentate is concentrated then microscopy techniques are used to detect, enumerate and characterize different microorganisms. These techniques include and are not limited to epifluorescent microscopy (EFM), scanning electron microscopy (SEM) and transmission electron microscopy (TEM). Epifluorescent microscopy is a technique used to enumerate the estimated total count of bacteria or viruses in a given environment and it relies on the ability of bound dyes/stains, to a cell's DNA or RNA, to produce fluorescent glows when excited with light at the appropriate wavelength and this allows cells below the limit or resolution of light to be visualized (Porter & Feig, 1980; Young, 1961). 4',6-Diamidino-2-Phenylindole, Dihydrochloride (DAPI) is a stain that interacts with the double-stranded DNA to emit a blue fluorescence by attaching at the minor groove of the double helix (Kepner & Pratt, 1994; Young, 1961). Acridine orange stain is able to bind to double-stranded and single-stranded DNA and also single-stranded RNA (Kepner & Pratt, 1994; Young, 1961). The stain emits a green fluorescent when it interacts with three base pairs of double-stranded DNA and emits a red fluorescent when it interacts with one phosphate group of single-stranded DNA or RNA (Kepner & Pratt, 1994;

Young, 1961). The debris normally stains with an orange colour (Young, 1961). The SYBR Green I is a stain that has a high affinity to bind to double-stranded DNA and the SYBR Gold stains is more sensitive than SYBR Green I and binds to both DNA and RNA (Tuma *et al.*, 1999; Zipper *et al.*, 2004). Besides being used for bacterial cell counts, both the SYBR Gold and Green I stains can be used for viral enumeration as they yield highly estimates of viral abundance in environmental water samples (Suttle & Fuhrman, 2010).

Flow cytometry, like the fluorescent microscopic techniques, is able to detect and quantify cells using fluorescent stains with certain wavelength ranges. The cells are stained with fluorescent stains and are passed through an interrogation point individually (Ambriz-Aviña *et al.*, 2014). Flow cytometers are able to differentiate cells by measuring and counting them according to their relative size which is important for complex environmental samples that have a consortium of microorganisms (Ambriz-Aviña *et al.*, 2014). The most widely used stains for enumeration are the SYTO stains as they have high sensitive and yield bright fluorescence (Lebaron *et al.*, 1998). Some of these stains like the SYTO-9 have minimal background noise signal which improves the accuracy of the cell counts (Lebaron *et al.*, 1998). For bacterial cells SYTO BC green fluorescent nucleic acid stain is used as it is highly specific form DNA in Gram-negative and Gram-positive bacteria and yields bright fluorescence (Lebaron *et al.*, 1998).

Electron Microscopy is a well-established method for the investigation of bacteria and bacteriophages and the variation of electron microscopy techniques include and are not limited to SEM and TEM (Swart *et al.*, 2010; van der Westhuizen *et al.*, 2013). Electron microscopes use an electrical beam to produce the image of the object (Demuth *et al.*, 1993; Selby, 1953). Both SEM and TEM techniques are used for the analysis of the morphological characteristics of microorganisms where SEM images are in 3D and TEM images are 2D (Swart *et al.*, 2010; van der Westhuizen *et al.*, 2013). SEM can detect particles with sizes as small at bacterial cells, but cannot detect those as small as bacteriophages (Essers, 2003; Swart *et al.*, 2010; van der Westhuizen *et al.*, 2013). TEM is used to detect both bacteria and bacteriophages and has been used widely for the morphological characterization of bacteriophages (Demuth *et al.*, 1993; Gentile & Gelderblom, 2014 ; Swart *et al.*, 2010; van der Westhuizen *et al.*, 2013).

Besides microscopic techniques there are sequenced based techniques used to preliminarily characterize the microbial diversity in a given sample from the environment of interest. An example of such techniques is the use of Sanger sequencing to sequence universally conserved genes or sequences of certain microorganisms like the 18S rRNA gene sequences

for eukaryotes and the 16S rRNA gene sequences for prokaryotes. The rRNA ribosomal unit has highly conserved regions which are separated with hyper variable stretches making it possible for PCR primer design (García-Martínez *et al.*, 1999). Sanger sequencing has been commonly used for prokaryotes and eukaryotes, where 16S and 18S rDNA gene sequences are PCR amplified from the genomic DNA of the sample, sequenced and used in the identification and phylogenetic classification of bacteria, archaea and eukarya (Diez *et al.*, 2001; Lane *et al.*, 1985). Unfortunately unlike the prokaryotes and eukaryotes phages do not have a universally conserved gene or sequence making it impossible to perform phylogenetic classification of phages (Maniloff, 1995). In 2002, Rowher and Edwards developed a phage proteomic tree based on 105 available complete phage genomes which highlights genes or sequence fragments that are conserved for each phage group or family and this enables groups of phages to be detected. Through the use of PCR and degenerate primers specific for the gene of interest, different phage groups can be detected.

3.1.2. Aims of chapter

The aims of this chapter were to

- Access deep subsurface fracture water samples and to complete the geochemical characterization.
- Concentrate the microorganisms in the fissure water using a 100 kDa TFF filter and separate the bacterial water fraction from the viral fraction using a 0.2 µm TFF filter.
- Enumerate the microbial biomass using DAPI and acridine orange staining and flow cytometry.
- Characterize the phages morphologically using transmission electron microscopy.
- Assess the microbial diversity of the fissure water using Sanger sequencing.
- Amplify phage genes (specific to T4 and T7 like phages) within the isolated viral DNA in order to characterize the phage diversity in the fissure water.

3.2. Methods and materials

3.2.1. Study site

The Star Diamonds mine (28° 19' S, 26°47' E) is part of the Theunissen district of the Free State province, approximately 40 km south of Welkom in South Africa. The mine consists mainly of Karoo sandstone and shale and exploits a series of kimberlite fissure segments. The borehole used for this study was an already existing borehole at a depth of approximately 640 m below surface. Fissure water samples were sampled twice 11 months apart between 2015 and 2016 from the same borehole.

3.2.2. Physiochemical analyses

On site geochemical parameters were measured using the ExStik[®]II EC500 probes (EXTECH[®] instruments) which included pH, salinity, total dissolved solids (TDS), temperature, electrical conductivity (EC) and oxidation reduction potential (ORP). The probes were calibrated using the procedures provided by the manufacturer.

Fissure water was collected in a 1 L sterile autoclaved Schott bottle to be used for geochemical analysis of the water at the Institute for Groundwater Studies (IGS) at the University of the Free State. The following parameters were measured EC, pH, total organic carbon (TOC) and TDS.

3.2.3. Concentration of water using tangential flow filtration

Fissure water was collected in two 220 L plastic drums (sterilized/washed prior to sampling with fissure water) from the borehole and transported to the surface where an on-site concentration of the microorganisms in the water using tangential flow filtration (TFF) was setup as shown in Figure 3.1. TFF was setup and run according to the procedure by Rohwer, (2005). The flow rate was determined as the time it takes 1 L of the water to run through the filter. Before the water was pumped through the TFF system, an aliquot of 45 mL was fixed with 3.7% formaldehyde for 30 minutes at room temperature and kept at -20°C after. The water was first concentrated using the 100 KDa (Amersham Biosciences, UFP-100-C-4A) (2x) viral filters at a pressure of approximately 50 kPa. This setup allows for particles bigger than 100 KDa to be concentrated which includes prokaryotes, eukaryotes and viruses. From the 220 L drums aliquots of 50 L in

sterile Nalgene Carboy bottles (Thermo Fischer Scientific) were used as reservoirs/retentates connected to the TFF setup. The water was concentrated from 50 L to approximately 3 L for each interval. Concentrated aliquots of 45 mL were fixed with 3.7% formaldehyde in the same manner as the unconcentrated aliquot. The concentrated water was transported back to the lab where the 0.2 μm (Amersham Biosciences, CFP-2-E-4A) TFF filter was used to separate the viruses from the prokaryotes/eukaryotes by excluding particles smaller than 0.2 μm (which includes viruses) from entering the membrane pores. A summary setup of the whole TFF process using both the 100 kDa and the 0.2 μm filters is shown in Figure 3.2. Aliquotes from both the filtrate and retentate were fixed with 3.7% formaldehyde.

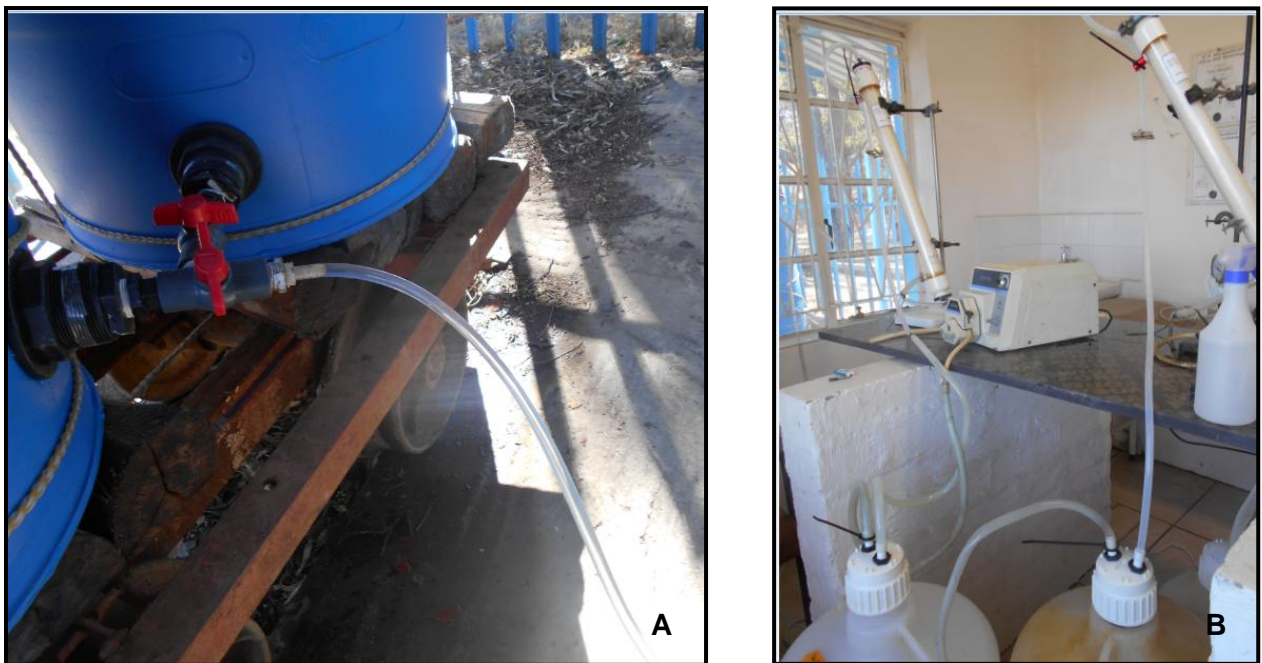


Figure 3.1: On-site tangential flow filtration setup. (A) Shows the taps and tubing on the 220 L drums that were used to fill the 50 L Nalgene Carboy drums. (B) Shows the TFF setup of the two 100 kDa filters.

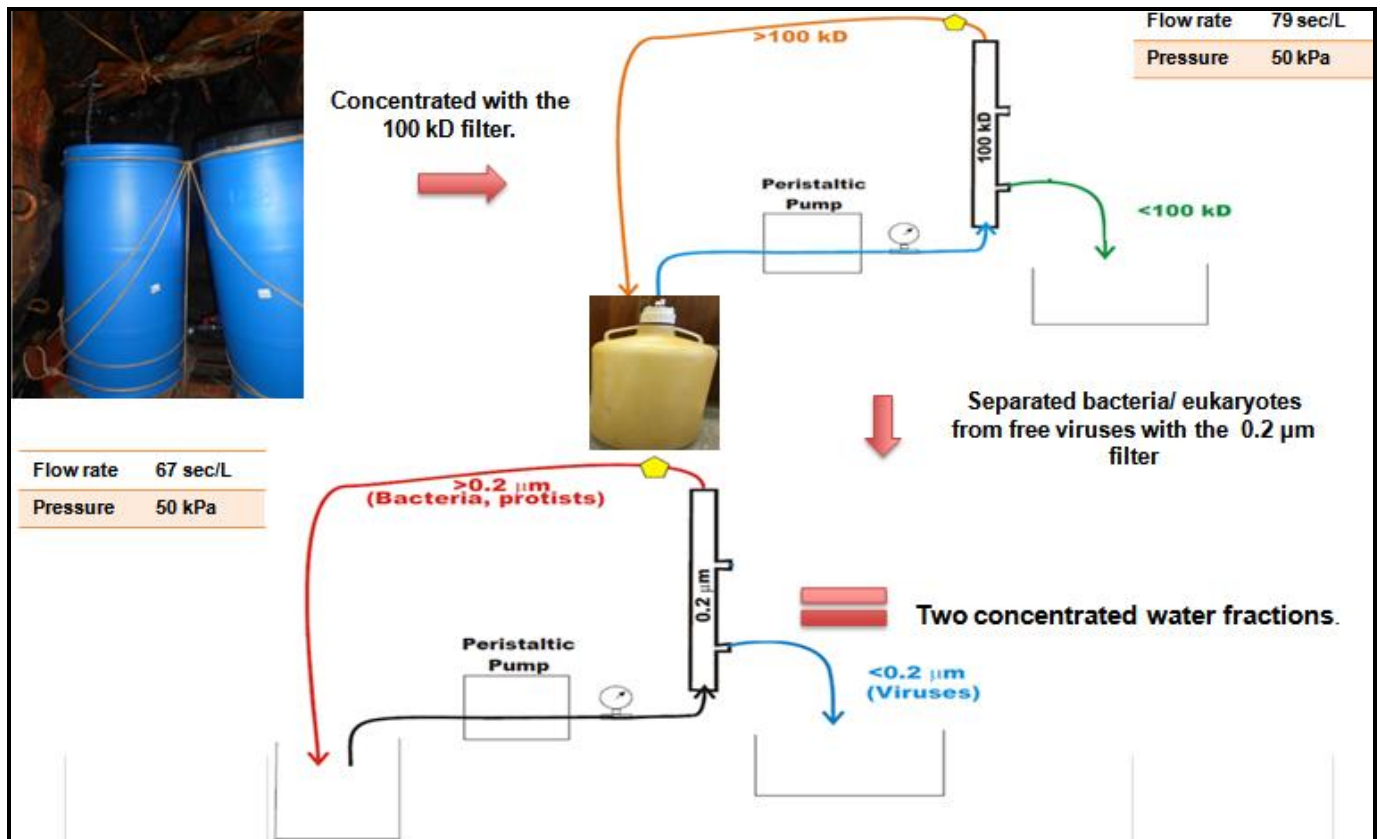


Figure 3.2: Diagram of the combined overall tangential flow filtration concentration method.

3.2.4. Microbial enumeration using DAPI (4',6-diamidino-2-phenylindole, dihydrochloride) staining and acridine orange staining

DAPI and acridine orange staining was done according to the procedure outlined by (Porter & Feig, 1980). The formaldehyde fixed samples were used to quantify the amount of biomass in the water before and after concentration. The sample were filtered through 0.22 μm DAPI filters (Millipore) using a 25 mm glass filter unit (Millipore). The samples were stained on a drop of DAPI solution (10 μg/mL) for 1 minute in the dark. Duplicate samples were stained using the acridine orange stain (100 μg/mL) for 5 minutes in the dark. After staining filters were rinsed and dried and counterstained with a 30 μL drop of the anti-fading solution on a slide cover slip, followed by visualization with the Axioskop light microscope (Zeiss) under the 100X oil immersion magnification using the DAPI filter set with a blue excitation light (G 365 nm) and images were captured using a Digital Sight DS-US (Nikon) camera. For acridine orange stains

the slides were viewed using the green fluorochromes filter set with a green excitation light (BP 450-490 nm) and a 100X magnification.

3.2.5. Microbial enumeration using flow cytometry

The samples that were fixed during sampling and TFF were used to quantify the amount of biomass in the water before and after concentration. The method outlined in the Bacterial counting kit (Invitrogen) was used. The samples (1 mL) were stained with 1 μ L of SYTO BC bacteria stain (Component A) and incubated at room temperature in the dark for 5 minutes. The microsphere standard suspension (Component B) was resuspended and 10 μ L was added to the sample. A control containing just the beads was prepared. An unstained sample was prepared as a second control. This control sample was used to test the specificity of the kit.

Samples were analysed using the BD FACS CaliburTM flow cytometer (BD Biosciences) at the Medical Faculty Haematology unit at the University of the Free State (Bloemfontein, RSA) trained and supported by Dr. Walter Janse van Rensburg. Before analyses, the control sample with just the beads was used as a reference in order to perform some instrumental adjustments. The instrumental adjustments were done to set gates in the areas in which the sample and beads would be expected so that they could be counted separately and to also estimate the area where the background noise would appear. For each sample it was set that flow cytometer stop counting the cell once the bead/microsphere count had reached 1000.

The number of bacteria per 10^{-6} mL of the sample can be calculated by dividing the number of bacterial signals by the bead signals within the gates. That count can be multiplied by 10^6 in order to get the count per mL.

3.2.6. Scanning electron microscopy (SEM) of the morphology of the bacterial cells

The TFF concentrated biomass samples (using the 100 kDa filter) were fixed with 2.5% glutaraldehyde for at least 3 hours at 4°C, followed by an hour fixation in 1% osmium tetroxide. The samples were dehydrated in a graded ethanol series of 50%, 70% and 95% for 20 minutes in each phase, followed by two changes of 100% for 1 hour. The samples were dried using a critical point dryer (Tousimis, Maryland, USA) at 31.5°C which was followed by mounting the

samples on stubs (Cambridge pin type 10 mm) by epoxy glue and gold coated (60 nm) with a Bio-Rad coater (United Kingdom). The samples were examined using the Shimadzu SSX-550 SEM (Kyoto, Japan). All sample preparations after fixing with glutaraldehyde and sample analyses was done at the Centre for Microscopy at the University of the Free State (Bloemfontein, RSA).

3.2.7. Microbial diversity assessment using Sanger sequencing

3.2.7.1. Genomic DNA isolation

The bacterial fraction (approximately 100 mL), previously concentrated and separated from the viral fraction by TFF, was filtered through a PES 0.22 µm filter (47 mm Supor[®] 200, Pall Corporation) using a 47 mm glass filter unit (Millipore). Cut-up filter pieces were added to tubes containing beads and the genomic DNA was isolated using the NucleoSpin[®] Soil kit (Macherey-Nagel) according to the manufacturer's instructions. The concentration and purity of the genomic DNA was analysed using the NanoDrop ONE^C Spectrophotometer (Thermo Scientific). The integrity of the genomic DNA was evaluated by agarose gel electrophoresis and visualized on a 0.8% (w/v) agarose gel in Tris-Acetate-EDTA (TAE) buffer (40 mM Tris, 40 mM acetic acid, 1mM EDTA, pH 7.4) with 0.6 µg/mL ethidium bromide (EtBr) (Merck Millipore). The 6x DNA loading dye (Thermo Scientific) was added to the genomic DNA and the MassRuler DNA Ladder Mix (Thermo Scientific) was used for amplicon size determination. Electrophoresis was performed at 90 V for 55 minutes in TAE buffer and the gel was visualized using the ChemiDoc XRS UV/Vis Gel Documentation system (Bio-Rad Laboratories, Inc.).

3.2.7.2. Polymerase chain reaction (PCR)

In order to assess the microbial diversity a polymerase chain reaction (PCR) was conducted on the isolated gDNA to amplify the Archaeal 16S (~600 bp), Bacterial 16S (~600 bp) and Eukaryal 18S (~600 bp) rRNA gene V3/V4 hypervariable region fragments using the universal oligonucleotide primers for each gene as indicated in Table 3.2.7.2 All the forward primers had a GC-clamp

Table 3.2.7.2: Oligonucleotide primers used for amplification of the V3/V4 hypervariable regions for Archaea, Bacteria and Eukarya.

Primer	Nucleotide sequence	V3/V4	Reference
Arc344F-GC	5'- GC-clamp –AC GGG GYG CAG CAG GCG CGA-3'	Archaea	Casamayor <i>et al.</i> , 2002
Arc934	5'-GTG CTC CCC CGC CAA TTC CT-3'	Archaea	Casamayor <i>et al.</i> , 2002
Bac341F-GC	5'- GC-clamp –CC TAC GGG AGG CAG CAG-3'	Bacteria	Muyzer <i>et al.</i> , 1993
Bac908R	5'-CCG TCA ATT CMT TTG AGT TT-3'	Bacteria	Muyzer <i>et al.</i> , 1993
Euk1AF-GC	5'- GC-clamp –CT GGT TGA TCC TGC CAG-3'	Eukarya	Diez <i>et al.</i> , 2001
Euk516R	5'-ACC AGA CTT GCC CTC-3'	Eukarya	Diez <i>et al.</i> , 2001

The PCR amplifications were performed using either the BioRad T100™ Thermo Cycler or the Mastercycler gradient Thermo Cycler (Eppendorf) with a total reaction of 50 µL. Each PCR reaction contained 1 µL template, 1 µL dNTPs (10 mM), 1 µL of each primer (10 µM), 1 µL BSA (10 µg/µL), 5 µL 10X buffer and 0.5 µL (5 U/µL) of *Taq* DNA polymerase (all reagents from New England BioLabs Inc.). The volume was adjusted to a final volume of 50 µL with sterile milliQ water. The optimized reaction parameters for the Archaea consisted of an initial denaturing step at 95°C (4 minutes), followed by 30 cycles of denaturing at 95°C (1 minute), annealing at 53°C (1 minute) and elongation at 72°C (1.5 minutes). The final elongation step was at 72°C (10 minutes). The optimized Bacterial amplification parameters consisted of an initial denaturing step at 95°C (5 minutes), followed by 25 cycles of denaturing at 95°C (45 seconds), annealing at 49°C (45 seconds) and elongation at 72°C (1 minute), with a final elongation at 72°C (10 minutes). The optimized Eukaryal amplification parameters consisted of an initial denaturing step at 95°C (4 minutes), followed by 30 cycles of denaturing at 95°C (1 minute), annealing at 56°C (1 minute) and elongation at 72°C (1.5 minutes), with a final elongation step at 72°C (10 minutes). The PCR products were separated on a 1% (w/v) and visualized under UV light.

3.2.7.3. Denaturing gradient gel electrophoresis

To evaluate the preliminary diversity profiles of the sample, the amplicons from the PCR were subjected to DGGE on a 7% polyacrylamide gel using the Dcode™ Universal Mutation Detection System (Bio-Rad Laboratories, Inc.) as described by Diez *et al.*, (2001). The stock solution included, 0% Urea-Formamide [40% acrylamide/bis (10 mL), 50X TAE (1 mL) and

MilliQ water to fill up to 50 mL] and 80% Urea-Formamide [40% acrylamide/bis (10 mL), 50X TAE (1 mL), formamide (16 mL), urea (16.8 g) and MilliQ water to fill up to 50 mL]. The gradient of urea/formamide denaturants (prepared from the above stock solutions) ranged from 35-70% for Archaea, 40-60% for Bacteria and 35-70% for Eukarya and each gradient volume was 12 mL with TEMED (8.4 μ L) and APS (84 μ L) added to each. The remaining PCR amplicons were used and separated in the gel at 100 V at 60°C for 17 hours. The gel was stained with 0.6 μ g/mL ethidium bromide (Merck Millipore) for 20 min, washed with distilled water and visualized under UV light using the ChemiDoc XRS Gel Documentation system (Bio-Rad Laboratories, Inc.).

3.2.7.4. Gel extraction, re-amplification and cloning into pGEM[®]-T Easy vector

From the diversity profiles selected DGGE bands were excised from the gel using the ChemiDoc XRS Gel Documentation system (Bio-Rad Laboratories, Inc.). The bands were eluted in 50 μ L autoclaved MilliQ water overnight at 37°C. The gel eluted bands (3-6 μ L depending on intensity of the band on the DGGE gel) were re-amplified using primers set up in Table 3.2.7.2 without the GC-clamp. The PCR was performed in a total reaction of 50 μ L using a T100[™] Thermal Cycler (Bio-Rad Laboratories, Inc.) as discussed in Section 3.2.7.2.

The purified PCR products were ligated into pGEM-T Easy vector (Figure 3.3) using 6 μ L of the purified PCR product, vector (50 ng), T4 DNA ligase (3 Weiss units/ μ L), 2X T4 ligase buffer (Fermentas) brought up to a final volume of 10 μ L with autoclaved MilliQ water. The reactions were left at room temperature for 1 hr and further incubated at 4°C overnight.

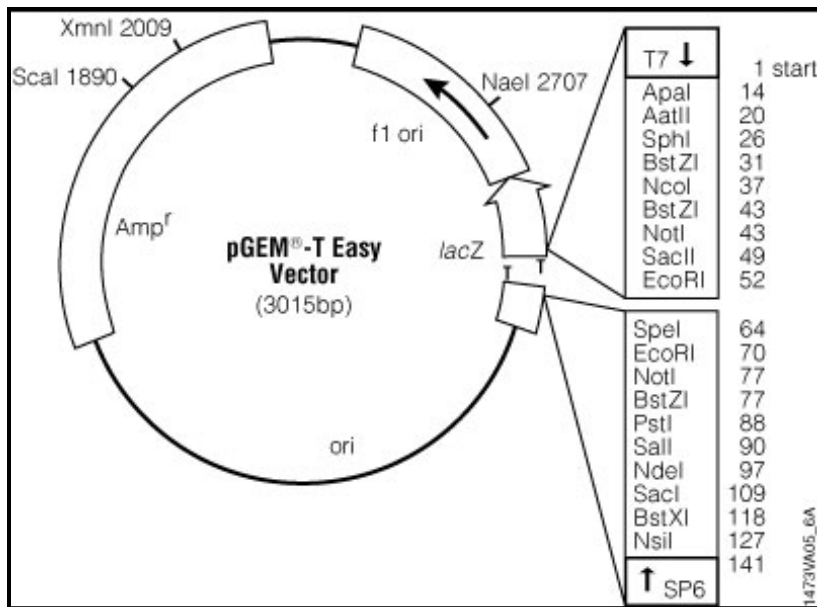


Figure 3.3: pGEM[®]-T Easy vector system circle map (Promega).

The ligated products were transformed into Top 10 *Escherichia coli* competent cells (Lucigen) in 1.5 mL Eppendorf tubes. The mixtures were placed on ice for 30 minutes, heat shocked for 45 seconds in an Eppendorf heating block and immediately placed on ice. LB media (900 uL made up at a ratio of 2 tryptone: 1 yeast extract: 2 NaCl) was added to each transformation mixture and incubate at 37°C for 1.5 hrs with shaking (~150 rpm). Positive clones were identified through selection on LB-plates containing ampicillin (30 µg/mL), IPTG (24 mg/mL) and X-gal (20 mg/mL). Plasmid isolation was performed using the Plasmid DNA Extraction Kit (Biospin) according to the instructions of the manufacturer. The vectors were digested using the restriction enzyme EcoRI and evaluated by agarose gel electrophoresis and visualized using the ChemiDoc XRS UV/Vis Gel Documentation system (Bio-Rad Laboratories, Inc.).

3.2.7.5. Sanger sequencing and analysis

The isolated plasmids were sequenced using Sanger sequencing. The ABI Prism[®] BigDye[™] Terminator Cycle Sequencing Ready Kit V.3.1 Cycle (Applied Biosystems) was used according to manufacturer's instructions. The composition for the sequencing reaction consisted of 5X dilution Buffer (2 µL), premix (1 µL), sequencing primer [3.2 pmol/uL] (1 µL), template (200-500 ng/µL) and Nano-pure water to fill up to a reaction volume of 10 µL. The M13 forward primer (5'-GTAAAACGACGGCCAGT-3') and the SP6 primer (5'-GCTATTTAGGTGACACTATAG-3') were used, and separate reactions were made for the forward and reverse primers. Amplifications cycling parameters consisted of an initial denaturing step at 96°C for 1 minute, followed by 25

cycles of denaturing at 96°C (10 seconds), annealing at 50°C (5 seconds) and elongation at 60°C (4 minutes).

The PCR products were purified using the EDTA/ethanol precipitation protocol for sequencing clean up. The sequencing reaction volume was adjusted to 20 µL, transferred to a 1.5 mL Eppendorf tube containing 125 mM EDTA (5 µL) and 60 µL absolute ethanol (Merck Millipore), mixed for 5 sec and precipitated at room temperature for 15 minutes. Centrifugation using a Centrifuge 5810R (Eppendorf) was done at 20 000 x *g* for 15 min at 4°C. Complete aspiration of the supernatant was done and the pellet washed using 70% (v/v) ethanol (1000 µL) and centrifuged once more (20 000 x *g*, 5 min, 4°C). The supernatant was removed by complete aspiration and the pellet was dried using a Concentrator 5301 (Eppendorf) for 5 min. The samples were stored at 4°C in the dark until sequencing was performed with a 3130xl Genetic Analyzer, HITACHI (Applied Biosystems) at the Department of Microbial, Biochemical and Food Biotechnology, University of the Free State.

The 16S and 18S rRNA gene sequences were quality filtered and assembled to create contigs using the Geneious software (V.7.1.7.9). The analysed 16S sequences were aligned against the life tree project (LTP) 119 rRNA database from SILVA (Pruesse *et al.*, 2007) using a variant (BLASTN) algorithm of the basic local alignment search tool (BLAST) (Zhang *et al.*, 2000).

3.2.8. Purifying viral-like particles

The concentrated viral fraction was applied to a stepwise cesium chloride gradient ultracentrifugation which was modified to that of Liu *et al.*, (2006). The cesium chloride gradient ultracentrifugation consisted of the following concentrations 1.15, 1.25 and 1.50 g/mL. The gradients (each 2 mL) were overlaid on each other in decreasing order with sufficient volume of the sample at the top. Centrifugation occurred at 22 000 rpm using the Beckman SW 32.1 Ti rotor (Beckman, Optima™ L-100XP BioSafe, USA) in thin wall ultracentrifuge tubes (Beckman, USA) at 4°C for 5 hrs. Approximately 3 mL of the viral-like particles were recovered between 1.25 and 1.5 g/mL. The recovered fraction was dialyzed overnight against phosphate buffered saline, PBS (pH 7.4) using Slide-A-Lyzer™ Dialysis Cassettes (10K MWCO) and the buffer was changed twice. The purified viral-like particles fraction was used for EFM and TEM.

3.2.9. Epifluorescent microscopy of free viral-like particles

EFM was done using a modified procedure to that outlined by (Suttle & Fuhrman, 2010). Aliquots (10 mL) of the viral and bacterial concentrate were fixed with 4% formaldehyde for one hour at 4°C. The samples were filtered through a 0.02 µm Anodisc filter (Millipore) using a 25 mm glass filter unit (Millipore). The Anodisc filters were placed on a petri dish with a drop of 5% SYBR Gold (5 µL diluted stock + 95 µL 0.02 µm filtered water) with the sample side facing down. The samples were stained for 15 minutes in the dark. After staining filters were rinsed and dried then counterstained with a 30 µL drop of the anti-fading solution on a slide cover slip, followed by visualization with the Axioskop light microscope (Zeiss) under the 100X oil immersion magnification using the green fluorochrome filter set with a green excitation (450-490) and images were captured using a Digital Sight DS-U2 (Nikon) camera.

3.2.10. Transmission Electron Microscopy (TEM)

Concentrated viral water samples were fixed with 2% glutaraldehyde at 4°C for at least 3 hours. Initially 10 µL of the sample was overlaid on carbon coated, formvar film copper grids (200 mesh, Agar Scientific). The suspension was allowed to dry on the grid, followed by negative staining with 3% uranyl acetate. The excess stain was removed using filter paper and the grid was allowed to air dry prior to examination using the Philips (FEI) CM100.

An additional modified TEM method was used where the volume of sample was increased and the use of modified ultracentrifuge tubes was utilized (Børshøj *et al.*, 1990). The tubes were prepared by applying epoxy to the bottom of the tube to create a flat surface. Carbon coated, formvar film copper grids (200 mesh) were placed at the top of the dry epoxy surface. Viral water samples (35 mL and unfixed) were added to the tubes with the epoxy and grids and ultracentrifuged at 25 000 rpm for 90 minutes at 4°C using the SW 32 Ti rotor (Beckman, Optima™ L-100XP BioSafe, USA). The grids were stained and examined as mentioned above.

All sample preparations and analyses after ultracentrifugation was done at the Centre of Microscopy at the University of the Free State (Bloemfontein, RSA).

3.2.11. Viral genomic DNA isolation

The viral fraction of the fissure water, previously concentrated by TFF was used to extract viral genomic DNA. Three different isolation methods were used.

The first method used was phenol/chloroform extraction using modified procedures to that of Lau *et al.*, (2014) and Sambrook & Russell, (2001). Lysis solution (100 mM NaCl, 10 mM Tris, 0.5% SDS, pH8.0) was added to the concentrated phage suspension along with proteinase K (20 µg/L). The sample was vortexed and incubated at 60°C for an hour. An equal volume of buffer-saturated phenol (at 50°C) was added to the sample and centrifuged (14 000 x *g*, 5 minutes, room temperature). The upper aqueous layer was added to a new tube and the phenol step was repeated. After the phenol step phenol: chloroform (1:1) was added to the aqueous layer and centrifuged (14 000 x *g*, 5 minutes, room temperature). A volume of 2.5 absolute ethanol was added to the volume of the sample along with a volume of one tenth of 3 M sodium acetate (pH 5.2). The sample was incubated on ice for 30 minutes and centrifuged (14 000 x *g*, 20 minutes, room temperature). The pellet was resuspended with 1 mL of 70% ethanol and centrifuged (14 000 x *g*, 2 minutes, room temperature). The pellet was dried and resuspended in sterile 300 µL 1 X TE buffer [1 M Tris pH 8.5 (10 mL), 0.5M EDTA (2 mL) in dH2O].

The second method was the formamide extraction where the concentrated phage suspensions were treated with DNase (5 µg/mL), RNase (50 µg/mL) and proteinase K (20 µg/mL) prior to isolation of the phage DNA. Viral DNA was then isolated from concentrates using the method as described by Sambrook & Russell, (2001). Deionized formamide was added at 1 volume of the phage suspension along with 0.1 volume of 2 M Tris (pH 8.5), and 0.5 volume of 0.5M EDTA (pH 8.0). The mixture was incubated a 65°C for 30 minutes and precipitated with 6 volumes of ice cold absolute ethanol for 30 minutes at -80°C. The sample was centrifuged at 10 000 x *g* for 10 minutes after which the pellet was re-suspended in 300 µL of 1X TE buffer [1 M Tris pH 8.0 (10 mL), 0.5M EDTA (2 mL) in dH2O]. Precipitation was performed using 6 µl NaCl and 750 µl absolute ethanol, followed by incubation for 30 minutes at -80°C. The sample was centrifuged as described above and the pellet was dried and resuspended in TE buffer.

The final method of extraction was using the NucleoSPIN® Soil kit (Macherey-Nagel). The viral fraction genomic DNA and or RNA extraction was performed using the NucleoSPIN® Soil kit as described by the manufacturer. The DNA extraction concentrations were measured for DNA and RNA on the Qubit™ (Invitrogen) using the Quant-IT™ dsDNA HS Assay kit and the Quant-IT™ RNA Assay kit (Both from Invitrogen) as described by the manufacturer.

3.2.12. Amplification of phage genes

In order to further characterize the possible types of free phages and prophages within the bacterial host genome in the fissure water, phage primers amplifying for either T4 or T7 like phages were used. PCR was conducted on the isolated viral DNA from the viral fraction of the water and the gDNA from the bacterial fissure water fraction using the primers in Table 3.2.12.

Table 3.2.12: Oligonucleotide primers used for the amplification of T4 and T7 phage genes.

Forward and reverse primers	Nucleotide sequence	Detects	Reference
MZIAbis	5'-GAT ATT TGI GGI CAG CCI-3'	g23 (major capsid protein) from T4 type phages.	Filée <i>et al.</i> , 2005
MZIA6	5'-CGC GGT TGA TTT CCA GCA TGA TTT C-3'		
T4g23F	5'-CTG ATC GCC TTC GAT ATT GCTGGT GTT CAG-3'		Mabizela, 2009
T4g23R	5'-GGG TTA ARA CCG ATY SCG TAA CGA G-3'		
HECTORPol29F	5'-GCA AGC AAC TTT ACT GTG G-3'	DNA polymerase from uncultured podophages	Breitbart <i>et al.</i> , 2004
HECTORPol711R	5'-TTC GTT GGT GTA TCT CTC G-3'		
HECTORPol500R	5'- TTC GTT GGT GTA TCT CTC G-3'		
1F	5'-GTT ACM GGG CAR MGA GTH GG-3'	Lysogenic phages using the integrase gene as the conserved loci	Balding <i>et al.</i> , 2005
1R	5'-ATG CCC GAG AAG AYG TTG AGC- 3'		
8F	5'-TGC TTA TAA CAC CCT GTT ACG TAT-3'		
8R	5'-CAG CCA CCA GCT TGC ATG ATC- 3'		
CPS3	5'-TGG TAYGTY GAT GGM AGA-3'	Gene encoding major capsid protein from Cyanophages	Zhong <i>et al.</i> , 2002
CPS4	5'-CAT WTC WTC CCA HTC TTC-3'		
CPS8	5'-AAA TAY TTD CCA ACA WAT GGA-3'		

The PCR amplifications were performed using the Mastercycler gradient Thermo Cycler (Eppendorf) with a total reaction of 25 µL. Each PCR reaction contained, 2X KAPA HiFi HotStart ReadyMix (0.3 M of each dNTP and 0.5 U of KAPA HiFi HotStart DNA polymerase), 0.75 µL of each primer (10 µM), 6 µL template and made up to the total volume with PCR-grade water. A gradient PCR amplification was performed for all the primer sets with an initial denaturing step

at 95°C (3 minutes), followed by 30 cycles of denaturing at 98°C (20 seconds), annealing with the following gradients for each set of primers [55°C, 56°C, 57°C, 59°C and 60°C] (15 seconds) and elongation at 72°C (1.5 minutes). The final elongation was at 72°C (1 minute). The PCR products were visualized under UV light on a 1% (w/v) agarose gel.

3.2.13. Cloning into the pSMART[®] HCKan vector and Sanger sequencing

The remaining PCR amplicons were loaded onto a 1% (w/v) agarose gel and the bands were excised from the gel under UV light on the ChemiDoc XRS Gel Documentation system (Bio-Rad Laboratories, Inc.). Then bands were purified using the DNA/RNA Extraction/Purification BioSpin Gel Extraction Kit (BioFlux) as per manufacturer's instructions.

The purified PCR products were first phosphorylated and then ligated into pSMART[®] HCKan vector (Figure 3.4) using 5 µL of the purified PCR product, vector (50 ng), T4 DNA ligase (3 Weiss units/µL), 2X T4 ligase buffer (Fermentas) and 1 µL 50% PEG. The final volume of the reaction was 10 µL. The reactions were left at room temperature for 1 hr and further incubated at 4°C overnight.

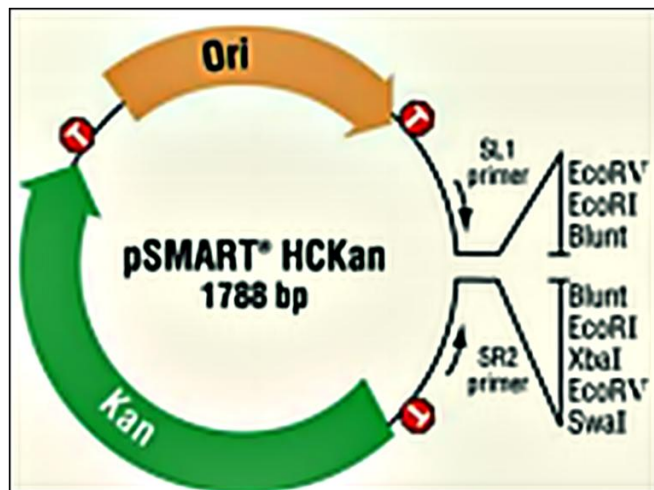


Figure 3.4: pSMART[®] HCKan vector system circle map (Lucigen).

The ligated products were transformed into Top 10 *Escherichia coli* competent cells (Lucigen) as outlined in Section 3.7.2.4. Positive clones were identified through selection on LB-plates containing kanamycin (30 µg/mL). Plasmid isolation was performed using the Plasmid DNA Extraction/Purification BioSpin Kit (BioFlux) according to the instructions of the manufacturer.

The vectors were digested using the restriction enzyme EcoRI and evaluated by agarose gel electrophoresis and visualized using the ChemiDoc XRS UV/Vis Gel Documentation system (Bio-Rad Laboratories, Inc.). The purified plasmids were sequenced as outlined in Section 3.7.2.5. using the SL1 (5'-CAGTCCAGTTACGCTGGAGTC-3') and SR2(5'-GGTCAGGTATGATTTAAATGGTCAGT-3') primers that were present on the vector.

3.3. Results and discussions

3.3.1. Physiochemical analysis

On and off-site analyses on the sampled borehole water was done as described in Section 3.2.2., for both the 2015 and 2016 sampling trips. Table 3.3.1 shows the results of the analyses of the samples taken 11 months apart.

Table 3.3.1: Analyses of the borehole water sampled in 2015 and 2016.

Parameter	29 July 2015	08 June 2016	IGS 2015 analysis	IGS 2016 analysis
pH	8.81	8.31	7.08	7.34
Temperature (°C)	29.2	28.2		
EC (mS/m)	312	476	234	230
Salinity (g/L)	2.06	3.32		
ORP (mV)	-122	-124		
TDS (g/L)	2.4	3.31	3.47	2.85
TOC (mg/L)			0.242	0.222

Legend: EC = electrical conductivity, ORP = oxidation reduction potential, TDS = total dissolved solids, TOC = total organic carbon.

In both samples, within the 11 months period, the borehole water had consistent results with regards to the parameters measured on and off-site. The pH remained alkaline with an average pH of 8.56 on-site and 7.21 off-site. The temperature remained moderate with an average of 28.7°C. These results are similar to those encountered by Erasmus, (2015) where she found the

temperature and pH to be at an average of 30.6°C and 8.4 respectively, in Star Diamonds mine and in the same borehole two-three years prior. Both sets of results from Star Diamonds mine are similar to those of shallower moderately saline fracture water in South African mines (Lin *et al.*, 2006; Magnabosco *et al.*, 2014). The EC and TDS values were moderate with measurement values on and off site between 2300-4760 $\mu\text{S}/\text{cm}$ and 2400-3470 mg/L respectively, and these values correlate with those of 45 South African mine water locations (Wolkersdorfer & Hubert, 2016) and also correlated with the TDS results (average of 2.7 g/L) observed previously in the same mine (Erasmus, 2015). TDS and EC measurements are generally directly proportional to each other, but are inversely proportional to the pH, as high EC or TDS results in low pH (Wolkersdorfer & Hubert, 2016). The TOC values were low with an average of 0.232 mg/L. When TOC values are lower the cell count is expected to be low as microorganisms rely on organic carbon for food and energy (Lin *et al.*, 2006). The salinity was moderate at an average of 2.69 g/L and the ORP had an average of -123 mV of which both were similar to those of Erasmus; (2015) (-155 mV and 2.6 g/L) therefore indicating a sustained reducing environment.

3.3.2. Microbial enumeration using DAPI (4',6-diamidino-2-phenylindole, dihydrochloride) staining and acridine orange staining

Microbial enumeration of the fissure water prior and after concentration using TFF, was performed using DAPI (Figure 3.5) and acridine orange staining (Figure 3.6). The estimated cell counts shown in Table 3.3.3 indicate higher cell counts for acridine orange compared to DAPI after concentration of the fissure water. This is due to the acridine stain being more suitable for environmental samples as it is able to distinguish between bacterial cells and detritus; thereby staining the cells green and the debris orange (Francisco *et al.*, 1973; Kepner & Pratt, 1994). The acridine stain had better visual structural resolution and fluorescence of the microorganisms was detected better making it easier to conduct cell counts. Concentrating the sample (Figure 3.5 B) resulted in more prominent macro and biofilm-like structures that made it difficult to perform cell counts. These macrostructures and biofilm-like structures were previously detected by Erasmus; (2015), who had sampled in the same mine. Even though these structures were not characterized and identified, they were indicative of the complex nature of the biofilms found in the subsurface. Flemming & Wingender, (2010) have previously reported that biofilms can consist of up to 90% extracellular polymeric substances, which could offer an explanation for the detection of macrostructures within the samples. To troubleshoot the samples were vortexed for

10 minutes prior to staining with DAPI and acridine orange in order to dislodge the cell from the biofilm-like structures.

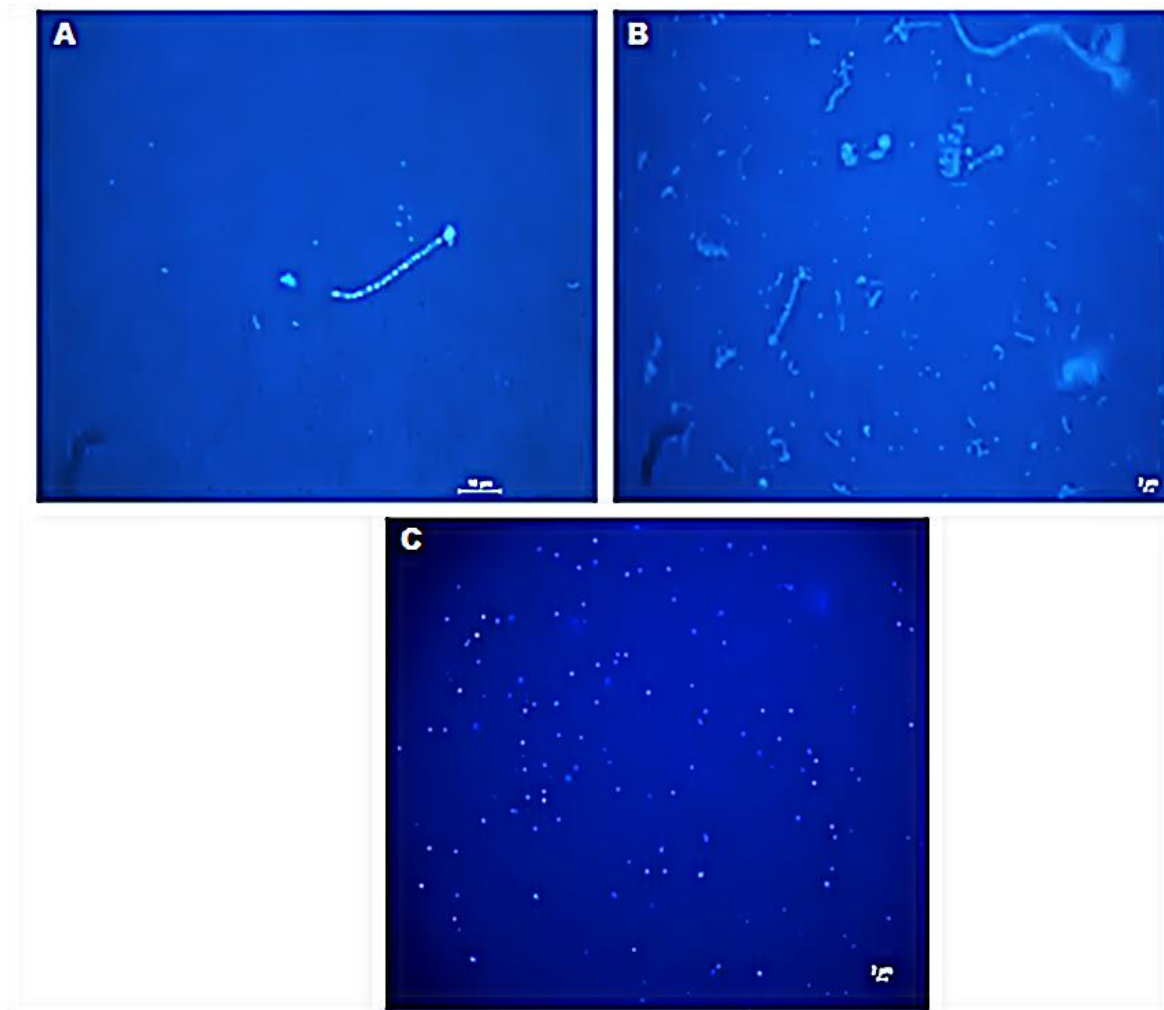


Figure 3.5: DAPI stain images of the fissure water from the Star Diamonds mine. (A) Sample before concentration showing a low count of stained cells in macro and biofilm-like structures. (B) Sample after concentration using the 100 kDa TFF filter showing a higher concentration of stained cells in macro and biofilm-like structures. (C) TFF (100 kDa filter) concentrated samples after vortexing for 10 minutes resulting in the observed dislodged cells from the biofilm-like structures. Scale for A is equal to 10 μm and 2 μm for B and C.

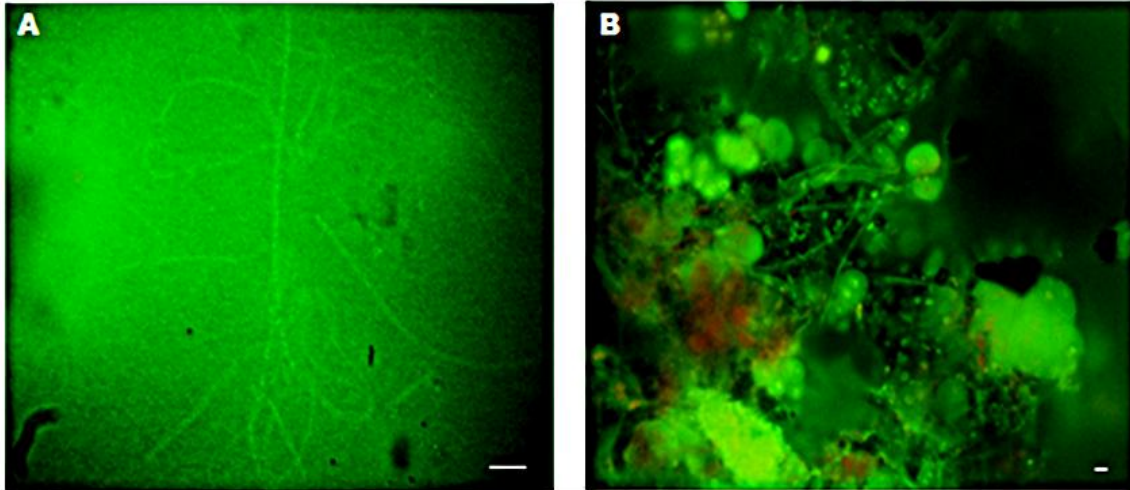


Figure 3.6: Acridine orange stain images of the fissure water from the Star Diamonds mine. (A) Sample before concentration with the TFF 100 kDa filter showing the macro-like structures. (B) Sample after concentration and vortexing for 10 minutes. The dislodged bacterial cells and macro-like structures are stained in green and the debris is stained in orange. Scale for A is equal to 10 μm and scale for B is equal to 2 μm .

3.3.3. Microbial enumeration using flow cytometry

Flow cytometry (Figure 3.7 and 3.8) was also used to enumerate the microorganisms in the fissure water prior and after TFF concentration using a stain specific for bacterial cells. The gating of areas and calculation of estimated cells was performed as indicated in Section 3.2.5. Table 3.3.3 indicates higher cell count for flow cytometry than of DAPI and acridine. Enumeration using flow cytometry is more accurate compared to DAPI and acridine staining as it is automated to counts per cell of all the cells in the given sample unlike DAPI and acridine staining where one has to manually count by estimation the fluorescence picked up by the microscope (Ambriz-Aviña *et al.*, 2014; Weaver, 2000). Table 3.3.2 shows the estimated cell counts from the 2015 and 2016 samples before and after concentration. Biofilm formation can easily obscure the number of cell events counted in a sample and this can be seen in the 2015 sample counts where the post concentration sample had less events counted than the pre-concentration sample (Kerstens *et al.*, 2015). To troubleshoot this, samples were vortexed for 10 minutes prior counting cells using flow cytometry in order to dissociate cells from the biofilm structure (Kerstens *et al.*, 2015).

Table 3.3.2: Cell counts using flow cytometry for both 2015 and 2016 samples.

Sample date	Pre-concentration (cells/mL)	Post concentration (cells/mL)	Post concentration and vortexing (cells/mL)
2015	8.7×10^6	6.2×10^6	18.9×10^6
2016	5.3×10^5	2.7×10^6	3.8×10^6

There is a distinct overall difference with regards to the cell counts when comparing the 2015 and 2016 samples with the 2015 sample containing a much larger amount of cells. The higher counts in the 2015 sample could have been due to the borehole being closed prior to the sampling in 2015. From previous research it has been shown that closed boreholes create biofilm structures which are dislodged due to high pressure of the water once the borehole is opened (Borgonie *et al.*, 2011; Borgonie *et al.*, 2015). From the 2015 sampling to the 2016 sampling the borehole was left to drain resulting in little biofilm formation inside the borehole. This phenomenon therefore offers a possible explanation to why the 2015 sample had higher cell counts and also considering that the majority of the microbial biomass in the deep terrestrial subsurface form biofilm structures due to the extreme environment (Toole *et al.*, 2000). Another explanation could have been due to some of the background noise shown in the dot plots in Figures 3.7 being incorporated into the gate for the sample counts therefore increasing the counts for the 2015 sample.

The dot plots from the 2016 sample (Figure 3.8) resulted in minimal background noise (it should be noted that each sample used a new bacterial counting kit and therefore the results suggested that background noise intensity is dependent on the batch kit used) decreasing the chances of incorporation of background noise into the sample count gate. This resulted in the 2016 sample counts being considered more accurate and the pre-concentrated cell counts of this sample correlated with those estimated by literature where it has been estimated that cell counts in the deep subsurface groundwater is between 10^3 - 10^6 cells/mL (Akob & Küsel, 2011; Borgonie *et al.*, 2011; Onstott *et al.*, 1998). Figure 3.8 A shows the unstained control sample which suggests that the background noise found under the bacterial cell count gate could be a result of degraded microspheres. Figure 3.8 C shows a dot plot of the water from the filtrate after filtering with the 100 kDa filter; this water was expected to have no bacterial or viral cells as it was considered clean water. The counts on this sample were very low and insignificant suggesting that the concentration of the microorganisms in the retentate fraction was successful. Figure 3.8 F shows a dot plot of the sample from the viral fraction. The counts in this sample were also very low and insignificant. As expected the viral and bacterial fractions were

successfully separated. The above controls suggested that the cell counts in the pre-concentration and post concentration samples were accurate and that the kit specifically stained bacterial cells.

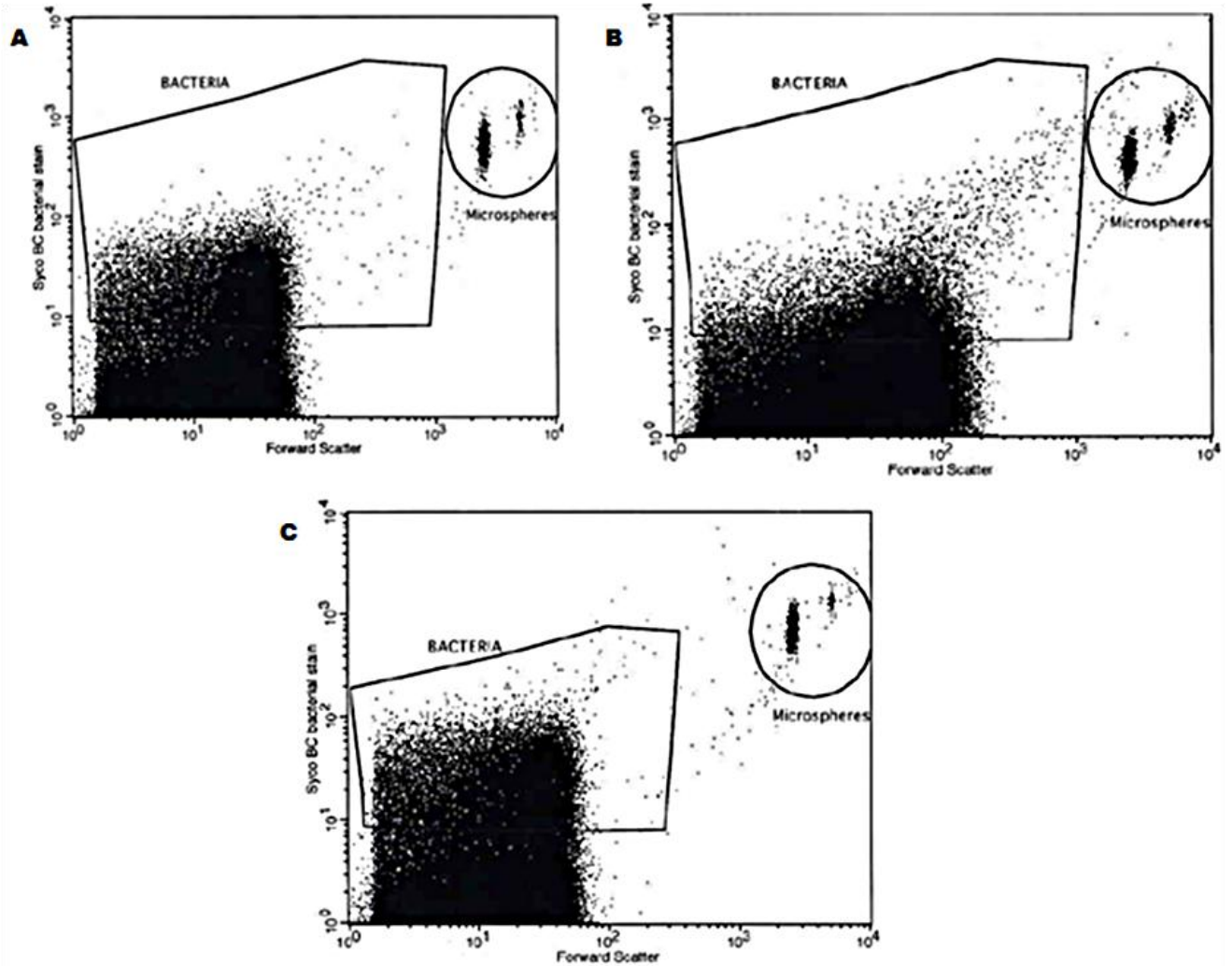


Figure 3.7: Cell counts using flow cytometry of the 2015 fissure water samples from the Star Diamonds mine. The dot plots include gated areas for the microspheres and bacterial cells. The background noise is represented as the area below the bacterial gate. (A) Sample pre-concentration. (B) Sample after concentration. (C) Sample after concentration and vortexing for 10 minutes.

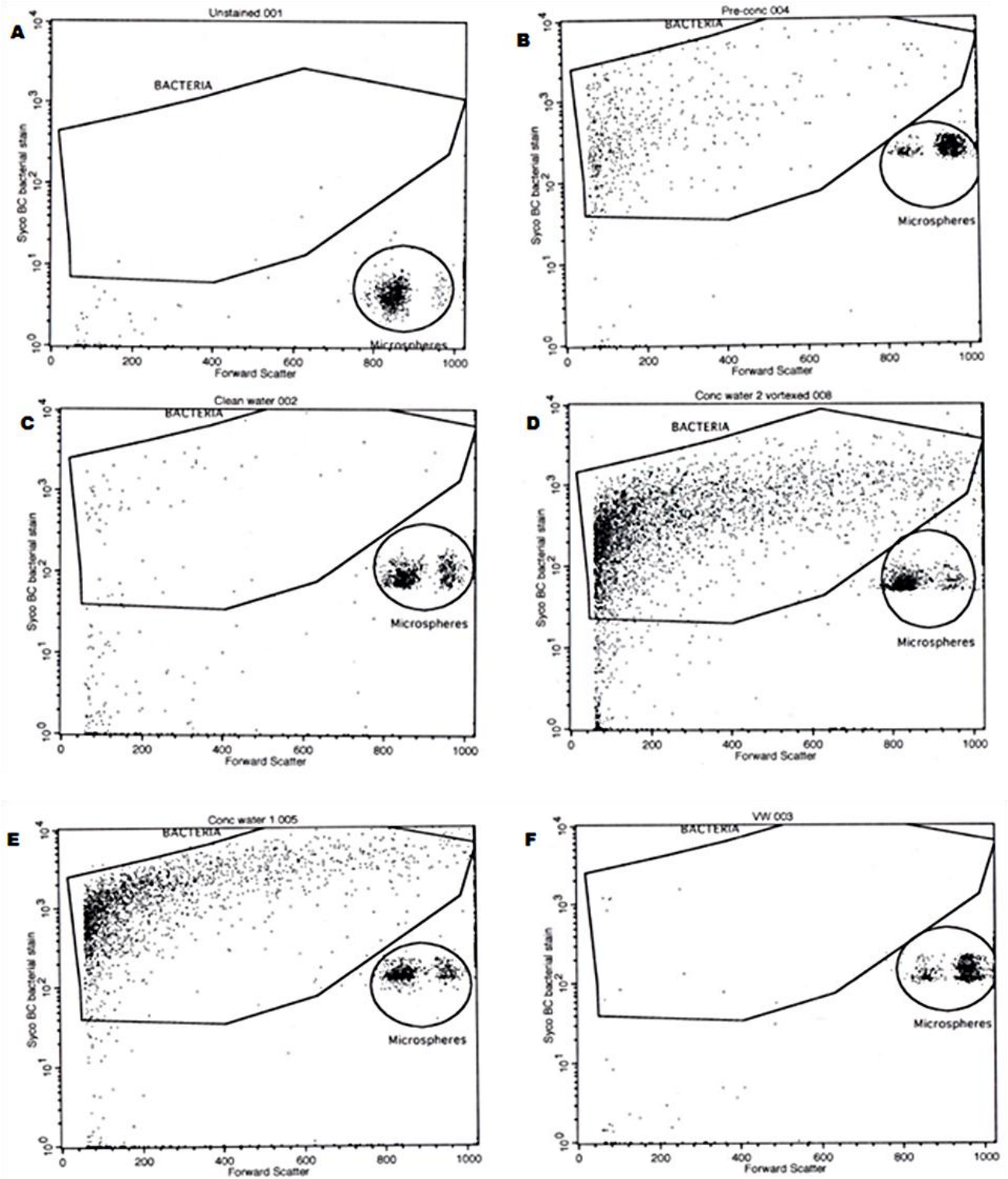


Figure 3.8: Cell counts using flow cytometry of the 2016 fissure water samples from the Star Diamonds mine. (A) Unstained control sample. (B) Sample pre-concentration. (C) Water from the filtrate of the TFF filter. (D) Sample after concentration and vortexing for 10 minutes. (E) Sample after concentration. (F) Sample from the viral fraction of the water.

The results of the estimated cell counts pre and post concentration in Table 3.3.3 for each enumeration method showed that the counts were 1 order of magnitude higher after concentration. The difference in magnitude after tangential flow filtration has been demonstrated in a previous study where an initial marine picoplankton cell count using EFM was 5×10^5 cells/mL and the post concentration was 4.7×10^{12} cells/mL (Giovannoni *et al.*, 1990). These cells were 7 orders of magnitude higher after concentration.

Table 3.3.3: Estimated cell counts using DAPI, acridine orange stains and flow cytometry

Enumeration Method (cells/mL)	Pre-concentration (cells/mL)	Post concentration (cells/mL)	Post concentration and vortexing (cells/mL)
DAPI stain	2000		1.6×10^4
Acridine orange stain	3685		4.1×10^4
Flow cytometry	5.3×10^5	2.7×10^5	3.8×10^6

3.3.4. Scanning electron microscopy of the morphology of the bacterial cells

Scanning electron microscopy was done in order to view the morphology of the bacterial cells within the concentrated bacterial fraction of the fissure water. Figure 3.9 A shows the different morphological structures (indicated in different colour) of the cells detected in the water. The results in Figure 3.9 A showed filamentous and coccobacillus morphologies. Figure 3.9 B shows one of the many biofilm-like structures indicated in red. The biofilm-like structures were more abundant compared to the free bacterial structures found in Figure 3.9 A. The long filamentous macrostructures shown in Figure 3.9 B are possibly the extracellular polymeric substances that are found to be part of the biofilms mentioned by Flemming and Wingender, (2010).

Biofilm high cell densities can be found in relation with the water filled fracture systems as they form close to the water filled fractures. Microbial communities form biofilms as a response to unfavourable environmental conditions such as low nutrient availability which is the case in deep subsurface environments (Toole *et al.*, 2000).

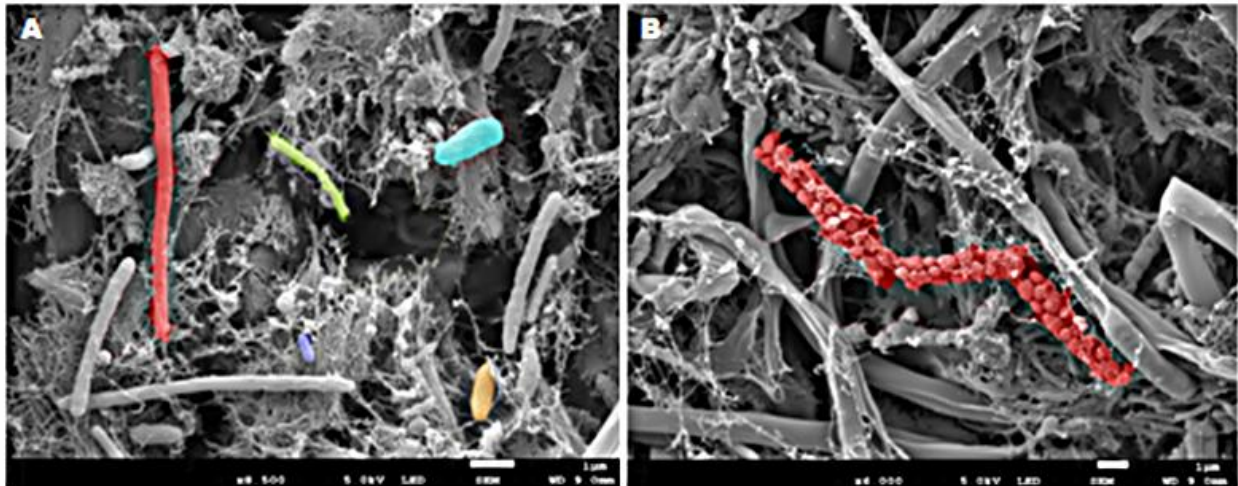


Figure 3.9: Scanning electron microscopy images of the fissure water from Star Diamonds mine. (A) Different bacterial morphological structures where the filamentous bacteria is indicated by the red and green while the coccobacillus is indicated by the blue, purple and orange bacteria. (B) A biofilm-like structure indicated in red. Scale bars are equal to 1 µm.

3.3.5. Microbial diversity assessments using Sanger sequencing

Genomic DNA was isolated from the concentrated bacterial fraction of the fissure water from Star Diamonds mine as discussed in Section 3.2.7.1. The isolation was successful with a band showing high integrity, but with some shearing indicated in Figure 3.10. The A260/A280 absorbance ratio was 1.7 indicating low protein or RNA contamination

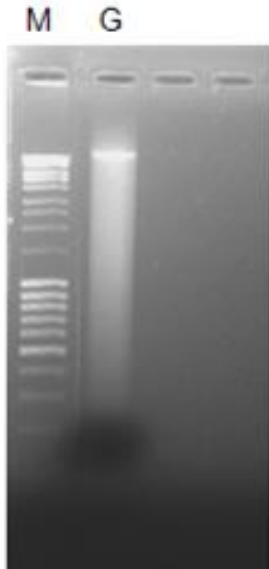


Figure 3.10: Isolated genomic from the bacterial fraction of the Star Diamonds mine fissure water visualized on an ethidium bromide-stained agarose gel 0.8% (w/v): lane M; MassRuler™ DNA ladder (Thermo Scientific), lanes G: genomic DNA

PCR amplification of the rRNA gene V3/V4 hypervariable regions were done as discussed in Section 3.2.7.2. Amplicons of expected size were successfully amplified for the Archaea (~600 bp), Bacteria (~600 bp), and Eukarya (~600 bp) (Figure 3.11).

The DGGE assessment was carried out as outlined in Section 3.2.7.3 where the PCR products were used to construct preliminary diversity profiles for Bacteria (Figure 3.12 a), Eukarya (Figure 3.12 b), and Archaea (Figure 3.12 c). The band migration profile for the Bacteria showed a broader diversity than that of the Archaea and Eukarya suggesting that the Eukarya and Archaea had lower microbial diversity within the studied environment.

The sequenced and cloned DGGE bands for the Bacteria are shown in Table 3.3.5. Unfortunately bands 1, 3 and 10 were cloned, but were not successfully sequenced. From analyses only two phyla were observed (Proteobacteria and Bacteroidetes). The Proteobacteria phylum was the most dominant (Gama- and Alphaproteobacteria). These results correlated to those of Erasmus, (2015) who had done microbial diversity studies on the same borehole as this study in the Star Diamonds mine. The results also correlated with subsurface diversity results obtained by other researchers (Labonté *et al.*, 2015; Lau *et al.*, 2014; Magnabosco *et al.*, 2014). Unfortunately after multiple attempts sequencing of the cloned DGGE bands from Eukarya and Archaea was not successful. The full diversity profile of an environmental sample cannot be done by evaluating a DGGE gel alone as the technique has its drawbacks. One of the

drawbacks is that different DNA sequences may have identical GC contents resulting in multiple species being misrepresented as one band (Muyzer *et al.*, 2004). Another drawback is that possible intra-specific or intra-isolate heterogeneity of rRNA genes can give rise to multiple banding patterns for one species (Duarte *et al.*, 2012). Therefore, whole metagenome sequencing is required for a full in depth analysis of the microbial diversity and has been done in Chapter 4.

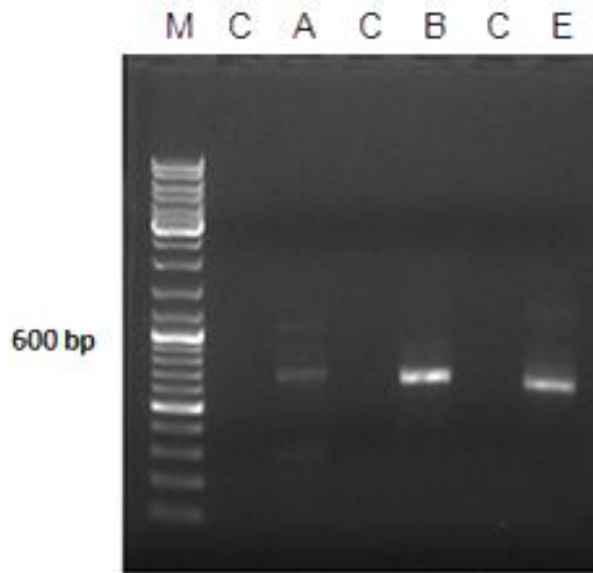


Figure 3.11: PCR amplicons of the rRNA gene fragments. The amplicons were visualized on an ethidium bromide-stained agarose gel 1% (w/v): lane M: GeneRuler™ DNA ladder (Thermo Scientific), lane C: negative control, lane A: Archaea, lane B: Bacteria and lane C: Eukarya.

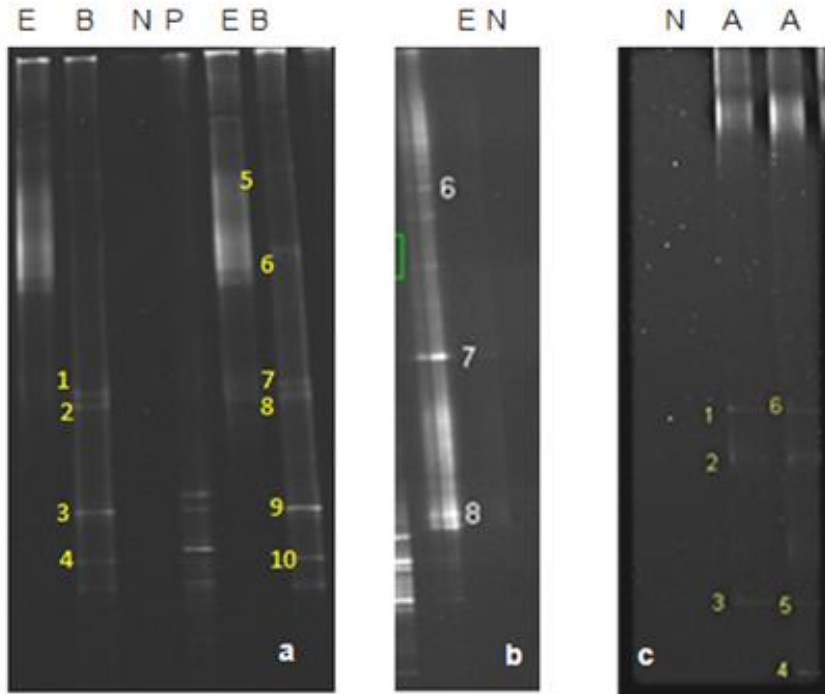


Figure 3.12: DGGE diversity profiles for (a) Bacteria, (b) Eukarya and (c) Archaea. Lane B: Bacteria, lane N: negative control, lane E: Eukarya and lane A: Archaea.

Table 3.3.5: Taxonomy of the pGEM[®]-T Easy cloned and sequenced bacteria DGGE bands.

DGGE band	Order	Phylum
2	Flavobacteriales	Bacteroidetes
4	Sphingomonadales	Alphaproteobacteria
5	Rhodobacterales	Alphaproteobacteria
6	Lysobacteriales	Gammaproteobacteria
7	Sphingomonadales	Alphaproteobacteria
8	Rhodobacterales	Alphaproteobacteria
9	Lysobacteriales	Gammaproteobacteria

3.3.6. Identification of viral-like particles using EFM and TEM

Viral like particles were detected as indicated in Sections 2.2.8. and 2.2.9. using EFM and TEM, which are both techniques that have been used widely in deep ocean environments to detect and enumerate free viral like particles. Figure 3.13 (B) shows the viral sample prior to CsCl gradient ultracentrifugation which purifies the viral-like particles from nanoparticles such as debris from cell material and DNA from lysed viral or bacterial cells (Vale *et al.*, 2010). The particles stained in this sample were most likely viral-like particles along with the above mentioned nanoparticles. Figure 3.13. C shows the purified viral-like particles after CsCl gradient ultracentrifugation of the sample. In the drop TEM method outlined in Section 3.2.8. no phages were observed as concentration were too low to detect phages. A modified method coupling TEM to ultracentrifugation was used as outlined in Section 3.2.8. The TEM analysis revealed *Podoviridae*-like phages shown in Figure 3.14. A and B and *Myoviridae*-like phages shown in Figure 3.14. C. The *Podoviridae*-like phages were approximately 59 nm in head diameter and were non enveloped and these morphological characteristics correlated with those of literature where T7-like *Podoviridae* are normally 60 nm in head diameter and are non-enveloped (Eydal, 2009; Schmidt & Schaechter, 2012). *Podoviridae* generally have short contractile tails, but those detected in this study did not present any tail structures. According to literature phages have the ability to lose their tails during sample preparation for TEM analysis, this could therefore explain the presence of non-tailed phages in this study (Suttle, 2005; Williamson *et al.*, 2012). The *Myoviridae*-like phages were approximately 113.3 nm in head diameter and were non-enveloped which correlated with results from literature where T4-like *Myoviridae* have head diameters between 50-110 nm that are non-enveloped (Eydal, 2009; Schmidt & Schaechter, 2012). The tail was short and non-contractile unlike those found in literature where the tail is long, but this can be explained by sample handling which could have resulted in tail disintegration (Suttle, 2005; Williamson *et al.*, 2012). The absence of *Siphoviridae*-like phages could have been owed to the fact that this type of phage is mainly lysogenic where they prefer to stay integrated in the hosts genome as temperate phages unlike the 2 above mentioned which are mostly lytic phages (Eydal, 2009; Suttle, 2005). The TEM results from this study are closely related to those shown by Mabizela, (2009), who was able to detect phages belonging to the families *Myoviridae* and *Siphoviridae* in a biofilm mat of a 176 kmbs deep gold mine from the Free State area (Beatrix, Sibanye Gold).

According to literature *Myoviridae*, *Podoviridae* and *Siphoviridae* have been found to exist in abundance in the marine environments as they have been isolated from surface and deep

marine environments and characterized according to morphology using TEM analysis (Breitbart & Rohwer, 2005; Liu *et al.*, 2006; Suttle, 2005; Zhang & Wang, 2010). These viral families are termed as tailed phages that have dsDNA and are bacteriophages meaning they infect bacteria (Anderson *et al.*, 2013). Phage families belonging to families *Siphoviridae* and *Myoviridae* have even been isolated and characterized in extreme temperature marine subsurface environments such as the hydrothermal fields and vents using TEM (Liu *et al.*, 2006; Zhang & Wang, 2010).

This is the first study in the South African deep subsurface mines to identify free phages using TEM in an isolated viral fissure water fraction, whereas Mabizela, (2009) isolated them in biofilm structures instead. It is also the first study to detect free viral particles in the South African diamond mine.

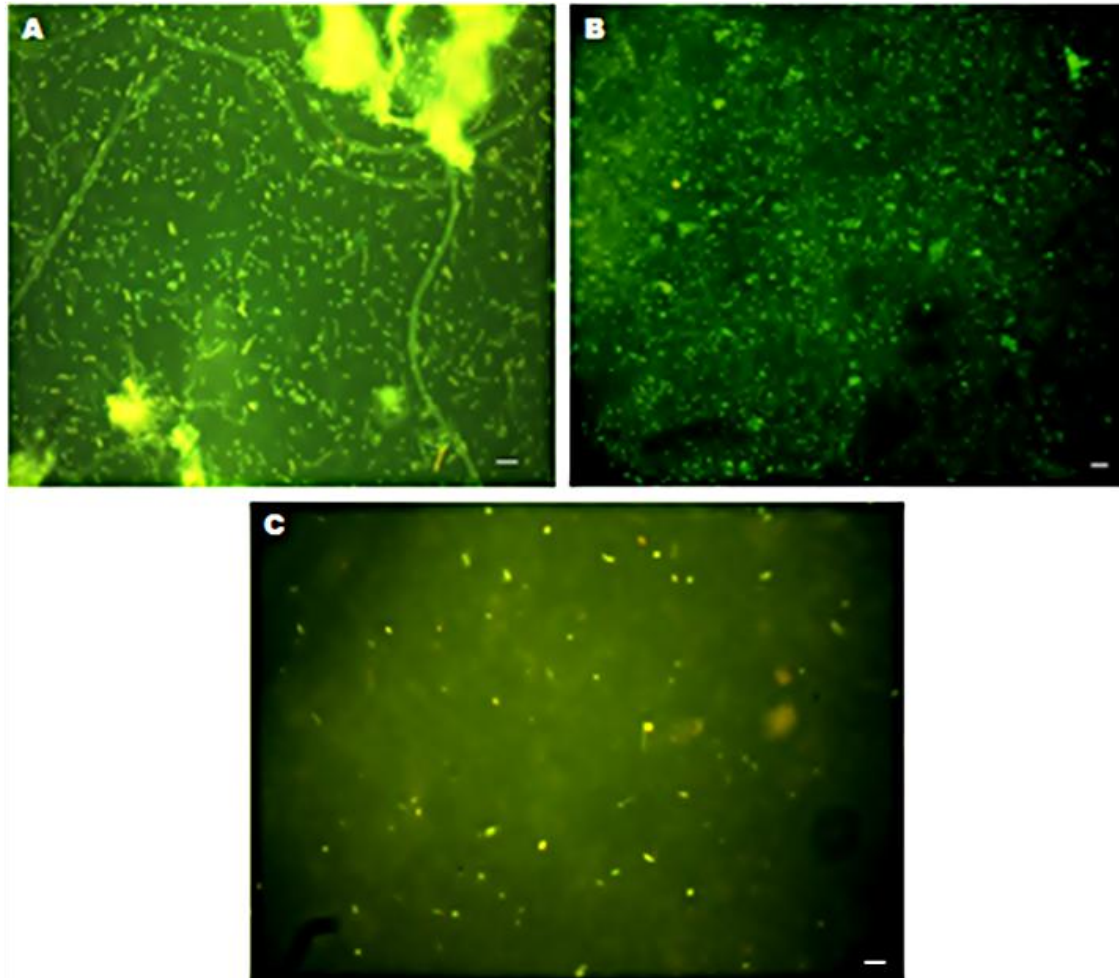


Figure 3.13: EFM images of the bacterial and viral fraction of the fissure water from Star Diamonds mine. (A) Bacterial fraction sample, (B) viral fraction sample prior to CsCl gradient ultracentrifugation and (C) viral fraction sample after CsCl gradient ultracentrifugation. Scale bar is equal to 2 μm for A and 1 μm for B and C.

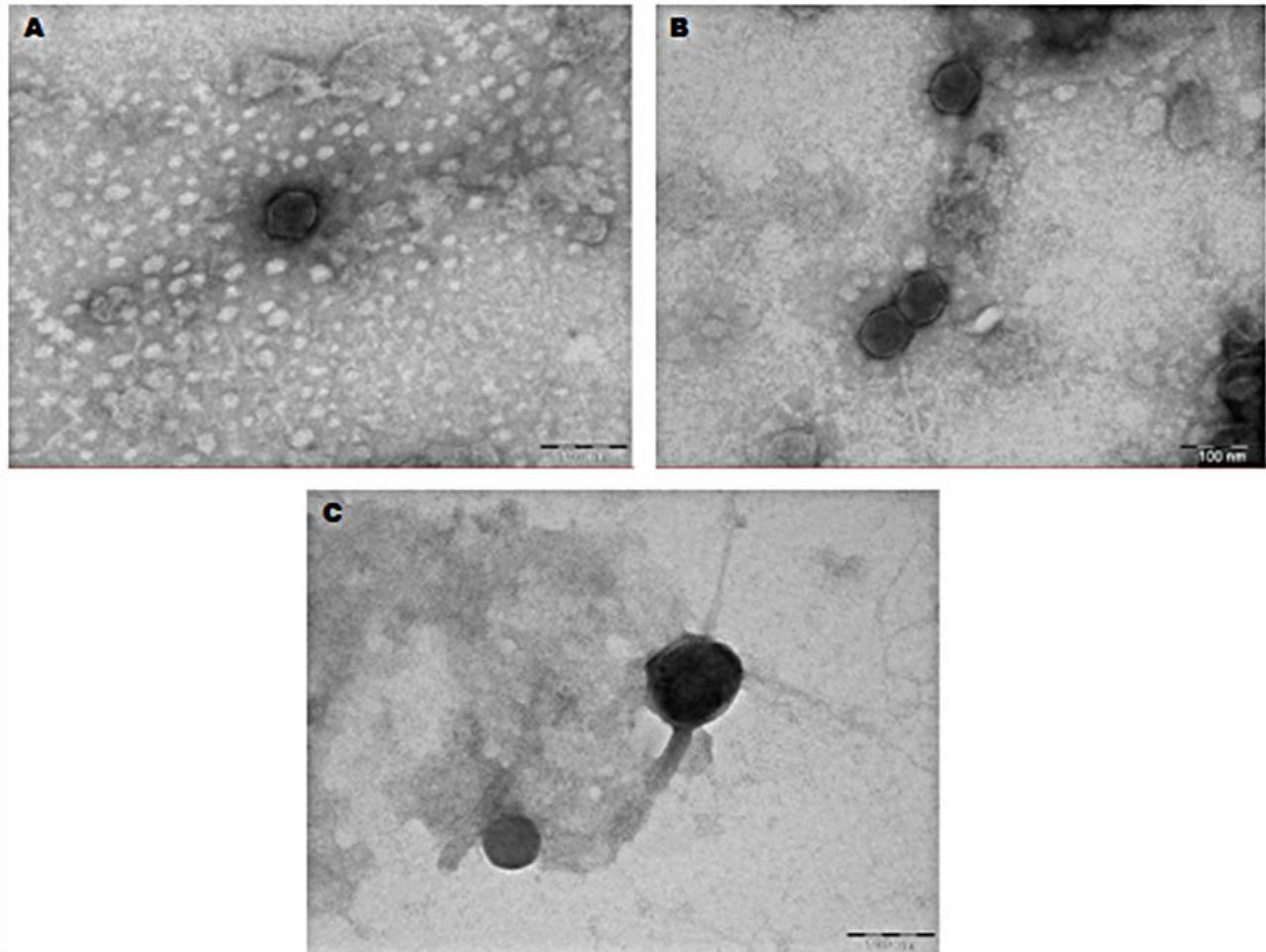


Figure 3.14: TEM images of the phage morphologies found in the fissure water from Star Diamonds mine. A and B are *Podoviridae*-like phages with head diameters of 59 nm and C is a *Myoviridae*-like phage with head diameter of 113.3 nm and a short non-contractile tale. Scale bars are equal to 100 nm.

3.3.7. Identification of phage genes using Sanger sequencing

The extraction of the virome from the viral fraction of the fissure water resulted in low concentrations of DNA and RNA. Shown in Table 3.3.7. are the concentrations from the extraction methods measured on the Qubit™. The phenol/chloroform extraction resulted in concentrations out of range meaning they were too low to be detected by the machine. The Formamide and NucleoSPIN® Soil kit extractions were able to detect RNA and only the NucleoSPIN® Soil kit was able to detect DNA at a concentration of 20 ng/μL. The virome extracted using the NucleoSPIN® Soil kit was evaluated for integrity by electrophoresis on a

0.8% agarose gel (Figure 3.15). The DNA yields shown in (Figure 3.15) Lane V indicate successful DNA extraction, but the integrity of the DNA is not high and had some shearing.

Table 3.3.7: DNA and RNA concentrations of the three extraction methods measured using the Qubit™.

Extraction method	Concentration for DNA and RNA
Phenol/chloroform	Below range for DNA and RNA
Formamide	5.17 ng/μL RNA Below range for DNA
NucleoSPIN Kit	20 ng/μL DNA and 11 ng/μL RNA



Figure 3.15: Extracted viral genomic DNA from the viral fraction of the Star Diamonds mine fissure water visualized on an ethidium bromide-stained agarose gel 0.8% (w/v): lane M; MassRuler™ DNA ladder (Thermo Scientific), lanes V: viral genomic DNA/virome.

The primer sets in Table 3.2.12 are DNA primers as they only amplify certain genes in dsDNA phages. From the viral DNA PCR amplifications using the primer sets, there was positive amplification for only the 1F/1R primers (Figure 3.16) which amplify the integrase gene of lysogenic phage at gradient temperatures 55°C and 56°C in lane 1 and 2 respectively. The

amplicons amplified had three sizes (2500, 1500 and 700 bp) which were all higher than the expected integrase gene size which is approximately 500 bp (Balding *et al.*, 2005). The low concentrations of viral DNA extracted and the difficulties in extracting it could be responsible for the lack of PCR amplification from the other primers. The gradient PCR amplification of the phage genes using the primers in Table 3.2.12 showed positive amplification for the 1F/1R, T4g23F/T4g23R and the CPS3/CPS8 primers. The T4g23F/T4g23R and the CPS3/CPS8 primers amplify the g23 (major capsid protein) from T4 type phages and the gene encoding major capsid proteins (g20) from Cyanophages respectively (Figure 3.17). All of the phage gene primer sets used in this study are universal/degenerate that have been used to detect phages in multiple environments thereby limiting their specificity for a specific environment. The latter could be an explanation for the multiple amplicons for each positive amplification primer set at each gradient temperature and also for the lack of PCR products for the HECTORPol29F/HECTORPol500R, HECTORPol29F/HECTORPol711R, 8F/8R and MZIABis/MZIA6 primer sets (Figure 3.17). The genomic DNA isolated from the bacterial fissure water fraction is of the whole microbial community in the sample therefore the multiple amplicons could be a result of different sizes of the same gene being amplified in different bacterial species. After multiple attempts the PCR amplicons were unsuccessfully cloned into the pSMART[®] HCKan vector.

A previous study done by Mabizela, (2009) using the same sets of primers, was only able to amplify the g23 (major capsid protein) from T4 type phages and the DNA polymerase from uncultured podophages in several mines within the Witwatersrand Basin. In the past few years environment phage studies have moved away from using universal primers to characterize phages in a specific environment and instead look at metagenomics studies as a tool to identify free phages and host genome integrated phages (prophages) and phage related genes as it gives broader more invasive sequencing (Anantharaman *et al.*, 2014; Dell'Anno *et al.*, 2015).

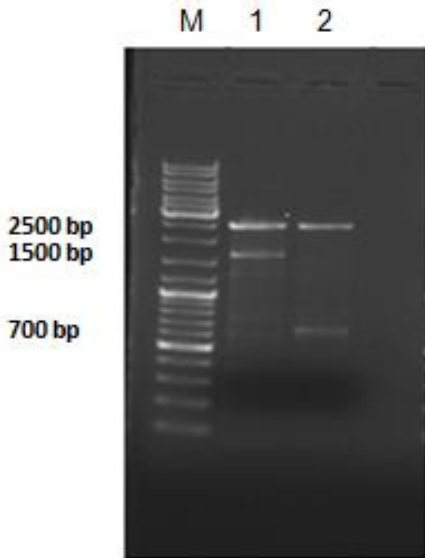


Figure 3.16: Gradient PCR amplification of the 1F/1R primers which amplify the conserved loci integrase gene of lysogenic phages. Amplification was performed on the viral genomic DNA extracted from the viral fraction of the Star Diamonds mine fissure water. The amplicons were visualized on an ethidium bromide-stained agarose gel 1% (w/v): lane M; GeneRuler™ DNA ladder (Thermo Scientific), lanes 1: amplification at an annealing temperature of 55°C and lane 2: at an annealing temperature of 56°C.

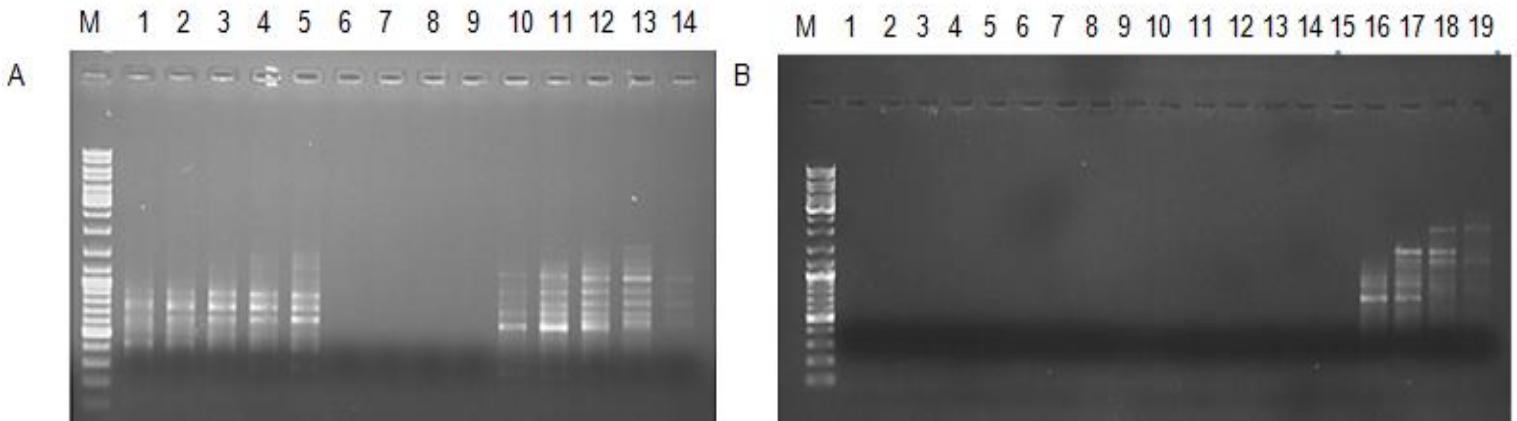


Figure 3.17: Gradient PCR amplification performed on the genomic DNA from the bacterial fraction of the Star Diamonds mine fissure water. The amplicons were visualized on an ethidium bromide-stained agarose gel 1% (w/v). (A): lane M; GeneRuler™ DNA ladder (Thermo Scientific), lanes 1-5 amplification using the 1F/1R primers which amplify the conserved loci integrase gene of lysogenic phage at gradient annealing temperatures 55-60°C respectively, lanes 6-9 8F/8R primers amplify the conserved loci integrase gene of lysogenic phage at gradient annealing temperatures 55-59°C respectively and lanes 10-14 T4g23F/T4g23R primers amplify the g23 (major capsid protein) from T4 type phages at gradient annealing temperatures 55-60°C respectively. (B): lane M; GeneRuler™ DNA ladder, lanes 1-5 MZIAbis/MZIA6 primers which amplify the g23 (major capsid protein) from T4 type phages at

gradient annealing temperatures 55-60°C respectively, lanes 6-10 HECTORPol29F/HECTORPol711R primers amplifying the DNA polymerase from uncultured podophages, lanes 11-15 HECTORPol29F/HECTORPol500R primers amplifying the DNA polymerase from uncultured podophages and lanes 16-19 CPS3/CPS8 primers amplifying the gene encoding major capsid proteins from Cyanophages.

3.4. Conclusions

A borehole 640 m below surface in the Star Diamonds mine was chosen as the study site and the fissure water sampled was characterized by its biogeochemical features. In both the 2015 and 2016 samples the water had moderate temperatures and close to alkaline pH values. The ORP values were consistent and suggested the environment sustained its reducing nature. Flow cytometry proved to be the most accurate method for cell enumeration prior and after concentration using TFF. The increased cell counts after TFF concentration, suggested that the concentration of the microorganisms was achieved. The pre-concentration flow cytometry cell count was 5.3×10^5 cells/mL which correlated with literature where it has been estimated that cell counts in the deep subsurface groundwater are between 1×10^3 - 1×10^6 cells/mL (Akob & Küsel, 2011; Borgonie *et al.*, 2011; Onstott *et al.*, 1998). The SEM images revealed that most of the bacterial communities were in biofilm structures of which was expected as bacteria prefer such structures in unfavourable conditions of which the deep subsurface is, due to its reducing nature and low nutrient availability (Toole *et al.*, 2000).

Successful PCR amplification of the rRNA gene V3/V4 hypervariable regions for Archaea, Bacteria and Eukarya was achieved. Sequencing results revealed two phyla (Proteobacteria and Bacteroidetes) present within the bacterial domain with the Proteobacteria phylum being the most dominant (Gamma- and Alphaproteobacteria). Unfortunately the sequencing of the cloned DGGE bands from Eukarya and Archaea was not successful after multiple attempts. The full diversity profile of an environmental sample cannot be done by evaluating a DGGE gel alone as the technique has its drawbacks. Therefore, whole metagenome sequencing is needed for a full in depth analysis of the microbial diversity and is outlined in Chapter 4.

The separation of the viral and bacterial fraction was achieved and this can be seen in the cell counts using flow cytometry of the viral fraction sample where there were low insignificant bacterial cell counts. Viral-like particle were observed using EFM and phages were identified using TEM where two double-stranded DNA phage families were observed. The phages belonged to the *Myoviridae* and *Podoviridae* as they correlated to the morphotypes in literature.

These phage families have been extensively studied and found in the deep subsurface and surface marine environments (Suttle, 2005). This is the first study to identify free viral particles in the fissure water of deep subsurface South African mines.

The 1F/1R primer set (amplifying the integrase gene) was the only positive amplification from the viral DNA. The genomic DNA from the bacterial fissure water fraction had positive amplification for the primers which amplify the g23 (major capsid protein) from T4 type phages and the gene encoding major capsid proteins (g20) from Cyanophages. All of the phage gene primer sets used in this study are universal/degenerate limiting their specificity for a specific environment and this could have been an explanation for the multiple amplicons for each positive amplification primer set at each gradient temperature and also for the lack of PCR products. Phage studies have moved away from using universal primers and instead look at metagenomics studies as a tool to identify free phages and host genome integrated phages (prophages) and phage related genes as it gives a broader more invasive analysis.

3.5. References

Akob, D. M. and Küsel, K. (2011). Where microorganisms meet rocks in the Earth's Critical Zone. *Biogeosciences*. **8**: 3531–3543.

Ambriz-Aviña, V., Contreras-Garduño, J. A. and Pedraza-Reyes, M. (2014). Applications of flow cytometry to characterize bacterial physiological responses. *Biomed Research International*. **2014**: 461941.

Anantharaman, K., Duhaime, M. B., Breier, J. A., Wendt, K. A., Toner, B. M. and Dick, G. J. (2014). Sulfur oxidation genes in diverse deep-sea viruses. *Science*. **344**: 757–760.

Anderson, R. E., Brazelton, W. J. and Baross, J. A. (2013). The deep virosphere : assessing the viral impact on microbial community dynamics in the deep subsurface. *Reviews in Mineral Geochemistry*. **75**: 649–675.

Balding C, Bromley S A, Pickup R W and Saunders J R. (2005). Diversity of phage integrases in *Enterobacteriaceae*: development of markers for environmental analysis of temperate phages. *Environmental Microbiology*. **7**: 1558–1567.

Bau, M., Romer, R. L., Luders, V. and Beukes, N. J. (1999). Pb , O , and C isotopes in silici ϕ ed Moodraai dolomite (Transvaal Supergroup , South Africa): implications for the composition of Paleoproterozoic seawater and ` dating ` the increase of oxygen in the Precambrian atmosphere. *Earth and Planetary Science Letters*. **174**: 43–57.

Børshem, K. Y., Bratbak, G. and Haldal, M. (1990). Enumeration and biomass estimation of planktonic bacteria and viruses by transmission electron microscopy. *Applied and Environmental Microbiology* **56**: 352-356.

Borgonie, G., Linage-Alvarez, B., Ojo, A., Mundle, S., Freese, L., Van Rooyen, C., Kuloyo, O., Albertyn, J., Pohl, C., Cason, E. D., Vermeulen, J., Pienaar, C., Litthauer, D., van Niekerk, H., van Eeden, J., Sherwood Lollar, B., Onstott, T. C. and van Heerden, E. (2015). Eukaryotic opportunists dominate the deep-subsurface biosphere in South Africa. *Nature Communications*. **6**: 1–12.

Borgonie, G., García-Moyano, A., Litthauer, D., Bert, W., Bester, A., van Heerden, E., Möller, C., Erasmus, M. and Onstott, T. C. (2011). Nematoda from the terrestrial deep subsurface of South Africa. *Nature*. **474**: 79–82.

Breitbart, M. and Rohwer, F. (2005). Here a virus , there a virus , everywhere the same virus ? *TRENDS in Microbiology*. **13(6)**: 276–284.

Breitbart, M., Wegley, L., Leeds, S., Rohwer, F. and Schoenfeld, T. (2004). Phage community dynamics in hot springs. *Applied and Environmental Microbiology*. **70(3)**: 1633–1640.

Casamayor, E.O., Massana, R., Benlloch, S., ØvreåS, L., Díez, B., Goddard, V.J., Gasol, J.M., Joint, I., Rodríguez-Valera, F. and Pedrós-Alió, C. (2002). Changes in Archaeal, Bacterial and Eukaryal assemblages along a salinity gradient by comparison of genetic fingerprinting methods in a multipond solar system. *Environmental Microbiology*. **4(6)**: 338-348

Chivian, D., Brodie, E. L., Alm, E. J., Culley, D. E., Dehal, P. S., Desantis, T. Z., Gihring, T. M., Lapidus, A., Lin, L., Lowry, S. R., Moser, D. P., Richardson, P. M., Southam, G., Wanger, G., Pratt, L. M., Andersen, G. L., Hazen, T. C., Brockman, F. J., Arkin, A. P. and Onstott, T.C. (2008). Environmental genomics reveals a single-species ecosystem deep within earth. *Science Reports*. **322**: 275–278.

Danovaro, R. (2000). Viral density and virus-to-bacterium ratio in deep-sea sediments of the Eastern Mediterranean. *Applied and Environmental Microbiology*. **66(5)**: 1857–1861.

Dell’Anno, A., Corinaldesi, C. and Danovaro, R. (2015). Virus decomposition provides an important contribution to benthic deep-sea ecosystem functioning. *PNAS*. E2014–E2019.

Demuth, J., Neve, H. and Witzel, K. (1993). Direct electron microscopy study on the morphological diversity of bacteriophage populations in Lake PluB3see. *Applied and Environmental Microbiology*. **59**: 3378–3384.

Díez B, Pedrós-Alió C, Marsh T L and Massana R, (2001). Application of denaturing gradient gel electrophoresis (DGGE) to study the diversity of marine picoeukaryotic assemblages and comparison of DGGE with other molecular techniques. *Applied and Environmental Microbiology*. **67**: 2942–2951.

Duarte, S., Cássio, F. and Pascoal, C. (2012). Denaturing gradient gel electrophoresis (DGGE) in microbial ecology – Insights from freshwaters. Chapter 3 in *Gel electrophoresis – Principles and Basics*.

Erasmus, M. (2015). *Geological Carbon Sequestration and its Influence on Subsurface Microbial Diversity and Metabolic Carbon Cycling*. (PhD thesis). University of the Free State.

Essers, E. (2003). (12) United States Patent.

Eydal, H. S. C. (2009). *Microbial ecology in deep granitic groundwater – activity and impact of viruses*.

Filée, J., Tétart, F., Suttle, C. A. and Krisch, H. M. (2005). Marine T4-Type bacteriophages, a ubiquitous component of the dark matter of the biosphere. *Proceedings of the National Academy of Sciences*. **102**: 12471-12476.

Flemming, H. and Wingender, J. (2010). The biofilm matrix. *Nature Reviews: Microbiology*. **8**: 623–33.

Francisco, D. E., Mah, R. A. and Rabin, A. C. (1973). Acridine orange-epifluorescence technique for counting bacteria in natural waters. *Transaction of the American Microscopical Society*. **92**: 416–421.

Garcia-Martinez, J., Acinas, S. G., Anton, A. I. and Rodriguez-Valera, F. (1999). Use of the 16S-23S ribosomal genes spacer region in studies of prokaryotic diversity. *Journal of Microbiological Methods*. **36**: 55–64.

Gentile, M. and Gelderblom, H. R. (2014). Electron microscopy in rapid viral diagnosis : an update. *New Microbiol ogica*. **37**: 403–422.

Giovannoni, S. J., DeLong, E. F., Schmidt, T. M. and Pace, N. R. (1990). Tangential flow filtration and preliminary phylogenetic analysis of marine picoplankton. *Applied and Environmental Microbiology*. **56**: 2572–2575.

Hambly, E. and Suttle, C. A. (2005). The virosphere , diversity , and genetic exchange within phage communities. *Current Opinion Microbiology*. **8**: 444–450.

Kepner, R. L. and Pratt, J. R. (1994). Use of fluorochromes for direct enumeration of total bacteria in environmental samples : Past and Present. *Microbiological Reviews*. **58(4)**: 603–615.

Kerstens, M., Boulet, G., Van kerckhoven, M., Clais, S., Lanckacker, E., Delputte, P., Maes, L. and Cos, P. (2015). A flow cytometric approach to quantify biofilms. *Folia Microbiologica*. **60**: 335–342.

Koster, W., Egli, T., Ashbolt, N., Botzenhart, K., Burlion, N., Endo, T., Grimont, P., Guillot, E., Mabilat, C., Newport, L., Niemi, M., Payement, P., Prescott, A., Renaud, P. and Rust, A. (2003). Analytical methods for microbiological water quality testing. In: *Assessing Microbial Safety of Drink Water*. 237–244.

Labonté, J. M., Field, E. K., Lau, M., Chivian, D., van Heerden, E., Wommack, K. E., Kieft, T. L., Onstott, T. C. and Stepanauskas, R. (2015). Single cell genomics indicates horizontal gene transfer and viral infections in a deep subsurface Firmicutes population. *Frontiers in Microbiology*. **6**: 1–11.

Lane, D J., Pace, B., Olson, G. J., Stahl, D. A., Sogin, M. L. and Pace, N. R. (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proceedings of the National Academy of Sciences*. **82**: 6955–6959.

Lau, M. C. Y., Cameron, C., Magnabosco, C., Brown, C. T., Schilkey, F., Grim, S., Hendrickson, S., Pullin, M., Lollar, B. S., van Heerden, E., Kieft, T. L. and Onstott, T. C. (2014). Phylogeny and phylogeography of functional genes shared among seven terrestrial subsurface metagenomes reveal N-cycling and microbial evolutionary relationships. *Frontiers in Microbiology*. **5**: 1–17.

Lebaron, P., Parthuisot, N. and Catala, P. (1998). Comparison of blue nucleic acid dyes for flow cytometric enumeration of bacteria in aquatic systems. *Applied and Environmental Microbiology*. **64**: 1725–1730.

Lin, L., Wang, P., Rumble, D., Lippmann-pipke, J., Boice, E., Pratt, L. M., Lollar, B. S., Brodie, E. L., Hazen, T. C., Andersen, G., DeSantis, T. Z., Moser, D. P., Kershaw, D. and Onstott, T. C. (2006). Long-term sustainability of a high-energy, low-diversity crustal biome. *Science*. **314**: 479–482.

Liu, B., Wu, S., Song, Q., Zhang, X. and Xie, L. (2006). Two novel bacteriophages of thermophilic bacteria isolated from deep-sea hydrothermal fields. *Current Microbiology*. **53**: 163–166.

Mabizela, L. (2009). *A metagenomic investigation of phage communities from South African mines* (PhD thesis). University of the Free State.

Magnabosco, C., Tekere, M., Lau, M. C. Y., Linage, B., Kuloyo, O., Erasmus, M., Cason, E., van Heerden, E., Borgonie, G., Kieft, T. L., Olivier, J. and Onstott, T. C. (2014). Comparisons of the composition and biogeographic distribution of the bacterial communities occupying South African thermal springs with those inhabiting deep subsurface fracture water. *Frontiers in Microbiology*. **5**: 1–17.

Maniloff J. (1995). Identification and classification of viruses that have not been propagated. *Archives of Virology*. **140**: 1515-1520.

Muyzer, G., De Waal, E.C. and Uitterlinden, A.G. (1993). Profiling of complex microbial populations by denaturing gradient Gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Applied and Environmental Microbiology*. **59(3)**: 695-700.

Muyzer, G., Brinkoff, T., Nübel, U., Santegoeds, C., Schäfer, H and Wawer, C. (2004). Denaturing gradient gel electrophoresis (DGGE) in microbial ecology. In: *Molecular microbial ecology Manual* (2nd ed.) *Kluwer Academic Publishers*.

Onstott, T. C., Phelps, T. J., Colwell, F. S., Ringelberg, D., White, D. C., Boone, D. R., Mckinley, J. P., Stevens, T. O., Long, P. E., Balkwill, D. L., Friffin, W. T. and Kieft, T. (1998). Observations pertaining to the origin and ecology of microorganisms recovered from the deep subsurface of Taylorsville Basin, Virginia. *Geomicrobiology Journal*. 15: 353–385.

Onstott, T. C., Ward, J., Lollar, B. S., Boice, E., Pratt, L. M., Pfiffner, S., Moser, D., Gihring, T., Kieft, T. L., Phelps, T.J., van Heerden, E., Litthaur, D., DeFlaun, M., Rothmel, R., Wanger, G. and Southam, G. (2006). The Origin and Age of biogeochemical trends in deep fracture water of the Witwatersrand Basin , South Africa. *Geomicrobiology Journal*. **23**, 369–414.

Petrusevski., Bolier, G., Van Breemen, A. N. and Alaerts, G. J. (1995). Tangential flow filtration: A method to concentrate freshwater algae. *Water Research*. **29(5)**: 1419–1424.

Porter, K. and Feig, Y. (1980). The use of DAPI for identifying aquatic microfloral. *Limnology and Oceanography*. **25(5)**: 943–948.

Pruesse, E., Quast, C., Knittel, K., Fuchs B.M., Ludwig, W., Peplies, J. and Glöckner, F.O. (2007). SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic. Acids Research*. **35(21)**: 7188-7196

Rohwer, F. and Edwards, R. (2002). The Phage Proteomic Tree: a genome-based taxonomy for phage. *Journal of Bacteriology*. **184**: 4529-35.

Rohwer, F. (2005). Tangential flow filtration, Rohwer Lab 1–15.

Sambrook, J. and Russell, D. W. (2001). *Molecular cloning: a laboratory manual*, (2rd ed.). *Cold Spring Harbor Laboratory Press*. Cold Spring Harbor, NY.

Schmidt, T. and Schaechter, M. (2012). Environmental viral pool: Topics in Ecology and Environmental Microbiology. *Academic Press*. 103

Selby, C. C. (1953). Microscopy.II. Electron Microscopy : A Review. *Cancer Research*. **13(11)**: 753–775.

Suttle, C. A. (2005). Viruses in the sea. *Nature*. **437**: 356–361.

Suttle, C. A. and Fuhrman, J. A. (2010). Enumeration of virus particles in aquatic or sediment samples by epifluorescence microscopy. *Limnology and Oceanography*. 145–153.

Swart, C. W., Swart, H. C., Coetsee, E., Pohl, C. H., Wyk, P. W. J. Van and Kock, J. L. F. (2010). 3-D architecture and elemental composition of fluconazole treated yeast asci. *Scientific Research and Essays* **5(22)**: 3411–3417.

Toole, G. O., Kaplan, H. B. and Kolter, R. (2000). Biofilm formation as microbial development. *Annual Reviews Microbiology*. **54**: 49–79.

Tuma, R. S., Beaudet, M. P., Jin, X., Jones, L. J., Cheung, C., Yue, S. and Singer, V. L. (1999). Characterization of SYBR Gold nucleic acid gel stain : A dye optimized for use with 300-nm ultraviolet transilluminators. *Analytical Biochemistry*. **268**: 278–288.

Vale, F. F., Correia, A. C., Matos, B., Nunes, J. F. M. and Matos, A. P. A. De. (2010). Applications of transmission electron microscopy to virus detection and identification. *Microscopy: Science, Technology and Applications and Education*. 128–136.

van der Westhuizen, W. A., Kock, J. L. F., Coetsee, E., van Wyk, P. W. J., Swart, H. C. and Bragg, R. R. (2013). Investigation of the final stages of a P4-like coliphage infection in *Escherichia coli* through scanning , transmission and nano-scanning-auger electron microscopy (NanoSAM). *Scientific Research and Essay*. **8(10)**: 382–387.

van Reis, R. D. (1993). Tangential flow filtration process and apparatus.

Vermaak, C. (2012). Water Monitoring Report for Star Diamonds (Theunissen).

Weaver, J. L. (2000). *Introduction to flow cytometry. Methods*.

Williamson, K. E., Helton, R. R. and Wommack, K. E. (2012). Bias in bacteriophage morphological classification by transmission electron microscopy due to breakage or loss of tail structures. **457**: 452–457.

Wolkersdorfer, C. and Hubert, E. (2016). Establishing a Total Dissolved Solids: Electrical Conductivity ratio for Mine Waters.

Young, M. R. (1961). Principles and Technique of Fluorescence Microscopy. *Quarterly Journal of Microscopical Science*. **102(4)**: 419–449.

Zhang, X. and Wang, Y. (2010). Genome analysis of deep-sea thermophilic phage D6E. *Applied and Environmental Microbiology*. **76(23)**: 7861–7866.

Zhang, Z., Schwartz, S., Wagner, L. and Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology*. **7**: 203-214.

Zhong, Y., Chen, F., Wilhelm, S. W., Poorvin, L. and Hodson, R. E. (2002). Phylogenetic diversity of marine cyanophage isolates and natural virus communities as revealed by sequences of viral capsid assembly protein gene *g20*. *Applied and Environmental Microbiology*. **8**: 1576–1584.

Zipper, H., Brunner, H., Bernhagen, J. and Vitzthum, F. (2004). Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications. *Nucleic Acids Research*. **32(12)**: e103.

CHAPTER 4

MICROBIAL DIVERSITY IDENTIFICATION OF PHAGE GENES AND HORIZONTAL GENE TRANSFERS

4.1. Introduction

Microorganisms isolated from deep subsurface environments, as previously mentioned, are predominantly unculturable due to their unique and extreme growth conditions, therefore leading to the use of culture-independent techniques such as genome sequencing and microscopic-based evaluation (Chivian *et al.*, 2008; Danovaro, 2000; Hambly & Suttle, 2005; Labonté *et al.*, 2015). There has been a shift in using DGGE and Sanger sequencing as the sole method for studying the microbial diversity profiles in environmental samples as this does not result in the full diversity profiles. As mentioned in Chapter 3, one of the drawbacks of DGGE is that different DNA sequences may have identical GC contents resulting in multiple species being misrepresented as one band (Muyzer *et al.*, 2004). The other is that possible intra-specific or intra-isolate heterogeneity of rRNA genes can give rise to multiple banding patterns for one species (Duarte *et al.*, 2012). The lack of full diversity profiles being achieved using these methods has led to the use of (NGS) techniques. These techniques allow for comprehensive sequencing as such one is able to identify and assemble genomes of species that are present at low frequencies in environmental samples (Gawad *et al.*, 2016; Hug *et al.*, 2016).

Some of the NGS techniques that have been widely used in diversity studies of environmental samples include Illumina 16S rRNA gene sequencing, Illumina shotgun whole metagenome sequencing and single cell genome sequencing. Illumina sequencing targeting the 16S rRNA gene is a high-throughput sequencing technique for archaeal and bacterial diversity (Klindworth *et al.*, 2013), this technique uses region specific primers either for bacteria or for the archaea (Sanschagrín & Yergeau, 2014). Due to the size of the end result data, the Illumina MiSeq is the preferred platform used for 16S rRNA gene sequencing (Caporaso *et al.*, 2012) since it is capable of producing long (301 bp) paired end reads. Forward and reverse reads can then be paired using specialized software to obtain contigs long enough for accurate genus level identification. Single cell genomics and shotgun whole metagenome sequencing are considered to be far more comprehensive NGS techniques as they are able to sequence the whole genome of a cell by performing shotgun sequencing. It is through the use of these techniques that draft and complete genomes of previously undetected phyla have been sequenced (Blainey, 2014; Hug *et al.*, 2016). Both these techniques are able to identify and assemble genomes of species that are present at low frequencies in environmental samples (Gawad *et al.*, 2016; Hug *et al.*, 2016). They are also genome based techniques that are able to provide information about the

metabolic potential and a variety of phylogenetically informative sequences that can be used to classify organisms (Hug *et al.*, 2016).

Single cell genomics performs shotgun whole genome sequencing on isolated cells of interest in the whole community of a given complex sample (Blainey, 2014). The workflow for this technique requires demanding sample preparations which include the isolation of the single cells using flow cytometry and fluorescence-activated cell sorting to detect cells with characteristics of interest (Blainey, 2014). A good example of cells with interesting characteristics are the Candidate Phyla Radiation lineage (previously mentioned in Chapter 1) which have relatively small genomes and most have restricted metabolic capabilities with many being inferred as symbionts (Hug *et al.*, 2016). The cells of interest are then lysed in order to free the genomic DNA which are then subjected to whole genome sequencing

In deep subsurface environments viruses in general have been shown to play a role in altering the biogeochemical cycles, microbial diversity profiles and their genetic contents (Anderson *et al.*, 2011; Prangishvili & Garrett, 2004). The role of phages in deep marine and terrestrial environments has been rarely considered and has therefore caused an interest in recent research (Anderson *et al.*, 2013; Breitbart & Rohwer, 2005; Labonté *et al.*, 2015; Prangishvili & Garrett, 2004). Thus far from research it has been shown that phages in deep marine environments prefer lytic lifestyles due to harsh environmental conditions being a limiting factor especially for phages that necessarily need to find a host (Anderson *et al.*, 2013). In the deep terrestrial subsurface viruses play a role in providing food and energy for the uninfected prokaryotes through viral shunts which are cell contents and biomass that is transformed into organic detritus (Danovaro *et al.*, 2008). As mentioned in Chapter 3 phage studies focusing on environmental samples have moved away from using universal primers to characterize phages in a specific environment and instead look at shotgun whole metagenome and single cell genome sequencing as techniques to study the phage and host interactions (Anantharaman *et al.*, 2014; Dell'Anno *et al.*, 2015).

Studies using shotgun whole metagenome sequencing in the deep mines of the Witwatersrand Basin in South Africa have revealed the bacteria and phage interactions in deep terrestrial environments. One of the most important bacteria-phage interaction mechanism is horizontal gene transfers (HGTs) where phages are able to alter the genetic content and expression of genes in the bacterial host through transduction (Thomas & Nielsen, 2005). By moving genetic material via HGT from one host to another the phages effectively alter the microbial evolution of

the microbial species (Angly *et al.*, 2006). Phage genes from prophages may improve the hosts fitness and these can be expressed during integration of viral genome into the hosts genome (Lindell *et al.*, 2005). Some examples of host-phage interactions found in South African mines include work done by Chivian and co-workers, (2008) on the *Candidatus Desulforudis audaxviator*, bacterium from the Mponeng mine (AngloGold Ashanti), which shared genes of archaeal origin in the nitrogen and carbon fixation pathways, cobalamin synthesis and sulfide reduction pathways and with motility and sporulation genes. These genes were believed to be products of HGT events from other bacteria and archaea and this was validated by the presence of transposase and clustered regularly interspaced palindromic repeats (CRISPRs) in its genome. Mabizela, (2009) sequenced the viral metagenome from samples in Beatrix (Sibanye Gold) which revealed 7 prophages and insertion sequences, believed to be transposase, in bacterial tRNA clusters. Labonté and co-workers, (2015) identified in their metagenome data partial genomes of Mu-like transposable phages, retrons, CRISPRs and lambda-like prophages in Firmicutes and Proteobacteria in the Tona (Anglo Gold Ashanti).

4.1.2. Aims of chapter

The aims of this chapter were to:

- Characterize the full microbial diversity using Illumina next generation sequencing techniques
- Annotate and analyze metagenome data contigs for the presence of phage genes and prophages.
- Predict and annotate phage genes and genes related to horizontal gene transfers within the binned bacterial metagenome data genomes.
- Speculate on the effect HGTs have on the deep subsurface bacterial communities in terms of evolution and survival.

4.2. Methods and materials

4.2.1. Bacterial diversity assessment using next generation 16S rRNA sequencing

4.2.1.1. Genomic DNA isolation

The bacterial fraction (approximately 100 mL) from the 2015 and 2016 samples, previously concentrated and separated from the viral fraction by TFF in Section 3.2.3. was filtered through a PES 0.22 µm filter (47 mm Supor[®] 200, Pall Corporation) using a 47 mm glass filter unit (Millipore). Cut-up filter pieces were added to tubes containing beads and the genomic DNA was isolated using the NucleoSpin[®] Soil kit (Macherey-Nagel) according to the manufacturer's instructions. The concentration and purity of the genomic DNA was analysed using the NanoDrop ONE[®] Spectrophotometer (Thermo Scientific). The integrity of the genomic DNA was evaluated by agarose gel electrophoresis and visualized on a 0.8% (w/v) agarose gel in Tris-Acetate-EDTA (TAE) buffer (40 mM Tris, 40 mM acetic acid, 1 mM EDTA, pH 7.4) with 0.6 µg/mL ethidium bromide (EtBr) (Merck Millipore). The genomic DNA was added to the 6X DNA loading dye (Thermo Scientific) and the MassRuler DNA Ladder Mix (Thermo Scientific) was used for amplicon size determination. Electrophoresis was performed at 90 V for 55 minutes in TAE buffer and the gel was visualized using the ChemiDoc XRS UV/Vis Gel Documentation system (Bio-Rad Laboratories, Inc.).

4.2.1.2. Sequencing of the targeted 16S rRNA of Bacteria using the Illumina MiSeq sequencer and analyses

The extracted gDNA from the 2015 and 2016 samples was used to conduct 16S rRNA metagenome sequencing at the Centre for Proteomic and Genomic Research (CPGR), Cape Town, South Africa, in order to further analyse the bacterial diversity within the metagenome. Prior to sequencing quality control was performed at the University of the Free State and at CPGR where the concentration of the DNA and purity was required to be ≥ 20 ng/µL and $A_{260/230} \geq 1.5$ respectively. The sequencing library was prepared by amplifying a ~460 bp region within the hypervariable V3/4 region of the 16S rRNA gene using region of interest-specific primers (Klindworth *et al.*, 2013) with overhang Illumina[®] adapter nucleotide sequences. The amplicons were purified using the Agencourt AMPure XP bead clean up kit (Beckman Coulter Genomics,

USA) and amplified using the Nextera XT Index kit (Illumina[®], San Diego, USA) to attach dual indices and Illumina[®] sequencing adapters. A final purification was done which was followed by quantification, normalization pooling and denaturing of the library. Subsequently the library was subjected to 2x 301 cycle sequencing using the MiSeq v3 reagent kit (Illumina[®], USA) on the Illumina[®] MiSeq.

The raw FASTQ files obtained from the 16S rRNA sequencing was analyzed using the QIIME pipeline as described by Caporaso and co-workers, (2010) with a few modifications to the protocol. Firstly, sequence quality control was assessed using PrinSeq-lite v0.20.4 where the data was trimmed to obtain an average quality score of ≥ 30 with a window of 7 nt and 3 nt step (Schmieder & Edwards 2011). All sequences shorter than 200 bp were filtered out and paired end reads merged using PEAR 0.9.6 (Zhang *et al.*, 2014). The demultiplex and quality filtering script was run on the QIIME pipeline and FASTA output files were analyzed. With the use of usearch against the RDP “Gold” database, the chimeric sequences were identified, these sequences were filtered out of the quality trimmed reads by using `identity_chimeric_seqs.py` and `filter_fasta.py` commands in QIIME (Edgar, 2010). Operational Taxonomic Units (OTUs) were clustered and taxonomy was assigned to representative OTUs using the `pick_open_reference_otus.py` script at 97% sequence similarity and this was run against the SILVA 119 database (Quast *et al.*, 2013).

4.2.2. Microbial diversity assessment using shotgun whole metagenome sequencing

4.2.2.1. Genomic DNA isolation

The genomic DNA from the 2016 sample was extracted and quantified as outlined in Section 4.2.1.1.

4.2.2.2. Illumina NextSeq 500 shotgun sequencing of the whole metagenome and analyses

The extracted gDNA from the 2016 sample was sent to the CPGR (Cape Town, South Africa) for shotgun whole metagenome sequencing. Prior to sequencing quality control was performed at the University of the Free State and at CPGR where the concentration of the DNA and purity was required to be ≥ 50 ng/uL and $A_{260/230} \geq 1.5$ respectively. The TruSeq Nano DNA Library Prep

protocol (Illumina) was used to construct libraries from the isolated DNA. To generate an insert size of 550 bp, the DNA was mechanically sheared using the M220 Focused-Ultrasonicator™ (Covaris) according to the recommendations outlined on the TruSeq Nano DNA Library Prep protocol (Illumina) which produced double stranded DNA fragments with 3' and 5' overhangs. The overhangs were converted into blunt ends followed by A-tailing. Subsequently a single indexed barcode was included by incorporating the TruSeq LT Adapter 19 as recommended by the KAPA Hyper Prep protocol (KAPA Biosystems) followed by clean-up using the Agencourt AMPure XP bead system (Beckman Coulter, USA). The library was quantified using the qPCR-based KAPA Library Quantification Kit (Illumina platform) prior to amplification using PCR at ~12 cycles. This was followed by a cleanup using the Agencourt AMPure XP bead system (Beckman Coulter, USA) which removed unligated adapter and adapter-dimer molecules. The fragments (~670 bp) were selected using the Agencourt AMPure XP system (Beckman Coulter, USA) and the library was validated using the KAPA Library Quantification Kit (Illumina® platform). The distribution of the fragment sizes was analyzed on the Agilent 2100 Bioanalyzer (Agilent Technologies, USA) using the DNA High-Sensitivity Kit (Agilent Genomics, USA). A pair-end sequencing run of the single-indexed library was performed on the Illumina NextSeq 500 System (Illumina®, USA) using the HiSeq v2 reagent kit (Illumina®, USA).

The Minia assembler was used to assemble the quality filtered reads into contigs as described by Chikhi and Rizk, (2013) with a k-mer value of 66. The optimum k-mer value was determined experimentally by running multiple assemblies with different k-mer values. The assembled contigs were uploaded to the metagenomics-RAST (MG-RAST) server for the full metagenomic diversity analysis and annotation (Meyer *et al.*, 2008). The metagenome data was analyzed using standard parameters and the results were visualized using KRONA software.

4.2.3. Preliminary identification and annotation of phage genes and prophages in the bacterial population

4.2.3.1. Identifying phage genes using VirSorter

The assembled contigs were uploaded to the automated viral signal detection pipeline VirSorter as described by Roux and co-workers, (2015). VirSorter accompanied by two reference databases (RefSeqABVir and Viromes) were used to predict viral sequences within and outside the host genome.

4.2.3.2. Annotating prophages using PHASTER

The contigs in nucleotide sequence files (FASTA format), were uploaded to the automated Phage Search Tool – Enhanced Release (PHASTER) and analyzed using standard parameters. The search tool was used to rapidly and accurately identify, annotate and graphically display prophage sequences within the bacterial genomes as described by Zhou and co-workers, (2011) and Arndt and co-workers, (2016). The National Center for Biotechnology Information (NCBI) and the prophage database were used as reference databases.

4.2.4. Binning of metagenome data and annotation of phage genes, prophages and genes related to horizontal gene transfers using RAST and Island Viewer 3

4.2.4.1 Binning of contigs using MetaBAT 2 and bin completeness and identifying contamination using CheckM

Draft genomes from the assembled contigs were predicted by metagenome binning using MetaBAT 2 (Metagenome Binning with Abundance and Tetra-nucleotide frequencies). MetaBAT 2 predicted the draft genomes and performed taxonomic assignment of each bin from the large scale metagenome data by following the protocol described by Kang and co-workers, (2015). In brief for each pair of contigs in the metagenome assembly MetaBAT 2 calculated their probabilistic distances based on tetranucleotide frequency (TNF) and abundance. The two distances were integrated into one composite distance and the pairwise distances form a matrix which is supplied to a clustering algorithm to bin contigs into genome bins. The TNF was used as sequence composition signatures as it has been shown previously that different microbial genomes have distinct TNF biases. The NCBI complete genomes were used as references to assign taxonomy to each bin by taking into consideration the TNFs of different genomes and calculating the likelihood of inter-and-intra-species Euclidean distance.

To estimate the completeness of each bin and to estimate and identify contaminating sequences CheckM was used (which is an automated method) as described by Parks and co-workers, (2015). The estimation and identification was done using marker genes that are specific to a genome's inferred lineage within a reference genome tree. The bins with high completeness, low contamination and likely to be infected by the phages found using VirSorter

and MG-RAST, had their contaminating sequences manually cleaned up from the bin's genomic sequence.

4.2.4.2 Annotation of the purified bin genomes using RAST

For the automatic annotation of the near complete and purified bin genomes the RAST pipeline was used as described by Aziz and co-workers, (2008). Briefly, the bin genomes were uploaded in FASTA format along with their taxonomic IDs. RAST analysis was performed with default settings. The called genes were annotated and given functions by comparison to the *FIGfams* (collection of protein families).

4.2.4.3 Prediction of genomic Islands/horizontal gene transfer related genes using Island Viewer 3

In order to detect the genomic islands (GIs) and horizontal gene transfer related genes within the genome of bins the web-based automated resource Island Viewer 3 was used as described by Dhillon and co-workers, (2015). GENBANK genome files of each bin were downloaded from RAST and uploaded to Island Viewer 3. Two sequence composition GI prediction methods SIGI-HMM and IslandPath-DIMOB were used along with a comparative GI prediction method IslandPick. SIGI-HMM identified codon usage bias with a hidden Markov model approach. IslandPath-DIMOB identified islands with dinucleotide bias and the presence of an associated mobility gene like integrase and transposase genes. IslandPick identified unique regions by comparing the specific bin genome against a closely related genome.

4.3. Results and discussions

4.3.1. 16S rRNA next generation sequencing of bacterial diversity assessment using the Illumina MiSeq next generation sequencer

Genomic DNA isolation was successful with a band showing high integrity and little to no shearing indicated in (Figure 4.1). The $A_{260/230}$ absorbance ratio was 1.52 indicating low protein or RNA contamination. The integrity of the genomic DNA in Lane G met the requirements

needed for Illumina MiSeq 16S rRNA sequencing as described in Section 4.2.1.2. After the quality filtering, clustering and taxonomy assignment using the QIIME pipeline, the taxonomy plots (Figure 4.2) were obtained. The phylum-level taxonomic distribution plots (Figure 4.2) show the bacterial distribution for the 2015 and 2016 samples. In both samples the Proteobacteria was the most abundant phylum (77.1 and 66.2% respectively). Bacteriodetes formed the second most dominant phylum in both samples (2015- 18.2%, 2016- 20.5%). The unassigned was the third most dominant group in both samples (2015-4.4%, 2016-2.6%) and this group of bacteria could represent the phyla that are low in frequencies within the environmental samples or have relatively smaller genomes and therefore need deeper sequencing NGS techniques such as single cell genome and shotgun whole metagenome sequencing in order to categorize them (Gawad *et al.*, 2016; Hug *et al.*, 2016). Some of the phyla like the Verrucomicrobia, Acidobacteria, Firmicutes, OD1 and Planctomycetes are only present in the 2016 data set; however a change in the microbial diversity profile is generally expected as the deep subsurface is not a closed system (Moser *et al.*, 2003; Purkamo *et al.*, 2017). Therefore, the microorganisms in the surrounding environments can influence the diversity in the deep subsurface. Microbial diversity shifts can also be influenced by chemical shifts in the environment for instance when there is abundant carbon sources, electron donors and electron acceptors the heterotrophic bacteria will be more abundant (Lippmann *et al.*, 2003; Moser *et al.*, 2003; Purkamo *et al.*, 2017). The diversity profiles at family level (Figure S4 in the Supplementary Section 4.5) show the Xanthomonadaceae family (from the class Gammaproteobacteria) to be the most abundant in the 2015 sample (40.7%) followed by the Flavobacteriaceae (from the phylum Bacteriodetes) and the Sphingomonadaceae (from the class Alphaproteobacteria) (18.1 and 6.5% respectively). In the 2016 sample the Thiotrichaceae (from the class Gammaproteobacteria) is the most abundant (32.3%) followed by the Flavobacteriaceae (13.2%) and the Erythobacteraceae (from the class Alphaproteobacteria) and Methylococcaceae (from the class Gammaproteobacteria) families both at 6.7%. The majority of the abundant families are part of the phylum Proteobacteria which was shown to be the most abundant at phylum level. The Ellin6075, Saprospiraceae, Caulobacteraceae, Verrucomicrobiaceae, Thiotrichaceae and Methylococcaceae families were only present in the 2016 sample.

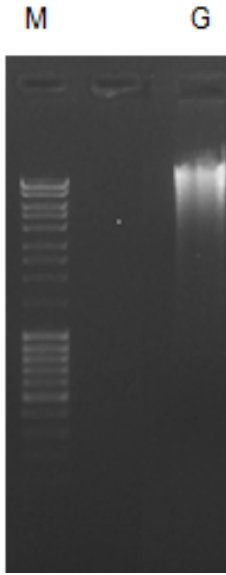


Figure 4.1: Isolated genomic for 16S rRNA next generation sequencing visualized on an ethidium bromide-stained agarose gel 0.8% (w/v): lane M; MassRuler™ DNA ladder (Thermo Scientific), lanes G: genomic DNA.

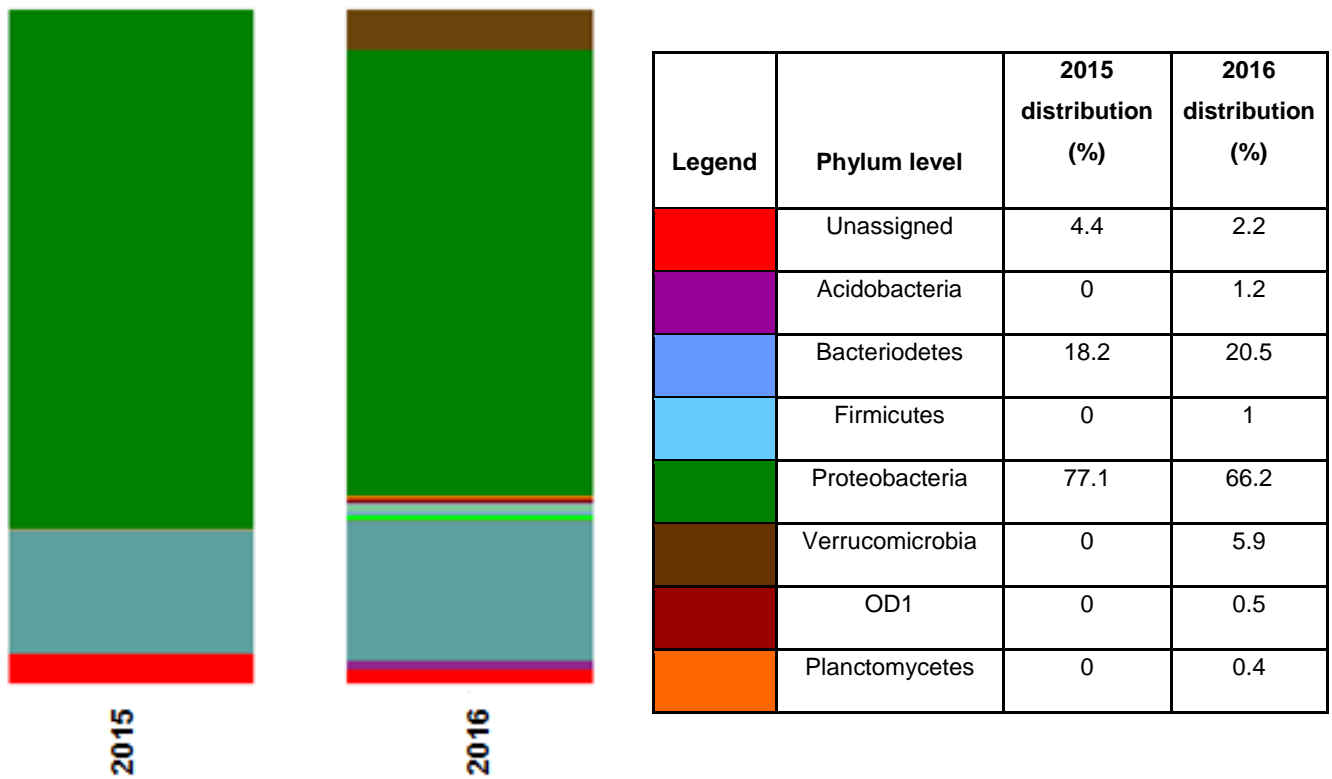


Figure 4.2: Phylum-level taxonomic distribution of the 2015 and 2016 samples respectively. The legend shows the phylum identification and also the percentage of Illumina tag composition represented by each phylum.

4.3.2. Microbial diversity studies using MG-RAST

Genomic DNA isolation was successful with a band showing high integrity and little to no shearing indicated in (Figure 4.3). The $A_{260/230}$ absorbance ratio was 1.86 indicating low protein or RNA contamination. The integrity of the genomic DNA in Lane G was slightly degraded but met the requirements needed for Illumina NextSeq shotgun whole metagenome sequencing as described in Section 4.2.2.2. Quality filtering resulted in 149 million high quality sequences. The sequences were assembled into contigs using Minia resulting in 338 783 contigs with an N50 score of 2306 meaning that more than half of the sequences were in contigs bigger than 2000 bp. The large data set presented with various difficulties related to computing power. Minia was the only assembler capable of assembling the data with the computational resources available. However, the assembly was judged good enough for further analysis. The contigs were uploaded onto the MG-RAST server for full metagenomic diversity analysis as described in Section 4.2.2.2.

Figure 4.4 shows the Krona diagram (drawn up from MG-RAST) of the microorganisms present in the metagenome. From observation all three domains in the tree of life (Archaea, Bacteria and Eukarya) are present along with Viruses. The domain Bacteria dominates making up 99% of the population. The second most dominant population is the Eukarya at 0.6% followed by Archaea at 0.4% and the lowest population is the Viruses at 0.06%. The low population percentages of Eukarya and Archaea further suggests that they are at low concentrations within the environments and further explains why it was difficult to sequence them using Sanger sequencing in Chapter 3.

The Archaeal domain (Figure 4.5) was dominated by the Euryarchaeota phylum followed by the Thermoprotei which is a class of the phylum Crenarchaeota. The diversity profile correlated with that described by Erasmus, (2015), who had previously done the same analysis on samples from the same mine and borehole. The results also corresponded with subsurface diversity profiles obtained by numerous researchers (Fry *et al.*, 2008; Kotelnikova, 2002; Newberry *et al.*, 2004; Takai *et al.*, 2001). The Bacterial domain (Figure 4.6) was dominated by the Proteobacteria (81%) phylum which consisted of mainly the classes Alpha-, Beta-, and Gammaproteobacteria. The Burkholderiales, Rhizobiales and Rhodobacterales were the most dominant within the phylum's orders. The second most abundant phylum was the Bacteroidetes (6%) followed by the Verrucomicrobia (3%), Actinobacteria (2%), Firmicutes (2%) and Planctomycetes (2%). These results correlated with those obtained in the Illumina MiSeq 16S rRNA sequencing and corresponded with those found by Erasmus, (2015). Other phyla such as

the Cyanobacteria, Acidobacteria, Nitrospirae, Spirochaetes, Chlorflexi and Deinococcus-Thermus amongst others were also present in percentages between 1-0.1 and these phyla were previously unrecognized using 16S rRNA sequencing. The bacterial diversity profile corresponded to that obtained by Erasmus, (2015) and those obtained by other researchers from South African deep mines (Chivian *et al.*, 2008; Labonté *et al.*, 2015; Lau *et al.*, 2014; Magnabosco *et al.*, 2014). The Eukaryal domain (Figure 4.7) was dominated by the Streptophyta (26%) phylum, followed by the Chordata, Ascomycota, Cnidaria, Arthropoda, Basidiomycota, Apicomplexa, Bacillariophyta and Nematoda respectively. The Nematoda were dominated by the class Chromadorea. The Eukarya diversity correlated to that of Erasmus, (2015). Two studies done by Borgonie and co-workers, (2011; 2015) were the first to detect the presence of Nematoda in the deep subsurface mines of South Africa. In both these studies 18S rRNA gene PCR and Sanger sequencing was used to study the diversity which resulted in limited diversity profiles compared to the analysis done in this study which sequenced the whole metagenome giving a more invasive diversity profile for the Eukarya specifically the Nematoda.

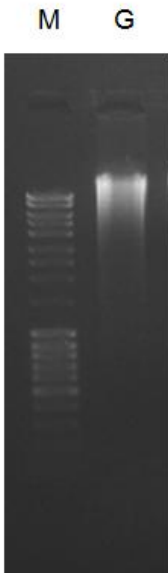


Figure 4.3: Isolated genomic for shotgun whole metagenome sequencing visualized on an ethidium bromide-stained agarose gel 0.8% (w/v): lane M; MassRuler™ DNA ladder (Thermo Scientific), lanes G: genomic DNA.

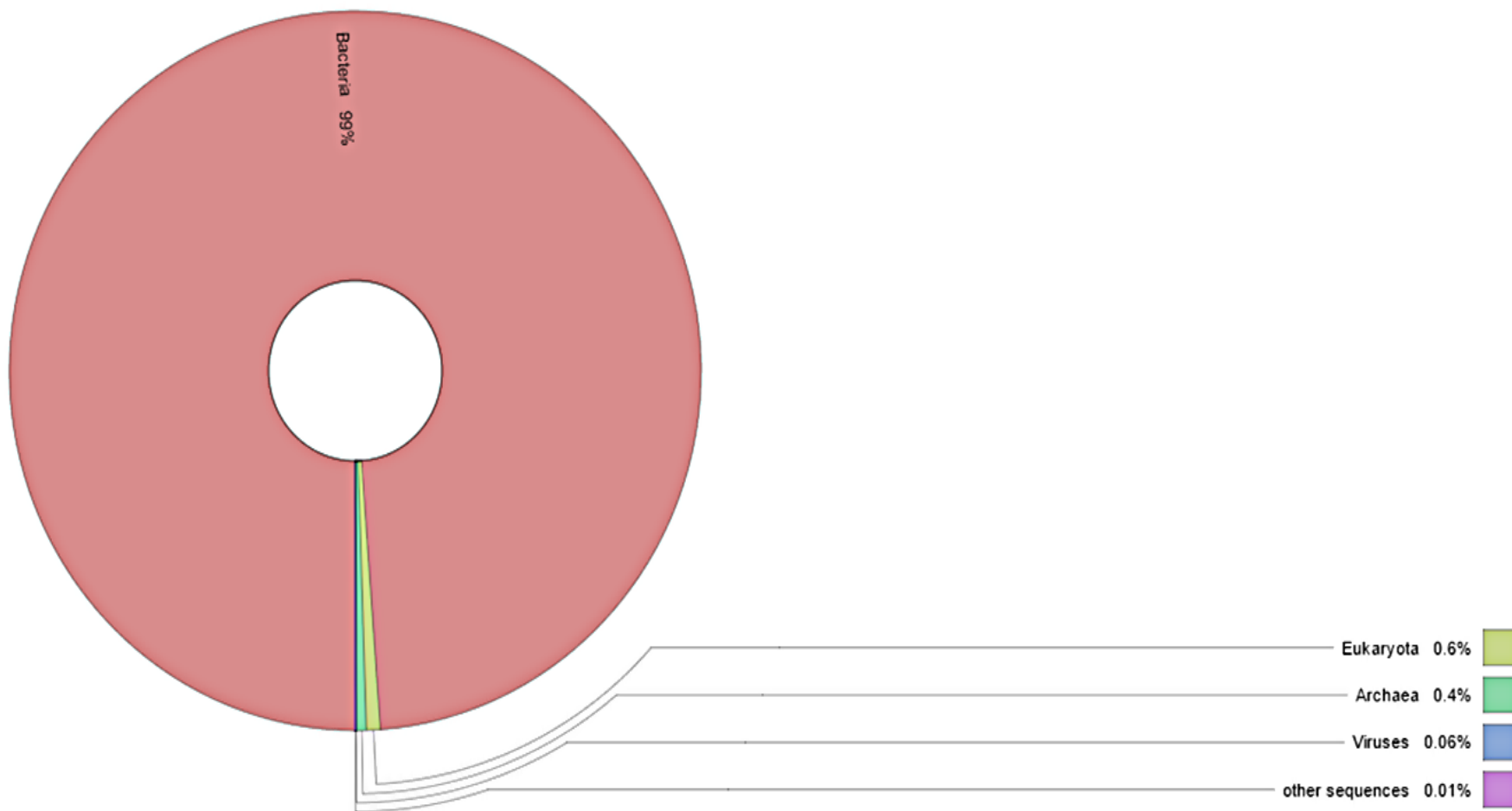


Figure 4.4: Krona diagram of the Metagenome microorganisms present from MG-RAST.

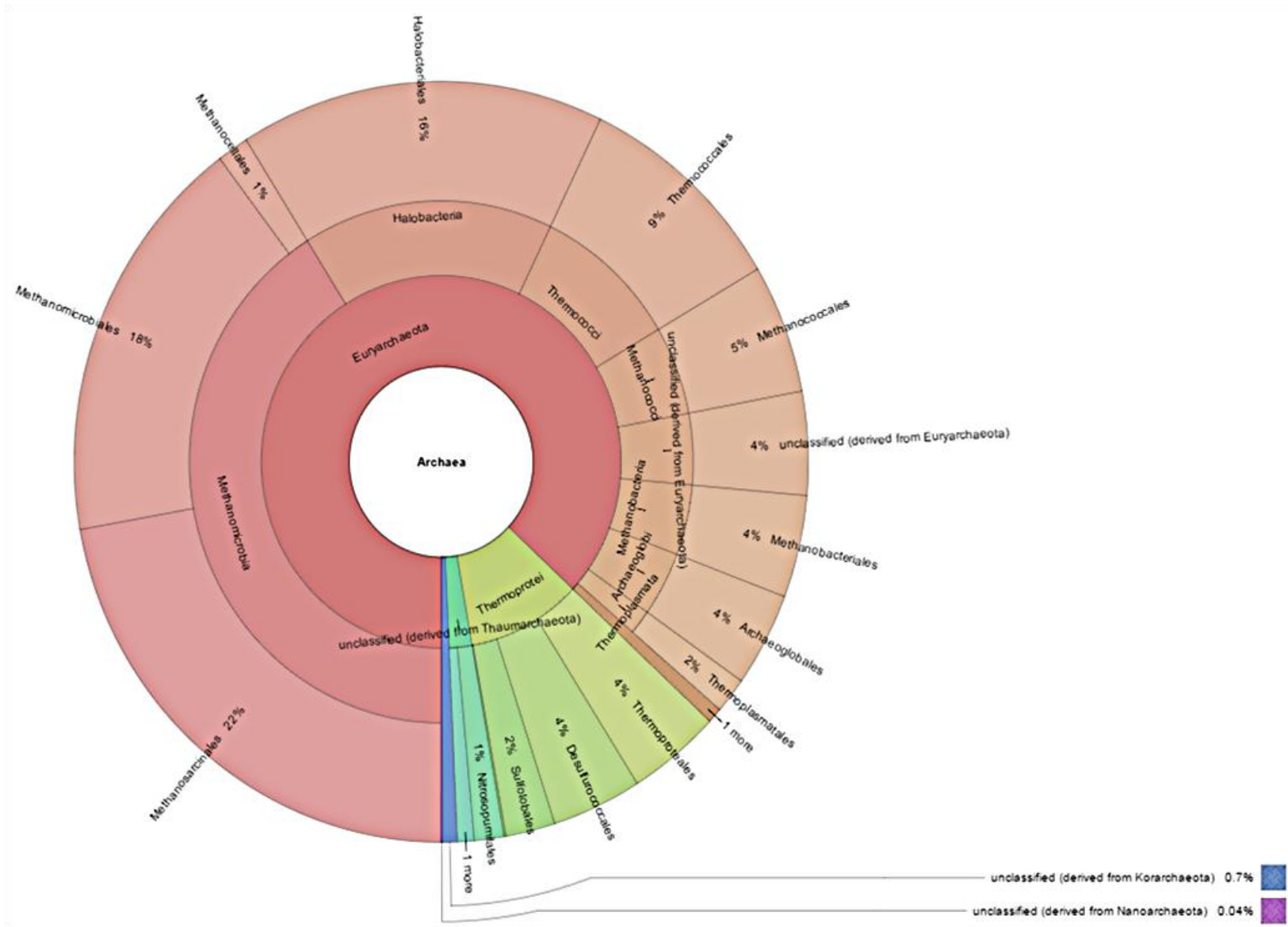


Figure 4.5: Krona diagram of the Archaea at order level from MG-RAST.

Microbial diversity identification of phage genes and horizontal gene transfers

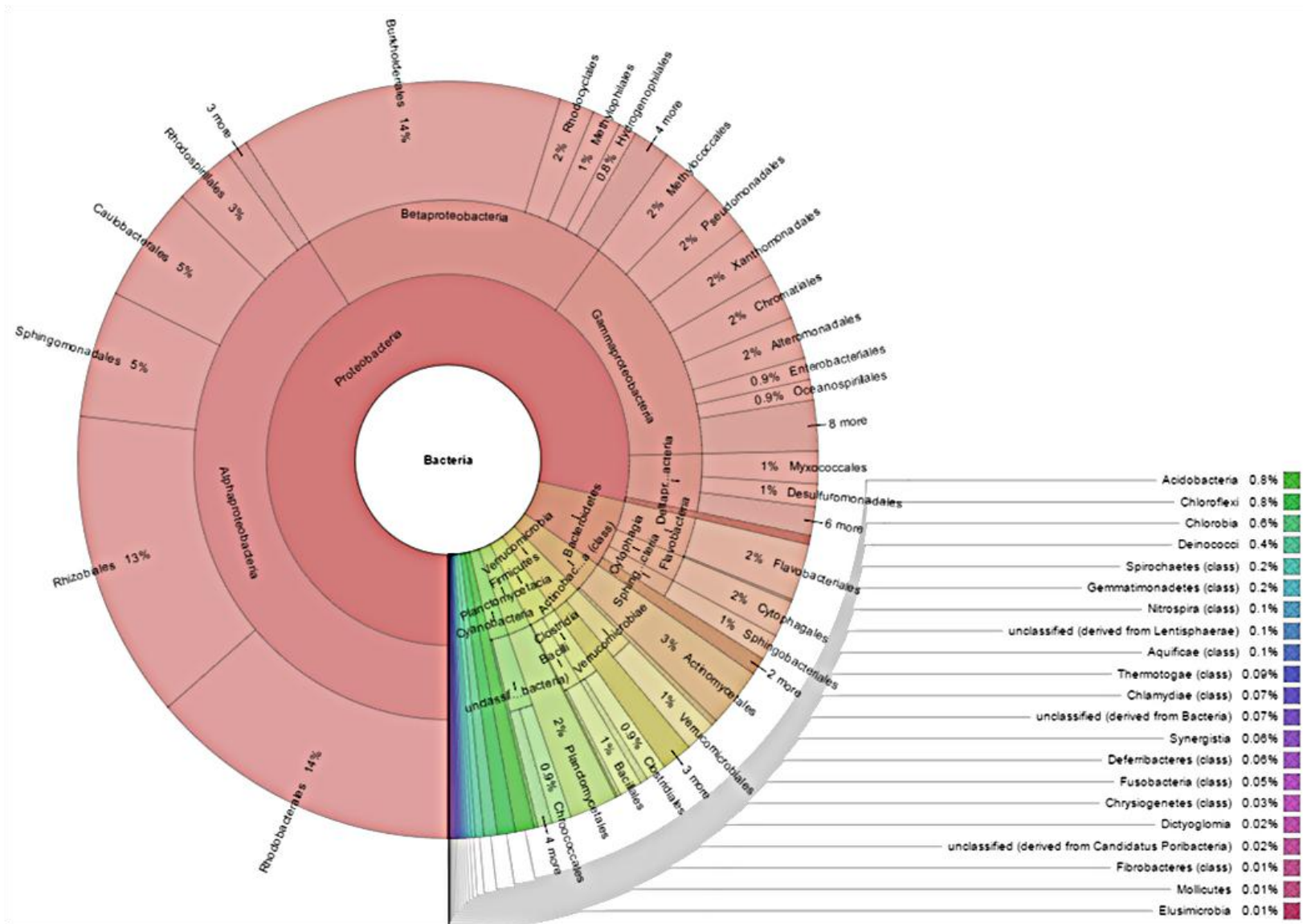


Figure 4.6: Krona diagram of the Bacteria at order level from MG-RAST.

Microbial diversity identification of phage genes and horizontal gene transfers

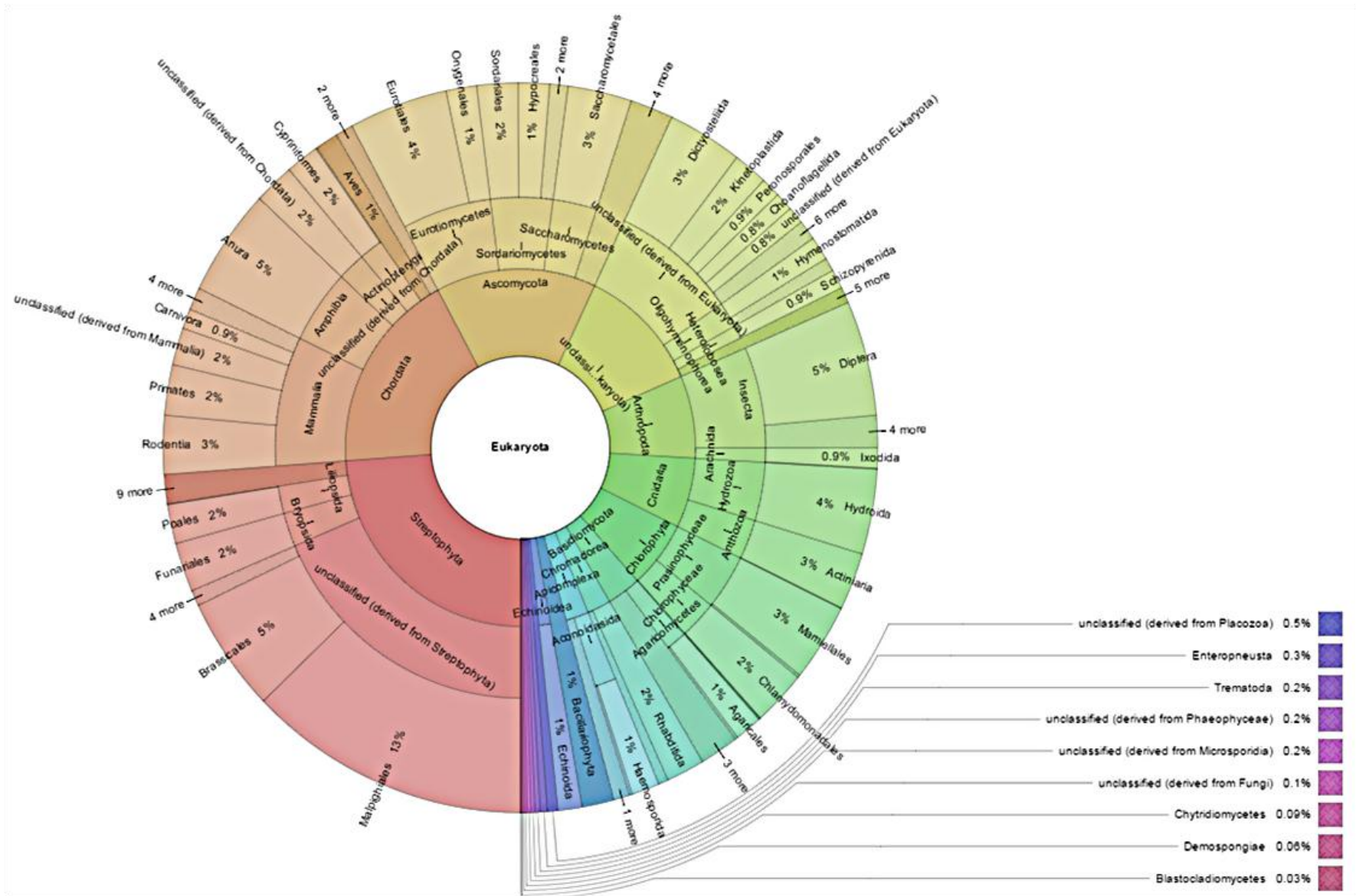


Figure 4.7: Krona diagram of the Eukarya at order level from MG-RAST.

The viruses detected in the metagenome belonged to the Caudovirales order (Figure 4.8). The *Siphoviridae* (46%) family was the most abundant followed by the *Myoviridae* (31%), *Podoviridae* (18%) and the unclassified virus families from the same order (4%). The presence of the *Myoviridae* and *Podoviridae* in the metagenome data sequencing further suggested that the phages characterized using TEM in Chapter 3 belonged to these families. The unclassified virus included the following families *Inoviridae*, *Iridoviridae*, *Microviridae* and *Phycodnaviridae*. The virus diversity profile resembled that of Dell'Anno and co-workers, (2015) where metagenomic analysis was done on DNA pools in the benthic deep-sea ecosystem.

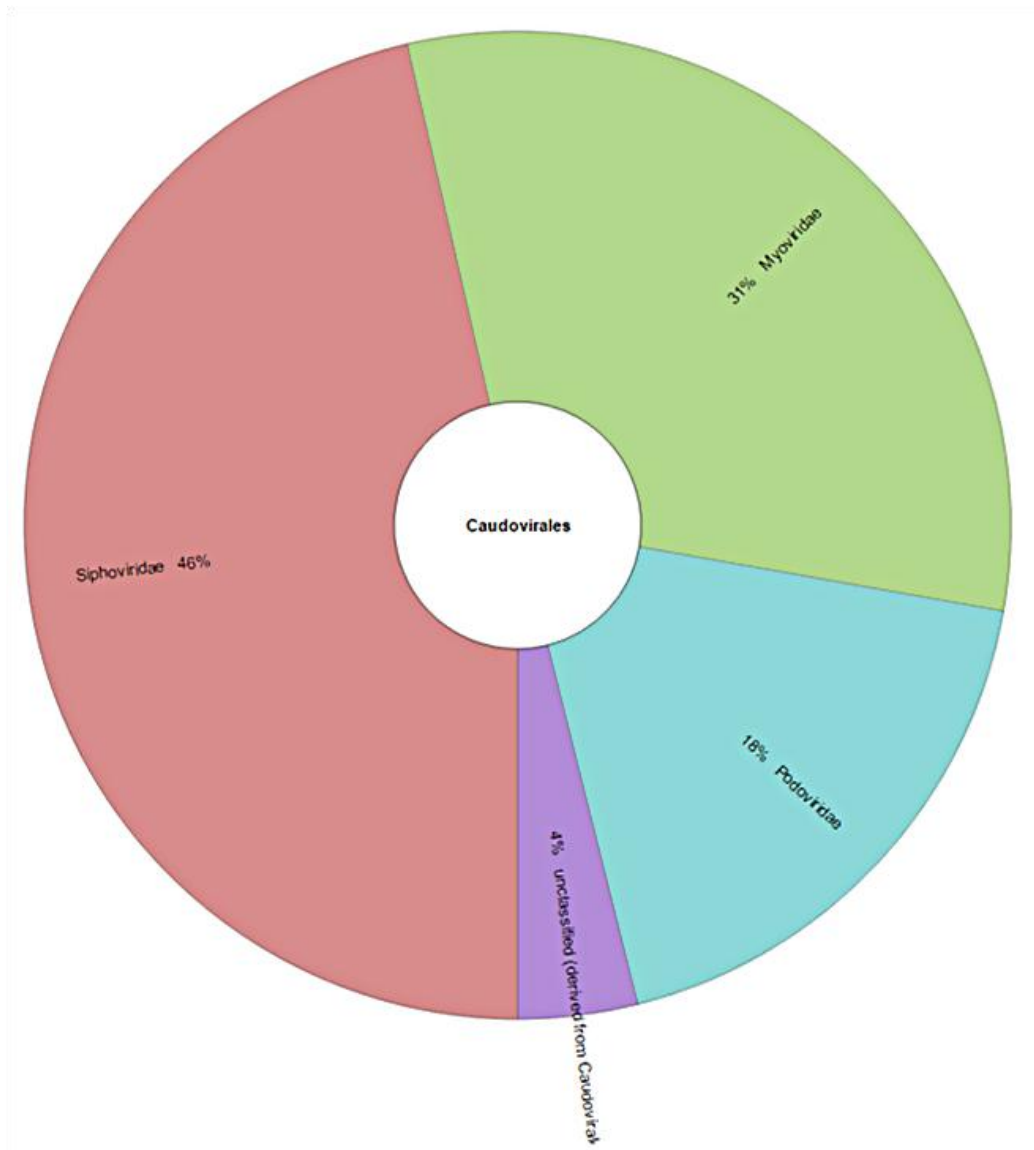


Figure 4.8: Krona diagram of the Caudovirales found using MG-RAST

4.3.3. Identification and annotation of phage genes and prophages in the bacterial population

The assembled contigs were uploaded to the automated viral signal detection pipeline VirSorter as described in Section 4.2.3.1. VirSorter outperforms most tools for fragmented metagenomic datasets as it identifies signals in assembled contigs as short as 3 kb and provides near perfect identification as it has >95% recall and 100% precision on contigs ≥ 10 kb (Roux *et al.*, 2015). VirSorter was able to identify approximately 122 viral sequences, which were grouped into three categories. Category 1 had 24 viral sequences while category 2 had 44 and category 3 had 54. Category 1 (most confident predictions) are regions with viral hallmark genes and viral enrichment genes, category 2 (likely predictions) are regions with either of the two types of genes and category 3 (possible predictions) are regions which lack both type of genes but have matrices such as enrichment in viral-like genes (Roux *et al.*, 2015). The phage genes shown in Table 4.3.3 are those from mostly category 1 and a few from category 2. These phage genes were detected along with the phage they originate from and the host which they are found in. Table 4.3.3 only has the phage genes of bacterial hosts that have been specifically identified at genus level in the metagenome data uploaded on MG-RAST. The majority of the phage hosts were from the phylum Proteobacteria which has been found to be the most dominant phyla in this environment and this also further suggests proof to the phenomenon that viruses prey on the most dominant host population in a given environment. Some of the genes detected were the integrase, transposase and gene transfer agents phage genes which are suggested to facilitate in horizontal gene transfers and are also marker genes for detecting possible regions that have undergone horizontal gene transfer (Juhas *et al.*, 2009).

The presence of prophages within the contigs was detected using the automated search tool PHASTER as described in Section 4.2.3.2. Only 6 of the annotated prophages were near complete with regards to annotation and are shown in Figure 4.9. The phage proteins annotated included hypothetical proteins, terminases, protease, portal proteins, fiber protein, plate protein, coat proteins, integrase, tail shaft and phage-like proteins. Only one integrase was found and this enzyme is used for the integration of phage genome to the host genome during infection. The presence of only one prophage with an integrase does not mean that the rest of the prophages do not possess the enzyme as mentioned above these prophages are incomplete. There was no detection of tRNAs which are target sites for prophage integration into the host genome and this is not alarming as the annotation is incomplete and phages do not always use tRNAs as target sites (Zhou *et al.*, 2011). Prophage may be dormant phages, but they can have

genes that are beneficial to the host for example allow the bacterial host to exist in new and extreme environments (Zhou *et al.*, 2011). Prophages in South African deep mines have been previously identified by Mabizela, (2009) and Labonté and co-workers, (2015).

Table 4.3.3: Phage genes from category 1 and 2 detected from the metadata contigs using VirSorter.

Phage gene	Phage	Host
Reverse transcriptase	phi3	<i>Pseudomonas</i>
DEAD box helicase	vB_PaeKakheti25	<i>Pseudomonas</i>
Portal protein	F10	<i>Pseudomonas</i>
Putative exonuclease	vB_PaeP_Tr60_Ab31	<i>Pseudomonas</i>
Transposase	MP42	<i>Pseudomonas</i>
Terminase large subunit	vB_AbaS_TRS1	<i>Acinetobacter</i>
Putative endolysin	phiAC-1	<i>Acinetobacter</i>
Terminase large subunit	11b	<i>Flavobacterium</i>
DNA modification methylase	P8625	<i>Prostheco bacter</i>
Integrase	KS9	<i>Burkholderiales</i>
Gene transfer agent	RDJL Phi 1	<i>Roseobacter</i>
addiction module antitoxin/toxin	BiPBO1	<i>Brucella</i>
Integrase	Cd	<i>Azospirillum</i>
ISBt3 transposase subunit protein	phiE125	<i>Burkholderia</i>
putative transposase OrfB protein of IS629	BP-4795	<i>Enterobacteria</i>
Transposase A	BalMu-1	<i>Bacillus</i>
DNA transposition protein B	vB_MhM_3927AP2	<i>Mannheimia</i>

Microbial diversity identification of phage genes and horizontal gene transfers

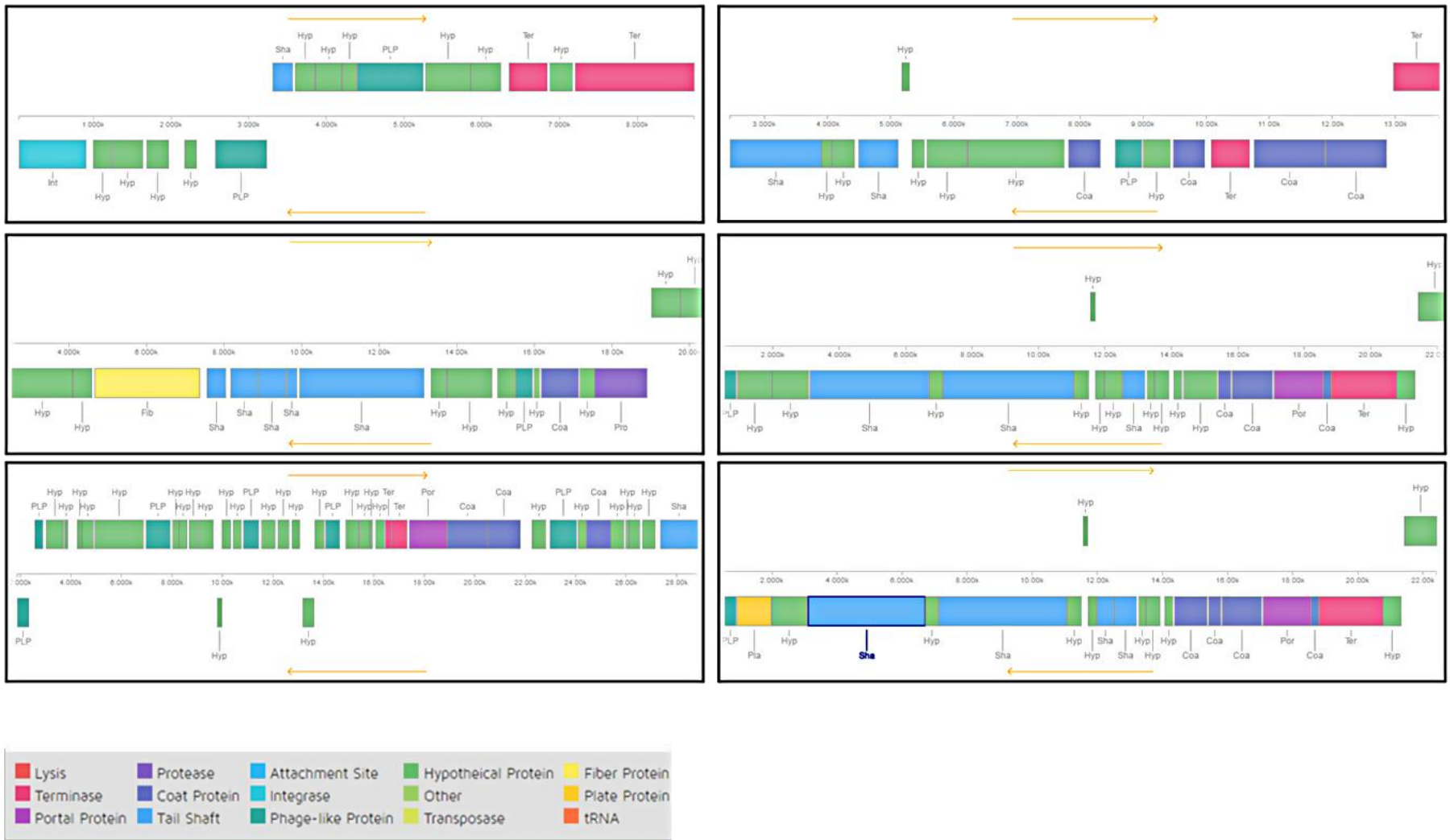


Figure 4.9: Incomplete prophage annotations of 6 prophages from the metagenome data contigs using PHASTER.

4.3.4. Annotation of phage genes and horizontal gene transfer related genes within the binned genomes using RAST and Island Viewer 3

Due to the large and complex metagenome data, contigs had to be binned in order to perform phage-host interactions on the individual organisms. The contigs were binned using MetaBAT 2 and checked for contamination and completeness as described in Section 4.2.4.1. The bins in Table 4.3.4.1 were chosen due to their high completeness, low contamination and likelihood to be infected by the phages detected using VirSorter and MG-RAST. These bins were manually cleaned up to remove the contaminating sequences which were detected by CheckM. Most of the bins had relatively high completeness (between 98-82%) and low contamination (0-3%) of which is expected as MetaBAT 2 outperforms other alternative binning methods in producing accurate high quality genome bins on very large assemblies consisting of millions of contigs (Kang *et al.*, 2015). The bins were mostly from the phylum Proteobacteria followed by the Bacterioidetes, Acidobacteria, Planctomycetes and Verrucomicrobia. These bin taxonomies at phylum level and at family level (with regards to the Comamonadaceae, Flavobacteriaceae, Xanthomonadaceae and Saprospiraceae families) correlated with those obtained in the diversity studies in Sections 4.3.1. and 4.3.2.

The purified bins were annotated using the RAST pipeline as described in Section 4.2.4.2. RAST was able to annotate the subsystem features of each bin genome and Table 4.3.4.2 consists of only the bins which had phage associated subsystems as part of their annotation (10 out of the 16 bins). The phage associated subsystem features identified the specific phage genes and their functionalities which are shown Table 4.3.4.2. Some of the phage associated features detected such as integrons, transposable elements, pathogenicity islands and phage introns are marker genes used to detect regions that have undergone horizontal gene transfers and some of these genes are products of horizontal gene transfers (Juhas *et al.*, 2009). Gene transfer agents were also detected which are just as important if not more than the above mention features, as they are small phage-like elements capable of packaging and transferring host DNA during the mechanism of horizontal gene transfer (Lang *et al.*, 2012). Transposable elements and integrases have been previously detected by researchers in the deep South African mines (Chivian *et al.*, 2008; Labonté *et al.*, 2015; Mabizela, 2009).

Table 4.3.4.1: Binned genomes from the metagenome contigs showing contamination and completeness percentages and taxonomy.

Bin	Completeness %	Contamination %	Taxonomy	Taxonomy level	Taxonomy ID
11	85.12	2.27	Saprospiraceae	Family	261514
12	93.58	3.18	Methylococcus	Genus	236501
14	82.28	13.86	Comamonadaceae	Family	165437
18	97.01	0.22	Acidobacteria	Phylum	194844
21	95.32	1.99	Comamonadaceae	Family	165437
22	89.21	9.63	Rhodobacteraceae	Family	157276
23	98.84	1.45	Flavobacteriaceae	Family	165436
26	88.3	0	Flavobacteriaceae	Family	165436
27	93.9	1.64	Xanthomonadaceae	Family	211441
28	93.52	0.93	Bacteria	Kingdom	2
32	63.09	1.3	Gammaproteobacteria	Class	1236
36	57.74	1	Burkholderiales	Order	208544
37	56.56	0.14	Sphingomonadales	Order	293405
40	53.32	0.12	Betaproteobacteria	Class	271085
42	90.54	0	Opitutales	Order	592834
43	52.01	0	Planctomycetaceae	Family	100233

Table 4.3.4.2: Bins annotation of phage associated subsystem features and the functions and specific phage genes identified.

Bin	Phage associated Subsystem features	Functions and specific phage genes found
11	Phage tail fiber proteins	Role in primary attachment of virion to host receptor
	Phage lysis module	Subsystem of phage lysis proteins
	Listeria pathogenicity island LIPI-1 extended	Genomic islands acquired by horizontal gene transfers that carry virulence genes.
12	Phage tail fiber proteins	Role in primary attachment of virion to host receptor
	Phage packaging machinery	Responsible for packing phage material: Phage terminase large subunit and phage portal protein
	Phage replication	DNA polymerase III alpha subunit and DNA helicase
	Transposable elements	Sequences moving within genome: TniB NTP-binding protein, TniA putative transposase, FIGfam050825, ComM-related protein, MG(2+) Chelatase family protein, FIGfam110555, segregation and condensation protein B
14	Integrans	Cassette-based systems upstream of the integrase gene promoting the collection of antibiotics into genomic regions that are associated with plasmids or transposable elements.
	Phage capsid protein	Structural protein: Phage major capsid protein
18	Phage packaging machinery	Phage terminase large subunit and phage portal protein
21	Integron	Cassette-based systems upstream of the integrase gene promoting the collection of antibiotics into genomic regions that are associated with plasmids or transposable elements.
	Phage replication	DNA polymerase III alpha subunit and DNA helicase
22	Phage packaging machinery	Phage portal protein
	Phage capsid protein	Phage major capsid protein
	Gene transfer agent	Small phage-like elements capable of packaging and transferring host DNA: terminase protein, prohead protease, FAD/FMN-containing dehydrogenase, NlpC/P60 family peptidase and host specific protein
23	Phage introns	Phage intervening sequences detected by host HNH homing endonuclease which confers mobility to introns within host genome
26	Phage introns	Phage intervening sequences detected by host HNH homing endonuclease which confers mobility to introns within host genome
28	Phage packaging	Responsible for assembly of packing machinery: Phage terminase large subunit
37	Phage capsid protein	Phage major capsid protein

In order to further understand the impact HGTs have on the host fitness and the biogeochemical cycles within the deep subsurface mines, Island Viewer 3 was used to upload and detect genomic islands in the bin genomes as described in Section 4.2.4.3. Genomic islands are clusters of genes that are suggested to have been acquired through HGTs (Dhillon *et al.*, 2015). They are of interest since they encode genes involved in medically and environmentally important adaptations, including antimicrobial resistance and virulence (Dhillon *et al.*, 2015). Therefore, the genes detected using Island Viewer 3 should theoretically have been acquired from horizontal origin.

The aligned bin genomes along with the genes detected using by Island Viewer 3 are shown in Figures 4.10- 4.20. The blue and orange blocks on the circular genome views represent the prediction methods used to predict specific genes within the GIs where blue represents IslandPath-DIMOB and orange SIGI-HMM (Dhillon *et al.*, 2015). The red blocks represent the GI genes that have been predicted by one or more of the three prediction methods within the integrated circular view (Dhillon *et al.*, 2015). Mobile elements were detected in bins 11, 12, 14, 21 and 27. Mobile elements/transposons have forged mutually beneficial relationships with genes that are associated with them which include the genes that encode for the following enzymes transposases, integrases and recombinases (Anderson *et al.*, 2013). The presence of such enzymes in a genome suggests that the organisms in that environment actively exchange genes (Anderson *et al.*, 2013). In bins 11 and 12 CRISPR-associated RAMP Cmr5 and Cmr6 (Figures 4.10 and 4.11) were detected. The presence of CRISPRs serves as identification of potential hosts for phages as they are immune systems used by archaea and bacteria to combat interfering genetic material from viruses and plasmids (Rath *et al.*, 2015). The CRISPR-associated (*cas*) genes are normally found adjacent to the CRISPR and these genes are required for the CRISPR activity (Rath *et al.*, 2015). The CRISPR-associated RAMP proteins cleave the RNA molecule of the interfering genetic material (Bikard & Marraffini, 2013). The detection of retron-type RNA-directed DNA polymerase in bins 11, 12 and 14 reveals the possible presence of retrons. Retrongs encode for reverse transcriptase and are thought to play a role in the evolution and fitness of bacteria (Labonté *et al.*, 2015). Labonté and co-workers, (2015) were the first to discover retrongs in the deep subsurface environments. Transposase were detected in bins 12, 14 and 23. These are enzymes that bind to the end of transposons and catalyze their mobility by cutting the transposon and integrating it into a new site on the genome. Transposase are best known for their role in HGT of antibiotic resistant gene and therefore play a role in the evolution and fitness of the host in harsh environments (Labonté *et*

al., 2015). As previously mentioned studies in the deep subsurface mines of South Africa have detected and predicted CRISPRs, mobile/transposable elements, transposase and retrons to be markers of possible phage-mediated HGTs (Chivian *et al.*, 2008; Labonté *et al.*, 2015; Mabizela, 2009).

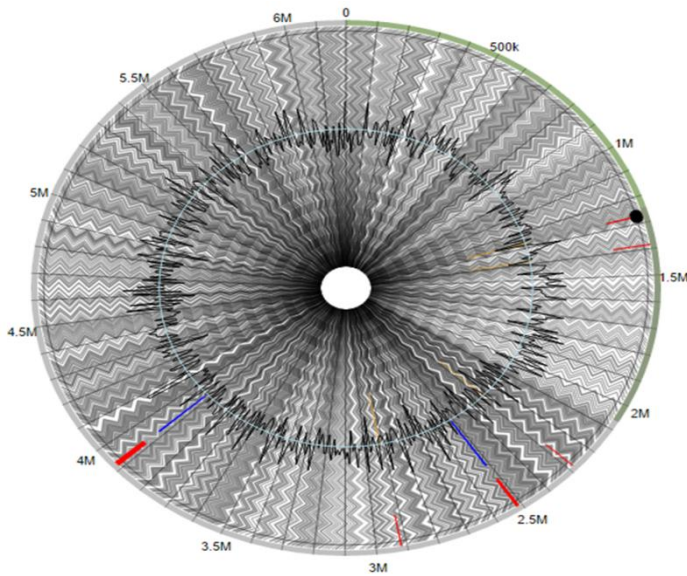
Some of the genes within the genomic islands are part of metabolic pathways and therefore confer novel capabilities for survivability to the hosts that they are transduced into via HGT. The gene that encodes for magnesium chelatase family protein/ComM related protein (Figure 4.10) was detected in bin 11. Magnesium chelatase is the insertion of Mg^{2+} into protoporphyrin and is exclusively found in bacteria that synthesize bacteriochlorophyll (phylum Cyanobacteria) (Papenbrock *et al.*, 2000). The presence of such genes in Bin 11 (family Saprospiraceae, phylum Bacterioidetes) further suggests that it was a product of HGT. Nitrogen regulatory protein P-II gene detected in bins 11 and 12 (Figures 4.10 and 4.11) is involved in the control of nitrogen metabolism in bacteria (Huergo *et al.*, 2013). The P-II protein is involved in the survivability of the bacteria in a given environment as it increases the nitrogen metabolism for nutrient uptake in a nitrogen starved environment (Huergo *et al.*, 2013). It can also decrease nitrogen metabolism when the nitrogen levels are too high as high levels of nitrogen can lead to the toxic death of bacterial cells (Huergo *et al.*, 2013). Bin 14 (Figure 4.12) had a mercury reductase gene amongst two other genes related to the cellular resistance to mercury. The reduction of mercury by bacteria is important in the biogeochemical cycling of extreme environments (Freedman *et al.*, 2012; Moller *et al.*, 2014). Bacteria are able to reduce mercury by using the mercuric reductase enzyme and this results in the detoxification of the environment (Moller *et al.*, 2014). The ability to reduce mercury has allowed such microbes to survive in mercury exposed environments (Freedman *et al.*, 2012). The cobalt-zinc-cadmium resistance gene found in bins 12 and 18 (Figure 4.13) encodes for the protein CzcA which is part of the cobalt, zinc and cadmium protein complex (CzcABC) mediates the resistance to all three metals in a given environment (Nies, 1995). All three metals have been detected in the water of the Star Diamonds mine (Erasmus, 2015) making the gene important in the biogeochemical cycling of the deep subsurface environment and is involved in bacterial host fitness. The gene for the sulfur oxidation protein SoxX detected in bin 22 (family Rhodobacteraceae) is found in many sulfur-oxidizing bacteria where it is part of the Sox pathway which oxidizes the sulfur containing molecule thiosulfate (Grabarczyk & Berks, 2017). The Rhodobacteraceae family is not considered sulfur oxidizers, but the data in this study suggests that some of the species in this

family have acquired the genes through HGT to oxidize sulfur as possible energy to survive in the deep subsurface.

Bin 14 had a genetic island with the universal stress gene (UspA) which is expressed in defense against superoxide-generating agents. Their expression is elevated in response to stress conditions such as starvation for carbon, nitrogen, phosphate, sulfate and amino acids and also due to exposure to heat, oxidant, metals, cycloserine, ethanol and antibiotics (Nachin *et al.*, 2005). Proteins related to exopolyphosphates like the exopolyphosphate-related protein gene in bin 27 (Figure 4.18) are involved in the regulation of stress responses and virulence (Malde *et al.*, 2014). The lack of these protein genes results in defects in motility, biofilm formation and nutrient survival during stress conditions (Malde *et al.*, 2014). Bins 12, 14, 21, 27 and 36 had ABC and multidrug transporter related genes detected in their genomic islands. ABC transporters are important for host fitness and evolvability as they take part in the uptake of nutrients or secretion of toxins (Moussatova *et al.*, 2008). The ABC-type multidrug transporters specifically confer multidrug resistance in bacterial cells by pumping diverse anti-cancer drugs and antibiotics into the extracellular spaces (Moussatova *et al.*, 2008). The Type IV pilin PilA are genes in bins 14 and 27 are involved in biofilm formation, bacterial motility and attachment and also serve in the transfer of DNA and virulence (Gorgel *et al.*, 2015). Bin 18 had the gene encoding for the Hsp20 protein. This gene is expressed in stress conditions (normally due to heat) and facilitate the folding of proteins during these conditions (Maaroufi & Tanguay, 2013). In this case the gene could be a result of HGT or could be specifically from a cyanophage that had infected the host as phages have been found to contain Hsp20 proteins. As previously mentioned phages can also confer their own genes on the host that can improve the host fitness (Maaroufi & Tanguay, 2013).

Some of these host survival genes within the genomic islands have been previously observed by Chivian and co-workers, (2008) in the Mponeng gold mine as previously mentioned. These genes were part of the nitrogen fixation, cobalamin synthesis and sulfide reduction pathways and motility and sporulation. The genes were believed to be products of HGT events and this was validated by the presence of transposase and CRISPRs. The presence of transposase and CRISPRs has been observed in the present study a well

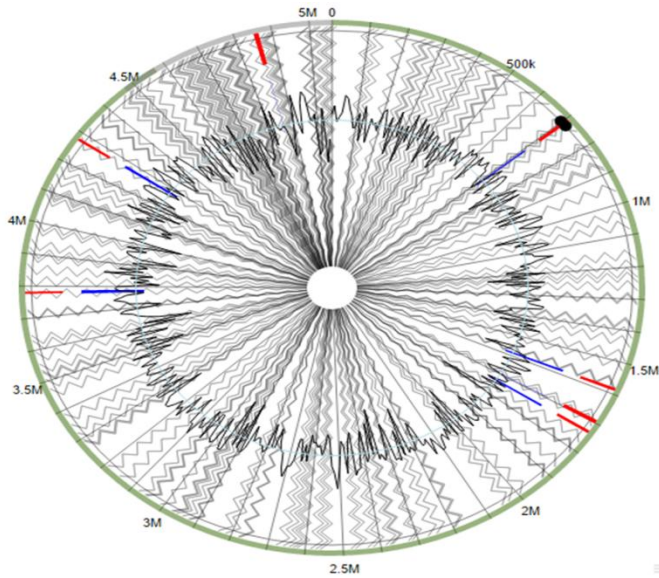
Microbial diversity identification of phage genes and horizontal gene transfers



Prediction method	Product
SIGI-HMM/IslandPath	27 hypothetical proteins
SIGI-HMM	Type I restriction-modification system, specificity subunit S
	Molybdopterin oxidoreductase subunit, predicted; chaperone protein HtpG
	Two component, sigma54 specific, transcriptional regulator, Fis family
IslandPath-DIMOB	Probable sugar transporter, motif=sugar transport protein signatures
	Glycosyl hydrolase, family 16
	Mobile element protein
	MG(2+) Chelatase family protein/ ComM-related protein
	Invasion plasmid antigen
	Phosphoribosylformylglycinamide synthase, glutamine
	CRISPR-associated RAMP Cmr5 and Cmr6, CRISPR-associated protein Cas 1
	Negative transcriptional regulator, DNA-binding response regulator
Retron-type RNA-directed DNA polymerase	

Figure 4.10: Circular genome view of bin 11 aligned with a reference genome. The colour blocks represent the prediction methods used where blue represents IslandPath-DIMOB and orange SIGI-HMM. The table represents the genomic island product genes along with their prediction methods.

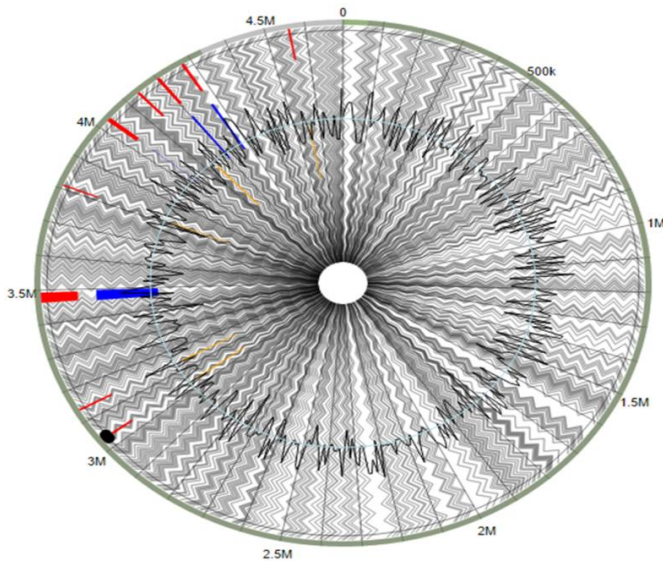
Microbial diversity identification of phage genes and horizontal gene transfers



Prediction method	Product
IslandPath-DIMOB	29 Hypothetical proteins
	3 Mobile elements, Tn5044 transposase, transposase,
	CRISPR-associated RAMP, Csx10 family
	Cation-transporting ATPase, Phosphate transport system regulatory protein PhoU
	S-adenosylmethionine synthetase
	Calcium/proton antiporter
	4-hydroxybenzoyl-CoA thioesterase family active site
	Tol biopolymer transport system, TolR protein, TolA protein, TolB protein precursor periplasmic protein
	Queusine Biosynthesis QueE Radical SAM, Queosine Biosynthesis QueC ATPase
	Nitrogen regulatory protein P-II
	Cobalt-zinc-cadium resistance protein CzcA; Cation efflux system protein CusA
	ABC transporter, ATP-binding/permease protein
	RelE/StbE replicon stabilization toxin
	Hydroxymethylpyrimidine phosphate synthase and a Methyl-accepting chemotaxis protein
	NADH:flavin oxidoreductase, Old Yellow Enzyme family and flavin reductase domain protein, FMN binding
	Molybdeum cofactor biosynthesis protein MoaC
	Dihydrofolate reductase
Retron-type RNA-directed DNA polymerase	

Figure 4.11: Circular genome view of bin 12 aligned with a reference genome. Table represents the genomic island product genes and prediction methods.

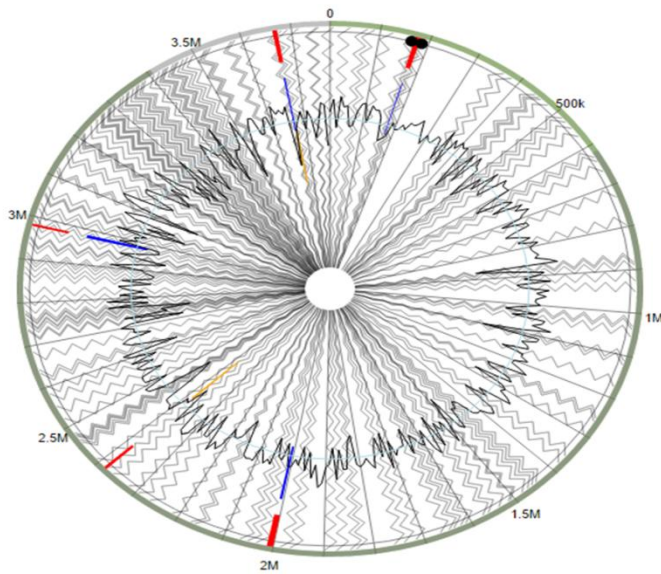
Microbial diversity identification of phage genes and horizontal gene transfers



Prediction method	Product
SIGI-HMM/IslandPath-DIMOB	12 Hypothetical proteins
SIGI-HMM	3-hydroxyacyl-CoA dehydrogenase, Benzyl-CoA oxygenase components A and B
SIGI-HMM/ IslandPath-DIMOB	UspA, Death on curing protein, Doc toxin
SIGI-HMM	Methyl-accepting chemotaxis protein I
SIGI-HMM	Phosphoenolpyruvate carboxykinase [GTP]
IslandPath-DIMOB	5 Type IV fimbrial biogenesis protein FimT, 2 Type IV pilus biogenesis protein Pile
	Malonyl CoA-acyl carrier protein transacylase
	Putative cytochrome P450 hydroxylase and a putative cytochrome P450
	Mobile element, transposase
	Retron-type RNA-directed DNA polymerase
SIGI-HMM	Ribonucleotide reductase transcriptional regulator NrdR and a transcriptional regulator, TetR family
	Phenylacetic acid degradation protein paal
	Quinone oxidoreductase
SIGI-HMM/IslandPath-DIMOB	2 Diguanylate cyclase/phosphodiesterase (GGDEF and EAL domains) with PAS/PAC sensors
IslandPath-DIMOB	ATPase component NikO of energizing module of nickel ECF transporter
	Putative metal chaperone, involved in Zn homeostasis, GTPase of COG0523 family
	ABC transporter permease protein, ATP-binding protein and periplasmic oligopeptide-binding protein OppA
ISIGI-HMM	Alkaline phosphatase
IslandPath-DIMOB	Mercuric reductase, Mercuric transport protein and periplasmic mercury (2+) binding protein
	Low-specific L-threonine aldolase
	Histidine utilization repressor

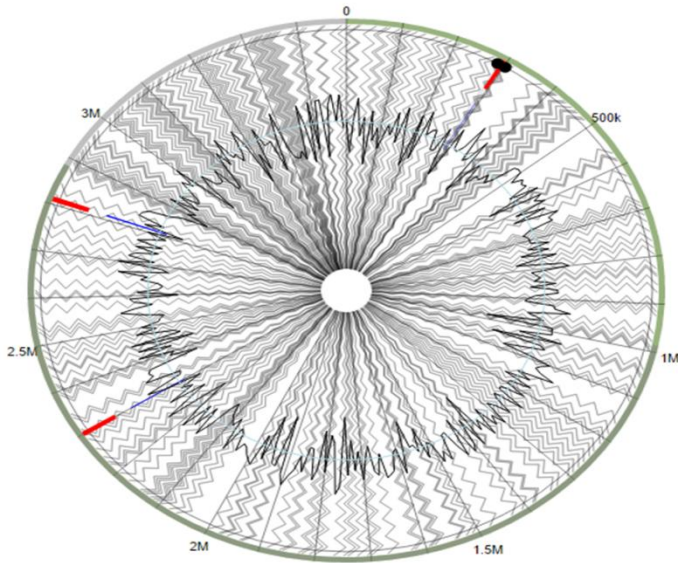
Figure 4.12: Circular genome view of bin 14 aligned with a reference genome. Table represents the genomic island product genes and prediction methods.

Microbial diversity identification of phage genes and horizontal gene transfers



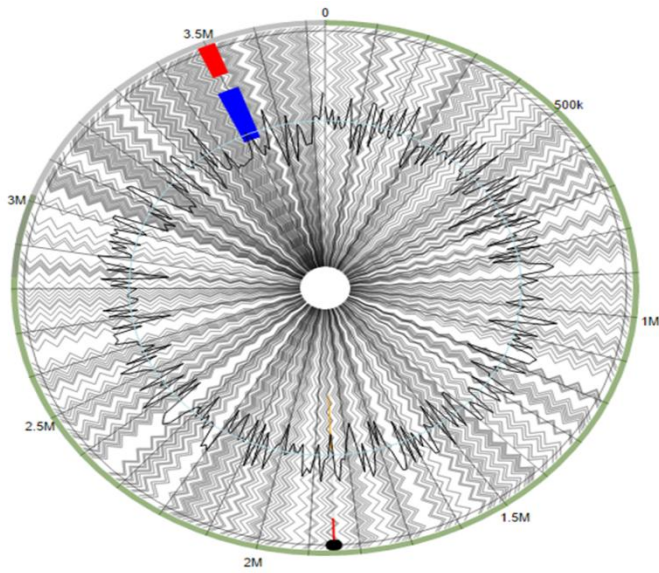
Prediction method	Product
IslandPath-DIMOB/SIGI-HMM	37 Hypothetical protein, 1 Hypothetical antitoxin
IslandPath-DIMOB	Homogentisate 1,2-dioxygenase
	Rel/StbE replicon stabilization toxin
	Osmotically inducible protein OsmC
	cAMP-binding proteins – catabolite gene activator and regulatory subunit of cAMP-dependent protein k
	2 Cytochrome d ubiquinol oxidase subunit
	Heat shock protein, Hsp20 family
	Ferredoxin, 2Fe-2S
	ClpB protein
	Lipoprotein-related protein
Integrase	
IslandPath-DIMOB/SIGI-HMM	2 Cobalt-zinc-cadium resistance
SIGI-HMM	2 Probable conserved integral membrane protein
IslandPath-DIMOB/SIGI-HMM	Nitrogen regulatory protein P-II

Figure 4.13: Circular genome view of bin 18 aligned with a reference genome. Table represents the genomic island product genes and prediction methods.



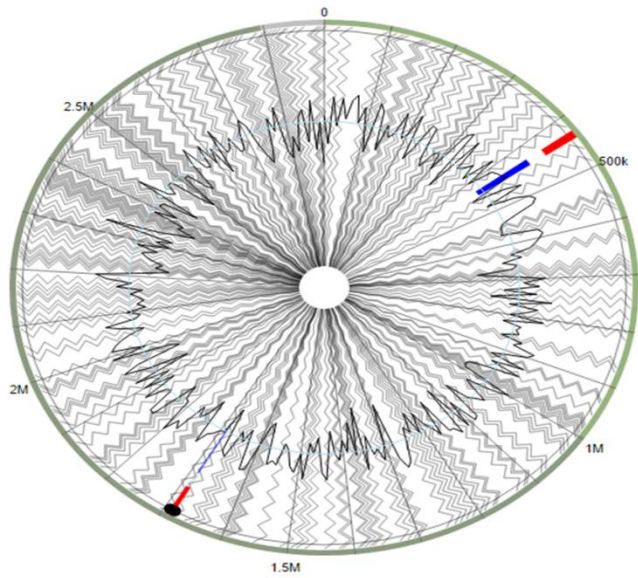
Prediction method	Product
IslandPath-DIMOB	VapB protein, VapC toxin protein
	8 Hypothetical proteins
	Putative periplasmic protein kinase ArgK and related GTPases of G3E family
	Orotate phosphoribosyltransferase
	Aspartyl-tRNA (Asn) amidotransferase subunit B at Glutamyl-tRNA (G1) amidotransferase
	Xanthine and CO dehydrogenases maturation factor, XdhC/CoxF family
	Mobile element protein
	ABC-type antimicrobial peptide transport system, permease component
	Peptide methionine sulfoxide reductase
	Shufflon-specific DNA recombinase
	tRNA pseudourine synthase A
	Long-chain-fatty-acid-CoA ligase
	Transcriptional regulator, CRP/Fnr family

Figure 4.14: Circular genome view of bin 21 aligned with a reference genome. Table represents the genomic island product genes and prediction methods.



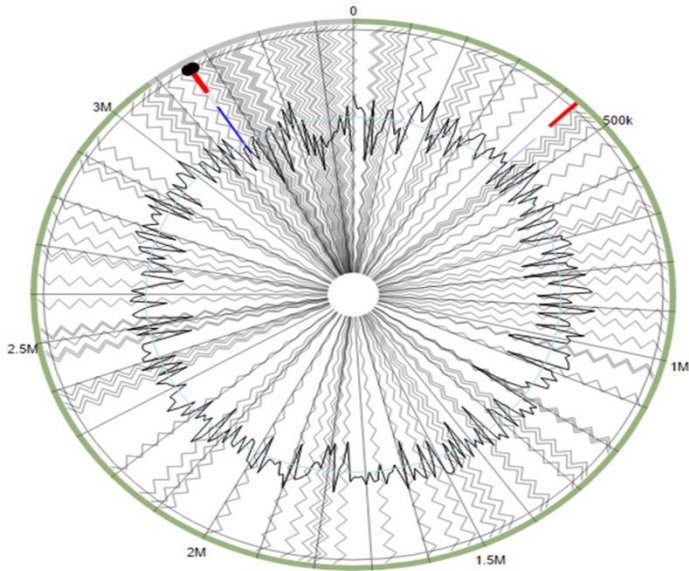
Prediction method	Product
SIGI-HMM	7 Hypothetical proteins
SIGI-HMM	Oxidoreductase, FAD/FMN-binding
IslandPath-DIMOB	Peptidase B
	LigA
	Molybdenum cofactor biosynthesis protein MoaB
	Soluble lytic murein transglycosylase precursor
	Uracil-DNA glycosylase, family 4
	Probable ploy (beta-D-mannuronate) O-acetylase
	Long-chain-fatty-acid-CoA ligase
	Probable tpr domain protein
	Diaminopimelate decarboxylase
	Segregation and condensation protein B
	Putative integrase/recombinase
	RepA
	Adenylate cyclase
	Putative lipoprotein
	RNA 5-methylaminomethyl-2-thiouridine synthase TusA
Sulfur oxidation protein SoxX	

Figure 4.15: Circular genome view of bin 22 aligned with a reference genome. Table represents the genomic island product genes and prediction methods.



Prediction method	Product
IslandPath-DIMOB	11 Hypothetical proteins
	Putative outer membrane protein involved in nutrient binding
	23S rRNA (guanosine-2-O-)-methyltransferase rimB
	Probable transposase
	LSU ribosomal proteins L4p and L3p
	SSU ribosomal proteins S10p, S7p and S12p
	Fic family protein
	Filamentation induced cAMP protein Fic
	HNH homing endonuclease
	Imidazolonepropionase

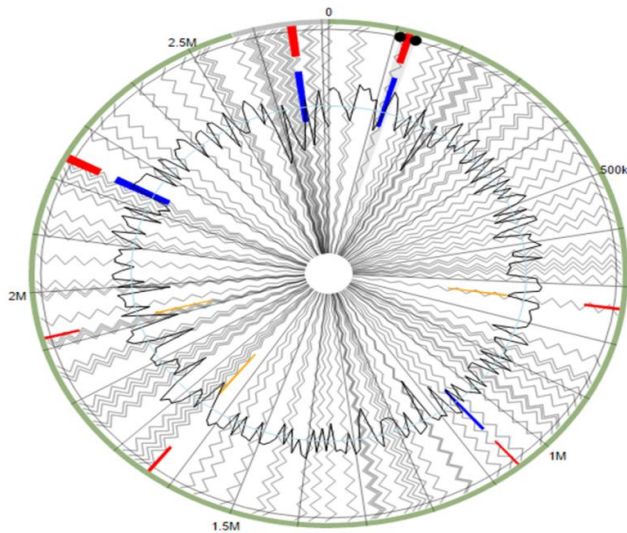
Figure 4.16: Circular genome view of bin 23 aligned with a reference genome. Table represents the genomic island product genes and prediction methods.



Prediction method	Product
IslandPath-DIMOB	8 Hypothetical proteins
	Butyryl-CoA dehydrogenase
	SSU ribosomal protein S21p
	Integrase, site-specific recombinase
	Ribosome hibernation protein YhbH
	LSU ribosomal proteins L11p, L1p, L10p and L7/L12
	Preprotein translocase subunit SecE
	Trasnscription antitermination protein NusG
	Translation elongation factor Tu
	Beta-lactamase class C and other penicillin binding proteins
	Beta-hexosaminidase
	DNA helicase

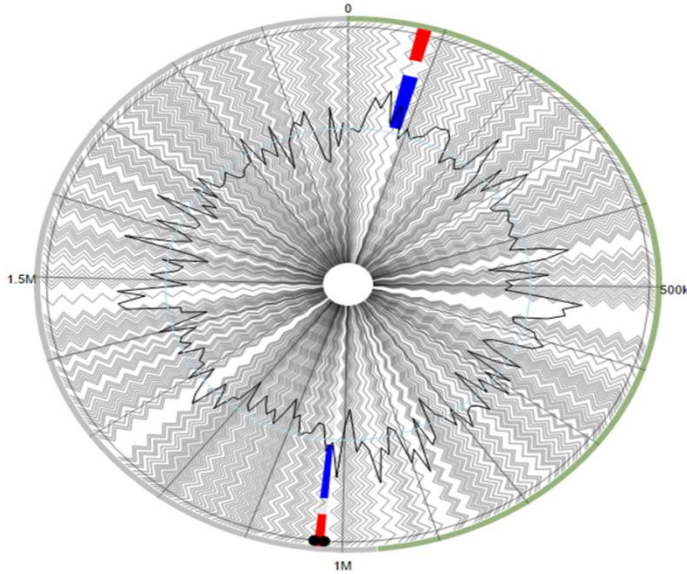
Figure 4.17: Circular genome view of bin 26 aligned with a reference genome. Table represents the genomic island product genes and prediction methods.

Microbial diversity identification of phage genes and horizontal gene transfers



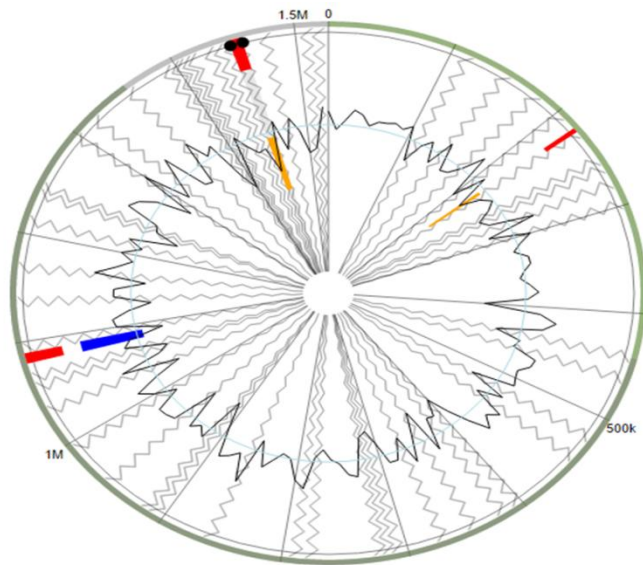
Prediction method	Product
Island Path-DIMOB/SIGI-HMM	16 Hypothetical proteins
IslandPath-DIMOB	Very-short-patch mismatch repair endonuclease (G-T specific)
	DNA-cytosine methyltransferase
	HipA protein
	Small GTP-binding protein domain
SIGI-HMM	4 Mobile element
SIGI-HMM	Aspartyl/Asparaginyl beta-hydroxylase
SIGI-HMM	TPR repeat, a TPR repeat-containing protein and a TPR domain/sulfotransferase domain protein
IslandPath-DIMOB	2 Type I restriction-modification system specificity subunit
	FIG188645: PAP/25A core domain: DNA polymerase, beta-like protein
	Exopolyphosphate-related protein
SIGI-HMM	Type IV pilin PilA
	4-amino-4-deoxy-L-arabinose transferase and related glycosyltransferases of PMT family
	Polysaccharide biosynthesis protein
	Related to dolichyl-phosphate mannose synthase
IslandPath-DIMOB	Ribosomal small subunit pseudouridine and a Ribosomal RNA small subunit methyltransferase C
	Gamma-glutamyl-petrescine oxidase
	Response regulator
	GII1812 protein
	ABC-type phosphate/phosphonate transport system periplasmic component like protein
	Predicted transcriptional regulator containing HTH domain and an uncharacterized domain

Figure 4.18: Circular genome view of bin 27 aligned with a reference genome. Table represents the genomic island product genes and prediction methods.



Prediction method	Product
IslandPath-DIMOB	Hypothetical protein
	2-amino-3-carboxymuconate 6-semialdehyde decarboxylase
	Integrase
	Error-prone, lesion bypass DNA polymerase V (UmuC) and error-prone repair protein (UmuD)
	ATP-dependent protease La Type I
	Transcriptional regulator, TetR family
	Lipid IVA 3-deoxy-D-manno-octulosonic acid transferase
	Periplasmic divalent cation tolerance protein cutA
	ABC-type multidrug transport system permease component
	ABC transporter, ATP-binding protein
	Preprotein translocase secY subunit
	DNA-directed RNA polymerase alpha subunit
	15 LSU ribosomal proteins
	9 SSU ribosomal proteins

Figure 4.19: Circular genome view of bin 36 aligned with a reference genome. Table represents the genomic island product genes and prediction methods.



Prediction method	Product
SIGI-HMM/IslandPath-DIMOB	14x Hypothetical proteins
IslandPath-DIMOB	Putative autotransporter protein
	Shikimate kinase
	Oxygen-insensitive NAD(P)K nitroreductase/ Dihydropteridine reductase
	Flagellin domain protein

Figure 4.20: Circular genome view of bin 42 aligned with a reference genome. Table represents the genomic island product genes and prediction methods.

4.4. Conclusions

The MG-RAST analysis of the metagenome data revealed the domain Bacteria to be the most dominant of the population followed by Eukarya and Archaea. The low population percentages of Eukarya and Archaea suggested that due to their low abundance their DNA fractions are also low making it difficult to sequence and analyze them using Sanger sequencing in Chapter 3. The Bacterial domain diversity profiles for the Illumina MiSeq 16S rRNA sequencing and the Illumina NextSeq shotgun whole metagenomic sequencing correlated with Proteobacteria being the most dominant phylum and this corresponded with the data from previous subsurface diversity studies (Chivian *et al.*, 2008; Erasmus, 2015; Labonté *et al.*, 2015; Lau *et al.*, 2014; Magnabosco *et al.*, 2014). As shown in many studies prior (Erasmus, 2015; Fry *et al.*, 2008; Kotelnikova, 2002; Newberry *et al.*, 2004; Takai *et al.*, 2001) the Euryarchaeota phylum followed by the Crenarchaeota dominated the Archaeal domain. The Streptophyta phylum dominated the Eukarya diversity profile and the Nematoda were dominated by the class Chromadorea. Compared to the work of Borgonie and co-workers, (2011; 2015) (who were the first to identify Nematoda in deep subsurface mines of South Africa) where 18S rRNA gene PCR and Sanger sequencing was used, the diversity profile for Eukarya in this current study was broader due to the use of whole metagenome sequencing which sequences deeper due to its sensitivity. All the diversity profiles corresponded with subsurface diversity profiles obtained by numerous researchers.

The viruses detected in the metagenome belonged to the Caudovirales order and the *Siphoviridae* family was most abundant followed by the *Myoviridae*, *Podoviridae* and the unclassified viral families from the same order. The presence of the *Myoviridae* and *Podoviridae* further suggested that the phages characterized using TEM in Chapter 3 belonged to these families. The virus diversity profile resembled that of Dell'Anno and co-workers, (2015) where metagenomic analysis was done on DNA pools in the benthic deep-sea ecosystem.

VirSorter was able to identify approximately 122 viral sequences. The phage sequences of phages infecting bacterial hosts (mostly from the phylum Proteobacteria) that had been specifically identified at genus level in the metagenome data uploaded on MG-RAST were picked. PHASTER was used to detect the 6 partially complete prophages and the phage proteins annotated included hypothetical proteins, terminases, protease, portal proteins, fiber protein, plate protein, coat proteins, integrase, tail shaft and phage-like proteins. Prophages in

South African deep mines have been previously identified by Mabizela, (2009) and Labonté and co-workers, (2015).

The binned contigs were chosen due to their high completeness, low contamination and likelihood to be infected by the phages detected using VirSorter and MG-RAST. The bins were mostly from the phylum Proteobacteria followed by the Bacteroidetes, Acidobacteria, Planctomycetes and Verrucomicrobia. RAST was able to detect phage associated subsystems in 10 of the 16 bins. Some of the phage associated features detected such as integrons, transposable elements, pathogenicity islands, phage introns and gene transfer agents are related to horizontal gene transfers. Most of the horizontal gene transfer related genes detected using VirSorter and RAST have been previously detected by researchers in the deep South African mines (Chivian *et al.*, 2008; Labonté *et al.*, 2015).

In order to assess genes acquired by HGT events in the bins, Island Viewer 3 was used to predict genes within genomic islands. The genes detected in this study are suggested to be markers for possible phage-mediated HGT events by previous studies (Chivian *et al.*, 2008; Labonté *et al.*, 2015; Mabizela, 2009) were the CRISPRs, mobile/transposable elements, transposase and retrons. Some of the genomic island genes encoded for magnesium chelatase family protein/ComM related protein; nitrogen regulatory protein P-II, mercury reductase, cobalt-zinc-cadmium resistance, sulfur oxidation protein SoxX, universal stress, exopolyphosphate-related protein, ABC and multidrug transporter related genes, Type IV pilin PilA and Hsp20 proteins. These genes conferred novel capabilities to the host (that they were transduced into via HGTs) for survivability and evolution in the extreme deep subsurface environments. The genes were part of the nitrogen fixation, cobalamin synthesis and sulfide reduction pathways and motility and sporulation and have been previously observed in the deep subsurface (Chivian *et al.*, 2008). Validation that the genes were products of HGT events was observed by the presence of transposase and CRISPRs.

It is important to note that the objectives of both studies by Chivian and co-workers, (2008) and Labonté and co-workers, (2015) was to assemble the bacterial populations into single genomes and analyze their genome contents. The discovery of phage related genes was by chance and not the objective. This therefore highlights the novelty of the current study as it is the first to specifically look at the bacteria-phage interactions within the South African mines and it is also the first to find viral infections within the Star Diamonds mine.

4.5. Supplement A

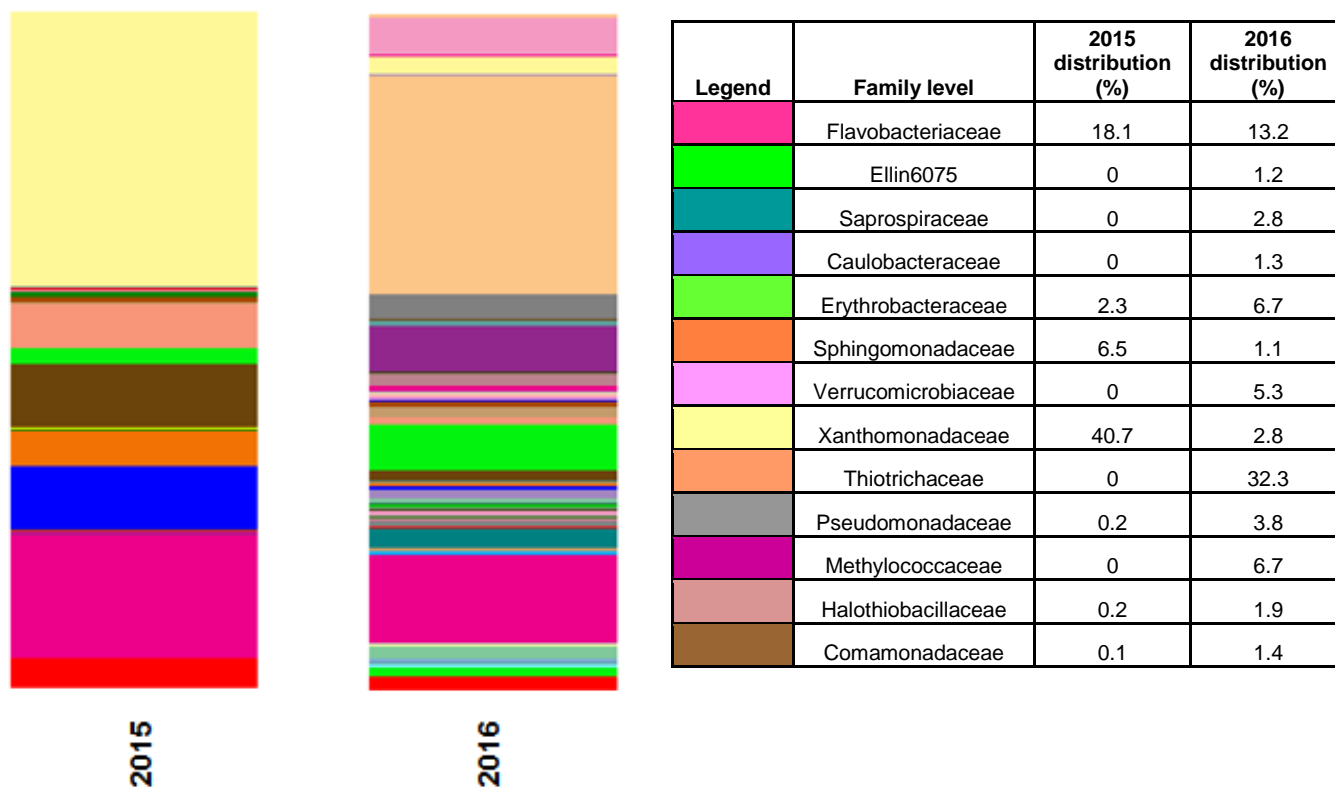


Figure S4: Family-level taxonomic distribution of the 2015 and 2016 samples respectively. The legend shows the family identification and also the percentage of Illumina tag composition represented by each family.

4.6. References

Anantharaman, K., Duhaime, M. B., Breier, J. A., Wendt, K. A., Toner, B. M. and Dick, G. J. (2014). Sulfur oxidation genes in diverse deep-sea viruses. *Science*. **344**: 757–760.

Anderson, R. E., Brazelton, W. J. and Baross, J. A. (2011). Is the genetic landscape of the deep subsurface biosphere affected by viruses? *Frontiers in Microbiology*. **2**: 1–16.

Anderson, R. E., Brazelton, W. J. and Baross, J. A. (2013). The deep virosphere: assessing the viral impact on microbial community dynamics in the deep subsurface. *Reviews in Mineral Geochemistry*. **75**: 649–675.

Angly, F. E., Felts, B., Breitbart, M., Salamon, P., Edwards, R. A., Carlson, C., Chan, A. M., Haynes, M., Kelley, S. Liu, H., Mahaffy, J. M., Mueller, J. E., Nulton, J., Olson, R., Parson,

R., Rayhawk, S., Suttle, C. A. and Rohwer, F. (2006). The marine viromes of four oceanic regions. *PLoS Biology*. **4(11)**: e368–378.

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y. and Wishart, D. S. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Research*. **44**: W16–W21.

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M. Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parello, B., Pusch, G. D., Stevens, R., Vasieva, O., Vonstein, V., Wilke, A. and Zagnitko, O. (2008). The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics*. **9(75)**: 1–15.

Bikard, D. and Marraffini, L. A. (2013). Control of gene expression by CRISPR-Cas systems. *F1000Prime Reports*. **5**, P5-47.

Blainey, P. C. (2014). The future is now: single-cell genomics of bacteria and archaea. *FEMS Microbial Review*. **37**: 1574–6976.

Borgonie, G., Linage-Alvarez, B., Ojo, A., Mundle, S., Freese, L., Van Rooyen, C., Kuloyo, O., Albertyn, J., Pohl, C., Cason, E. D., Vermeulen, J., Pienaar, C., Litthauer, D., van Niekerk, H., van Eeden, J., Sherwood Lollar, B., Onstott, T. C. and van Heerden, E. (2015). Eukaryotic opportunists dominate the deep-subsurface biosphere in South Africa. *Nature Communications*. **6**: 1–12.

Borgonie, G., García-Moyano, A., Litthauer, D., Bert, W., Bester, A., van Heerden, E., Möller, C., Erasmus, M. and Onstott, T. C. (2011). Nematoda from the terrestrial deep subsurface of South Africa. *Nature*. **474**: 79–82.

Breitbart, M. and Rohwer, F. (2005). Here a virus , there a virus , everywhere the same virus ? *TRENDS in Microbiology*. **13(6)**: 276–284.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pena A. G., Goodrich, J. K., Gordon J. I., Huttley, G. A., Kelley, S.T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C.A., McDonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J., Yatsuneko, T.,

Zaneveld, J. and Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*. **7**: 335–336

Caporaso, J. G., Lauber, C. L., Walters, W. a, Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S. M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J. A., Smith, G. and Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *International Society of Microbial Ecology Journal*. **6**: 1621–1624.

Chikhi, R. and Rizk, G. (2013). Space-efficient and exact de Bruijn graph representation based on a Bloom filter. *Algorithms for Molecular Biology*. **8(22)**: 1–9.

Chivian, D., Brodie, E. L., Alm, E. J., Culley, D. E., Dehal, P. S., Desantis, T. Z., Gihring, T. M., Lapidus, A., Lin, L., Lowry, S. R., Moser, D. P., Richardson, P. M., Southam, G., Wanger, G., Pratt, L. M., Andersen, G. L., Hazen, T. C., Brockman, F. J., Arkin, A. P. and Onstott, T.C. (2008). Environmental genomics reveals a single-species ecosystem deep within earth. *Science Reports*. **322**: 275–278.

Danovaro, R. (2000). Viral density and virus-to-bacterium ratio in deep-sea sediments of the Eastern Mediterranean. *Applied and Environmental Microbiology*. **66(5)**: 1857–1861.

Danovaro, R., Dell’Anno, A., Corinaldesi, C., Magagnini, M., Noble, R., Tamburini, C. and Weinbauer, M. (2008). Major viral impact on the functioning of benthic deep-sea ecosystems. *Nature*. **454**: 1084–1088.

Dell’Anno, A., Corinaldesi, C. and Danovaro, R. (2015). Virus decomposition provides an important contribution to benthic deep-sea ecosystem functioning. *PNAS*. E2014–E2019.

Dhillon, B. K., Laird, M. R., Shay, J. A., Winsor, G. L., Lo, R., Nizam, F., Pereira, S. K., Waglechner, N., McArthur, A. G., Langille, M. G. I. and Brinkman, F. S. L. (2015). IslandViewer 3: More flexible, interactive genomic island discovery, visualization and analysis. *Nucleic Acids Research*. **43**: W104–W108.

Duarte, S., Cássio, F and Pascoal, C. (2012). Denaturing gradient gel electrophoresis (DGGE) in microbial ecology – Insights from freshwaters. Chapter 3 in *Gel electrophoresis – Principles and Basics*.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST, *Bioinformatics* **26(19)**: 2460-2461

Erasmus, M. (2015). *Geological Carbon Sequestration and its Influence on Subsurface Microbial Diversity and Metabolic Carbon Cycling*. (PhD thesis). University of the Free State.

Freedman, Z., Zhu, C. and Barkay, T. (2012). Mercury resistance and mercuric reductase activities and expression among chemotrophic thermophilic aquificae. *Applied and Environmental Microbiology*. **78**: 6568–6575.

Fry, J. C., Parkes, R. J., Cragg, B. A., Weightman, A. J. and Webster, G. (2008). Prokaryotic biodiversity and activity in the deep seafloor biosphere. *FEMS Microbiology Ecology*. **66**: 181–196.

Gawad, C., Koh, W. and Quake, S. R. (2016). Single-cell genome sequencing : current state of the science. *Nature Reviews*. **17**: 175–188.

Gorgel, M., Ulstrup, J. J., Bøggild, A., Jones, N. C., Hoffmann, S. V, Nissen, P. and Boesen, T. (2015). High-resolution structure of a type IV pilin from the metal-reducing bacterium *Shewanella oneidensis*. *BMC Structural Biology*. **15(4)**: 1–17.

Grabarczyk, D. B. and Berks, B. C. (2017). Intermediates in the Sox sulfur oxidation pathway are bound to a sulfane conjugate of the carrier protein SoxYZ. *PLoS One* **12**, e0173395.

Hambly, E. and Suttle, C. A. (2005). The virosphere , diversity , and genetic exchange within phage communities. *Current Opinion Microbiology*. **8**: 444–450.

Huergo, L. F., Chandra, G. and Merrick, M. (2013). PII signal transduction proteins: Nitrogen regulation and beyond. *FEMS Microbiology Reviews* **37**: 251–283.

Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., Butterfield, C. N., Hermsdorf, A. W., Amano, Y., Ise, K., Suzuki, Y., Dudek, N., Reiman, D. A., Finstad, K. M., Amundson, R., Thomas, B. C. and Banfield, J. F. (2016). A new view of the tree of life. *Nature Microbiology* **1**: 1–6.

Juhas, M., Van Der Meer, J. R., Gaillard, M., Harding, R. M., Hood, D. W. and Crook, D. W. (2009). Genomic islands: Tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiology Reviews*. **33**: 376–393.

Kang, D. D., Froula, J., Egan, R. & Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ*. **3**: e1165.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M. and Glöckner, F. O. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*. **41**: 1–11.

Kotelnikova, S. (2002). Microbial production and oxidation of methane in deep subsurface. *Earth-Science Reviews*. **58**: 367–395.

Labonté, J. M., Field, E. K., Lau, M., Chivian, D., van Heerden, E., Wommack, K. E., Kieft, T. L., Onstott, T. C. and Stepanauskas, R. (2015). Single cell genomics indicates horizontal gene transfer and viral infections in a deep subsurface Firmicutes population. *Frontiers in Microbiology*. **6**: 1–11.

Lang, A. S., Zhaxybayeva, O. and Beatty, J. T. (2012). Gene transfer agents: phage-like elements of genetic exchange. *Nature Reviews Microbiology*. **10(7)**: 472–82.

Lau, M. C. Y., Cameron, C., Magnabosco, C., Brown, C. T., Schilkey, F., Grim, S., Hendrickson, S., Pullin, M., Lollar, B. S., van Heerden, E., Kieft, T. L. and Onstott, T. C. (2014). Phylogeny and phylogeography of functional genes shared among seven terrestrial subsurface metagenomes reveal N-cycling and microbial evolutionary relationships. *Frontiers in Microbiology*. **5**: 1–17.

Lindell, D., Jaffe, J. D., Johnson, Z. I., Church, G. M. and Chisholm, S. W. (2005). Photosynthesis genes in marine viruses yield proteins during host infection. *Nature*. **438**: 86–89.

Lippmann, J., Stute, M., Torgersen, T., Moser, D. P., Hall, J. A., Lin, L., Borcsik, M., Bellamy, R. E. S. and Onstott, T. C. (2003). Dating ultra-deep mine waters with noble gases and ³⁶Cl, Witwatersrand Basin, South Africa. *Geochimica et Cosmochimica Acta*. **67**: 4597–4619.

Maaroufi, H. and Tanguay, R. M. (2013). Analysis and phylogeny of small heat shock proteins from marine viruses and their cyanobacteria host. *PLoS One*. **8**: e81207.

Mabizela, L. (2009). *A metagenomic investigation of phage communities from South African mines* (PhD thesis). University of the Free State.

Magnabosco, C., Tekere, M., Lau, M. C. Y., Linage, B., Kuloyo, O., Erasmus, M., Cason, E., van Heerden, E., Borgonie, G., Kieft, T. L., Olivier, J. and Onstott, T. C. (2014). Comparisons of the composition and biogeographic distribution of the bacterial communities occupying South African thermal springs with those inhabiting deep subsurface fracture water. *Frontiers in Microbiology*. **5**: 1–17.

Malde, A., Gangaiah, D., Chandrashekhar, K., Pina-Mimbela, R., Torrelles, J. B. and Rajashekara, G. (2014). Functional characterization of exopolyphosphatase/guanosine pentaphosphate phosphohydrolase (PPX/GPPA) of *Campylobacter jejuni*. *Virulence*. **5**: 521–33.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J. and Edwards R. A. (2008). The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BioMed Central Bioinformatics*. **9**: 386.

Moller, A. K., Barkay, T., Hansen, M. A., Norman, A., Hansen, L. H., Sorensen, S. J., Boyd, E. S. and Kroer, N. (2014). Mercuric reductase genes (*merA*) and mercury resistance plasmids in High Arctic snow, freshwater and sea-ice brine. *FEMS Microbiology Ecology*. **87**: 52–63.

Moser, D.P., Onstott, T.C., Fredrickson, J.K., Brockman, F.J., Balkwill, D.L., Drake, G.R., Pfiffner, S.M., White, D.C., Takai, K., Pratt, L.M., Fong, J., Sherwood Lollar, B., Slater, G., Phelps, T.J., Spoelstra, N., DeFlaun, M., Southam, G., Welty, A.T., Baker, B.J. and Hoek, J. (2003). Temporal shifts in the geochemistry and microbial community structure of an ultradeep mine borehole following isolation. *Geomicrobiology. Journal*. **20**: 517–548.

Moussatova, A., Kandt, C., O'Mara, M. L. and Tieleman, D. P. (2008). ATP-binding cassette transporters in *Escherichia coli*. *Biochimica Biophysica Acta - Biomembranes*. **1778**: 757–1771.

- Muyzer, G., Brinkoff, T., Nübel, U., Santegoeds, C., Schäfer, H and Wawer, C.** (2004). Denaturing gradient gel Electrophoresis (DGGE) in microbial ecology. In: *Molecular microbial ecology Manual* (2nd ed.) *Kluwer Academic Publishers*.
- Nachin, L., Nannmark, U. and Nyström, T.** (2005). Differential roles of the universal stress proteins of *Escherichia coli* in oxidative stress resistance, adhesion, and motility. *Journal of Bacteriology*. **187**: 6265–6272.
- Newberry, C. J., Webster, G., Cragg, B.A., Parkes, R. J., Weightman, A. J. and Fry, J. C.** (2004). Diversity of prokaryotes and methanogenesis in deep subsurface sediments from the Nakai Trough, Ocean Drilling Program Leg 190. *Environment Microbiology*. **6**: 274–287.
- Nies, D. H.** (1995). The cobalt, zinc, and cadmium efflux system CzcABC from *Alcaligenes eutrophus* functions as a cation-proton antiporter in *Escherichia coli*. *Journal of Bacteriology*. **177**: 2707–2712.
- Papenbrock, J., Mock, H.-P. P., Tanaka, R., Kruse, E. and Grimm, B.** (2000). Role of magnesium chelatase activity in the early steps of the tetrapyrrole biosynthetic pathway. *Plant Physiology*. **122**: 1161–1169.
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. and Tyson, G. W.** (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*. **25**: 1043–55.
- Prangishvili, D. and Garrett, R. A.** (2004). Exceptionally diverse morphotypes and genomes of crenarchaeal hyperthermophilic viruses. *Biochemical Society Transaction*. **32**: 204–208.
- Purkamo, L., Bomberg, M., Nyssonen, M., Ahonen, L., Kukkonen, I. and Itavaara, M.** (2017). Response of deep subsurface microbial community to different carbon sources and electron acceptors during ~2 months incubation in microcosms. *Frontiers in Microbiology*. **8(232)**: 1–14.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Scheweer, T., Yara, P., Peplies, j. and Glockner, F. O.** (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*. **41**: D590–D596.

Rath, D., Amlinger, L., Rath, A. and Lundgren, M. (2015). The CRISPR-Cas immune system: biology, mechanisms and applications. *Biochimie*. **117**: 119–28.

Roux, S., Enault, F., Hurwitz, B. L. and Sullivan, M. B. (2015). VirSorter: mining viral signal from microbial genomic data. *PeerJ*. **3**: e985.

Sanschagrin, S. and Yergeau, E. (2014). Next-generation sequencing of 16S ribosomal RNA gene amplicons. *Journal of Visualized Experiments*. **90**: e51709.

Schmieder, R. and Edwards, R. (2011). Quality control and pre-processing of metagenomic datasets. *Bioinformatics*. **27**: 863–864.

Suttle, C. A. (2005). Viruses in the sea. *Nature*. **437**: 356–361.

Takai, K., Moser, D., Deflaun, M., Onstott, T. C. and Fredrickson, J. K. (2001). Archaeal diversity in waters from deep South African gold mines. *Applied and Environmental Biology*. **67**: 5750–5760.

Thomas, C. M. and Nielsen, K. A. (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature Reviews in Microbiology*. **3**: 711–721.

Zhang J., Kobert K., Flouri T. and Stamatakis A. (2014). PEAR: a fast and accurate Illumina paired-end reAd mergeR. *Bioinformatics*, **30**: 614–620.

Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J. and Wishart, D. S. (2011). PHAST: A Fast Phage Search Tool. *Nucleic Acids Research*. **39**: W347–W352.

CHAPTER 5

SUMMARY

5.1. Summary

Viruses (more specifically bacteriophages which infect both bacteria and archaea) are the most abundant microorganisms in the ecosystem. In extreme environments such as the deep subsurface, bacteriophages play an important role in the altering of biogeochemical cycles and in the evolvability and survival of their hosts. Unfortunately even in their abundance, little is known about phages and their interactions with their hosts and this is due to the fact that 99% of the microorganisms in the environment (especially in the deep subsurface) cannot be cultured. Instead the use of culture-independent techniques such as microscopy and next generation sequencing based techniques. Recent research done on microbial diversities in deep subsurface environments prefers the use of metagenome platforms that have 100% sequence coverage as it allows for comprehensive sequencing as such one is able to identify and assemble genomes of species that are present at low frequencies in environmental samples. These techniques also allow for the detection of phage genes and HGT events within the genomes of the host, thereby giving insight into the host-phage interactions in these extreme environments. Access to the microbial matter in deep terrestrial subsurface is through mines and several in South African deep mines have identified phage genes and HGT events within genomes of indigenous bacteria. Microscopic techniques such as EFM and TEM have been used to detect and morphologically characterize phages (respectively). The characterization of phages in extreme environments using TEM has only been published for deep marine environments, hot springs and solfataric fields and not for the deep mine fissure water. Therefore, phages in the deep terrestrial environment have only been studied through the use of metagenomic sequencing techniques.

In this study the main objectives were to identify phage genes and the presence of horizontally acquired genes and their effect on the deep subsurface bacterial communities in terms of evolution and survival. Sampling and concentration (using tangential flow filtration) of fracture water from Star Diamonds mine, Fronteirs mining resulted in samples eligible for microscopic analysis (using SEM, EFM and TEM) and sequencing using whole metagenome sequencing to identify the phage related genes. SEM analysis revealed most of the bacterial communities are in biofilm structures and this is expected due to the unfavourable conditions of the environment. The free phages in the viral fraction fracture water were detected using EFM and the TEM analysis identified phages belonging to the *Myoviridae* and *Podoviridae* families which have been previously studied and found in the marine environments. This study is the first to identify

free viral particles in the fissure water of deep subsurface South African mines. It is also the first study to detect free viral particles in the South African diamond mine.

Whole metagenome sequencing of fracture water from Star Diamonds mine, Fronteir mining was done in order to achieve the above stated main objectives. All three domains of life (Bacteria, Archaea and Eukarya) were present in the fracture water and dominated by bacteria from the phylum Proteobacteria. The phages detected in the metagenome belonged to the order Caudovirales with the *Siphoviridae* family being the most abundant. The presence of the *Myoviridae* and *Podoviridae* families further confirmed the results from the phage characterization using TEM. Previous studies in deep subsurface samples from South Africa have discovered viral infections mostly in Firmicutes, which dominate older fracture water at deeper depths. In this study the majority of the phage sequences identified using VirSorter were from phages that infect hosts from the phylum Proteobacteria which is the most abundant phyla in the fracture water. This study therefore provides valuable insight into other host microorganisms such as the Proteobacteria which generally dominate in younger fracture water at shallower depths. Partially complete prophages were detected and annotated and the presence of prophages in South African deep mines has been previously identified by two researchers. The presence of prophages suggests that phages in this environment can be both lytic (as observed with the detection of free phages using TEM) and lysogenic, but the low abundance of phages detected using TEM and EFM suggests that they prefer lysogenic infections.

The CRISPRs, mobile/transposable elements, transposase and retrons detected within the binned metgenome data are suggested to be markers for possible phage mediated HGT events by previous researchers. This study further identified genes suggested to be products of HGT events that were part of the nitrogen fixation, cobalamin synthesis and sulfide reduction pathways and motility and sporulation. These genes conferred novel capabilities to the host (that they were transduced into via HGTs) for survivability and evolution in the extreme deep subsurface environment.

This is the first study to specifically look at the bacteria-phage interactions within the subsurface mines of South Africa and it is also the first to find viral infections within the Star Diamonds mine. As future research, phylogenetic analyses would need to be done in order to further confirm the identified HGT events.